Data Augmentation for medical imaging: a Systematic Literature Review

Data Augmentation for medical imaging: a Systematic Literature Review

Fabio Garcea[a], Alessio Serra, Fabrizio Lamberti[a], Lia Morra[a]

[a]*Dipartimento di Automatica e Informatica, Politecnico di Torino, C.so Duca degli Abruzzi, 24, Torino, 10129, Italy*

Abstract

Recent advances in Deep Learning have largely benefited from larger and more diverse training sets. However, collecting large datasets for medical imaging is still a challenge due to privacy concerns and labeling costs. Data augmentation makes it possible to greatly expand the amount and variety of data available for training without actually collecting new data. Data augmentation techniques range from simple yet surprisingly effective transformations such as cropping, padding, and flipping, to complex generative models. Depending on the nature of the input and the visual task, different data augmentation strategies are likely to perform differently. For this reason, it is conceivable that medical imaging requires specific augmentation strategies that generate plausible data samples and enable effective regularization of deep neural networks. Data augmentation can also be used to augment specific classes that are underrepresented in the training set, e.g., to generate artificial lesions. The goal of this systematic literature review is to investigate which data augmentation strategies are used in the medical domain and how they affect the performance of clinical tasks such as classification, segmentation, and lesion detection. To this end, a comprehensive analysis of more than 300 articles published in recent years (2018-2022) was conducted. The results highlight the effectiveness of data augmentation across organs, modalities, tasks, and dataset sizes and suggest potential avenues for

future research.

Keywords: data augmentation, deep learning, medical imaging, generative adversarial networks, MRI

---

1. Introduction

Deep learning (DL) has established state-of-the-art performance in many areas of computer vision and pattern recognition, including medical image analysis (1; 2). In order to successfully build well-generalizing deep neural networks (DNNs), we need, in most of the cases, a large dataset to avoid overfitting such large-capacity learners. Collecting a sufficient number of samples has become a significant obstacle that makes deep neural networks quite challenging to apply in medical image analysis, since the acquisition of high-quality reference standards requires a large amount of time, money, and human resources (3; 1). Moreover, most manually annotated datasets are imbalanced, with some specific classes that are often underrepresented. To mitigate the problem of limited medical training sets, several data augmentation techniques have been developed in the literature to generate synthetic training examples.

Data augmentation is a strategy allows to increase the diversity and size of data available for training models, without actually collecting new data (4). This is achieved by applying several image manipulation techniques on the original data or by creating new samples that employ other models. Data augmentation can be used to reduce class imbalance and dataset biases, avoid overfitting and, in general, increase the variability of the data. Another well-known property of data augmentation is that it promotes learning invariance with respect to transformations of the input data that should not affect the output. Thus, it provides a convenient way to learn invariance without designing the model architecture to be equivariant or invariant (5).

Different data augmentation techniques are likely to be more or less effective depending on the type of input, downstream visual task and application field. In the literature, there are many reviews and surveys that analyze different data augmentation techniques in different fields, such as (6), which presents existing methods for data augmentation, promising developments, and high-level guidelines for implementing data augmentation on RGB images. However, given the unique properties of medical images, it is likely that data augmentation should be tailored for this specific domain (7). To the best of our knowledge, existing surveys in the medical domain cover specific subfields: (8) as an example highlights the most promising research directions for synthesizing high-quality artificial brain tumor examples with the goal of improving the generalization capabilities of deep models. Considering the large number of data augmentation techniques applied in the most different subfields of medical image analysis, it is not trivial to identify the most appropriate one for a specific task.

Inspired by these premises, in this paper, we present a systematic literature review exploring the data augmentation techniques used in the context of medical image analysis. In particular, we identify, analyze, and summarize techniques categorizing them by the visual task to which they are applied and by other medical-specific factors such as pathology, organ, or data modality. As a result of our analysis, we provide a detailed comparison of the usage of different data augmentation techniques in the medical domain, as well as a discussion on their potential impacts on the performance of deep learning models. Moreover, we identify the current knowledge gaps, also by comparing with the corresponding literature in the non-medical domain. The objectives of our review can be summarized with the following research questions:

- **RQ1**: What are the most common study designs to present and assess data

augmentation methods in the medical domain? Are there studies comparing the effect of different data augmentation models on downstream tasks?

- **RQ2**: What types of data augmentation are used in the medical domain?
- **RQ3**: What are their effects on the performance of deep learning-based methods for medical image analysis?
- **RQ4**: Which data augmentation methods have not been explored in the medical domain?

2. Background

In this section, we first propose a taxonomy for data augmentation strategies used in the medical domain, illustrated in Fig. 1, and provide a brief introduction to each transformation family. Following previous reviews (6; 9), we divided data augmentation techniques into two broad categories: *transformation of original data*, detailed in Section 2.1 and *generation of artificial data*, detailed in Section 2.2.

The rest of this section introduces concepts that apply transversely to many augmentation techniques. In Section 2.3, specific challenges arising from the volumetric nature of many medical modalities are discussed. In Section 2.4, the concept of learnable data augmentation is introduced, which aims at learning to combine and adapt multiple transformations and their hyper-parameters to a specific dataset or task. Finally, we briefly mention how data augmentation can be exploited at test time to increase accuracy and reduce uncertainty (Section 2.5).

2.1. Transformation of original data

The first category of data augmentation strategies apply one or more image manipulation techniques to existing samples. They range from the simplest affine trans-

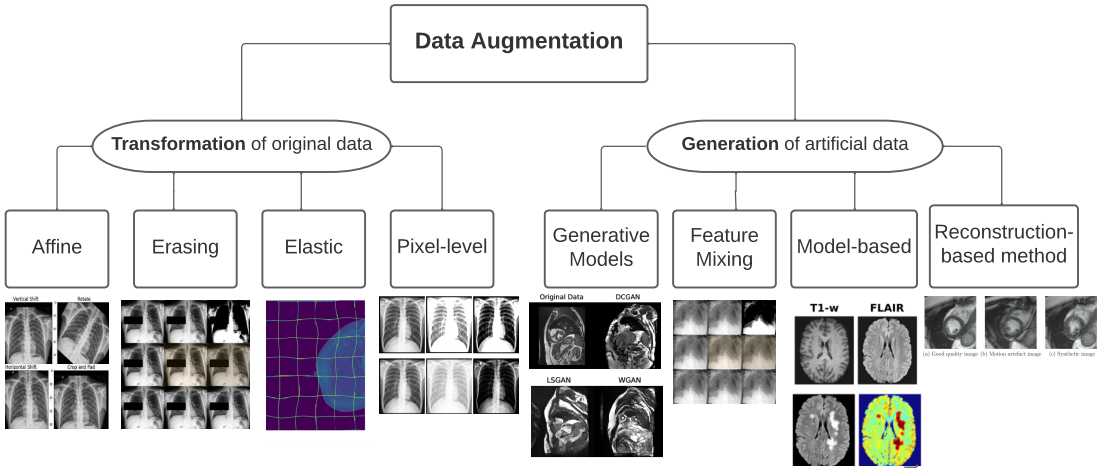


Figure 1: Data augmentation for medical imaging - a taxonomy.

formations (such as flipping or rotating an image) to pixel-level and elastic distortions. Most of the transformations included in this category are relatively easy to implement and are either built-in in deep learning frameworks or provided by easily integrated general purpose libraries (10). More recently, frameworks and libraries specifically targeting the medical domain, such as the Medical Open Network for AI (MONAI), have been proposed (11; 12). It must be observed that, being based on transformations of the original samples, these techniques cannot increase the generalization capabilities of the network beyond the initial training population, and tend to generate highly correlated samples.

2.1.1. Affine transformation

Affine transformations refer to a class of geometrical transformations that preserve lines and parallelism, but not necessarily distances and angles. This behavior is guaranteed by the constraint imposed during the transformation, which typically preserves the aspect ratio of the image along one or more axes of symmetry. Trans-

formations in this group are translation, rotation, flipping, scaling, cropping, and shearing. All these geometrical transformations can be expressed in the form of matrices and thus can be combined to obtain composed affine transformations. Examples of affine transformations are reported in Fig. A.1.

2.1.2. Erasing transformation

Erasing transformations select a region of an image and replace the selected pixels with a fixed pixel intensity value or random noise. This transformation was originally proposed in the RGB domain to increase robustness to occlusions (which however are not present in the medical domain) and to force DNNs to avoid simplified detection patterns. It also reduces spurious correlations between the object of interest and other features, usually due to dataset biases. An example of an erasing transformation is reported in Fig. A.2.

2.1.3. Elastic transformation

Elastic transformation operate by applying a spatial deformation field to an image. Differently from affine transformations, elastic transformations do not impose any constraint on the preservation of col-linearity and aspect ratio, and thus can introduce local shape variation into an organ or lesion. These transformations can be used, for instance, to increase the robustness of segmentation algorithms (8). Elastic transformation may also simulate image deformations due to breathing and patient movements, thus making DNNs more robust to such common artifacts (13; 14). A critical aspect to consider is that unconstrained elastic transformation may yield highly unrealistic samples, although the impact on the performance of downstream tasks is still being debated (8).

An important class of elastic transformation are those based on *diffeomorphism*, i.e. a mapping between two differentiable and invertible mainfolds (8). Diffeomorphic

mapping can be used to transform a source image I to a target image J. By using an actual sample to guide the elastic transformation, it is more likely to obtain a plausible and realistic augmentation. An example of elastic transformation is shown in Fig. A.3.

2.1.4. Pixel-level transformation

Pixel-level transformations change the pixel values in order to modify image characteristics such as brightness, contrast, saturation and noise. Most medical imaging modalities are grayscale, and hence color-based transformations are not common. Pixel-level transformations are useful to increase DNNs robustness across different scanners and imaging protocols, which may affect the pixel distribution. Examples of pixel-level transformations are shown in Fig. A.4.

2.2. Generation of artificial data

Generating artificial or synthetic samples can yield more diverse and challenging samples, thus overcoming the limitations of transformation-based techniques. Generative networks, and specifically generative adversarial networks, are today the most common approach to medical image synthesis (15). Yet, artificial images can be generated also through feature mixing or leveraging ad-hoc modelling strategies tailored to a specific medical imaging task or modality. These techniques allow for a greater variety, at the expense of increased computational requirements and complexity. Artificial samples may also not reflect the visual characteristics or distribution of true samples.

2.2.1. Generative (Adversarial) Networks

Generative models have been used to generate realistic images (16; 17; 18), speech (19; 20), text (21; 22) and more. Generative models include variational autoen-

coders (23; 24; 25), generative adversarial networks (26; 23; 15), and, more recently, diffusion-based models (27; 28).

Today, the most common generative architecture is the Generative Adversarial Network (GAN), introduced by (26). GANs are composed by two components, namely a generator and a discriminator. While the generator tries to trick the discriminator into believing that the fake data is authentic, the discriminator attempts to discern if the image is real or synthetic. The two networks contest with each other in a zero-sum game, where the gain of one is the loss of the other. The success of GANs can be attributed to their exceptional visual fidelity, especially when compared to variational autoencoders.

A thorough analysis of GAN applications in medical imaging, including commonly used architectures and losses, is available in (15). We report some examples of synthetic images generated by GANs in Fig. A.5. It is worth mentioning that GANs allow for controlled image generation (29) that is particularly useful in the context of data augmentation. For instance, GANs conditioned on a label or a segmentation map can be used to generate synthetic lesions or, more generally, to balance a dataset by augmenting underrepresented groups (30). Image translation architectures, such as CycleGANs (31) have been used for cross-domain medical image synthesis, which allows transferring samples from modalities for which data is relatively abundant (e.g., CT) to more expensive or less widespread modalities (e.g., MRI) (32). It can also be used to hallucinate missing modalities for multi-parametric sequences, such as MRI (33).

However, leveraging GANs for data augmentation is tricky. Besides the complexity and computational cost of generating high resolution, realistic images, GANs are prone to mode collapse (34), in which the generator is prone to yield very similar examples and thus has limited impact on generalizability, and have been shown to

hallucinate image features that mimic or mask the presence of lesions (35).

In very recent years, following seminal work by Ho and colleagues (36), several diffusion-based generative models have been proposed achieving extraordinary results in photorealistic image synthesis, including unconstrained and text-to-image synthesis. These models generate samples by gradually removing noise from a signal and, compared to GANs, have more desirable properties in terms of distribution coverage, ease of training and scalability. Now that diffusion-models are now achieving competitive results compared to GANs (27), it is reasonable to expect that, in the future, this class of model will start to complement the use of GANs in the medical domain as well.

2.2.2. Feature mixing methods

Feature mixing methods combine two or more samples from the original dataset to create a new one. Mixing images together by averaging their pixel values or cutting and pasting some parts of an image into another may represent counter-intuitive approaches from a human perspective. In fact, generated images may not look meaningful to human observers (6). However, these techniques were shown to enhance generalization capabilities of DNNs, and increase robustness to adversarial examples (37). The mix-up technique, shown in Fig. A.7, is one of the most used feature mixing methods in the medical field.

2.2.3. Model-based methods

Under the umbrella of model-based techniques we include a variety of physically or biologically inspired models to generate new images, or modify existing ones. These transformations can be performed both using DNNs (not necessarily GANs) or traditional image processing techniques, such as shape modelling, shape deformation or image blending. Since medical datasets are often very skewed towards negative

cases, the most common model-based techniques involve the synthesis and injection of artificial lesions in otherwise healthy subjects. For instance, these techniques have been used to simulate multiple sclerosis lesions in brain MR images (38) or to add cancer signs to breast mammography images (39). An example of these model-based data augmentation is shown in Figure A.8.

Other techniques incorporate physiological time progression models, such as reproducing the effect of aging on organs or simulating disease progression and regression (40). Finally, some model-based techniques exploit prior knowledge on human anatomy and on images formation process to generate random variations representative of inter-subject variability, or to simulate common acquisition artifacts to increase the robustness of image analysis algorithms (41). Unlike GAN-based techniques, these techniques typically require minimal or no training, and directly incorporate anatomical constraints on the generated images. However, they can only be applied to specific organs for which mathematical models are available.

2.2.4. Reconstruction-based methods

Three-dimensional medical imaging modalities are reconstructed using a mathematical process that transforms the raw data collected by the image scanner into a three-dimensional volume that can be inspected by the radiologist. For example, image reconstruction in CT involves generating tomographic images from X-ray projections acquired at many different angles. Reconstruction algorithms and their settings influence the signal-to-noise ratio, resolution, presence of artifacts, and general quality of the reconstructed images. While the majority of data augmentation techniques operate directly on the reconstructed images, a few techniques have been proposed that operate directly on the raw data space and then reconstruct the distorted images. For example, in (42) realistic motion artifacts are simulated by corrupting the

raw data using Gaussian blurring, and then a segmentation network is trained on increasingly distorted data. These techniques may be relatively complex to apply, as they require access to the raw data acquired by the scanner (often not available in a clinical setting), and implementing the entire reconstruction algorithm. A subset of these techniques involves reconstructing 2D images (e.g., X-rays) from 3D volumes (e.g, CT scans), in order to simulate different acquisition angles and views (43). An example is shown in Figure A.6.

2.3. 2D vs. 3D data augmentation

One of the characteristics of modalities like CT and MRI is their volumetric nature. Nonetheless, not all DNNs employed in the medical domain employ 3D convolutions (44; 7). Many authors, especially toward the beginning of the deep learning era, have used 2D convolutions, by focusing on individual slices: 2D models have lower memory and computational footprint, and also allow the use of backbones pre-trained on ImageNet. 3D convolutions have the advantage of combining information from adjacent slices, and achieve higher performance if a suitable large dataset is available (45). An alternative is to use 2.5D convolutions, slices captured along three perpendicular planes (axial, coronal, and sagittal) mimic RGB images and are input to the three channels of ImageNet-pretrained models. Some image transformations (most notably, affine transformations) are widely available for 2D images, but may require extensions to 3D, which may not be straightforward or computationally efficient. Examples of 3D-aware data augmentations include 3D GANs (46), multiplanar image synthesis (47) and 3D affine transformations (48).

2.4. Learnable data augmentation

Learnable data augmentation is a recent subfield of deep learning research that studies approaches that can reduce the human effort required when selecting and

validating a set of data augmentation techniques (49). The main idea is to discover automatically an optimal data augmentation strategy for a specific task. Aside from generative models, which we covered in Section 2.2.1, the most common approach is to learn an optimal data augmentation policy. Techniques based on this approach require the concurrent training of two networks, one for learning how to solve a task and the second to learn how to augment the data for the first one. The most common approach is augmentation policy learning or autoaugment, which learns the best policy to maximize a network performance by combining a list of known transformations (such as affine, pixel-level, and so forth) (50). The optimal policy can be determined by a neural network, trained using reinforcement learning (50), evolutionary algorithms (51), adversarial training (52) or parametric statistical models (53).

2.5. Train-Time & Test-Time Data Augmentation

Although most of the studies in this survey focus on improving data at training time, test time augmentation (TTA) has also been explored to increase DNN accuracy and robustness. In test-time augmentation, multiple transformations of a given input are generated, and then the DNN predictions are averaged to obtain a more robust estimate. TTA has been shown to increase the accuracy of a DNN (6), and can also be used to estimate the aleatoric uncertainty of the predictions (54; 55). Unlike epistemic uncertainty, which refers to the DNN parameters, aleatoric uncertainty reflects noise or randomness in the input image, and thus cannot be explained or reduced by increasing the training data (54).

3. Methodology

In this section we describe the methodology adopted to perform the systematic literature review. The analysis was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) guidelines.

3.1. Search strings and Study selection

The search was conducted on PubMed and on the search engine of the main publishers (ACM, IEEE, and Elsevier) in the fields of computer science, medical imaging and biomedical engineering. We searched for the keywords "data augmentation" and "medical imaging", as shown in Table B.7. We considered only studies published between January 2018 and July 2022.

We defined inclusion and exclusion criteria on parameters such as language, publication date, type of study, modality, context, and task. We selected only primary studies written in English from 2018 to 2022, including both proceedings and journal papers, but excluding preprints. We included papers related to the main medical imaging modalities: X-Ray (XR), Magnetic Resonance (MR), Computed Tomography (CT), Positron Emission Tomography (PET) and Ultrasound (US); we excluded papers in the dermatology, ophthalmology, endoscopy, and histopathology domains. We also excluded studies that do not use data augmentation or for which the full text was not available. The downstream tasks included in the review were classification, detection, and segmentation (in other words, we selected papers in which the efficacy of data augmentation was tested on classification, lesion detection or segmentation DNNs). Moreover, we specifically selected only papers that used DNNs to perform any of the downstream tasks selected, to increase the homogeneity of the selected studies. The selection was done in two steps: first, we selected papers based on their

titles and abstracts, and then performed a more in depth analysis on the full-texts. Inclusion and exclusion criteria are summarized in Table B.8.

3.2. Data extraction

To answer **RQ1**, we classified all retrieved studies in four categories: **Type 1** (standard data augmentation techniques are used, impact on the performance of the downstream task not assessed); **Type 2** (novel data augmentation techniques are proposed, impact on the performance of the downstream task not assessed); **Type 3** (performances of the downstream task with and without data augmentation are compared) and **Type 4** (performances of the downstream task with several data augmentation strategies are compared). Each paper was assigned to the highest number category (i.e., a paper that proposes a novel data augmentation technique, and compare its impact on a classification task against standard data augmentation, would be classified as Type 4, rather than Type 2). In addition, only papers in categories 3 and 4 were used to answer the question **RQ3**. The resulting categorization and an interpretation legend are reported in Table B.9.

3.3. Data analysis

The papers were reported grouping into five groups (*brain*, *lung*, *heart*, *breast* and *others*) according to the target anatomical district. The impact of data augmentation on the performance of downstream tasks was estimated by calculating the relative increase in performance with and without data augmentation, and the distribution of the relative increase (overall and by organ) was calculated using boxplots. The same technique was used to calculate the effect of different data augmentation techniques, focusing in particular on studies in which complex data augmentation tecniques were compared against the simplest transformations (affine).

4. Results

The selection process was conducted in August 2022. The preliminary set initially contained 919 papers, but after the removal of 45 duplicates, the set was reduced to 874 papers. After applying the inclusion and exclusion criteria reported in Table B.8 to the titles and abstracts, the set of articles was reduced to 493 candidates. In particular, we discarded 53 studies that did not belong to the medical domain, 47 reviews or surveys, 74 studies that did not use any data augmentation, and 207 studies classified as out of scope (i.e., they did not meet the inclusion criteria related to the imaging modality or downstream task). Additional 10 studies were included through other sources (e.g., based on the authors' expertise, citations, or experts' suggestions). The full-texts of the remaining 503 papers were analyzed in detail; we further discarded 141 articles that did not meet inclusion criteria and selected the remaining 362 as candidates for the literature review. Abbreviations used in this section are reported in Table 1. The control flow diagram of the selection process is shown in Fig. 2.

4.1. RQ1 & RQ2: Which data augmentation techniques have been studied in the medical domain?

Each article was assigned to a category according to the methodology described in Section 3. As it can be noticed by analyzing the distribution shown in Fig. 3a, more than half of the papers, roughly 54% (196/362), quantitatively evaluate the effect of at least one data augmentation technique, while only 13% (47/362) propose a new technique without evaluating its impact on multiple downstream tasks.

The distribution of different data augmentation techniques is shown in Fig. 3b. The most used type of data augmentation is affine transformations, being used in

Abbreviation	Meaning	Abbreviation	Meaning
(*)	Test Time Augmentation	M-AF	Mix Affine
ACC	Accuracy	MD	Mean Distance (mm)
ACS	Acute Coronary Sindrome	MG	Mammography
AF	Affine	MOD	Model-based
ALZ	Alzheimer	MR	Magnetic Resonance
AUC	Area under the ROC Curve	MS	Multiple Sclerosis
BIF	Bifurcation	NMD	Neuromuscolar Diseases
C	Classification	NOD	Nodule
CAN	Cancer	OS	Osteoporosis
CAS	Coronary Angioscopy	OSD	One Stage Detector
COV	COVID-19	PACC	Pixel Accuracy
CPM	Competition Performing Metric	PAR	Parkinson
CT	Computed Tomography	pGAN	custom GAN
CY	Cyst	PNE	Pneumothorax
D	Detection	POL	Polyp
DICE	DICE Score	PSNR	Peak Signal-to-Noise Ratio
EL	Elastic	PX	Pixel-level
ER	Erasing	REC	Reconstruction-based
F1	F1 Score	S	Segmentation
FM	Feature Mixing	S-AF	Single Affine
FPI	False Per Image	SNS	Sensitivity
FRC	Fractures	SP	Specificity
FROC	Free-Response Receiver Operating Characteristic	TPR	True Positive Rate
GAN	Generative Adversial Network	TUB	Tubercolosis
GN	Generative Network	URS	Urinary Stones
H	Healthy	US	Ultrasound scan
HAND	Hand Injuries and Disorders	wGAN	Wasserstein GAN
iGAN	Info GAN	XR	X-Ray scan
L	Lesion	OTH	Other
TF	Transformer	LRN	Learnable data augmnetation

Table 1: List of the abbreviations used in Section 4

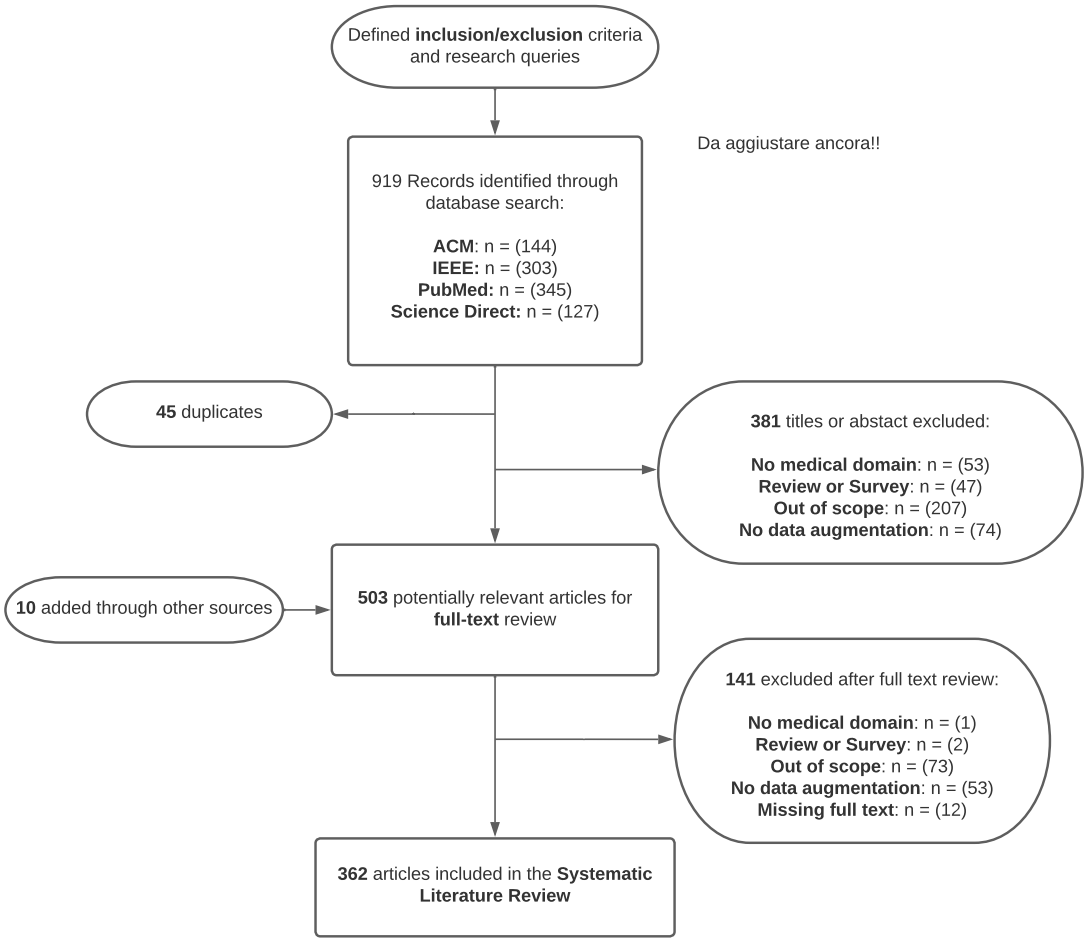
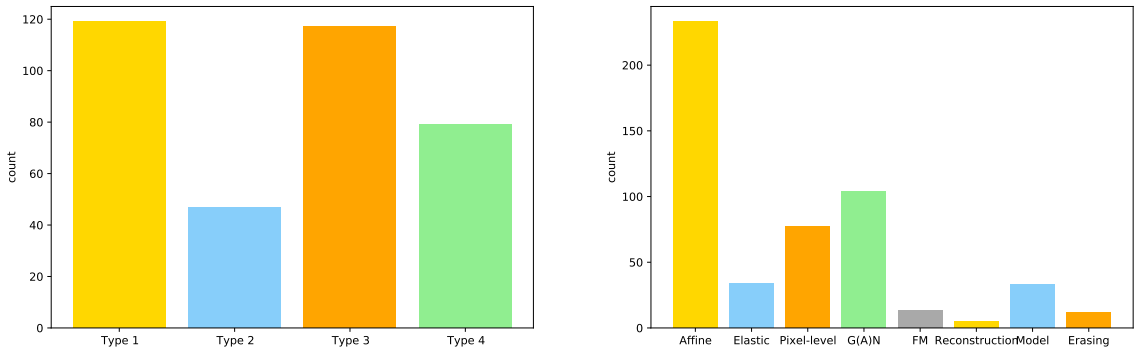


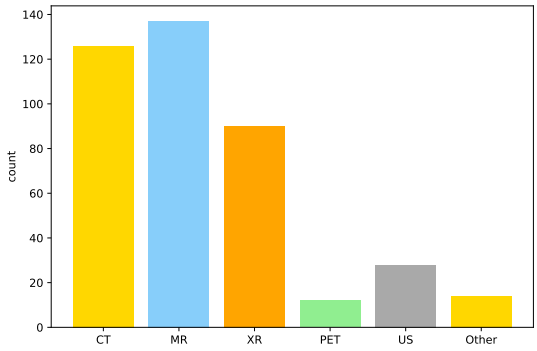
Figure 2: Systematic Literature Review Control Flow Diagram

64% (233/362) of the studies. This result was expected, as this type of data augmentation is fairly effective and extremely easy to implement. GAN-based and pixel-level transformations follow, appearing in 29% (104/362) and 22% (77/362) of the studies, respectively. Pixel-level transformations are popular, as they enhance generalization to different scanners and acquisition protocols (56). Model-based methods and elastic transformation are used in approximately 9% of the studies (33/362 and 34/362,

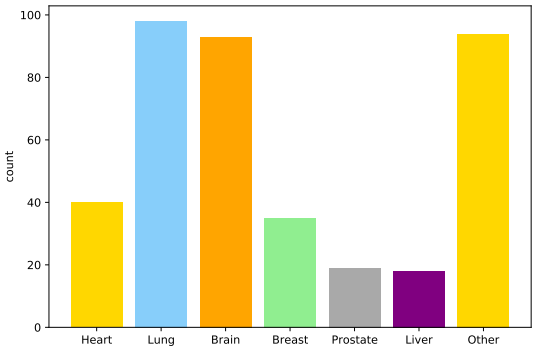


(a)

(b)



(c)



(d)

Figure 3: Distribution of studies by (a) study design, (b) data augmentation category, (c) modality and (d) organ.

respectively), and the remaining techniques only represent 8% (30/362) of our sample (42; 43; 57; 58; 59; 60).

As shown in Fig. 3c, CT, MR and XR (including mammography) are the most common modalities in our sample, reflecting their widespread adoption in clinical practice. Under *other*, we grouped less common modalities such as *Coronary Angioscopy (CAS)* (61), *Myocardial perfusion imaging (MPI)* (62) and *Digital breast tomosynthesis (DBT)* (63). We identified a few studies that leveraged *cross-modality*

image synthesis for data augmentation, e.g., setting MR as the source and CT as the target (64), CT as the source and MR as the target (65), and bidirectional style transfer between the two modalities (66).

Lastly, in Fig. 3d, we report the distribution of the studies with respect to different organs. As it can be noticed, the most studied organs are lungs and brain with 98 and 93 papers respectively. Under the category *other* we grouped organs included in less than 10 studies, such as the musculoskeletal system (arm, calcaneus, knee, hand, humerus, hip, maxillary, neck, pelvis, skull, shoulder, spine), specific organs in the abdomen (colon, kidney, lymph node, pancreas, stomach, rectum), the pituitary membrane, teeth, thyroid, and the urinary conduct.

Studies were further grouped by type of data augmentation and (i) organ, (ii) modality, (iii) type of paper, and (iv) downstream task (related bar plots are reported in the Appendix Appendix C). In general, there does not appear to be a qualitative correlation between organ, modality, and downstream task and the distribution of the related data enhancement strategies; affine transformations represent the most common technique in all cases. In most cases, type 1 papers include affine and pixel-level data augmentation techniques, whereas type 2, 3 and 4 papers tend to be biased towards more complex strategies, such as generative or model-based techniques. This particular result may be a side-effect of our selection process, which sought to highlight novel data augmentation strategies emerging in the literature. The results are consistent with a previous study by Nalepa et al., who compared the data augmentation strategies employed in the BraTS challenge, mostly affine or pixel-level transformations (9).

4.2. RQ3: What is the effect of data augmentation on the performance of the downstream task?

To evaluate the effect of different data augmentation techniques, we only selected type 3 and type 4 papers that make a comparison with other data augmentation strategies or without data augmentation. Papers for which performance information could not be extracted (14 papers) were excluded from the quantitative analysis. We report the results separately for papers that include experiments on different organs or tasks, and hence the total number of records is 206 out of 196 papers.

The main characteristics of these studies are reported in the following sections, grouped by organ. For each study, we report the modality, the target disease, the type of DA used, the characteristics of the downstream task, and the performance metrics used. Whenever available, we report the percentage increase in performance with DA with respect to the experiments in which DA was not used (DA vs. no-DA). For type 4 papers, we selected the simplest form of data augmentation (typically affine transformations) as the baseline, the best performing data augmentation strategy as the proposed DA strategy, and report the relative difference in performance of the proposed vs. baseline DA. The use of relative differences allows us to compare studies based on different metrics.

4.2.1. Brain

Neuroimaging is used to assess (directly or indirectly) the structure and function of the nervous system, primarily with computed tomography (CT), magnetic resonance imaging (MR), and positron emission tomography (PET). The main pathologies analyzed in the different studies are Alzheimer's disease, Multiple Sclerosis, Parkinson's disease, and brain tumor.

The most common downstream tasks in brain imaging, as reported in Table

2, are classification (18/49) and segmentation (30/49). Classification is useful in determining the stage of progression in brain disease, such as in (67) for *glioma grading*. Segmentation, on the other hand, finds application both in healthy subjects and in the quantitative study of different brain diseases (68). We excluded from the analysis one paper (69) in which data augmentation was used to make the network robust to adversarial attacks, instead of improving the its generalization ability.

Every data augmentation technique used in these studies improves performance compared to those who do not use data augmentation. The highest relative increase is achieved by (30) with an increase of 26.56%, and the highest relative improvement with respect to baseline DA is achieved by (70) with an increase in performance of 10.36%.

Model-based techniques (9/39) and G(A)Ns (18/49) are used in more than half the selected studies. It is possible that the relatively stable structure of the brain, and the fact that accurate anatomical models have been developed over the course of many years, facilitate the application of these techniques. Many model-based techniques often involve the injection of synthetic lesions into otherwise healthy brains, such as the introduction of synthetic MS lesions (38), brain tumors (71) or microbleeds (72). Model-based techniques have also been used to simulate common image acquisition artifacts such as intensity inhomogeneity (73), with the aim of making DNNs more robust to artifacts. This is a unique feature of DNNs which can learn features that are invariant to the presence of specific artifacts, if exposed to them during training, shifting attention from removing such artifacts from the images (e.g., through normalization techniques) to selecting more realistic and varied training data.

In (74), the authors propose a framework based on GANs to create structural synthetic brain networks in multiple sclerosis. This paper shows that generative

models are competitive with respect to DA strategies that operate in the feature domain, such as Synthetic Minority Oversampling Technique (SMOTE) (75).

A few authors have proposed to use data augmentation techniques in order to train DNNs in few-shot learning scenarios by leveraging otherwise unlabelled data, thus pushing forward the boundaries of data augmentation. For instance, (76) and (70) used model-based and generative data augmentation techniques, respectively, to learn spatial and appearance transformations to match images of real healthy subjects to one or more reference atlases: since segmentation is known for the atlas, the images generated through the proposed methodology can be automatically labeled.

4.2.2. Heart

Cardiac imaging refers to non-invasive imaging of the heart using primarily ultrasound (US), magnetic resonance imaging (MR) and computed tomography (CT). Studies related to this organ are reported in Table 3; the most common downstream task is segmentation (15/21), followed by classification (4/21) and detection (2/21).

As detailed in Table 3, data augmentation techniques used in these studies usually improve performance with respect to the no-DA baseline and the other DA baselines. In interpreting the results, attention must be paid to the baseline performance, which in some cases is very low (114), likely reflecting poor hyper-parameter optimization and perhaps inflating comparison. Best results are achieved by Spectral augmentation (115) (DICE: 0.891 to 0.811) and BigAug (56) (DICE: 0.858 to 0.914). Some of the less performing studies, on the other hand, employ transformations that are not suited to this specific organ. For instance, horizontal flip was found to reduce performance, except in the very small data scenario (116), likely due to the fact that the heart has a specific orientation. Several GAN variants have been proposed for heart imaging, including *SpeckleGAN* (117), *ScreenGAN* (118) or *CycleGAN* (65).

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	w/o DA -¿ Proposed	Baseline -¿ Proposed
(77)	C	CNN	MR	CAN	M-AF	S-AF	ACC	0.64	0.77	0.79	23.44%	2.60%
(78)	C	CNN	CT	H	AF	-	ACC	0.73	-	0.923	26.44%	-
(57)	C	CNN	PET	PAR	MOD	-	ACC	0.876	-	0.882	0.68%	-
(79)	C	CNN	MR	CAN	AF	-	ACC	0.976	-	1	2.46%	-
(80)	C	CNN	MR	CAN	AF	-	ACC	0.923	-	0.983	6.50%	-
(81)	C	CNN	PET-MR	ALZ	GAN	-	ACC	0.67	-	0.74	10.45%	-
(82)	C	CNN	MR	CAN	AF - PX	-	ACC	0.874	-	0.907	3.78%	-
(83)	C	CNN	MR	CAN	pGAN	GAN	SNS	0.689	0.74	0.789	14.51%	6.62%
(84)	C	CNN	MR	ALZ	AF - PX - ER	-	ACC	0.834	-	0.888	6.47%	-
(74)	C	AU + RF	MR	MS	GAN	-	F1	0.656	-	0.81	23.48%	-
(85)	D	Yolo-3	MR	CAN	pGAN	GAN	SNS	0.83	-	0.91	9.64%	-
(86)	S	U-Net	MR	MS	AF	-	DICE	0.803	-	0.866	7.85%	-
(87)	S	U-Net	MR	MS	MOD	AF	CPM	-	0.77	0.79	-	2.60%
(88)	S	AMRUNet	MR	CAN	FM	-	DICE	0.771	-	0.787	2.08%	-
(89)	S	U-Net	MR	CAN	GAN	AF - PX	DICE	-	0.841	0.851	-	1.19%
(90)	S	U-Net	MR	CAN	AF	-	DICE	0.636	-	0.701	10.22%	-
(71)	S	U-Net	MR	CAN	GAN - AF	AF	DICE	0.623	0.644	0.653	4.82%	1.40%
(91)	S	U-Net	MR	CAN	GAN	-	DICE	0.71	-	0.736	3.66%	-
(30)	S	U-Net	MR	CAN - ALZ	GAN	-	DICE	0.64	-	0.81	26.56%	-
(92)	S	U-Net	MR	L	GAN	-	DICE	0.438	-	0.55	25.57%	-
(90)	S	U-Net	MR	L	AF	-	DICE	0.427	-	0.46	7.73%	-
(73)	S	U-Net	MR	H	MOD-AF	AF	DICE	-	0.876	0.892	-	1.83%
(93)	S	Custom	MR-CT	H	pGAN	GAN	DICE	-	0.783	0.81	-	3.45%
(94)	S	U-Net	MR	H	EL	AF	DICE	-	0.884	0.906	-	2.49%
(68)	S	U-Net	PET	H	GAN	-	DICE	0.872	-	0.889	1.95%	-
(95)	S	Custom	MR	H	PX	-	DICE	0.805	-	0.82	1.86%	-
(96)	S	U-Net	MR	H	GN	-	DICE	0.83	-	0.88	6.02%	-
(40)	S	U-Net	CT	CAN	MOD	AF	DICE	0.852	0.859	0.875	2.70%	1.86%
(76)	S	U-Net	MR	H	MOD - AF - PX	AF - PX	DICE	0.76	0.775	0.815	7.24%	5.16%
(97)	S	U-Net	MR	CAN	AF - EL	-	DICE	0.763	-	0.8	4.85%	-
(38)	S	U-Net	MR	MS	MOD	-	DICE	0.57	-	0.63	10.53%	-
(98)	S	U-Net	MR	CAN	GAN	-	DICE	0.72	-	0.82	13.89%	-
(99)	C	CNN	MR	ALZ	AF - PX - ER	-	ACC	0.786	-	0.92	17.05%	-
(100)	C	CNN	MR	OTH	AF - PX	-	AUC	0.967	-	0.987	2.07%	-
(101)	S	Custom	MR	CAN	AF	-	DICE	0.9	-	0.93	3.33%	-
(102)	C	CNN	MR	CAN	AF - PX	-	ACC	0.9	-	0.94	4.44%	-
(103)	S	U-Net	MR	H	FM	-	DICE	0.713	-	0.717	0.56%	-
(104)	C	CNN	MR	ALZ	GAN	-	ACC	0.806	-	0.863	7.07%	-
(105)	S	U-Net	MR	MS	PX	-	DICE	0.677	-	0.646	-4.58%	-
(106)	S	3D U-Net	DWI	OTH	AF-MOD-GAN	-	DICE	0.632	-	0.708	12.03%	-
(72)	C	CNN	MR	ALZ	MOD	AF	AUC	0.984	0.993	0.995	1.12%	0.20%
(107)	S	U-Net	MR	CAN	GAN - MOD	AF - EL	DICE	0.784	0.801	0.79	0.77%	-1.37%
(70)	S	3D U-Net	MR	H	GAN	EL - MOD	DICE	-	0.743	0.82	-	10.36%
(108)	C	3D CNN	PET	ALZ	AF - PX	AF	ACC	-	0.395	0.352	-	-10.89%
(109)	S	3D U-Net	MR	CAN	FM	AF	DICE	-	0.853	0.864	-	1.29%
(110)	S	U-Net	MR	CAN	GAN	AF	ACC	0.9127	0.9366	0.9424	3.25%	0.62%
(111)	C	CNN	MR	CAN	GAN	AF	ACC	-	0.888	0.92	-	3.60%
(112)	C	CNN	MR	CAN	FM	AF	ACC	-	0.953	0.9676	-	1.53%
(113)	S	3D U-Net	MR	CAN	AF - PX - EL	FM	DICE	0.705	0.714	0.727	3.12%	1.82%

Table 2: Comparison on **49** studies on data augmentation techniques for brain imaging. See Table 1 for an explanation of the acronyms.

This section also includes one of the few studies that evaluated the performance of *reconstruction-based* methods (42). The authors presented an automatic spatio-temporal DNNs for cardiac motion artifact data which was trained by corrupting

data in the k-space and then reconstructing the images: this strategy was able to simulate different levels of realistic motion-related artifacts (e.g., breathing artifacts).

Finally, one of the studies investigated the role of data augmentation in a federated learning setting (119). Interestingly, the authors observed that in a federated training scheme, the impact of data augmentation was lower than in the standard centralized setting, in which the data is shared on a central server. Specifically, in the federated learning setting, AUC was 0.746 ±0.001 without DA, 0.791 ± 0.001 for basic (affine) transformations, 0.766 ± 0.001 for shape transformations, and 0.731 ± 0.009 for shape and intensity transformations. In the centralized setting, the corresponding performance was generally lower and with higher variance, with AUCs equal to 0.732 ± 0.008 (w/o DA), 0.759 ± 0.016 (affine), 0.764 ± 0.022 (shape), and 0.776 ± 0.008 (shape and intensity), respectively.

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	w/o DA -¿ Proposed	Baseline -¿ Proposed
(120)	S	U-Net	CT	Many	AF	-	DICE	0.911	-	0.927	1.76%	-
(121)	S	CNN	MR	H	PX	AF - EL	DICE	-	0.652	0.894	-	37.12%
(114)	S	CNN	MR - US	H	AF - PX - EL	GAN	DICE	0.188	0.575	0.85	352.13%	47.83%
(122)	S	U-Net	MR - CT	H	MOD	FM	DICE	0.332	0.652	0.747	125.00%	14.57%
(117)	S	U-Net	US	H	pGAN	GAN	DICE	0.824	0.833	0.846	2.67%	1.56%
(65)	S	U-Net	MR - CT	H	GAN	AF	DICE	0.613	0.676	0.712	16.15%	5.33%
(58)	D	CNN	CT	ACS	AF - FM	-	AUC	0.808	-	0.867	7.30%	-
(42)	D	3D CNN	MR	Many	AF - PX - REC	AF	AUC	0.581	0.674	0.735	26.51%	9.05%
(61)	C	CNN	CAS	Many	GAN	-	AUC	0.77	-	0.81	5.19%	-
(123)	C	CNN	CT	Many	AF - ER	-	ACC	0.74	-	0.903	22.03%	-
(118)	C	CNN	US	H	GAN	-	ACC	0.7	-	0.85	21.43%	-
(124)	S	3D nn-U-Net	US	H	GAN		DICE	0.822	-	0.823	0.12%	-
(125)	S	3D U-Net	CT	H	GAN	AF	DICE	0.876	0.878	0.893	1.94%	1.71%
(115)	S	Custom	CT	H	REC		DICE	0.811	-	0.891	9.86%	-
(126)	S	3D U-Net	MR - CT	H	pGAN	GAN	DICE	-	0.687	0.847	-	-
(127)	S	U-Net	MR - CT	H	EL	PX	DICE	0.703	0.721	0.731	3.99%	1.39%
(119)	C	3D CNN	MR	Many	AF - PX - EL	AF	DICE	0.746	0.791	0.731	-2.01%	-7.59%
(52)	S	U-Net	MR	H	AF - PX - EL (**)	FM	DICE	-	0.782	0.808	-	3.38%
(56)	S	3D U-Net	MR	H	AF - PX - EL	AF	DICE	0.919	0.913	0.914	-0.54%	0.11%
(56)	S	3D U-Net	US	H	AF - PX - EL	AF	DICE	0.858	0.910	0.921	7.34%	1.21%
(116)	S	U-Net	CT	H	AF		DICE	0.750	-	0.732	-2.38%	-

Table 3: Comparison of **21** studies on data augmentation techniques for heart imaging. See Table 1 for an explanation of the acronyms.

4.2.3. Lung

Lung imaging refers to non-invasive imaging of the lungs using primarily magnetic resonance imaging (MR), computed tomography (CT), and positron emission tomography (PET). Due to the fact that lung cancer is one of the leading causes of cancer-related deaths and due to the COVID-19 pandemic, the lung is probably the most studied and analyzed organ in this entire SLR. In lung imaging, the most common downstream task is classification (37/51), whereas detection (5/51) and segmentation (9/51) are less common, as shown in Table 4.

With few exceptions, data augmentation consistently improved performance with respect to the w/o DA baseline and the other DA baseline. In one study, data augmentation was associated with a a -14.45% decrease in COVID-19 classification performance (128). It is possible that the affine transformations employed in this study (flip, translation, and scaling) do not yield images consistent with the actual distribution at test time. However, this remains an isolated case.

Overall, 20/51 papers are based on transformations of existing data, including affine transformations, pixel-level transformations and, less frequently, random erasing and elastic deformations. The remainder of the studies are based on GANs or model-based techniques. Although most generative models are single-modality, some studies have also proposed cross-modality synthesis. An example of generative model is a tumor-aware unsupervised cross-domain generative network which translates images from CT to MRI while preserving tumor details using an ad-hoc loss (32). The synthetic MRI data were then used to train a U-Net segmentation model in a semi-supervised fashion leveraging segmented images already available in the CT domain.

As mentioned in the introduction, data augmentation techniques can be employed to build invariance with respect to the transformation used. Some authors have questioned whether the invariance could be encoded directly into the DNN architecture.

A fitting example is invariance to rotation, which can be hard-coded into steerable convolution filters, bypassing the need and even surpassing the performance of rotational data augmentation (129; 130). However, it is hard to imagine that such strong inductive biases can be exploited for all relevant transformations.

Finally, a few studies have employed TTA (131; 132), observing an improvement in performance (DICE coefficient increased from 0.941 to 0.979 and 0.901 to 0.902, respectively). As introduced in Section 2.5, TTA augmentation can yield increases in performance.

4.2.4. Breast

Breast imaging is a sub-speciality of diagnostic radiology that involves imaging of the breast for cancer screening or diagnostic purposes. The most common modality is mammography (MG), complemented by breast US and DBT. In the reviewed studies (Table 5), 14/19 analyze whole-breast or patch-level classification, 3/19 lesion detection and 2/19 lesion segmentation. As in previous organs, all studies report an increase in performance when using DA with respect to non-DA.

Affine and pixel-level transformations are the most common data augmentations reported. Also elastic transformations are applicable to the breast. One study (172) compared elastic to affine transformations, decreasing the False Positives per Image (FPI) of about 25.66% in a lesion detection task.

Interestingly, several model-based techniques have been proposed in breast imaging. A first technique is based on synthetic images produced using a virtual three-dimensional anthropomorphic phantoms (39). The virtual phantom is generated using a procedural analytic model in which the main anatomical structures (including fat and glandular tissues, ductal tree, vasculature, and ligaments) are stochastically generated within a predefined breast volume. Artificial masses were inserted

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	w/o DA -¿ Proposed	Baseline -¿ Proposed
(128)	C	CNN	XR	COV	AF	-	ACC	0.868	-	0.745	-14.17%	-
(133)	C	CNN	CT - XR	PNE	PX	-	ACC	0.928	-	0.988	6.47%	-
(134)	C	CNN	CT	CAN	AF	-	ACC	0.9	-	1	11.11%	-
(135)	C	CNN	XR	TUB	AF	-	ACC	0.78	-	0.8	2.56%	-
(59)	C	CNN	XR	COV - PNE	AF - ER - FM	-	ACC	0.787	-	0.837	6.35%	-
(136)	C	CNN	XR	PNE	GAN - PX - AF	-	ACC	0.795	-	0.91	14.47%	-
(137)	C	CNN - RNN	CT - XR	COV	AF - GAN	-	AUC	0.91	-	0.99	8.79%	-
(131)	C	CNN	XR	COV	AF (*)	-	ACC	0.941	-	0.979	4.04%	-
(138)	C	CNN	XR	PNE	AF - PX	AF - PX	ACC	0.84	0.86	0.945	12.50%	9.88%
(139)	C	CNN	CT	CAN	MOD	-	ACC	0.5	-	0.786	57.20%	-
(46)	C	CNN	CT	CAN	AF - PX - GAN	AF - PX	ACC	0.5	0.761	0.825	65.00%	8.41%
(129)	C	CNN	CT	CAN	AF	-	ACC	0.842	-	0.942	11.88%	-
(47)	C	CNN	CT	CAN	AF - GAN	-	AUC	0.635	-	0.812	27.87%	-
(140)	C	CNN	CT	NOD	MOD	-	CPM	0.77	-	0.79	2.60%	-
(141)	C	CNN	CT	NOD	pGAN	wGAN	ACC	-	0.923	0.95	-	2.93%
(142)	C	CNN	XR	PNE	AF	AF - PX	ACC	0.74	0.79	0.85	14.86%	7.59%
(143)	C	PSSPNN	CT	COV	AF - PX	-	F1	0.939	-	0.958	2.02%	-
(144)	C	CNN	CT	NOD	AF	AF (*)	ACC	-	0.848	0.88	-	3.77%
(145)	C	CNN	CT	CAN	pGAN	wGAN	ACC	0.342	0.547	0.577	68.71%	5.48%
(146)	C	CNN	CT	H	MOD	AF	SNS	-	0.839	0.858	-	2.26%
(60)	C	CNN	CT	NOD	AF - FM	-	ACC	0.694	-	0.74	6.63%	-
(147)	C	CNN	CT	CAN	AF	-	AUC	0.598	-	0.655	9.53%	-
(148)	C	RNN	XR	TUB	AF - PX	-	ACC	0.471	-	0.928	97.03%	-
(149)	C	CNN	CT	NOD	AF - EL	wGAN	ACC	-	0.57	0.65	-	14.04%
(150)	C	CNN	XR	Many	GAN	-	ACC	0.662	-	0.957	44.56%	-
(151)	C	CNN	CT	CAN	MOD	AF	ACC	-	0.83	0.92	-	10.84%
(152)	D	3D CNN	CT	NOD	AF	-	SNS	0.75	-	0.9	20.00%	-
(153)	D	-	CT	NOD	AF	-	ACC	0.697	-	0.788	13.06%	-
(154)	D	3D CNN	CT	NOD	AF	-	SNS	0.923	-	0.99	7.26%	-
(155)	D	R-CNN	CT	NOD	GAN	-	CPM	0.518	-	0.55	6.18%	-
(132)	S	U-Net	XR	PNE	AF - PX - EL (*)	-	DICE	0.9011	-	0.9023	0.13%	-
(156)	S	U-Net	MR - CT	CAN	GAN	AF	DICE	-	0.63	0.7	-	11.11%
(32)	S	U-Net	CT	CAN	pGAN	GAN	DICE	0.55	0.63	0.8	45.45%	26.98%
(40)	S	U-Net	XR	NOD	MOD	AF	DICE	0.725	0.794	0.809	11.59%	1.89%
(157)	C	CNN	CT	COV	AF - PX		ACC	0.921	-	0.944	2.59%	-
(158)	S	3D U-Net	CT	COV	AF - EL - PX		DICE	0.756	-	0.761	0.66%	-
(159)	C	3D CNN	CT	COPD	GAN		ACC	0.597	-	0.629	5.36%	-
(160)	C	CNN	XR	PNE	GAN		ACC	0.795	-	0.903	13.52%	-
(161)	S	U-Net	CT	Many	MOD		DICE	0.948	-	0.947	-0.07%	-
(162)	C	Custom	CT	NOD	GAN		AUC	0.851	-	0.921	8.19%	-
(163)	C	CNN	CT	COV	GAN		AUC	0.935	-	0.96	2.67%	-
(164)	C	CNN	CT	NOD	GAN		ACC	0.851	-	0.879	3.19%	-
(165)	C	CNN	CT	COV	AF - GAN	AF	ACC	-	0.929	0.9	-	-3.14%
(166)	D	RetinaNet	XR	NOD	MOD - AF - PX	AF	SNS	-	0.494	0.52	-	5.26%
(167)	S	U-Net	MRI	COPD	AF - GAN	AF	DICE	0.955	0.961	0.965	1.05%	0.42%
(168)	C	CNN	XR	PNE	GAN - AF	AF	ACC	0.805	0.815	0.83	3.11%	1.84%
(169)	C	CNN	XR	PNE	GAN	AF	AUC	0.941	0.951	0.962	2.23%	1.10%
(170)	C	CNN	CT	CAN	GAN		ACC	0.342	-	0.577	68.71%	-
(113)	S	3D U-Net	CT	CAN	AF - PX - EL	FM	DICE	0.564	0.581	0.639	13.30%	9.98%
(116)	S	U-Net	CT	CAN	AF		DICE	0.983	-	0.985	0.26%	-
(171)	C	CNN	XR	Many	GAN		ACC	0.93	-	0.95	2.15%	-

Table 4: Comparison of **51** studies on data augmentation techniques for lung imaging. See Table 1 for an explanation of the acronyms.

according to a previously defined and validated simulation model. Then, breast compression is simulated using a finite element model, and a Monte Carlo-based X-Ray transport simulation code was exploited to project the voxelized 3D phantoms into realistic-looking synthetic mammograms. Thanks to the relative simplicity of the breast organ, virtual phantoms are also increasingly used in image acquisition and reconstruction, denoising, and to conduct virtual trials (173). Likewise, their role in

training DNNs is likely to increase. Other authors have proposed a method called stochastic evolution (SE) to mimic the irregular deterioration and healing processes of the diseased tissue according to the direction of local distortion (40).

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	w/o DA - ¿ Proposed	Baseline - ¿ Proposed
(174)	C	CNN	MG	CAN	AF	-	ACC	0.77	-	0.82	6.49%	-
(175)	C	CNN	MG	CAN	GAN	-	ACC	0.782	-	0.87	11.25%	-
(176)	C	CNN	US	CAN	AF	-	AUC	0.94	-	0.96	2.13%	-
(177)	D	OSD	MG	CAN	AF - EL - PX	-	TPR	0.861	-	0.913	6.04%	-
(178)	S	U-Net	MG	CAN	AF	-	DICE	0.922	-	0.951	3.15%	-
(179)	C	CNN	MG	CAN	AF - PX	-	ACC	0.782	-	0.836	6.91%	-
(180)	C	CNN	MG	CAN	AF	-	ACC	0.71	-	0.859	20.99%	-
(181)	C	CNN	MG	CAN	AF - PX	-	ACC	0.721	-	0.938	29.00%	-
(172)	C	CNN	MG	CAN	EL	AF	FPI	-	2.934	2.181	-	-25.66%
(182)	C	CNN	MG	CAN	AF - GAN	AF	ACC	0.699	0.88	0.94	34.48%	6.82%
(183)	C	CNN	CT	CAN	GAN	AF	AUC	-	0.741	0.869	-	17.27%
(184)	D	R-CNN	MG	CAN	GAN	AF	AUC	0.151	0.159	0.172	13.91%	8.18%
(185)	C	CNN	US	CAN	pGAN	iGAN	ACC	-	0.887	0.904	-	1.92%
(39)	D	R-CNN	CT	CAN	MOD	-	SNS	0.802	-	0.833	3.87%	-
(40)	S	U-Net	CT	CAN	MOD	AF	DICE	0.547	0.643	0.666	21.76%	3.58%
(186)	C	CAPSULE	MG	CAN	AF		ACC	0.828	-	0.849	2.54%	-
(187)	C	CNN	US	CAN	MOD - ER		AUC	0.767	-	0.799	4.19%	-
(188)	C	CNN	MG	CAN	GAN - AF		AUC	-	0.835	0.889	-	6.49%
(112)	C	CNN	MG	CAN	FM		ACC	0.62	-	0.721	16.32%	-
(171)	C	CNN	US	CAN	GAN		ACC	0.863	-	0.891	3.20%	-

Table 5: Comparison of **19** studies on data augmentation techniques for breast imaging. See Table 1 for an explanation of the acronyms.

4.2.5. Other Organs

In Table 6 we collected studies conducted on less studied organs, including less frequent diseases such as detection of urinary stones in (189), classification of cyst and tumors of both jaws in (190) or classification of calcaneal fractures in (191). The most common organs *liver* and *prostate* with 18/66 and 16/66 studies, respectively. Despite a few exceptions (192; 193; 194; 195), data augmentation consistently improves performance in all organs.

One study uses train-time data augmentation to randomly increase the size of the training set and test-time data augmentation to estimate the uncertainty of the predictions (196). Both strategies contribute to increase the performance with

respect w/o DA baseline. Model-based techniques were proposed to generate several types of images, such as spine (197) and urinary tract images (189).

In one study (198), a positive effect of augmentation was observed for the *"severe"* cases, whereas for the *"moderate"* cases there were no significant improvements, suggesting that the benefit of data augmentation may depend on the specific subpopulation.

It is worth highlighting the few studies in this SLR that investigate learnable data augmentation strategies (199; 200; 201). For instance, the optimal policy can be learned through reinforcement learning, choosing from a set of transformations that include flipping, rotation, cropping, elastic deformation, zooming, noise, and brightness adjustment (199). The authors observed an increase in performance with respect to a fixed data augmentation policy on kidney stone segmentation, but the results are rather preliminary and more studies are needed to assess the applicability to a wider range of organs and modalities.

Another interesting study investigates the application of three-dimensional transformations on various organs (48). The same model is trained across different views, which forces the network to use the same weights to capture structures at different viewing-angles. The results show that combining and augmenting three different views could further boost performance.

4.2.6. Effect of data augmentation on deep learning performance

In this section, we summarize the distribution of the relative increase in performance associated to the use of different data augmentation techniques. A total of 185 papers were included in this analysis, corresponding to different organs (Heart: 16, Brain: 39, Lung: 43, Breast: 16, Other: 71).

As shown in Figure 4, the median increase in performance associated to data

Paper	Organ	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	w/o DA -¿ Proposed	Baseline -¿ Proposed
(195)	Abdomen	D	YOLO	CT	H	GAN	-	MD	8.66	-	7.95	-8.20%	-
(48)	Abdomen	S	U-Net	MR	H	AF	-	DICE	0.897	-	0.9	0.33%	-
(202)	Abdomen	S	3D U-Net	CT	H	EL	AF - EL - PX	DICE	0.84	0.85	0.86	2.38%	1.18%
(202)	Abdomen	S	3D U-Net	CT	H	EL	AF - EL - PX	DICE	0.84	0.85	0.86	2.38%	1.18%
(194)	Abdomen	S	U-Net	CT	H	AF		ACC	0.954		0.946	-0.84%	-
(194)	Abdomen	S	U-Net	CT	H	AF		ACC	0.954		0.946	-0.84%	-
(203)	Arm	S	U-Net	US	H	EL	AF	PACC	0.98	0.998	0.999	1.94%	0.10%
(204)	Arteries	S	U-Net	IVOCT	OTH	REC	AF	DICE	0.615	0.696	0.756	22.93%	8.62%
(204)	Arteries	S	U-Net	IVOCT	OTH	REC	AF	DICE	0.615	0.696	0.756	22.93%	8.62%
(116)	Bladder	S	U-Net	CT	H	AF		DICE	0.904	-	0.914	1.11%	-
(205)	Bones	C	TF	XR	FRC	ER	AF	ACC		0.786	0.805	-	2.35%
(205)	Bones	C	TF	XR	FRC	ER	AF	ACC		0.786	0.805	-	2.35%
(206)	Colon	D	3D FCN	CT	POL	AF - EL	AF	FROC	-	0.9	0.97	-	7.78%
(207)	Colon	C	3D CNN	CT	POL	GN	GAN	AUC	0.811	0.846	0.881	8.63%	4.14%
(200)	Eye	S	U-Net	OCT	H	MOD - FM - LRN	MOD - LRN	DICE	0.724	0.733	0.752	3.87%	2.59%
(208)	Head	D	YoloX	CT	FRC	AF -FM	AF	AP	0.688	0.692	0.698	1.45%	0.87%
(208)	Head	D	YoloX	CT	FRC	AF -FM	AF	AP	0.688	0.692	0.698	1.45%	0.87%
(209)	Hip	C	Custom	CT	FRC	FM		ACC	0.519	-	0.977	88.19%	-
(209)	Hip	C	Custom	CT	FRC	FM		ACC	0.519	-	0.977	88.19%	-
(210)	Humerus	S	U-Net	US	H	GAN - AF - PX	AF - PX	DICE	0.58	0.605	0.66	13.79%	9.09%
(199)	Kidney	S	U-Net	CT	CAN	AF - ER - PX - LRN	AF	DICE	0.749	0.832	0.84	12.15%	0.96%
(48)	Kidney	S	U-Net	MR	H	AF	-	DICE	0.935	-	0.954	2.03%	-
(211)	Kidney	S	U-Net	CT	H	GAN - AF	AF	DICE	0.92	0.94	0.944	2.61%	0.43%
(212)	Kidney	S	U-Net	US	H	MOD	-	DICE	0.93	-	0.942	1.29%	-
(213)	Knee	S	U-Net	MR	H	GAN		DICE	0.836	-	0.847	1.34%	-
(213)	Knee	S	U-Net	MR	H	GAN		DICE	0.836	-	0.847	1.34%	-
(214)	Liver	C	CNN	CT	CAN	GAN	AF	SNS	-	0.797	0.865	-	8.53%
(215)	Liver	S	U-Net	CT	CAN	AF - FM	-	DICE	0.915	-	0.945	3.28%	-
(41)	Liver	S	3D U-Net	CT	H	MOD	-	ACC	0.861	-	0.883	2.56%	-
(216)	Liver	C	CNN	CT	CAN	GAN	GAN + AF	ACC	0.802	-	0.855	6.61%	-
(48)	Liver	S	U-Net	MR	H	AF	-	DICE	0.961	-	0.962	0.10%	-
(171)	Liver	S	CNN	CT	CAN	GAN		DICE	0.912	-	0.941	3.25%	-
(217)	Liver	C	CNN	CT	L	GAN	AF	ACC	-	0.786	0.857	-	9.03%
(218)	Liver	C	Custom	MR	CAN	AF		AUC	0.799	-	0.826	3.38%	-
(218)	Liver	C	Custom	MR	CAN	AF		AUC	0.799	-	0.826	3.38%	-
(219)	Liver	S	U-Net	CT	H	AF - PX	-	DICE	0.896	-	0.915	2.12%	-
(219)	Liver	S	U-Net	CT	CAN	AF - PX	-	DICE	0.659	-	0.702	6.53%	-
(201)	Liver	S	U-Net	MR	CAN	AF - EL - PX - LRN	AF	DICE		0.603	0.635	-	5.31%
(201)	Liver	S	U-Net	MR	CAN	AF - EL - PX - LRN	AF	DICE		0.603	0.635	-	5.31%
(220)	Liver	C	CNN	CT	CAN	GAN - AF	AF	ACC		0.84	0.853	-	1.55%
(220)	Liver	C	CNN	CT	CAN	GAN - AF	AF	ACC		0.84	0.853	-	1.55%
(113)	Liver	S	3D U-Net	CT - MR	CAN	AF - PX - FM	FM	DICE	0.492	0.298	0.66	34.15%	121.48%
(113)	Liver	S	3D U-Net	CT - MR	CAN	AF - PX - FM	FM	DICE	0.492	0.298	0.66	34.15%	121.48%
(211)	Liver	S	U-Net	CT	H	GAN - AF	AF	DICE	0.944	0.941	0.947	0.32%	0.64%
(221)	Muscle	S	U-Net	CT	H	AF - PX	-	DICE	0.833	-	0.915	9.84%	-
(198)	Muscle	S	U-Net	MR	NMD	GAN - EL	EL	DICE	0.84	0.87	0.88	4.76%	1.15%
(48)	Pancreas	S	U-Net	MR	H	AF	-	DICE	0.86	-	0.864	0.47%	-
(13)	Pelvis	S	U-Net	MR	Many	EL	-	PACC	0.724	-	0.841	16.16%	-
(41)	Prostate	S	3D U-Net	MR	H	MOD	-	ACC	0.742	-	0.868	16.98%	-
(222)	Prostate	S	U-Net	MR	CAN	AF - GAN	-	DICE	0.678	-	0.738	8.85%	-
(52)	Prostate	S	U-Net	MR	H	AF - EL - PX - LRN	MX	DICE	-	0.706	0.733	-	3.80%
(52)	Prostate	S	U-Net	MR	H	AF - EL - PX - LRN	MX	DICE	-	0.706	0.733	-	3.80%
(223)	Prostate	S	U-Net	MR - PET - CT	CAN	AF - FM	-	DICE	0.776	-	0.819	5.54%	-
(224)	Prostate	C	CNN	MR	CAN	AF		ACC	0.802	-	0.85	6.00%	-
(224)	Prostate	C	CNN	MR	CAN	AF		ACC	0.802	-	0.85	6.00%	-
(225)	Prostate	S	U-Net	MR	H	MOD	-	DICE	0.813	-	0.863	6.15%	-
(40)	Prostate	S	U-Net	XR	H	MOD	AF	DICE	0.785	0.81	0.848	8.03%	4.69%
(226)	Prostate	S	AMS	MR	CAN	GAN	-	DICE	0.858	-	0.882	2.80%	-
(116)	Prostate	S	U-Net	CT	CAN	AF		DICE	0.697	-	0.731	4.93%	-
(227)	Prostate	S	U-Net	US	CAN	MOD		DICE	0.906	-	0.908	0.17%	-
(227)	Prostate	S	U-Net	US	CAN	MOD		DICE	0.906	-	0.908	0.17%	-
(228)	Prostate	C	Custom	MR	CAN	GAN	AF	ACC	-	0.815	0.892	-	9.45%
(56)	Prostate	S	3D U-Net	MR	CAN	AF - EL - PX		DICE	0.896	0.908	0.913	1.90%	0.55%
(56)	Prostate	S	3D U-Net	MR	CAN	AF - EL - PX		DICE	0.896	0.908	0.913	1.90%	0.55%
(229)	Rectal	S	MRSN	MR	H	AF - PX	-	DICE	0.938	-	0.943	0.53%	-
(229)	Rectal	S	MRSN	MR	CAN	AF - PX	-	DICE	0.732	-	0.742	1.37%	-
(116)	Rectum	S	U-Net	CT	CAN	AF		DICE	0.718	-	0.753	4.97%	-
(197)	Spine	S	U-Net	CT	H	MOD	-	PACC	0.743	-	0.948	27.59%	-
(193)	Spine	S	2D U-Net	MR	OTH	AF - EL		DICE	0.921	-	0.913	-0.87%	-
(193)	Spine	S	2D U-Net	MR	OTH	AF - EL		DICE	0.921	-	0.913	-0.87%	-
(48)	Spleen	S	U-Net	MR	H	AF	-	DICE	0.944	-	0.944	0.00%	-
(211)	Spleen	S	U-Net	CT	H	GAN - AF	AF	DICE	0.884	0.89	0.919	3.96%	3.26%
(230)	Spleen	S	U-Net	CT	OTH	AF - GAN	AF	DICE	-	0.488	0.53	-	8.56%
(230)	Spleen	S	U-Net	CT	OTH	AF - GAN	AF	DICE	-	0.488	0.53	-	8.56%
(196)	Teeth	S	U-Net	XR	H	AF	AF (*)	ACC	0.947	-	0.948	0.02%	-
(190)	Teeth	C	CNN	XR	CAN - CY	AF - PX	-	AUC	0.86	-	0.94	9.30%	-
(192)	Teeth	C	CNN	XR	H	AF - PX	-	AUC	0.89	-	0.88	-1.12%	-
(231)	Teeth	C	CNN	XR	OS	AF	-	ACC	0.925	-	0.98	5.95%	-
(232)	Teeth	C	CNN	XR	CAN	AF	-	ACC	0.699	-	0.904	29.33%	-
(233)	Thyroid	C	CNN	US	CAN	GAN	AF	ACC	0.701	0.756	0.915	30.53%	21.03%
(234)	Thyroid	C	CNN	US	CAN	GAN + AF + PX	AF + PX	ACC	0.845	0.887	0.953	12.75%	7.46%
(234)	Thyroid	C	CNN	US	CAN	GAN + AF + PX	AF + PX	ACC	0.845	0.887	0.953	12.75%	7.46%
(189)	Urinary Tract	D	U-Net	XR	URS	MOD	-	F1	0.561	-	0.603	7.49%	-
(235)	Whole body	C	Custom	SPECT	Many	AF - GAN	AF	ACC	0.691	0.775	0.729	5.44%	-5.95%
(235)	Whole body	C	Custom	SPECT	Many	AF - GAN	AF	ACC	0.691	0.775	0.729	5.44%	-5.95%

Table 6: Comparison of **66** studies on data augmentation techniques for other organs, not covered by previous tables. See Table 1 for an explanation of the acronyms.

augmentation is around 10% across all organs. Studies related to the heart, lung and breast report a higher benefit associated to data augmentation than brain and other organs. This difference could be explained by various factors: better data augmentation techniques are available for organs that are more heavily researched; datasets are typically smaller; some organs/modalities are susceptible to higher inter-subject or inter-scanner variability, thus benefitting more from data augmentation.

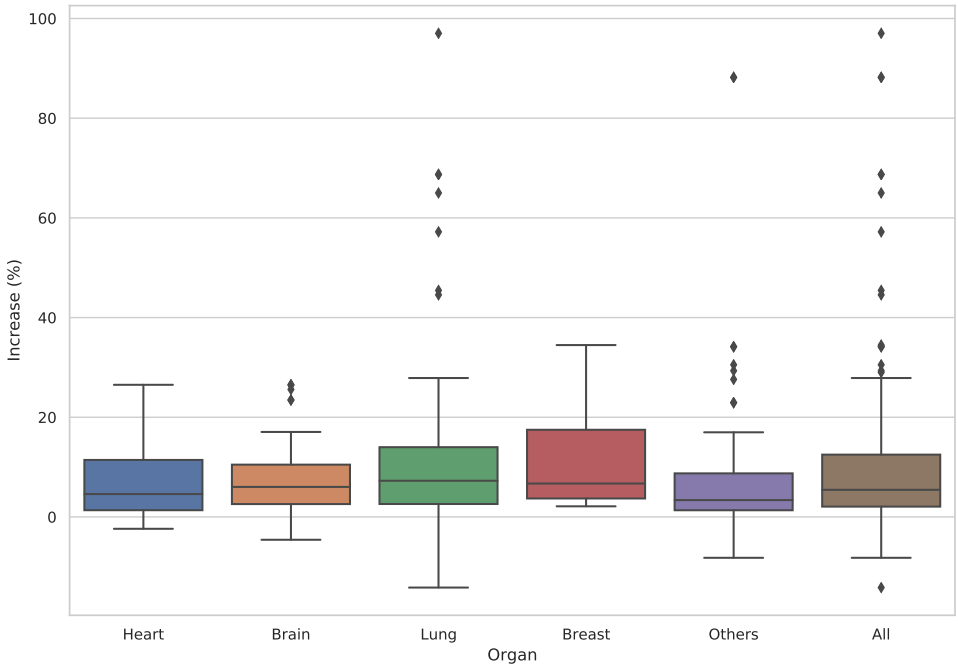


Figure 4: Box plot of the relative performance increase associated to data augmentation (DA vs. no-DA) for each organ.

We further evaluated the relative performance benefit that can be ripped by exploiting more complex data augmentation techniques. For this analysis, we split data augmentation techniques in two categories: affine and non-affine (other). We

choose affine transformations as the baseline given their simplicity, ease of use and widespread adoption in the medical domain. Methods that combine multiple data augmentation techniques were tagged with the most complex transformation (in reverse order: GAN, model-based, reconstruction-based, feature mixing, erasing, pixel-level, affine).

Figure 5 reports the distribution of the relative performance increase for three groups of papers: affine DA vs. no-DA (41 studies), non-affine DA vs. no-DA (144 studies) and non-affine DA vs. affine DA (56 studies). Results show that simple affine transformations offer a substantial performance advantage with respect to not using DA at all. More complex techniques yield higher performance, at the expense of greater complexity. While these distributions must be interpreted with great caution, since we are pooling very heterogeneous studies without adjustments, they overall suggest that "simple" data augmentation techniques should always be tried before moving to more complex solutions.

Finally, in Figure 6 we report the average relative increase in performance associated with the use of DA for classification (78), segmentation (93), and detection (14) tasks. Overall, classification tasks seem to benefit more from data augmentation than segmentation tasks. However, these results should be interpreted with caution as tasks are not evenly distributed across organs and modalities, and segmentation tasks are often associated with volumetric (3D) modalities, which are more complex to augment than 2D images.

We concluded this analysis by plotting the relative increase in performance with respect to the size of the training set, on a logarithmic scale (reported in Figure 7). Only studies comparing the results with and without DA (182 studies) were included in this graph. More than half of the datasets are small dataset with less than 1000 samples. However, correlation between relative performance increase and training

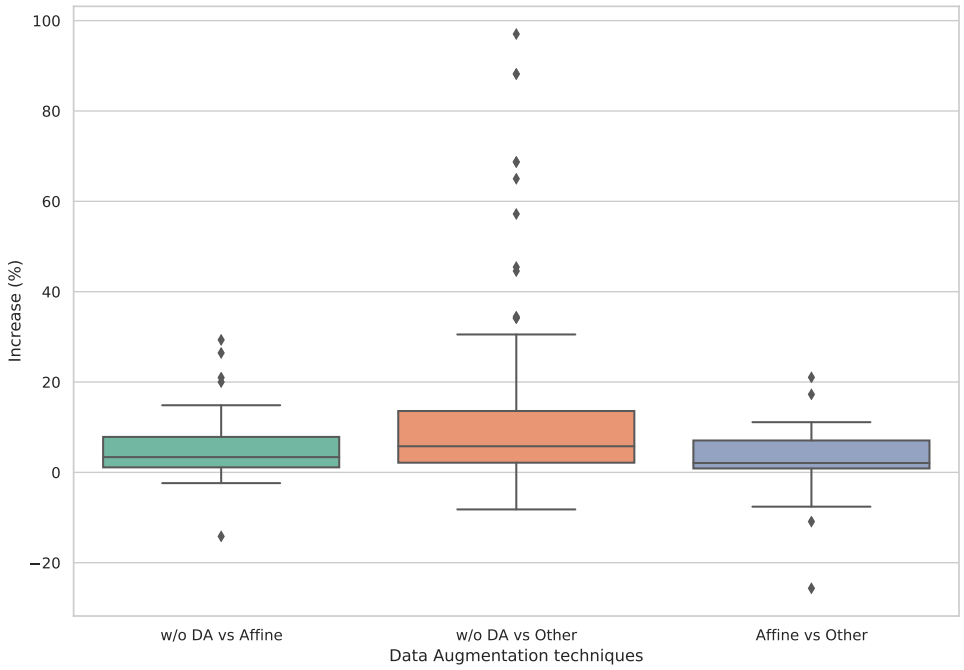


Figure 5: Box plot of the relative performance increase associated to different types of data augmentation (affine and non-affine).

set size is poor ($-5.11e-05$), suggesting that DA can be beneficial regardless of the size of the training set. We attribute this to two phenomena: i) data augmentation promotes learned invariances that, in turn, enable more effective generalization, and ii) data augmentation is often used to balance the distribution of the training set, typically by generating additional lesion cases. In the case of severe data imbalance, the number of examples from minority classes is likely insufficient even in medium- to large-scale datasets.

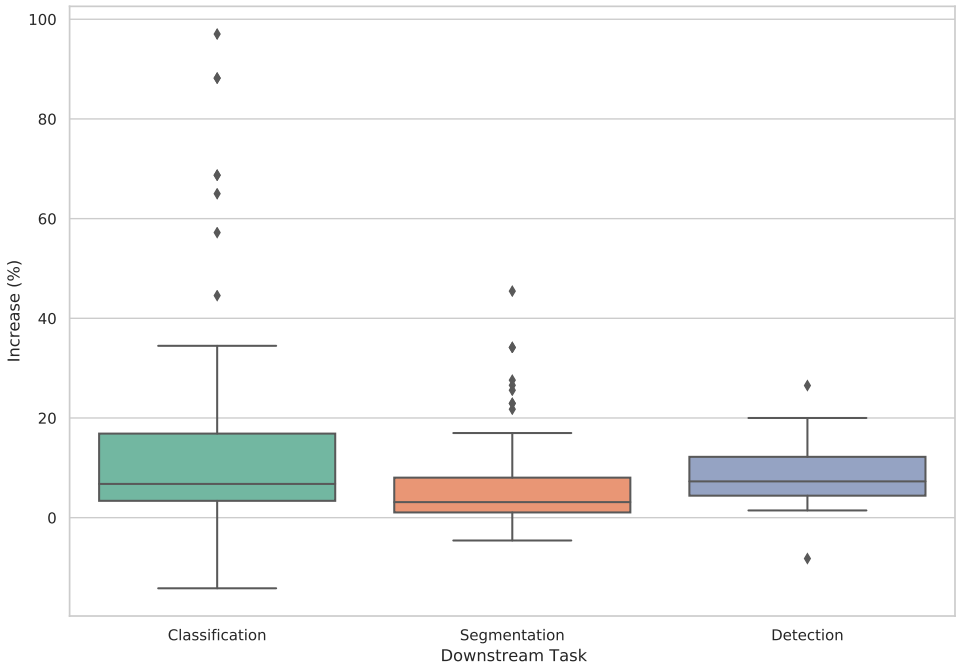


Figure 6: Box plot of the relative performance increase associated to data augmentation (DA vs. no-DA) for each task.

4.3. RQ4: Which data augmentation methods have not been explored in the medical domain?

To answer RQ4, we compared the results of our SLR with previous literature review tackling the more general computer vision field (6). Most of the techniques adopted in general computer vision have found widespread adoption in the medical domain, with a few notable exceptions: random erasing, feature mixing, and learnable data augmentation. These differences could be attributed to publication bias, since authors are more likely to publish techniques that yield positive rather than negative results. For instance, random erasing was introduced to make DNNs

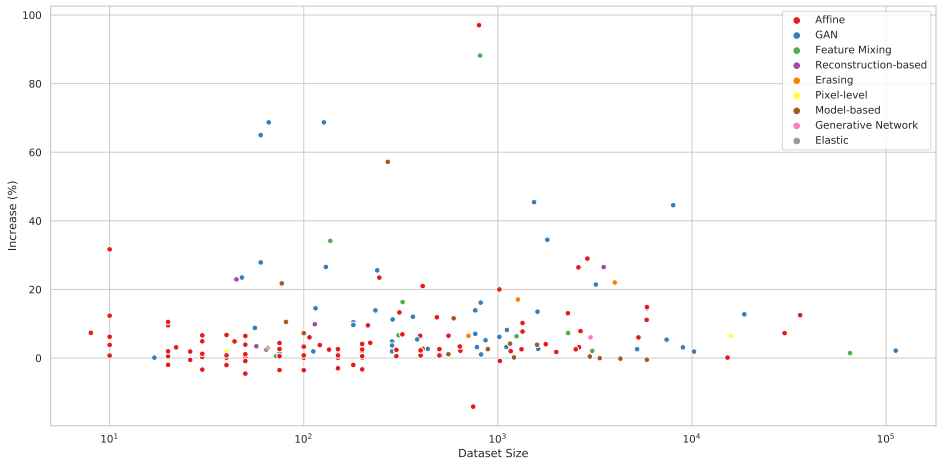


Figure 7: Scatter plot of percentage increase associated to data augmentation with respect to the training set size (in logarithmic scale). (Best viewed in color)

robust to occlusions, which however do not occur in medical imaging. Hence, it is conceivable that their use is not particularly effective in this domain.

On the other hand, a more widespread adoption of learnable DA techniques could in principle discover policies that are uniquely tailored to the medical domain. However, many techniques that have been proposed, such as those based on reinforcement learning, are computationally intensive, as they require training multiple networks to determine which transformations are more challenging at a given stage of training and should be selected with higher probabilities. This fact, coupled with the fact that medical images are typically bigger than most RGB datasets, may have made these techniques less appealing for the medical imaging community.

5. Comparison with previous research

We have compared our findings with those of Nalepa et al., who reviewed the data augmentation techniques used in 20 submissions to the Multimodal Brain Tumor Segmentation Challenge (BraTS 2018) challenge (9). Participants in the BraTS challenge received multimodal MRI data of brain tumor patients, all co-registered to a common anatomical reference. The task is to build a supervised segmentation model, which is then tested on unseen data released during the testing phase. Although most proposed augmentations are single-modal, they can be easily applied to each co-registered series, thus yielding multi-modal samples.

Briefly summarized, Nalepa and colleagues found that 75% of the submissions included affine transformations (mostly horizontal flip); all but one paper employed 2D, rather than 3D transformations. Pixel-level transformations were less common, but appeared in the three top-performing solutions (236; 237; 238). Interestingly, one of the submissions experimentally observed that simpler networks combined with carefully designed data augmentation outperformed more complex solutions (236). Only one submission used generative models (a combination of synthetic images and cycle-GANs). In the small number of submissions that reported performance with and without data augmentation, the relative DICE score increased between -5% to +7%, and the relative Hausdorff distance decreased between -86% to +26%. For comparison, in the 49 studies related to brain image analysis (analyzed in detail in Section 4.2.1), one out of two papers described generative models or model-based synthetic image generation techniques, sometimes in combination with affine and pixel-level transformations. Studies focusing on segmentation techniques reported a relative increase in DICE score between 2% and 26%.

The comparison between Nalepa's findings and ours highlights the increasing role

that GANs are playing in the data augmentation landscape. Indeed, the present SLR covers articles published from 2018 (when the BraTS challenge analyzed in (9) took place) to 2022. On the other hand, articles that evaluate the impact of data augmentation are likely to analyze more complex techniques (and thus more interesting, from an editorial perspective). It is possible, however, that practical adoption of generative models is lagging behind, given the significant gap in implementation time and computational resources with respect to transformations that are available out of the box from most deep learning frameworks.

Another study that provides key insights to integrate and complement of our findings is that of Zhang et al. (239), who investigated the performance of BigAug on different modalities and organs. BigAug is a sequence of n stacked transformations, selected from a pool of options that the authors divide in three categories: *image quality*, *appearance* and *spatial configuration*. Image quality and appearance include pixel-level transformations: noise insertion, blurriness and sharpness simulate the effect of varying image quality, whereas brightness, contrast and intensity perturbation emulate the properties of different scanners and imaging protocols. Spatial configuration transformations, on the other hand, include both affine and elastic transformations that simulate inter- and intra-patient differences due to, e.g., positioning in the scanner.

The main conclusion of this study, as reported also in Tables 3 and 6, is that BigAug outperforms any individual transformation, which is line with the findings of this review. Additionally, the impact of each individual transformation is thoroughly investigated on datasets from two modalities (MR and US). The results suggest an important correlation between the modality and the optimal transformations. In the case of MRI, transformations such as sharpening, contrast enhancement, brightness, and intensity perturbation appear to be the most beneficial, whereas in the case of

US, the most effective transformations are scaling, brightness, blurring, and contrast. Thus, in the case of MRI, generalizing to different intensity distributions appears to be more important than accounting for shape variations, which on the other hand are most prominent in US.

In addition, Zhang et al. prove that the benefits of data augmentation persist, and are even amplified, when the trained model is transferred to a different target dataset, such as one acquired in a different center (239). As an example, in the case of prostate segmentation in MR images, the model trained without data augmentation has a good performance on the source dataset (DICE 0.896), which drops significantly on the three target datasets (DICE 0.604, 0.580 and 0.768, respectively). On the other hand, the model trained with BigAug, despite having similar performance on the source dataset (DICE 0.913), generalizes significantly better to different target datasets (DICE 0.802, 0.850 and 0.865, respectively). When averaged across all experiments, BigAug slightly outperformed the baseline on the source dataset (mean DICE 0.916 vs. 0.891), but showed much stronger generalization (mean DICE 0.80 vs. 0.498). In conclusion, even when data augmentation shows minor improvements on the source dataset (which is the setting analyzed in this SLR), further benefits could be obtained in terms of cross-domain generalization. These properties should and could be further investigated in modalities and organs not covered by Zhang et al. (239).

6. Discussion, limitations, and key findings

In this SLR, we have analyzed more than 300 articles that proposed and / or evaluated data augmentation techniques for modalities such as CT, MR, and XR imaging. Our key findings can be summarized as follows:

- Although affine and pixel-level transformations remain the most adopted transformations, papers published between 2018 and 2022 have substantially shifted their attention towards the generation of synthetic data for training, either based on generative models or, for selected organs for which computational models are available, through model-based computer-simulated imagery.

- Traditional data augmentation techniques, based on transformation of existing data, and generative techniques serve two different purposes. Thus, practitioners exploiting generative models should consider including also conventional transformations in their data pipelines. In particular, traditional data augmentation plays a fundamental role in making DNNs robust to variations introduced by the acquisition system, as well as to inter- and intra-patient variability. Among traditional data augmentation, pixel-level transformations appear to be under-utilized with respect to their potential.

- Generative techniques, on the other hand, are more flexible and can be used to balance training data sets by generating cases with specific characteristics (149; 30), inserting artificial lesions (155; 72; 30), transferring data across domains (156) and alleviating the need for expensive data labeling, as labels can be co-generated with the corresponding image (76).

- Learnable data augmentation and feature mixing, despite showing promising results in other domains, have been scarcely investigated in medical imaging. Recent advances in generative models, that have led to photographic images of unprecedented visual fidelity (27), have also untapped potential in the medical domain.

- Data augmentation can be particularly effective in more complex training set-

ting, such as one-shot/few-shot learning (70; 76; 32; 52), cross-domain transfer learning (239), and federated learning (119). Authors proposing new techniques for data augmentation should consider including experiments on cross-domain generalization, as done in (239).

- The uptake of more complex data augmentation strategies could be facilitated by the availability of open-source tools, including code and pre-trained models. Many authors tend to adopt transformations available in standard deep learning libraries, which however do not fully support the needs of medical imagery. For instance, adapting even standard affine transformations for on-the-fly data augmentation of volumetric 3D images is not trivial (239). Recently, dedicated medical imaging platforms, such as MONAI (12), provide tailored data augmentation pipelines, including several 3D transformations.

- In the future, available resources could and should extend to generative models and synthetic imagery, to facilitate their adoption. As an example, open-source tools are already available to support *in silico* virtual clinical trials for regulatory purposes (240), and may in the future extend to other uses as well. Synthetic datasets will likely play a prominent role in future medical imaging deep learning research. For instance, NVIDIA and King's College London have recently announced the release of a large-scale dataset containing more than 100,000 synthetic MRI images, generated by a combination of variational auto-encoder and transformers (241).

Despite the large number of papers included in the analysis, our study has two potential limitations:

- Like all SLRs, our results may reflect publication bias. Since we primarily

included papers that specifically reported performance with and without data augmentation, we may risk overestimating the practical relevance of generative and model-based techniques that, while undoubtedly beneficial in terms of performance, are much more complex and computationally expensive to implement than more traditional data transformations. Papers that mentioned the use of data augmentation without investigating its impact on downstream performance predominantly made use of standard affine and pixel-level transformations. As mentioned before, the gap between research and practice could be alleviated by the development of open-source AI platforms and tools.

- We attempted to quantitatively combine results from heterogeneous studies, which may not be directly comparable due to different datasets used or experimental conditions. In this respect, it is encouraging to see how our results align with previous studies which compared a wide range of data augmentation strategies on the same dataset, as detailed in Section 5. In the future, more controlled experiments should compare different data augmentation strategies on the same datasets, especially considering how the results may vary across modalities.

7. Conclusions

In this SLR, we have analyzed more than 300 articles that proposed and / or evaluated data augmentation techniques for different imaging modalities (e.g., CT, MR, XR, MG and functional imaging), organs (e.g., brain, heart, lung, breast, liver, prostate and many others), and tasks (classification, segmentation and detection). Overall, we found consistent benefits, across all organs, modalities, and tasks, with the use of data augmentation, from the simplest affine transformations to the most

complex generative models. Through extensive qualitative and quantitative analysis, as well as comparison with previous surveys and experimental studies, we draw a comprehensive and up-to-date picture of the state of the art of this lively field of research, as well as highlight current research gaps and directions of improvement.

References

[1] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[2] Berkman Sahiner, Aria Pezeshk, Lubomir M Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny H Cha, Ronald M Summers, and Maryellen L Giger. Deep learning in medical imaging and radiation therapy. *Medical physics*, 46(1):e1–e36, 2019.

[3] Marc D Kohli, Ronald M Summers, and J Raymond Geis. Medical image data and datasets in the era of machine learning: Whitepaper from the 2016 C-MIMI meeting dataset session. *Journal of digital imaging*, 30(4):392–399, 2017.

[4] Suorong Yang, Weikang Xiao, Mengcheng Zhang, Suhan Guo, Jian Zhao, and Furao Shen. Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*, 2022.

[5] Shuxiao Chen, Edgar Dobriban, and Jane H. Lee. Invariance reduces variance: Understanding data augmentation in deep learning and beyond. *CoRR*, abs/1907.10905, 2019.

[6] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, Jul 2019.

[7] Lia Morra, Silvia Delsanto, and Loredana Correale. *Artificial intelligence in medical imaging: From theory to clinical practice*. CRC Press, 2019.

[8] Jakub Nalepa, Michal Marcinkiewicz, and Michal Kawulok. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13:83, 2019.

[9] Jakub Nalepa, Michal Marcinkiewicz, and Michal Kawulok. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13:83, 2019.

[10] Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020.

[11] Fernando Pérez-García, Rachel Sparks, and Sebastien Ourselin. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236, 2021.

[12] MONAI Consortium. Monai: Medical open network for ai, June 2022. If you use this software, please cite it using these metadata.

[13] Zuoyu Yan, Liangcai Gao, Zhi Tang, and Xinpeng Zhang. A non-local based segmentation method for pelvic mr images. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1265–1267, Nov 2019.

[14] Pablo Ribalta Lorenzo, Jakub Nalepa, Barbara Bobek-Billewicz, Pawel Wawrzyniak, Grzegorz Mrukwa, Michal Kawulok, Pawel Ulrych, and Michael P. Hayball. Segmenting brain tumors from FLAIR MRI using fully convolutional neural networks. 176:135–148. Place: Ireland.

[15] Richard Osuala, Kaisar Kushibar, L. Garrucho, Akis Linardos, Zuzanna Szafranowska, Stefan Klein, Ben Glocker, O. Díaz, and Karim Lekadir. A review of generative adversarial networks in cancer imaging: New applications, new solutions. *ArXiv*, abs/2107.09543, 2021.

[16] Yu Shen, Junbang Liang, and Ming C Lin. Gan-based garment generation using sewing pattern images. In *European Conference on Computer Vision*, pages 225–247. Springer, 2020.

[17] L Abady, M Barni, A Garzelli, and B Tondi. Gan generation of synthetic multispectral satellite images. In *Image and Signal Processing for Remote Sensing XXVI*, volume 11533, pages 122–133. SPIE, 2020.

[18] Nir Diamant, Dean Zadok, Chaim Baskin, Eli Schwartz, and Alex M Bronstein. Beholder-gan: Generation and beautification of facial images with conditioning on their beauty level. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 739–743. IEEE, 2019.

[19] Sefik Emre Eskimez, Dimitrios Dimitriadis, Robert Gmyr, and Kenichi Kumanati. Gan-based data generation for speech emotion recognition. In *INTERSPEECH*, pages 3446–3450, 2020.

[20] Santiago Pascual, Antonio Bonafonte, and Joan Serra. Segan: Speech enhancement generative adversarial network. *arXiv preprint arXiv:1703.09452*, 2017.

[21] Md Haidar, Mehdi Rezagholizadeh, et al. Textkd-gan: Text generation using knowledge distillation and generative adversarial networks. In *Canadian conference on artificial intelligence*, pages 107–118. Springer, 2019.

[22] Liqun Chen, Shuyang Dai, Chenyang Tao, Haichao Zhang, Zhe Gan, Dinghan Shen, Yizhe Zhang, Guoyin Wang, Ruiyi Zhang, and Lawrence Carin. Adversarial text generation via feature-mover's distance. *Advances in Neural Information Processing Systems*, 31, 2018.

[23] Harshvardhan GM, Mahendra Kumar Gourisaria, Manjusha Pandey, and Siddharth Swarup Rautaray. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, 38:100285, 2020.

[24] Xiao Wang and Han Liu. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and wasserstein gan. *Journal of Process Control*, 85:91–99, 2020.

[25] Xinyu Chen, Jiajie Xu, Rui Zhou, Wei Chen, Junhua Fang, and Chengfei Liu. Trajvae: A variational autoencoder model for trajectory generation. *Neurocomputing*, 428:332–339, 2021.

[26] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[27] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.

[28] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47–1, 2022.

[29] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.

[30] Hoo-Chang Shin, Neil A. Tenenholtz, Jameson K. Rogers, Christopher G. Schwarz, Matthew L. Senjem, Jeffrey L. Gunter, Katherine P. Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In Ali Gooya, Orcun Goksel, Ipek Oguz, and Ninon Burgos, editors, *Simulation and Synthesis in Medical Imaging*, pages 1–11, Cham, 2018. Springer International Publishing.

[31] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.

[32] Jue Jiang, Yu-Chi Hu, Neelam Tyagi, Pengpeng Zhang, Andreas Rimner, Gig S. Mageras, Joseph O. Deasy, and Harini Veeraraghavan. Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 777–785, Cham, 2018. Springer International Publishing.

[33] Anmol Sharma and Ghassan Hamarneh. Missing mri pulse sequence synthesis using multi-modal generative adversarial network. *IEEE transactions on medical imaging*, 39(4):1170–1183, 2019.

[34] Paolo Soda. Evaluating gans in medical imaging. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings*, volume 13003, page 112. Springer Nature, 2021.

[35] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation, 2018.

[36] Jonathan Ho, Ajay Jain, and Peter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

[37] Alfred Laugros, Alice Caplier, and Matthieu Ospici. Addressing neural network robustness with mixup and targeted labeling adversarial training. In *European Conference on Computer Vision*, pages 178–195. Springer, 2020.

[38] Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. Multiple sclerosis lesion synthesis in mri using an encoder-decoder u-net. *IEEE Access*, 7:25171–25184, 2019.

[39] Kenny H. Cha, Nicholas A. Petrick, Aria X. Pezeshk, Christian G. Graff, Diksha Sharma, Andreu Badal, and Berkman Sahiner. Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning. *Journal of Medical Imaging*, 7(1):1 – 9, 2019.

[40] Hong Liu, Haichao Cao, Enmin Song, Guangzhi Ma, Xiangyang Xu, Renchao Jin, Tengying Liu, Lei Liu, Daiyang Liu, and Chih-Cheng Hung. A new data augmentation method based on local image warping for medical image segmentation. 48(4):1685–1696. Place: United States.

[41] Zhixian Tang, Kun Chen, Mingyuan Pan, Manning Wang, and Zhijian Song. An augmentation strategy for medical image processing based on statistical shape model and 3d thin plate spline for deep learning. *IEEE Access*, 7:133111–133121, 2019.

[42] Ilkay Oksuz, Bram Ruijsink, Esther Puyol-Antón, James R. Clough, Gastao Cruz, Aurelien Bustin, Claudia Prieto, Rene Botnar, Daniel Rueckert, Julia A. Schnabel, and Andrew P. King. Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning. ISSN: 1361-8423 1361-8415 Journal Abbreviation: Med Image Anal Pages: 136-147 Publication Title: Medical image analysis.

[43] Bastian Bier, Florian Goldmann, Jan-Nico Zaech, Javad Fotouhi, Rachel Hegeman, Robert Grupp, Mehran Armand, Greg Osgood, Nassir Navab, Andreas Maier, and Mathias Unberath. Learning to detect anatomical landmarks of the pelvis in x-rays from arbitrary views. 14(9):1463–1473.

[44] Jihye Yun, Jinkon Park, Donghoon Yu, Jaeyoun Yi, Minho Lee, Hee Jun Park, June-Goo Lee, Joon Beom Seo, and Namkug Kim. Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Medical image analysis*, 51:13–20, 2019.

[45] Zongwei Zhou, Vatsal Sodha, Md Mahfuzur Rahman Siddiquee, Ruibin Feng, Nima Tajbakhsh, Michael B Gotway, and Jianming Liang. Models genesis: Generic autodidactic models for 3d medical image analysis. In *International conference on medical image computing and computer-assisted intervention*, pages 384–393. Springer, 2019.

[46] Yuya Onishi, Atsushi Teramoto, Masakazu Tsujimoto, Tetsuya Tsukamoto, Kuniaki Saito, Hiroshi Toyama, Kazuyoshi Imaizumi, and Hiroshi Fujita. Investigation of pulmonary nodule classification using multi-scale residual network enhanced with 3dgan-synthesized volumes. 13(2):160–169. Place: Japan.

[47] Yuya Onishi, Atsushi Teramoto, Masakazu Tsujimoto, Tetsuya Tsukamoto, Kuniaki Saito, Hiroshi Toyama, Kazuyoshi Imaizumi, and Hiroshi Fujita. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. 15(1):173–178. Place: Germany.

[48] Yuhua Chen, Dan Ruan, Jiayu Xiao, Lixia Wang, Bin Sun, Rola Saouaf, Wensha Yang, Debiao Li, and Zhaoyang Fan. Fully automated multiorgan segmentation in abdominal magnetic resonance imaging with deep neural networks. 47(10):4971–4982.

[49] Joseph Lemley and Peter Corcoran. ¡italic¿deep learning for consumer devices and services 4—¡/italic¿a review of learnable data augmentation strategies for improved training of deep neural networks. *IEEE Consumer Electronics Magazine*, 9(3):55–63, 2020.

[50] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data, 2019.

[51] Kosaku Fujita, Masayuki Kobayashi, and Tomoharu Nagao. Data augmentation using evolutionary image processing. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–6, 2018.

[52] Chen Chen, Chen Qin, Cheng Ouyang, Zeju Li, Shuo Wang, Huaqi Qiu, Liang Chen, Giacomo Tarroni, Wenjia Bai, and Daniel Rueckert. Enhancing MR image segmentation with realistic adversarial data augmentation. *Medical Image Analysis*, 82:102597, November 2022.

[53] Søren Hauberg, Oren Freifeld, Anders Boesen Lindbo Larsen, John W.

Fisher III au2, and Lars Kai Hansen. Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation, 2016.

[54] Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338:34–45, 2019.

[55] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Robust chest CT image segmentation of COVID-19 lung infection based on limited data. *Informatics in Medicine Unlocked*, 25:100681, January 2021.

[56] Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J Wood, Holger Roth, Andriy Myronenko, Daguang Xu, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE transactions on medical imaging*, 39(7):2531–2540, 2020.

[57] Javier Barbero-Gómez, Pedro-Antonio Gutiérrez, Víctor-Manuel Vargas, Juan-Antonio Vallejo-Casas, and César Hervás-Martínez. An ordinal cnn approach for the assessment of neurological damage in parkinson's disease patients. *Expert Systems with Applications*, 182:115271, 2021.

[58] Ran Liu, Yanzhen Zhang, Yangting Zheng, Yaqiong Liu, Yang Zhao, and Lin Yi. Automated detection of vulnerable plaque for intravascular optical coherence tomography images. 10(4):590–603. Place: United States.

[59] Mizuho Nishio, Shunjiro Noguchi, Hidetoshi Matsuo, and Takamichi Murakami. Automatic classification between COVID-19 pneumonia, non-COVID-

19 pneumonia, and the healthy on chest x-ray image: combination of data augmentation methods. 10(1):17532.

[60] Wei Zhao, Jiancheng Yang, Bingbing Ni, Dexi Bi, Yingli Sun, Mengdi Xu, Xiaoxia Zhu, Cheng Li, Liang Jin, Pan Gao, Peijun Wang, Yanqing Hua, and Ming Li. Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3d deep learning. 8(7):3532–3543.

[61] Toru Miyoshi, Akinori Higaki, Hideo Kawakami, and Osamu Yamaguchi. Automated interpretation of the coronary angioscopy with deep convolutional neural networks. 7(1).

[62] Ioannis D. Apostolopoulos, Nikolaos D. Papathanasiou, Trifon Spyridonidis, and Dimitris J. Apostolopoulos. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. 23(2):125–132. Place: Greece.

[63] Xiaofei Zhang, Yi Zhang, Erik Y. Han, Nathan Jacobs, Qiong Han, Xiaoqin Wang, and Jinze Liu. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. 17(3):237–242. Place: United States.

[64] Falk Schwendicke, Karim Elhennawy, Sebastian Paris, Philipp Friebertshäuser, and Joachim Krois. Deep learning for caries lesion detection in near-infrared light transillumination images: A pilot study. 92:103260. Place: England.

[65] Agisilaos Chartsias, Thomas Joyce, Rohan Dharmakumar, and Sotirios A. Tsaftaris. Adversarial image synthesis for unpaired multi-modal cardiac data. In Sotirios A. Tsaftaris, Ali Gooya, Alejandro F. Frangi, and Jerry L. Prince,

editors, *Simulation and Synthesis in Medical Imaging*, pages 3–13, Cham, 2017. Springer International Publishing.

[66] Lukas Fetty, Mikael Bylund, Peter Kuess, Gerd Heilemann, Tufve Nyholm, Dietmar Georg, and Tommy Löfstedt. Latent space manipulation for high-resolution medical image synthesis via the stylegan. *Zeitschrift für Medizinische Physik*, 30(4):305–314, 2020.

[67] Ying Zhuge, Holly Ning, Peter Mathen, Jason Y. Cheng, Andra V. Krauze, Kevin Camphausen, and Robert W. Miller. Automated glioma grading on conventional MRI images using deep convolutional neural networks. 47(7):3044–3053. Place: United States.

[68] Yi Sun, Peisen Yuan, and Yuming Sun. Mm-gan: 3d mri data augmentation for medical image segmentation via generative adversarial networks. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 227–234, Aug 2020.

[69] Xiaolong Liu, Yusheng Hong, Qifang Yin, and Shuo Zhang. DnT: Learning Unsupervised Denoising Transformer from Single Noisy Image. In *Proceedings of the 4th International Conference on Image Processing and Machine Vision*, IPMV '22, pages 50–56, New York, NY, USA, March 2022. Association for Computing Machinery.

[70] Devavrat Tomar, Behzad Bozorgtabar, Manana Lortkipanidze Guillaume Vray, Mohammad Saeed Rad, and Jean-Philippe Thiran. Self-Supervised Generative Style Transfer for One-Shot Medical Image Segmentation. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1737–1747, January 2022. ISSN: 2642-9381.

[71] Sunho Kim, Byungjai Kim, and HyunWook Park. Synthesis of brain tumor multicontrast MR images for improved data augmentation. 48(5):2185–2198. Place: United States.

[72] Saba Momeni, Amir Fazlollahi, Paul Yates, Christopher Rowe, Yongsheng Gao, Alan Wee-Chung Liew, and Olivier Salvado. Synthetic microbleeds generation for classifier training without ground truth. *Computer Methods and Programs in Biomedicine*, 207:106127, August 2021.

[73] N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. A. Viergever, M. J. N. L. Benders, and I. Išgum. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. 64:77–89. Place: Netherlands.

[74] Berardino Barile, Aldo Marzullo, Claudio Stamile, Françoise Durand-Dubief, and Dominique Sappey-Marinier. Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. *Computer Methods and Programs in Biomedicine*, 206:106113, July 2021.

[75] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, jun 2002.

[76] Amy Zhao, Guha Balakrishnan, Frédo Durand, John V. Guttag, and Adrian V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8535–8545, June 2019.

[77] Shikar Rajcomar, Anban W. Pillay, and Edgar Jembere. Paired augmentation

for improved image classification using neural network models. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6, Dec 2020.

[78] Abdulaziz Namozov and Young Im Cho. An improvement for medical image analysis using data enhancement techniques in deep learning. In *2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT)*, pages 1–3, Sep. 2018.

[79] R. Meena Prakash and R. Shantha Selva Kumari. Classification of mr brain images for detection of tumor with transfer learning from pre-trained cnn models. In *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pages 508–511, March 2019.

[80] Ahammed Muneer K. V, V. R. Rajendran, and Paul Joseph K. Glioma tumor grade identification using artificial intelligent techniques. 43(5):113. Place: United States.

[81] Shengye Hu, Wen Yu, Zhuo Chen, and Shuqiang Wang. Medical image reconstruction using generative adversarial network for alzheimer disease assessment with class-imbalance problem. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 1323–1327, Dec 2020.

[82] Muhammad Sajjad, Salman Khan, Khan Muhammad, Wanqing Wu, Amin Ullah, and Sung Wook Baik. Multi-grade brain tumor classification using deep cnn with extensive data augmentation. *Journal of Computational Science*, 30:174–182, 2019.

[83] Zhenghua Xu, Chang Qi, and Guizhi Xu. Semi-supervised attention-guided

cyclegan for data augmentation on medical images. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 563–568, Nov 2019.

[84] Matej Kompanek, Martin Tamajka, and Wanda Benesova. Volumetrie data augmentation as an effective tool in mri classification using 3d convolutional neural network. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 115–119, June 2019.

[85] Changhee Han, Kohei Murao, Tomoyuki Noguchi, Yusuke Kawata, Fumiya Uchiyama, Leonardo Rundo, Hideki Nakayama, and Shin'ichi Satoh. Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, pages 119–127. Association for Computing Machinery.

[86] Amish Kumar, Oduri Narayana Murthy, Shrish, Palash Ghosal, Amritendu Mukherjee, and Debashis Nandi. A dense u-net architecture for multiple sclerosis lesion segmentation. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pages 662–667, Oct 2019.

[87] Ava Assadi Abolvardi, Len Hamey, and Kevin Ho-Shon. Registration based data augmentation for multiple sclerosis lesion segmentation. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–5, Dec 2019.

[88] Mengli Sun, Jiajun Wang, and Zheru Chi. Brain tumor segmentation based

on amrunet++ neural network. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 1920–1924, Dec 2020.

[89] Xiangchuan Gao, Lei Ma, Jin Jin, Junmin Li, Zhenxia Ma, Yunkai Zhai, and Xingwang Li. Glioma segmentation strategies in 5g teleradiology. In *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 1–6, April 2020.

[90] Charley Gros, Andreanne Lemay, and Julien Cohen-Adad. Softseg: Advantages of soft versus binary training for image segmentation. *Medical Image Analysis*, 71:102038, 2021.

[91] Qingyun Li, Zhibin Yu, Yubo Wang, and Haiyong Zheng. TumorGAN: A multi-modal data augmentation framework for brain tumor segmentation. 20(15).

[92] Wenshan Wu, Yuhao Lu, Ravikiran Mane, and Cuntai Guan. Deep learning for neuroimaging segmentation with a novel data augmentation strategy. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1516–1519, July 2020.

[93] Xu Chen, Chunfeng Lian, Li Wang, Hannah Deng, Tianshu Kuang, Steve H. Fung, Jaime Gateno, Dinggang Shen, James J. Xia, and Pew-Thian Yap. Diverse data augmentation for learning image segmentation with cross-modality annotations. *Medical Image Analysis*, 71:102060, July 2021.

[94] Lin Teng, Hang Li, and Shahid Karim. DMCNN: A deep multiscale convolutional neural network model for medical image segmentation. 2019:8597606.

[95] Quentin Delannoy, Chi-Hieu Pham, Clément Cazorla, Carlos Tor-Díez, Guillaume Dollé, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Pas-

sat, and François Rousseau. SegSRGAN: Super-resolution and segmentation using generative adversarial networks - application to neonatal brain MRI. 120:103755. Place: United States.

[96] Mehran Pesteie, Purang Abolmaesumi, and Robert N. Rohling. Adaptive augmentation of medical data using independently conditional variational auto-encoders. *IEEE Transactions on Medical Imaging*, 38(12):2807–2820, Dec 2019.

[97] Jakub Nalepa, Grzegorz Mrukwa, Szymon Piechaczek, Pablo Ribalta Lorenzo, Michal Marcinkiewicz, Barbara Bobek-Billewicz, Pawel Wawrzyniak, Pawel Ulrych, Janusz Szymanek, Marcin Cwiek, Wojciech Dudzik, Michal Kawulok, and Michael P. Hayball. Data augmentation via image registration. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4250–4254, Sep. 2019.

[98] Yuanqi Du, Quan Quan, Hu Han, and S. Kevin Zhou. Semi-Supervised Pseudo-Healthy Image Synthesis via Confidence Augmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, March 2022. ISSN: 1945-8452.

[99] Fan Zhang, Bo Pan, Pengfei Shao, Peng Liu, Alzheimer's Disease Neuroimaging Initiative, Australian Imaging Biomarkers Lifestyle flagship study of ageing, Shuwei Shen, Peng Yao, and Ronald X. Xu. A Single Model Deep Learning Approach for Alzheimer's Disease Diagnosis. *Neuroscience*, 491:200–214, May 2022.

[100] Kaishin W. Tanaka, Carlo Russo, Sidong Liu, Marcus A. Stoodley, and Antonio Di Ieva. Use of deep learning in the MRI diagnosis of Chiari malformation type I. *Neuroradiology*, 64(8):1585–1592, August 2022.

[101] Abhishek Bal, Minakshi Banerjee, Rituparna Chaki, and Punit Sharma. An efficient brain tumor image classifier by combining multi-pathway cascaded deep neural network and handcrafted features in MR images. *Medical & Biological Engineering & Computing*, 59(7-8):1495–1527, August 2021.

[102] Amjad Rehman Khan, Siraj Khan, Majid Harouni, Rashid Abbasi, Sajid Iqbal, and Zahid Mehmood. Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification. *Microscopy Research and Technique*, 84(7):1389–1399, July 2021.

[103] Qi Lu, Wan Liu, Zhizheng Zhuo, Yuxing Li, Yunyun Duan, Pinnan Yu, Liying Qu, Chuyang Ye, and Yaou Liu. A transfer learning approach to few-shot segmentation of novel white matter tracts. *Medical Image Analysis*, 79:102454, July 2022.

[104] Clément Chadebec, Elina Thibeau-Sutre, Ninon Burgos, and Stéphanie Allassonnière. Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–18, 2022. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[105] Reda Abdellah Kamraoui, Vinh-Thong Ta, Thomas Tourdias, Boris Mansencal, José V. Manjon, and Pierrick Coup. DeepLesionBrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation. *Medical Image Analysis*, 76:102312, February 2022.

[106] Moritz Platscher, Jonathan Zopes, and Christian Federau. Image translation for medical image generation: Ischemic stroke lesion segmentation. *Biomedical Signal Processing and Control*, 72:103283, February 2022.

[107] Abhishek Singh Sambyal, Narayanan C Krishnan, and Deepti R Bathula. Towards Reducing Aleatoric Uncertainty for Medical Imaging Tasks. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, March 2022. ISSN: 1945-8452.

[108] Ahsan Bin Tufail, Nazish Anwar, Mohamed Tahar Ben Othman, Inam Ullah, Rehan Ali Khan, Yong-Kui Ma, Deepak Adhikari, Ateeq Ur Rehman, Muhammad Shafiq, and Habib Hamam. Early-Stage Alzheimer's Disease Categorization Using PET Neuroimaging Modality and Convolutional Neural Networks in the 2D and 3D Domains. *Sensors (Basel, Switzerland)*, 22(12):4609, June 2022.

[109] Yu Wang, Yarong Ji, and Hongbing Xiao. A data augmentation method for fully automatic brain tumor segmentation. *Computers in Biology and Medicine*, 149:106039, October 2022.

[110] Debadyuti Mukherkjee, Pritam Saha, Dmitry Kaplun, Aleksandr Sinitca, and Ram Sarkar. Brain tumor image generation using an aggregation of GAN models with style transfer. *Scientific Reports*, 12(1):9141, June 2022.

[111] Andrés Anaya-Isaza and Leonel Mera-Jiménez. Data augmentation and transfer learning for brain tumor detection in magnetic resonance imaging. *IEEE Access*, 10:23217–23233, 2022.

[112] Rishi Raj, Jimson Mathew, Santhosh Kumar Kannath, and Jeny Rajan. Crossover based technique for data augmentation. *Computer Methods and Programs in Biomedicine*, 218:106716, May 2022.

[113] Yukihiro Nomura, Shouhei Hanaoka, Tomomi Takenaga, Takahiro Nakao, Hi-

saichi Shibata, Soichiro Miki, Takeharu Yoshikawa, Takeyuki Watadani, Naoto Hayashi, and Osamu Abe. Preliminary study of generalized semiautomatic segmentation for 3D voxel labeling of lesions based on deep learning. *International Journal of Computer Assisted Radiology and Surgery*, 16(11):1901–1913, November 2021.

[114] Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J. Wood, Holger Roth, Andriy Myronenko, Daguang Xu, and Ziyue Xu. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Transactions on Medical Imaging*, 39(7):2531–2540, July 2020.

[115] Pierre-Jean Lartaud, David Hallé, Arnaud Schleef, Riham Dessouky, Anna Sesilia Vlachomitrou, Philippe Douek, Jean-Michel Rouet, Olivier Nempont, and Loïc Boussel. Spectral augmentation for heart chambers segmentation on conventional contrasted and unenhanced CT scans: an in-depth study. *International Journal of Computer Assisted Radiology and Surgery*, 16(10):1699–1709, October 2021.

[116] Takafumi Nemoto, Natsumi Futakami, Etsuo Kunieda, Masamichi Yagi, Atsuya Takeda, Takeshi Akiba, Eride Mutu, and Naoyuki Shigematsu. Effects of sample size and data augmentation on U-Net-based automatic segmentation of various organs. *Radiological Physics and Technology*, 14(3):318–327, September 2021.

[117] Lennart Bargsten and Alexander Schlaefer. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. 15(9):1427–1436.

[118] Yuxin Gong, Yingying Zhang, Haogang Zhu, Jing Lv, Qian Cheng, Hongjia Zhang, Yihua He, and Shuliang Wang. Fetal congenital heart disease echocardiogram screening based on DGACNN: Adversarial one-class classification combined with video transfer learning. 39(4):1206–1222. Place: United States.

[119] Akis Linardos, Kaisar Kushibar, Sean Walsh, Polyxeni Gkontra, and Karim Lekadir. Federated learning for multi-center imaging diagnostics: a simulation study in cardiovascular disease. *Scientific Reports*, 12(1):3551, March 2022.

[120] Changling Li, Xiangfen Song, Hang Zhao, Li Feng, Tao Hu, Yuchen Zhang, Jun Jiang, Jianan Wang, Jianping Xiang, and Yong Sun. An 8-layer residual u-net with deep supervision for segmentation of the left ventricle in cardiac CT angiography. 200:105876. Place: Ireland.

[121] Georgios Simantiris and Georgios Tziritas. Cardiac mri segmentation with a dilated cnn incorporating domain-specific constraints. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1235–1243, Oct 2020.

[122] Krishna Chaitanya, Neerav Karani, Christian F. Baumgartner, Anton Becker, Olivio Donati, and Ender Konukoglu. Semi-supervised and task-driven data augmentation. In Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao, editors, *Information Processing in Medical Imaging*, pages 29–41, Cham, 2019. Springer International Publishing.

[123] Nils Gessert, Matthias Lutz, Markus Heyder, Sarah Latus, David M. Leistner, Youssef S. Abdelwahed, and Alexander Schlaefer. Automatic plaque detection in IVOCT pullbacks using convolutional neural networks. 38(2):426–434. Place: United States.

[124] Cristiana Tiago, Andrew Gilbert, Ahmed Salem Beela, Svein Arne Aase, Sten Roar Snare, Jurica Šprem, and Kristin McLeod. A Data Augmentation Pipeline to Generate Synthetic Labeled Datasets of 3D Echocardiography Images Using a GAN. *IEEE Access*, 10:98803–98815, 2022. Conference Name: IEEE Access.

[125] Pierre-Jean Lartaud, Claire Dupont, David Hallé, Arnaud Schleef, Riham Dessouky, Anna Sesilia Vlachomitrou, Jean-Michel Rouet, Olivier Nempont, and Loïc Boussel. A conventional-to-spectral CT image translation augmentation workflow for robust contrast injection-independent organ segmentation. *Medical Physics*, 49(2):1108–1122, February 2022.

[126] Leo Segre, Or Hirschorn, Dvir Ginzburg, and Dan Raviv. Shape-Consistent Generative Adversarial Networks for Multi-Modal Medical Segmentation Maps. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, March 2022. ISSN: 1945-8452.

[127] Feng Yang, Fangxuan Liang, Liyun Lu, and Mengxiao Yin. Dual attention-guided and learnable spatial transformation data augmentation multi-modal unsupervised medical image segmentation. *Biomedical Signal Processing and Control*, 78:103849, September 2022.

[128] Tuan D Pham. A comprehensive study on classification of covid-19 on computed tomography with pretrained convolutional neural networks. *Scientific reports*, 10(1):1–8, 2020.

[129] Vincent Andrearczyk, Julien Fageot, Valentin Oreiller, Xavier Montet, and Adrien Depeursinge. Local rotation invariance in 3d CNNs. 65:101756. Place: Netherlands.

[130] Liu Yang and Rudrasis Chakraborty. An "augmentation-free" rotation invariant classification scheme on point-cloud and its application to neuroimaging. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 713–716, April 2020.

[131] Saleh Albahli and Waleed Albattah. Detection of coronavirus disease from x-ray images using deep learning and transfer learning algorithms. 28(5):841–850.

[132] Ayat Abedalla, Malak Abdullah, Mahmoud Al-Ayyoub, and Elhadj Benkhelifa. 2st-unet: 2-stage training model using u-net for pneumothorax segmentation in chest x-rays. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, July 2020.

[133] Xin Li, Fan Chen, Haijiang Hao, and Mengting Li. A pneumonia detection method based on improved convolutional neural network. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, volume 1, pages 488–493, June 2020.

[134] Rajesh Kumar, WenYong Wang, Jay Kumar, Ting Yang, Abdullah Khan, Wazir Ali, and Ikram Ali. An integration of blockchain and ai for secure data sharing and detection of ct images for the hospitals. *Computerized Medical Imaging and Graphics*, 87:101812, 2021.

[135] Mostofa Ahsan, Rahul Gomes, and Anne Denton. Application of a convolutional neural network using transfer learning for tuberculosis detection. In *2019 IEEE International Conference on Electro Information Technology (EIT)*, pages 427–433, May 2019.

[136] Vedant Bhagat and Swapnil Bhaumik. Data augmentation using generative adversarial networks for pneumonia classification in chest xrays. In *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pages 574–579, Nov 2019.

[137] Ahmed Sedik, Abdullah M. Iliyasu, Basma Abd El-Rahiem, Mohammed E. Abdel Samea, Asmaa Abdel-Raheem, Mohamed Hammad, Jialiang Peng, Fathi E. Abd El-Samie, and Ahmed A. Abd El-Latif. Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. 12(7).

[138] Yaqi Wang, Lingling Sun, and Qun Jin. Enhanced diagnosis of pneumothorax with an improved real-time augmentation for imbalanced chest x-rays data based on dcnn. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(3):951–962, May 2021.

[139] Tuan D. Pham. Geostatistical simulation of medical images for data augmentation in deep learning. *IEEE Access*, 7:68752–68763, 2019.

[140] Octavio E. Martinez Manzanera, Sam Ellis, Vasileios Baltatzis, Arjun Nair, Loic Le Folgoc, Sujal Desai, Ben Glocker, and Julia A. Schnabel. Patient-specific 3d cellular automata nodule growth synthesis in lung cancer without the need of external data. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 5–9, April 2021.

[141] Hitesh Tekchandani, Shrish Verma, and Narendra Londhe. Performance improvement of mediastinal lymph node severity detection using GAN and inception network. 194:105478. Place: Ireland.

[142] Jakub Garstka and Michał Strzelecki. Pneumonia detection in x-ray chest images based on convolutional neural networks and data augmentation methods. In *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 18–23, Sep. 2020.

[143] Shui-Hua Wang, Yin Zhang, Xiaochun Cheng, Xin Zhang, and Yu-Dong Zhang. Psspnn: Patchshuffle stochastic pooling neural network for an explainable diagnosis of covid-19 with multiple-way data augmentation. *Computational and Mathematical Methods in Medicine*, 2021:6633755, Mar 2021.

[144] Marysia Winkels and Taco S. Cohen. Pulmonary nodule detection in CT scans with equivariant CNNs. 55:15–26. Place: Netherlands.

[145] Ryo Toda, Atsushi Teramoto, Masakazu Tsujimoto, Hiroshi Toyama, Kazuyoshi Imaizumi, Kuniaki Saito, and Hiroshi Fujita. Synthetic CT image generation of shape-controlled lung cancer using semi-conditional InfoGAN and its applicability for type classification. 16(2):241–251. Place: Germany.

[146] Akinyinka O. Omigbodun, Frederic Noo, Michael McNitt-Gray, William Hsu, and Scott S. Hsieh. The effects of physics-based data augmentation on the generalizability of deep neural networks: Demonstration on nodule false-positive reduction. 46(10):4563–4574. Place: United States.

[147] Euijoon Ahn, Ashnil Kumar, Michael Fulham, Dagan Feng, and Jinman Kim. Unsupervised domain adaptation to classify medical images using zero-bias convolutional auto-encoders and context-based feature augmentation. *IEEE Transactions on Medical Imaging*, 39(7):2385–2394, July 2020.

[148] Ojasvi Yadav, Kalpdrum Passi, and Chakresh Kumar Jain. Using deep learn-

ing to classify x-ray images of potential tuberculosis patients. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2368–2375, Dec 2018.

[149] Qingfeng Wang, Xuehai Zhou, Chao Wang, Zhiqin Liu, Jun Huang, Ying Zhou, Changlong Li, Hang Zhuang, and Jie-Zhi Cheng. Wgan-based synthetic minority over-sampling technique: Improving semantic fine-grained classification for lung nodules in ct images. *IEEE Access*, 7:18450–18463, 2019.

[150] Talib Iqball and M. Arif Wani. X-ray images dataset augmentation with progressively growing generative adversarial network. In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 93–97, March 2021.

[151] Tuan D. Pham. Classification of benign and metastatic lymph nodes in lung cancer with deep learning. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 728–733, Oct 2020.

[152] Aria Pezeshk, Sardar Hamidian, Nicholas Petrick, and Berkman Sahiner. 3-d convolutional neural networks for automatic detection of pulmonary nodules in chest ct. *IEEE Journal of Biomedical and Health Informatics*, 23(5):2080–2090, Sep. 2019.

[153] Zhehao He, Wang Lv, and Jian Hu. A simple method to train the AI diagnosis model of pulmonary nodules. 2020:2812874.

[154] Man Tan, Fa Wu, Bei Yang, Jinlian Ma, Dexing Kong, Zengsi Chen, and Dan Long. Pulmonary nodule detection using hybrid two-stage 3d CNNs. 47(8):3376–3388. Place: United States.

[155] Changhee Han, Yoshiro Kitamura, Akira Kudo, Akimichi Ichinose, Leonardo Rundo, Yujiro Furukawa, Kazuki Umemoto, Yuanzhong Li, and Hideki Nakayama. Synthesizing diverse lung nodules wherever massively: 3d multi-conditional gan-based ct image augmentation for object detection. In *2019 International Conference on 3D Vision (3DV)*, pages 729–737, Sep. 2019.

[156] Jue Jiang, Yu-Chi Hu, Neelam Tyagi, Pengpeng Zhang, Andreas Rimner, Joseph O. Deasy, and Harini Veeraraghavan. Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets. 46(10):4392–4404.

[157] Yudong Zhang, Suresh Chandra Satapathy, Li-Yao Zhu, Juan Manuel Górriz, and Shuihua Wang. A Seven-Layer Convolutional Neural Network for Chest CT-Based COVID-19 Diagnosis Using Stochastic Pooling. *IEEE Sensors Journal*, 22(18):17573–17582, September 2022. Conference Name: IEEE Sensors Journal.

[158] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks. *IEEE Access*, 10:66467–66480, 2022. Conference Name: IEEE Access.

[159] Li Sun, Junxiang Chen, Yanwu Xu, Mingming Gong, Ke Yu, and Kayhan Batmanghelich. Hierarchical Amortized GAN for 3D High Resolution Medical Image Synthesis. *IEEE journal of biomedical and health informatics*, 26(8):3966–3975, August 2022.

[160] Zheng-Zheng Guo, Li-Xin Zheng, De-Tian Huang, Tan Yan, and Qiu-Ling Su. RS-FFGAN:Generative adversarial network based on real sample feature fusion

for pediatric CXR image data enhancement. *Journal of Radiation Research and Applied Sciences*, 15(4):100461, December 2022.

[161] Joana Sousa, Tania Pereira, Inês Neves, Francisco Silva, and Hélder P. Oliveira. The Influence of a Coherent Annotation and Synthetic Addition of Lung Nodules for Lung Segmentation in CT Scans. *Sensors (Basel, Switzerland)*, 22(9):3443, April 2022.

[162] Ioannis D. Apostolopoulos, Nikolaos D. Papathanasiou, and George S. Panayiotakis. Classification of lung nodule malignancy in computed tomography imaging utilising generative adversarial networks and semi-supervised transfer learning. *Biocybernetics and Biomedical Engineering*, 41(4):1243–1257, October 2021.

[163] Zonggui Li, Junhua Zhang, Bo Li, Xiaoying Gu, and Xudong Luo. COVID-19 diagnosis on CT scan images using a generative adversarial network and concatenated feature pyramid network with an attention mechanism. *Medical Physics*, 48(8):4334–4349, August 2021.

[164] Qiuli Wang, Xiaohong Zhang, Wei Zhang, Mingchen Gao, Sheng Huang, Jian Wang, Jiuquan Zhang, Dan Yang, and Chen Liu. Realistic Lung Nodule Synthesis With Multi-Target Co-Guided Adversarial Mechanism. *IEEE transactions on medical imaging*, 40(9):2343–2353, September 2021.

[165] Usman Asghar, Muhammad Arif, Khurram Ejaz, Dragos Vicoveanu, Diana Izdrui, and Oana Geman. An Improved COVID-19 Detection using GAN-Based Data Augmentation and Novel QuNet-Based Classification. *BioMed Research International*, 2022:8925930, 2022.

[166] Minki Chung, Seo Taek Kong, Beomhee Park, Younjoon Chung, Kyu-Hwan Jung, and Joon Beom Seo. Utilizing Synthetic Nodules for Improving Nodule Detection in Chest Radiographs. *Journal of Digital Imaging*, 35(4):1061–1068, August 2022.

[167] Xuzhe Zhang, Elsa D. Angelini, Fateme S. Haghpanah, Andrew F. Laine, Yanping Sun, Grant T. Hiura, Stephen M. Dashnaw, Martin R. Prince, Eric A. Hoffman, Bharath Ambale-Venkatesh, Joao A. Lima, Jim M. Wild, Emlyn W. Hughes, R. Graham Barr, and Wei Shen. Quantification of lung ventilation defects on hyperpolarized MRI: The Multi-Ethnic Study of Atherosclerosis (MESA) COPD study. *Magnetic Resonance Imaging*, 92:140–149, October 2022.

[168] Saman Motamed, Patrik Rogalla, and Farzad Khalvati. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Informatics in Medicine Unlocked*, 27:100779, January 2021.

[169] Xiaoli Qin, Francis Minhthang Bui, Ha H. Nguyen, and Zhu Han. Learning From Limited and Imbalanced Medical Images With Finer Synthetic Images From GANs. *IEEE Access*, 10:91663–91677, 2022. Conference Name: IEEE Access.

[170] Ryo Toda, Atsushi Teramoto, Masashi Kondo, Kazuyoshi Imaizumi, Kuniaki Saito, and Hiroshi Fujita. Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation. *Scientific Reports*, 12(1):12867, July 2022.

[171] Xiaocong Chen, Yun Li, Lina Yao, Ehsan Adeli, Yu Zhang, and Xianzhi Wang. Generative adversarial U-Net for domain-free few-shot medical diagnosis. *Pattern Recognition Letters*, 157:112–118, May 2022.

[172] Eduardo Castro, Jaime S. Cardoso, and Jose Costa Pereira. Elastic deformations for data augmentation in breast cancer mass detection. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 230–234, March 2018.

[173] Bruno Barufaldi, Predrag R Bakic, David D Pokrajac, Miguel A Lago, and Andrew DA Maidment. Developing populations of software breast phantoms for virtual clinical trials. In *14th International Workshop on Breast Imaging (IWBI 2018)*, volume 10718, pages 481–489. SPIE, 2018.

[174] Ines Domingues, Pedro H. Abreu, and Joäo Santos. Bi-rads classification of breast cancer: A new pre-processing pipeline for deep models training. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1378–1382, Oct 2018.

[175] Shrinivas D Desai, Shantala Giraddi, Nitin Verma, Puneet Gupta, and Sharan Ramya. Breast cancer detection using gan for limited labeled dataset. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 34–39, Sep. 2020.

[176] Bashir Zeimarani, Marly Guimaraes Fernandes Costa, Nilufar Zeimarani Nurani, Sabrina Ramos Bianco, Wagner Coelho De Albuquerque Pereira, and Cicero Ferreira Fernandes Costa Filho. Breast lesion classification in ultrasound images using deep convolutional neural network. *IEEE Access*, 8:133349–133359, 2020.

[177] Haichao Cao, Shiliang Pu, Wenming Tan, and Junyan Tong. Breast mass detection in digital mammography based on anchor-free architecture. 205:106033. Place: Ireland.

[178] Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, Clifford Yang, and Jinbo Bi. Convolutional neural network for automated mass segmentation in mammography. In *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, pages 1–1, Oct 2018.

[179] Peng Shi, Chongshu Wu, Jing Zhong, and Hui Wang. Deep learning from small dataset for bi-rads density classification of mammography images. In *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, pages 102–109, Aug 2019.

[180] Shahbaz Siddeeq, Jiyun Li, Hafiz Muhammad Ali Bhatti, Arslan Manzoor, and Umar Subhan Malhi. Deep learning RN-BCNN model for breast cancer BI-RADS classification. In *2021 The 4th International Conference on Image and Graphics Processing*, ICIGP 2021, pages 219–225. Association for Computing Machinery.

[181] Lucas M. Valério, Daniel H. A. Alves, Luigi F. Cruz, Pedro H. Bugatti, Claiton de Oliveira, and Priscila T. M. Saito. Deepmammo: Deep transfer learning for lesion classification of mammographic images. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 447–452, June 2019.

[182] Dhivya S, Mohanavalli S, Karthika S, Shivani S, and Mageswari R. Gan based data augmentation for enhanced tumor classification. In *2020 4th International*

Conference on Computer, Communication and Signal Processing (ICCCSP), pages 1–5, Sep. 2020.

[183] Chisako Muramatsu, Mizuho Nishio, Takuma Goto, Mikinao Oiwa, Takako Morita, Masahiro Yakami, Takeshi Kubo, Kaori Togashi, and Hiroshi Fujita. Improving breast mass classification by shared data with domain transformation using a generative adversarial network. *Computers in Biology and Medicine*, 119:103698, 2020.

[184] Tianyu Shen, Kunkun Hao, Chao Gou, and Fei-Yue Wang. Mass image synthesis in mammogram with contextual information based on gans. *Computer Methods and Programs in Biomedicine*, 202:106019, 2021.

[185] Ting Pang, Jeannie Hsiu Ding Wong, Wei Lin Ng, and Chee Seng Chan. Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification. *Computer Methods and Programs in Biomedicine*, 203:106018, May 2021.

[186] Khaoula Belhaj Soulami, Naima Kaabouch, and Mohamed Nabil Saidi. Breast cancer: Classification of suspicious regions in digital mammograms based on capsule network. *Biomedical Signal Processing and Control*, 76:103696, July 2022.

[187] Huayu Wang, Yixin Hu, Yao Lu, Jianhua Zhou, and Yongze Guo. The uncertainty of boundary can improve the classification accuracy of BI-RADS 4A ultrasound image. *Medical Physics*, 49(5):3314–3324, May 2022.

[188] Qiu Guan, Yizhou Chen, Zihan Wei, Ali Asghar Heidari, Haigen Hu, Xu-Hua Yang, Jianwei Zheng, Qianwei Zhou, Huiling Chen, and Feng Chen. Medical

image augmentation for lesion detection using a texture-constrained multichannel progressive GAN. *Computers in Biology and Medicine*, 145:105444, June 2022.

[189] Wongsakorn Preedanan, Itsuo Kumazawa, Toshiaki Kondo, and Ishioka Junichiro. Urinary stones segmentation in abdominal x-ray images based on u-net deep learning model and data augmentation techniques. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pages 118–123, Oct 2020.

[190] Odeuk Kwon, Tae-Hoon Yong, Se-Ryong Kang, Jo-Eun Kim, Kyung-Hoe Huh, Min-Suk Heo, Sam-Sun Lee, Soon-Chul Choi, and Won-Jin Yi. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. 49(8):20200185.

[191] Nurya Aghnia Farda, Jiing-Yih Lai, Jia-Ching Wang, Pei-Yuan Lee, Jia-Wei Liu, and I.-Hui Hsieh. Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. 52(3):616–624. Place: Netherlands.

[192] Nikolaos Kyventidis and Christos Angelopoulos. Intraoral radiograph anatomical region classification using neural networks. 16(3):447–455. Place: Germany.

[193] E. O. Wesselink, J. M. Elliott, M. W. Coppieters, M. J. Hancock, B. Cronin, A. Pool-Goudzwaard, and K. A. Weber Ii. Convolutional neural networks for the automatic segmentation of lumbar paraspinal muscles in people with low back pain. *Scientific Reports*, 12(1):13485, August 2022.

[194] Kuen-Jang Tsai, Chih-Chun Chang, Lun-Chien Lo, John Y. Chiang, Chao-

Sung Chang, and Yu-Jung Huang. Automatic segmentation of paravertebral muscles in abdominal CT scan by U-Net: The application of data augmentation technique to increase the Jaccard ratio of deep learning. *Medicine*, 100(44):e27649, November 2021.

[195] Maryam Hammami, Denis Friboulet, and Razmig Kechichian. Cycle gan-based data augmentation for multi-organ detection in ct images via yolo. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 390–393, Oct 2020.

[196] Thorbjørn Louring Koch, Mathias Perslev, Christian Igel, and Sami Sebastian Brandt. Accurate segmentation of dental panoramic radiographs with u-nets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 15–19, April 2019.

[197] Xu Yin, Yan Li, Xu Zhang, and Byeong-Seok Shin. Medical image augmentation using image synthesis with contextual function. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6, Oct 2019.

[198] Michael Gadermayr, Kexin Li, Madlaine Müller, Daniel Truhn, Nils Krämer, Dorit Merhof, and Burkhard Gess. Domain-specific data augmentation for segmenting MR images of fatty infiltrated human thighs with neural networks. 49(6):1676–1683. Place: United States.

[199] Tiexin Qin, Ziyuan Wang, Kelei He, Yinghuan Shi, Yang Gao, and Dinggang Shen. Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation. In *ICASSP 2020 - 2020 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1419–1423, May 2020.

[200] Yuxin Ma, Shuo Wang, Yang Hua, Ruhui Ma, Tao Song, Zhengui Xue, Heng Cao, and Haibing Guan. Perceptual Data Augmentation for Biomedical Coronary Vessel Segmentation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–12, 2022. Conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[201] Wenxuan He, Min Liu, Yi Tang, Qinghao Liu, and Yaonan Wang. Differentiable Automatic Data Augmentation by Proximal Update for Medical Image Segmentation. *IEEE/CAA Journal of Automatica Sinica*, 9(7):1315–1318, July 2022. Conference Name: IEEE/CAA Journal of Automatica Sinica.

[202] Xiaokun Liang, Na Li, Zhicheng Zhang, Jing Xiong, Shoujun Zhou, and Yaoqin Xie. Incorporating the hybrid deformable model for improving the performance of abdominal CT segmentation via multi-scale feature fusion network. *Medical Image Analysis*, 73:102156, October 2021.

[203] Yonatan Nozik, Laura A. Hallock, Daniel Ho, Sai Mandava, Chris Mitchell, Thomas Hui Li, and Ruzena Bajcsy. Openarm 2.0: Automated segmentation of 3d tissue structures for multi-subject study of muscle deformation dynamics. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 982–988, 2019.

[204] Chao Li, Haibo Jia, Jinwei Tian, Chong He, Fang Lu, Kaiwen Li, Yubin Gong, Sining Hu, Bo Yu, and Zhao Wang. Comprehensive Assessment of Coronary Calcification in Intravascular OCT Using a Spatial-Temporal Encoder-Decoder

Network. *IEEE Transactions on Medical Imaging*, 41(4):857–868, April 2022. Conference Name: IEEE Transactions on Medical Imaging.

[205] Yuxuan Mu, He Zhao, Jia Guo, and Huiqi Li. MSRT: Multi-Scale Spatial Regularization Transformer For Multi-Label Classification in Calcaneus Radiograph. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, March 2022. ISSN: 1945-8452.

[206] Yizhi Chen, Yacheng Ren, Ling Fu, Junfeng Xiong, Rasmus Larsson, Xiaowei Xu, Jianqi Sun, and Jun Zhao. A 3d convolutional neural network framework for polyp candidates detection on the limited dataset of ct colonography. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 678–681, 2018.

[207] Tomoki Uemura, Janne J. Näppi, Yasuji Ryu, Chinatsu Watari, Tohru Kamiya, and Hiroyuki Yoshida. A generative flow-based model for volumetric data augmentation in 3d deep learning for computed tomographic colonography. 16(1):81–89.

[208] Gwiseong Moon, Seola Kim, Woojin Kim, Yoon Kim, Yeonjin Jeong, and Hyun-Soo Choi. Computer Aided Facial Bone Fracture Diagnosis (CA-FBFD) System Based on Object Detection Model. *IEEE Access*, 10:79061–79070, 2022. Conference Name: IEEE Access.

[209] Yiren Chen, Kunjin He, Bo Hao, Yiping Weng, and Zhengming Chen. FractureNet: A 3D Convolutional Neural Network Based on the Architecture of m-Ary Tree for Fracture Type Identification. *IEEE Transactions on Medical Imaging*, 41(5):1196–1207, May 2022. Conference Name: IEEE Transactions on Medical Imaging.

[210] Asaduz Zaman, Sang Hyun Park, Hyunhee Bang, Chul-Woo Park, Ilhyung Park, and Sanghyun Joung. Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images. 15(6):931–941. Place: Germany.

[211] Veit Sandfort, Ke Yan, Perry J. Pickhardt, and Ronald M. Summers. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. 9(1):16884.

[212] Shi Yin, Qinmu Peng, Hongming Li, Zhengqiang Zhang, Xinge You, Katherine Fischer, Susan L. Furth, Gregory E. Tasian, and Yong Fan. Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. 60:101602.

[213] Nawaf Waqas, Sairul Izwan Safie, Kushsairy Abdul Kadir, Sheroz Khan, and Muhammad Haris Kaka Khel. DEEPFAKE Image Synthesis for Data Augmentation. *IEEE Access*, 10:80847–80857, 2022. Conference Name: IEEE Access.

[214] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 289–293, April 2018.

[215] Thanh-Nghia Truong, Vu-Duy Dam, and Thanh-Sach Le. Medical images sequence normalization and augmentation: Improve liver tumor segmentation from small dataset. In *2018 3rd International Conference on Control, Robotics and Cybernetics (CRC)*, pages 1–5, Sep. 2018.

[216] Peng Chen, Yuqing Song, Deqi Yuan, and Zhe Liu. Feature fusion adversarial learning network for liver lesion classification. In *Proceedings of the ACM Multimedia Asia*, MMAsia '19, pages 1–7. Association for Computing Machinery.

[217] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.

[218] Fei Gao, Kai Qiao, Bin Yan, Minghui Wu, Linyuan Wang, Jian Chen, and Dapeng Shi. Hybrid network with difference degree and attention mechanism combined with radiomics (H-DARnet) for MVI prediction in HCC. *Magnetic Resonance Imaging*, 83:27–40, November 2021.

[219] Lin Han, Yuanhao Chen, Jiaming Li, Bowei Zhong, Yuzhu Lei, and Minghui Sun. Liver segmentation with 2.5d perpendicular unets. *Computers Electrical Engineering*, 91:107118, 2021.

[220] Hansang Lee, Haeil Lee, Helen Hong, Heejin Bae, Joon Seok Lim, and Junmo Kim. Classification of focal liver lesions in CT images using convolutional neural networks with lesion information augmented patches and synthetic data augmentation. *Medical Physics*, 48(9):5029–5046, September 2021.

[221] Yingying Liu, Ji Zhou, Shiyao Chen, and Lei Liu. Muscle segmentation of l3 slice in abdomen ct images based on fully convolutional networks. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–5, Nov 2019.

[222] Alvaro Fernandez-Quilez, Steinar Valle Larsen, Morten Goodwin, Thor Ole

Gulsrud, Svein Reidar Kjosavik, and Ketil Oppedal. Improving prostate whole gland segmentation in t2-weighted mri with synthetically generated data. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1915–1919, April 2021.

[223] Christina Gsaxner, Peter M. Roth, Jürgen Wallner, and Jan Egger. Exploit fully automatic low-level segmented PET data for training high-level deep learning algorithms for the corresponding CT data. 14(3):e0212550.

[224] Ruqian Hao, Khashayar Namdar, Lin Liu, Masoom A. Haider, and Farzad Khalvati. A Comprehensive Study of Data Augmentation Strategies for Prostate Cancer Detection in Diffusion-Weighted MRI Using Convolutional Neural Networks. *Journal of Digital Imaging*, 34(4):862–876, August 2021.

[225] Davood Karimi, Golnoosh Samei, Claudia Kesch, Guy Nir, and Septimiu E. Salcudean. Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. 13(8):1211–1219. Place: Germany.

[226] Xiaozhi Ma, Dongdong Xie, Jing Fang, and Shu Zhan. Segmentation of prostate peripheral zone based on multi-scale features enhancement. In *Proceedings of the Third International Symposium on Image Computing and Digital Medicine*, ISICDM 2019, pages 349–353. Association for Computing Machinery.

[227] Xuanang Xu, Thomas Sanford, Baris Turkbey, Sheng Xu, Bradford J. Wood, and Pingkun Yan. Shadow-Consistent Semi-Supervised Learning for Prostate Ultrasound Segmentation. *IEEE Transactions on Medical Imaging*, 41(6):1331–1345, June 2022. Conference Name: IEEE Transactions on Medical Imaging.

[228] Houqiang Yu and Xuming Zhang. Synthesis of prostate MR images for classification using capsule network-based GAN model. 20(20).

[229] Joohyung Lee, Ji Eun Oh, Min Ju Kim, Bo Yun Hur, and Dae Kyung Sohn. Reducing the model variance of a rectal cancer segmentation network. *IEEE Access*, 7:182725–182733, 2019.

[230] Yuyin Zhou, David Dreizin, Yan Wang, Fengze Liu, Wei Shen, and Alan L. Yuille. External Attention Assisted Multi-Phase Splenic Vascular Injury Segmentation With Limited Data. *IEEE Transactions on Medical Imaging*, 41(6):1346–1357, June 2022. Conference Name: IEEE Transactions on Medical Imaging.

[231] Jae-Seo Lee, Shyam Adhikari, Liu Liu, Ho-Gul Jeong, Hyongsuk Kim, and Suk-Ja Yoon. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. 48(1):20170344.

[232] Zijia Liu, Jiannan Liu, Zijie Zhou, Qiaoyu Zhang, Hao Wu, Guangtao Zhai, and Jing Han. Differential diagnosis of ameloblastoma and odontogenic keratocyst by machine learning of panoramic radiographs. 16(3):415–422.

[233] Guohua Shi, Jiawen Wang, Yan Qiang, Xiaotang Yang, Juanjuan Zhao, Rui Hao, Wenkai Yang, Qianqian Du, and Ntikurako Guy-Fernand Kazihise. Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification. 196:105611. Place: Ireland.

[234] Ruixuan Zhang, Wenhuan Lu, Xi Wei, Jialin Zhu, Han Jiang, Zhiqiang Liu, Jie Gao, Xuewei Li, Jian Yu, Mei Yu, and Ruiguo Yu. A Progressive Gen-

erative Adversarial Method for Structurally Inadequate Medical Image Data Augmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(1):7–16, January 2022. Conference Name: IEEE Journal of Biomedical and Health Informatics.

[235] Qiang Lin, Chuangui Cao, Tongtong Li, Zhengxing Man, Yongchun Cao, and Haijun Wang. dSPIC: a deep SPECT image classification network for automated multi-disease, multi-lesion diagnosis. *BMC medical imaging*, 21(1):122, August 2021.

[236] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No new-net. In *International MICCAI Brainlesion Workshop*, pages 234–244. Springer, 2018.

[237] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018.

[238] Richard McKinley, Raphael Meier, and Roland Wiest. Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 456–465. Springer, 2018.

[239] Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J. Wood, Holger Roth, Andriy Myronenko, Daguang Xu, and Ziyue Xu. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. 39(7):2531–2540.

[240] Diksha Sharma, Christian G Graff, Andreu Badal, Rongping Zeng, Purva

Sawant, Aunnasha Sengupta, Eshan Dahal, and Aldo Badano. In silico imaging tools from the victre clinical trial. *Medical physics*, 46(9):3924–3928, 2019.

[241] Ryan Morrison. Ai-generated brain scans highlight potential for synthetic healthcare data. https://techmonitor.ai/technology/ai-and-automation/synthetic-healthcare-data-brain-scans-nvidia, note = Accessed: 2022-06-27.

[242] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David Mong, Safwan Halabi, Jesse Sandberg, Ricky Jones, David Larson, Curtis Langlotz, Bhavik Patel, Matthew Lungren, and Andrew Ng. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:590–597, 07 2019.

[243] Jitesh Seth, Rohit Lokwani, Viraj Kulkarni, Aniruddha Pant, and Amit Kharat. Reducing labelled data requirement for pneumonia segmentation using image augmentations, 02 2021.

Appendix A. Examples of data augmentation techniques

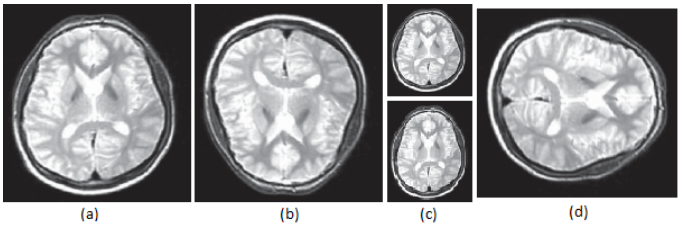


Figure A.1: Example of affine transformations applied to a brain magnetic resonance image. Respectively: (a) original image (b) vertical flipping (c) scaling (d) horizontal flipping

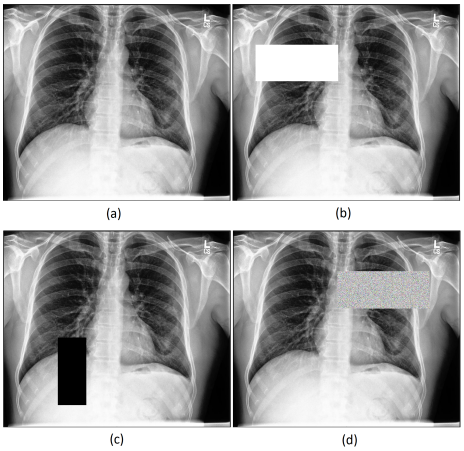


Figure A.2: Example of random erasing transformation applied to chest x-ray image from CheXpert (242). Respectively: (a) original image (b) white erasing value (c) black erasing value (d) random erasing value

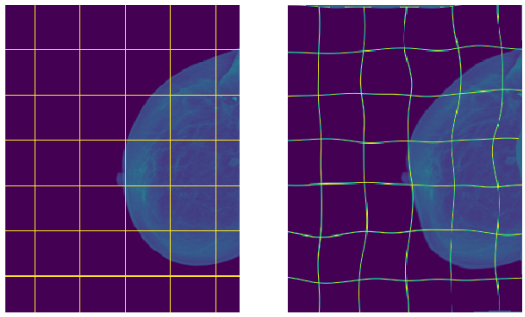


Figure A.3: Example of elastic transformation applied to a mammography.A Gaussian filter is applied after the distortion. (Image credit: (172))

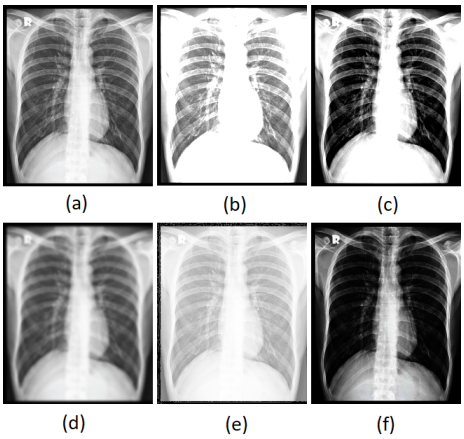


Figure A.4: Example of pixel-level transformations applied to a chest x-ray image. Respectively: (a) original image (b) brightness (c) contrast (d) blur (e) light gamma correction (f) dark gamma correction

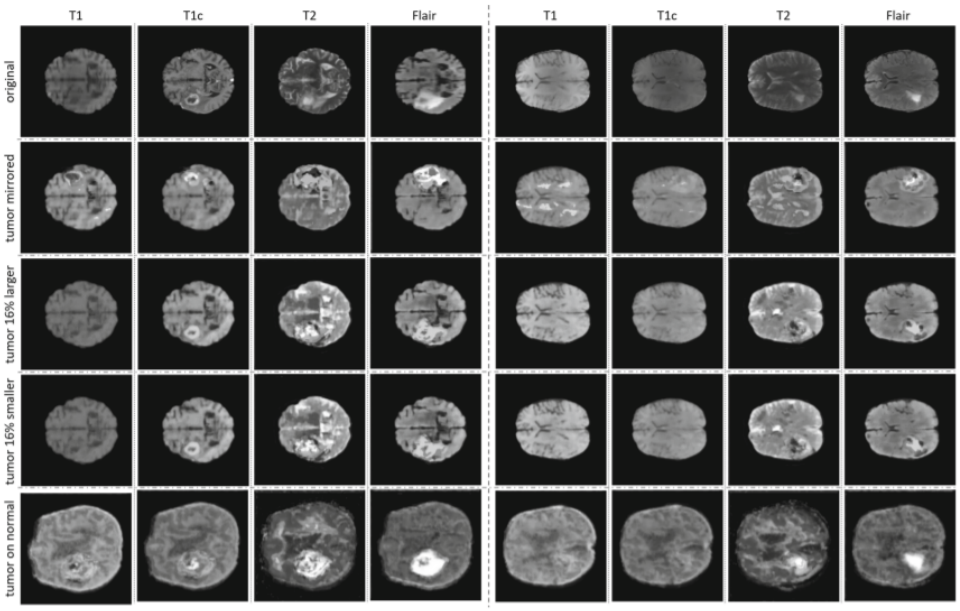
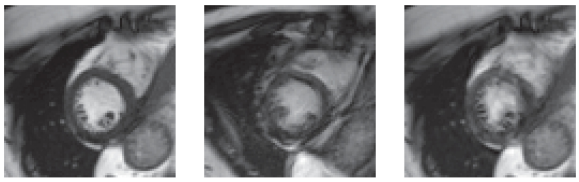


Figure A.5: Example of GAN generated images. The first row depicts the real images on which the synthetic tumors were based. Generated images without adjustment of the segmentation label are shown in the second row. Examples of generated images with various adjustments to the tumor segmentation label are shown in the third through fifth rows. The last row depicts examples of synthetic images where a tumor label is placed on a tumor-free brain label from the ADNI data set. (Image credit: (30))



(a) Good quality image (b) Motion artefact image (c) Synthetic image

Figure A.6: Example of reconstruction-based method: a good quality cine CMR image (a), an image with blurring motion artifacts (b), and a k-space corrupted image (c). The k-space corruption process is able to simulate realistic motion-related artifacts (Image credit: (42))

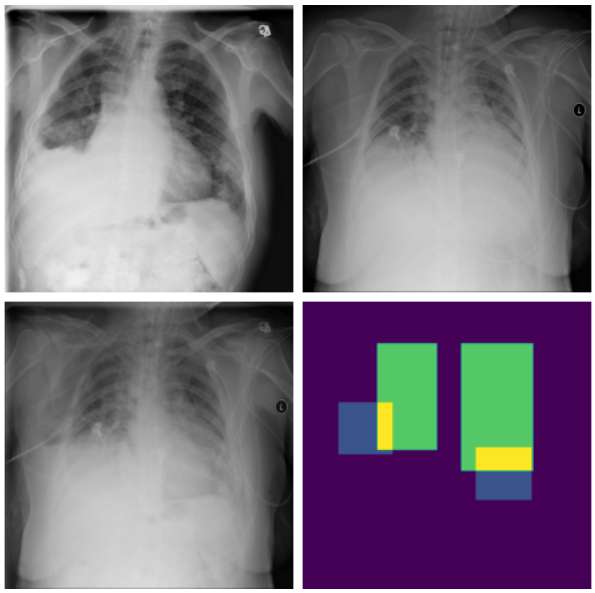


Figure A.7: Example of mix-up generated images. In the first row there original chest x-rays, in the bottom left there is the result of mix-up augmentation and in the bottom right the corresponding mask.(Image credit: (243))

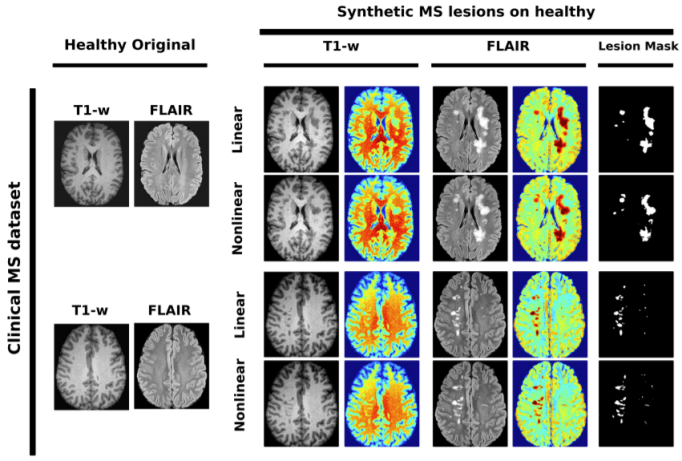


Figure A.8: Example of model-based method: synthetic multiple sclerosis lesions generated on a healthy subject. Slices are also displayed using jet color maps to visually enhance the intensities (Image credit: (38))

Appendix B. Supplementary material: methods

Digital Library	Search string
ACM Digital Library	[All: "data augmentation"] AND [All: "medical imaging"]
IEL-IEEE	("Abstract":Data Augmentation) AND ("Abstract":Medical Imaging)
PubMed	("data augmentation") AND ("imaging")
Science Direct	Title, abstract, keywords: ("Data Augmentation") AND ("Medical")

Table B.7: Digital libraries with respective URL and search string.

	Inclusion criteria	Exclusion criteria
Language	English	Any other languages
Time window	January 2018 to July 2022	Published before 2018
Study	Original research	Review, Survey, Commentaries, Opinion paper, Abstract
Data Augmentation	Used	Not used
Full text	Available	Unavailable
Modality	CT, XR, MR, PET, US	Dermatology, Ophthalmology, Histopathology
Context	Deep Learning	Any other type of machine learning
Downstream task	Classification, Detection, Segmentation	Any other tasks

Table B.8: Inclusion and exclusion criteria.

Column Name	Description	Possible Values
Source(*)	Digital library which the study came from	ACM, IEEE, PubMed, ScienceDirect
Pubblication Type(*)	Type of publication	Conference paper, Journal article
Year(*)	Year of publication	-
Author(*)	List of authors	-
Title(*)	Title of the study	-
Abstract(*)	Abstract of the study	-
Language(*)	Language of the study	English
Exclusion Reason (1)	Reason of exclusion in selection process	No D.A., Out of scope, No Medical domain
Deep Learning architecture	Deep Learning architecture used in the study	-
Task	Task object of experiment	Classification, Detection, Segmentation
Data-set size	Number of images in original dataset	-
Modality	Imaging modalities	CT, MR, XR, US, PET
Organ	Organs involved in the study	-
Pathology	Pathologies involved in the study	-
D.A. Objective	Purpose of Data Augmentation implementation	Balance Dataset, Random
Train/Test	D.A. applied to train or test phase	Train, Test
D.A. Type	Data augmentation techniques used in the study	-
D.A. Details	Brief description of the D.A. used	-
Paper Type	Type of paper according to our division in the next chapter	1, 2, 3, 4
Results	All the results obtained in relative task.	-
Exclusion Reason (2)	Reason of exclusion in extraction process	No D.A., Out of scope, No Medical domain, Missing text

Table B.9: Description and possible values of data extracted from the papers

Appendix C. Supplementary material: results

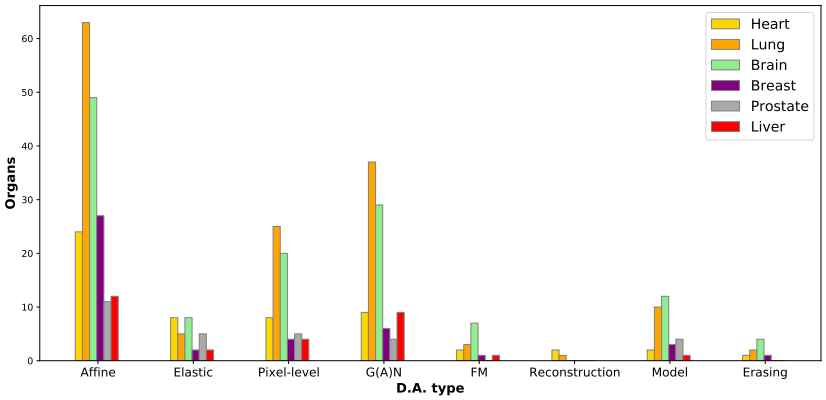


Figure C.9: Number of different data augmentation techniques for each organ

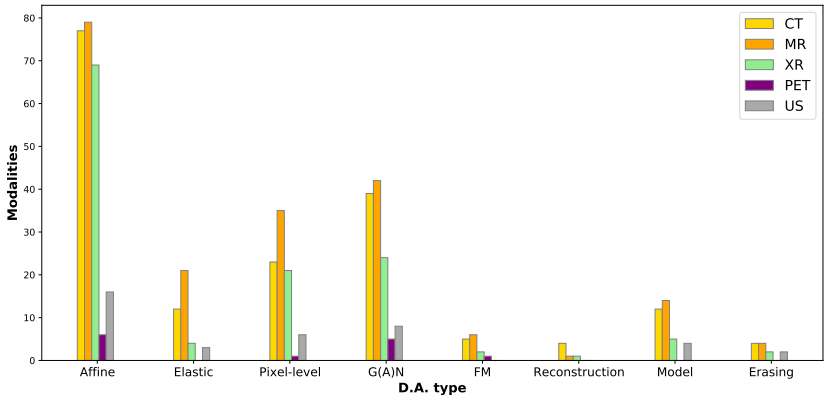


Figure C.10: Number of different modality for each data augmentation technique

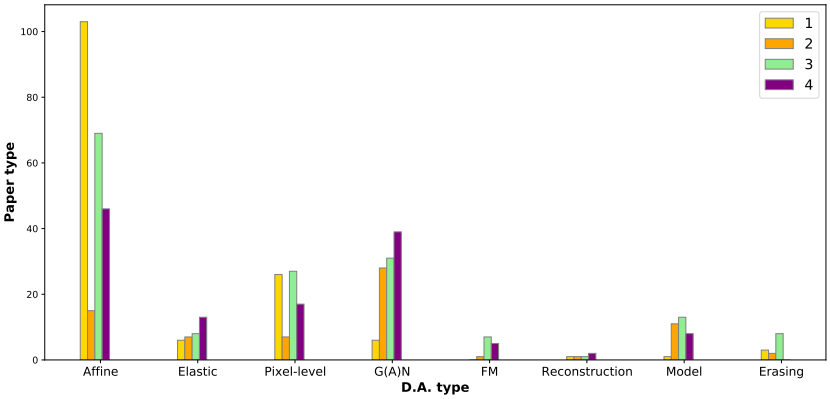


Figure C.11: Number of different data augmentation techniques for each type of paper

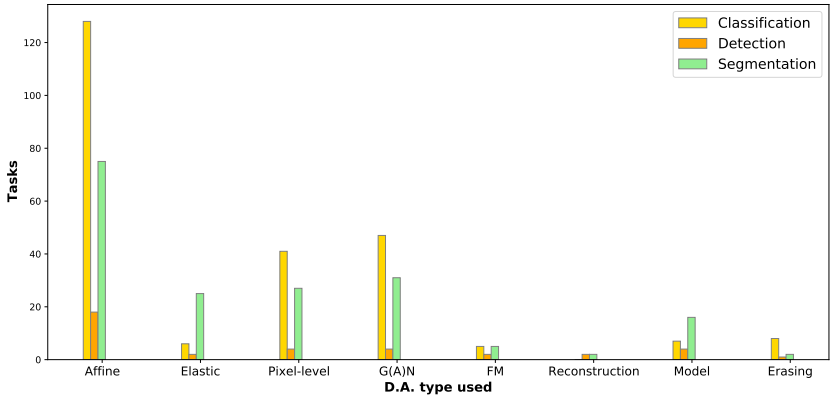


Figure C.12: Number of different data augmentation techniques for each task