POLITECNICO DI TORINO

Master's Degree
in Data Science And Engineering

Academic Year 2020-2021

Master's Degree Thesis

Data augmentation for medical image analysis: a Systematic Literature Review

Supervivors
prof. Fabrizio LAMBERTI
prof. Lia MORRA

Candidate
Alessio SERRA

December 2021

To my family,
to my friends,
to everyone who always
believed in me.

Summary

Recent advances in deep learning models have been largely attributed to the quantity and diversity of data gathered in recent years. Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks for image analysis, and in more recent years adaptive techniques, such as GAN-based or Model-based approaches, have been proposed to increase the effectiveness of data augmentation strategies.

Different data augmentation strategies are likely to perform differently depending on the type of input and visual task. For this reason, it is conceivable that medical imaging may require specific augmentation strategies that produce plausible data samples and allow effective regularization of deep learning model.

Data augmentation may also be used to enhance specific classes that are underrepresented in the training set, e.g., to generate artificial lesion samples. The goal of this thesis is to do an extensive systematic literature review in order to answer the following research questions: (i) which study designs are used to evaluate the effect of data augmentation, (ii) what types of data augmentation are used in the medical domain, (iii) what are their effects on the performance of deep learning-based methods for medical image analysis, (iv) what types of data augmentation are not adopted in the medical domain?

To answer these research questions, 273 papers published in reputable venues have been retrieved and analyzed to highlight trends in recent literature.

The key findings of the systematic literature review have been complemented by practical experiments on a chest x-ray data set, particularly to explore lesser-used data augmentation techniques in the literature.

Acknowledgements

I would like to express my gratitude to my supervisors, prof. Lia Morra and prof. Fabrizio Lamberti, who assisted and guided me throughout this thesis. It has been a long and tough journey and it has been a privilege to do it together.

I would like to thank my 3 families: my blood family, my Priolo's family, and my Turin's family. Without everything we have been through over these years, I probably would not be the person I am today. Thus, simply thank you.

Lastly, I would like to thank my parents, Loriana and Umberto, for EVERYTHING they have done for me since that Monday, February 24, 1997. You have allowed all my dreams to come true. I'll be lucky if I become even half of the parents you are.
Thank you with all my heart.

Contents

List of Tables

List of Figures

Sic Parvis Magna
Sir Francis Drake, 1587

Chapter 1

Introduction

1.1 Deep Learning in Medical Domain

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain, called artificial neural networks. It has established the state-of-the-art in many areas of computer vision and pattern recognition including medical imaging and medical image analysis. In order to successfully build well-generalized deep models, we need a large data set to avoid overfitting of such large-capacity learners. It has become a significant obstacle which makes deep neural networks quite challenging to apply in the medical image analysis field where acquiring high-quality ground-truth data requires time, money and human resources.

Moreover, the majority of manually annotated data sets are imbalanced, with some specific classes that are often underrepresented. To mitigate the problem of limited medical training sets, data augmentation techniques, which generate synthetic training examples, are being actively developed in the literature.

Nowadays, there are so many data augmentation techniques and strategies that is not trivial anymore to understand which is the best choice to reach the best results. Another important aspect to evaluate is the medical consistency, because every synthetic training example has to respect the human anatomy to not fall in a wrong evaluation that could be dangerous for the patients.

1.2 Data Augmentation

Data augmentation is a strategy that enables practitioners to significantly increase the diversity and the size of data available for training models, without actually collecting new data, using different image manipulation technique techniques or creating new ones that employ other models. By way of illustration, some techniques are showed in Figure 1.1.

Data augmentation can be performed both during the training phase, in order to reduce the class imbalance, reduce the *overfitting* or reduce the data-set bias, both during the testing phase in order to increase the accuracy or to measure the uncertainty of the model. In addition, data augmentation can generate new images from the original data set or can generate new images from scratch using machine learning.

Another important property of data augmentation is that it promotes learning invariance to transformations that should not affect the outcome. Data augmentation provides a way to exploit invariance without designing the architecture in order to be equivariant or invariant [23].

Therefore, Data Augmentation can improve the performance of their models and expand limited datasets to take advantage of the capabilities of big data.

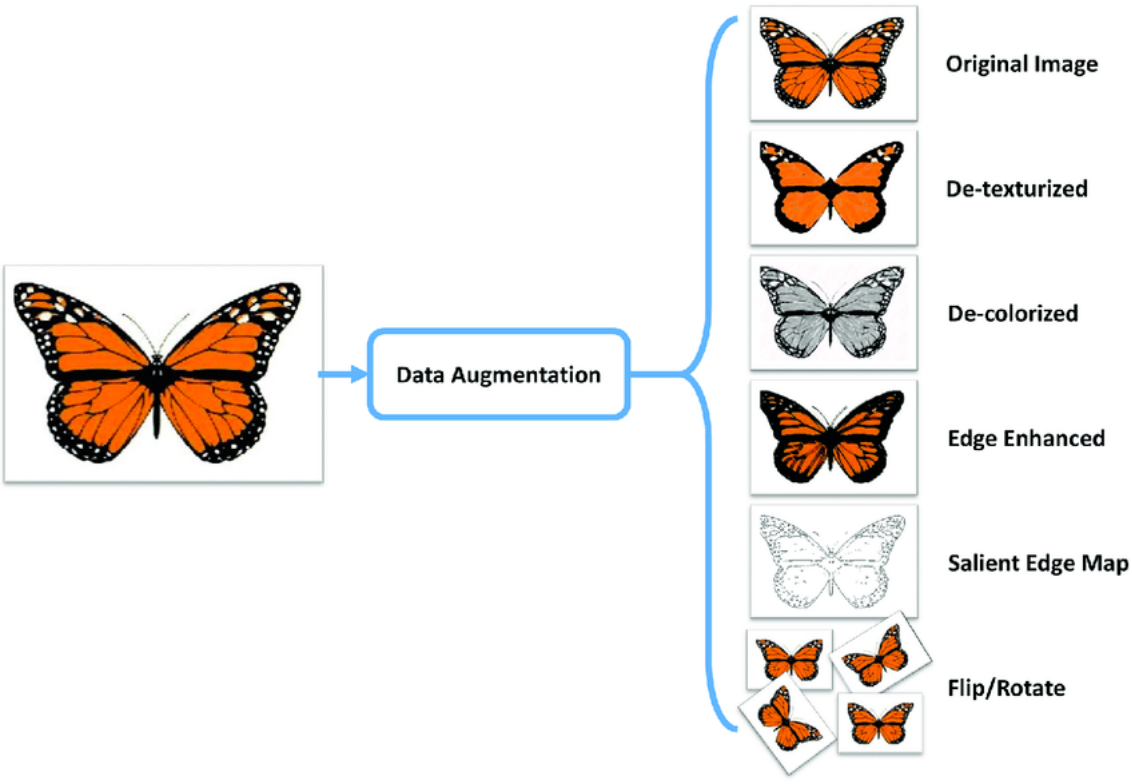


Figure 1.1. Different Data Augmentation techniques (Image credit: [5])

1.3 Systematic Literature Review's motivation

Different data augmentation strategies are likely to perform differently depending on the type of input, visual task and research field. In the literature there are so many reviews and surveys that analyze different data augmentation techniques in the most disparate fields, such as [117] which present existing methods for data augmentation, promising developments, and meta-level decisions for implementing data augmentation across the entire image domain.

There are also studies that analyze only narrow and too specific medical fields, for example [83] which highlights the most promising research directions to follow in order to synthesize high-quality artificial brain-tumor examples which can boost the generalization abilities of deep models.

It is conceivable that medical imaging may require *specific* augmentation strategies that produce plausible data samples and allow effective regularization of deep learning models, since the biases distancing the training data from the testing data are generally more complex.

Thus, there are no studies that provide a general overview of the application of data augmentation for medical imaging analysis . For this reason, I have decided to write this systematic literature review.

In this systematic literature review, I explore the data augmentation world for medical image analysis to identify, analyze, and summarize the best techniques, strategies and ideas to apply in the different medical tasks, such as classification, detection and segmentation, for each pathology, organ or data modality.

Chapter 2

Data augmentation techniques

2.1 Data Augmentation Taxonomy

Since there are different data augmentation techniques, we proposed the following taxonomy, which is adapted from [117] to account for methodologies that are uniquely tailored to the medical imaging domain. At the upper level, the two main groups are *transformation of original data* and *generation of artificial data*. Subsequently, in the second level, the first group is divided into *affine*, *erasing*, *elastic* and *pixel-level* transformations and the second group is divided into *GAN-based*, *feature mixing*, *model-based* and *reconstruction-based* transformations.
The complete taxonomy is shown in Figure 2.1 and each group and transformation is described in the next sections.

2.2 Transformation of original data

The transformations belonging to this group perform image manipulation on the original data set. The affine transformations like cropping, flipping, or translation are the easiest, simplest and most popular types of data augmentation transformations, there are several libraries that allow you to perform them by writing a few lines of code, including the famous TensorFlow and Keras. For these reasons, it is also tried out in the majority of the papers, both for finding new combinations of data augmentation and for comparison with new generation of artificial data-based method.

Other transformations belonging to this group are pixel-level, erasing and elastic techniques. They could be more complex to implement with respect to the affine, but they offer a wider customization spectrum of action for different medical applications.

Get into the details, the different transformations of this group differ as follows:

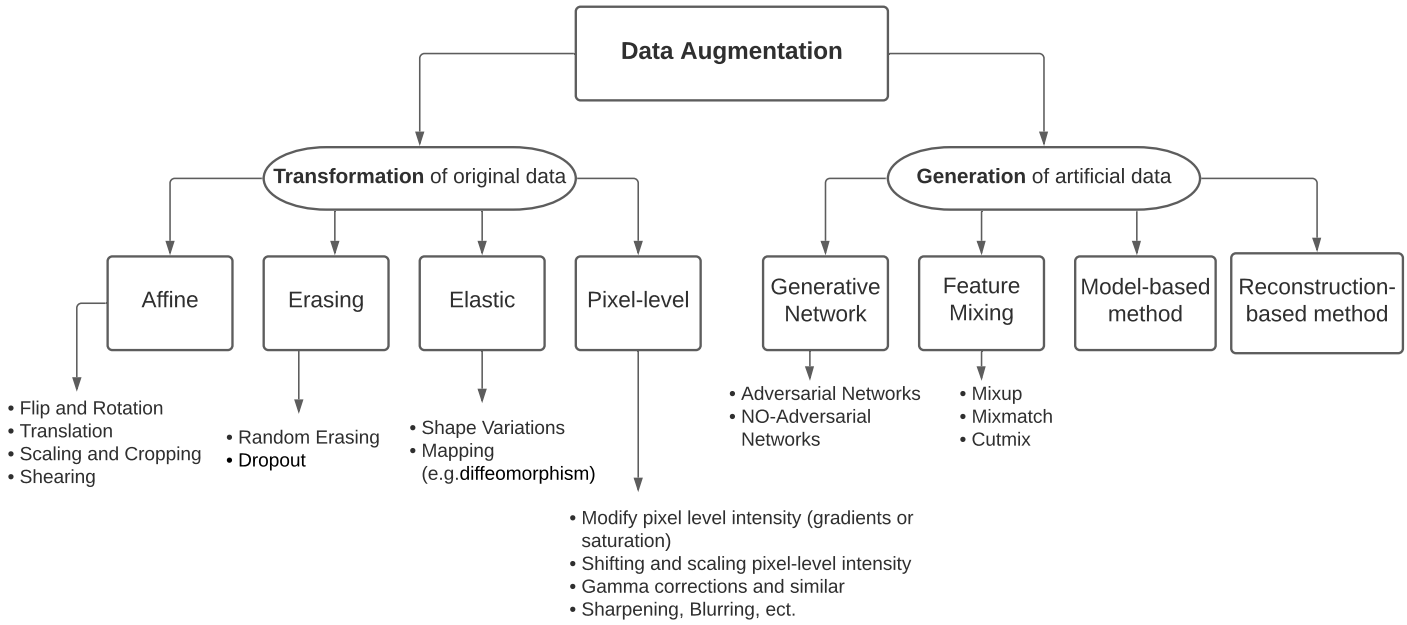


Figure 2.1. Data augmentation taxonomy

2.2.1 Affine transformation

Affine transformations refer to a class of geometrical transformations which preserve lines and parallelism, but not necessarily distances and angles. [ref: `https://mathworld.wolfram.com/AffineTransformation.html`]

This kind of preservation is possible because the transformations take place respecting one or more axes of symmetry within the image. The transformation along an axis is typically expressed through a matrix that contains the coordinates along the abscissa and ordinates or the angles of rotation.
Note that, in the medical domain, there are 2-dimensional or 3-dimensional images. For example, the mathematical 2-dimensional transformations are the following:

- **Translation**: a function that moves every point with a constant distance in a specified direction. In the transformation matrix, t_x specifies the shift along the x axis and t_y specifies the shift along the y axis.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mathrm{t}_x & \mathrm{t}_y & 1 \end{bmatrix}$$

- **Rotation and Flipping**: circular transformation around a point or an axis. The value q in the transformation matrix specifies the angle of rotation in the range $[1°, 359°]$. Note that when apply a rotation with $q = 90°$ or with $q = 180°$ this technique is known as **flipping**, respectively *horizontal flipping* and *vertical flipping*.

$$\begin{bmatrix} \cos(q) & \sin(q) & 0 \\ -\sin(q) & \cos(q) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- **Scaling**: a linear transformation that enlarges or shrinks the image by a scale factor which is the same in all directions. In the transformation matrix, s_x specifies the scale factor along the x axis and s_y specifies the scale factor along the y axis.

$$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

It is important to note that since the transformations are defined by matrices, it is possible to combine the matrices to obtain combined affine transformations.
Example of this kind of transformations are showed in Figure 2.2.

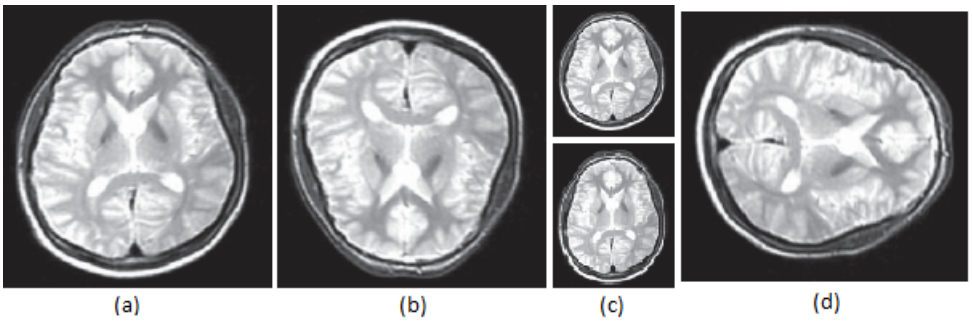


Figure 2.2. Example of affine transformations applied to a brain magnetic resonance image. Respectively: (a) original image (b) vertical flipping (c) scaling (d) horizontal flipping

2.2.2 Erasing transformation

Erasing transformation selects a region in an image and erases its pixels by substituting their intensity with other values. Generally, the region erased is a rectangle or a circle and it is chosen randomly to better generalize the training. The erasing values could be expressed with a single value of the gray scale or with a tuple of length 3, that is used to erase R, G, B channels respectively, or with random values.
This type of transformation could seem very counterintuitive for data augmentation, but in a certain task, such as *face recognition*, it is very helpful because the model is forced to not create simplified detection patterns.
Example of erasing transformation is showed in Figure 2.3.

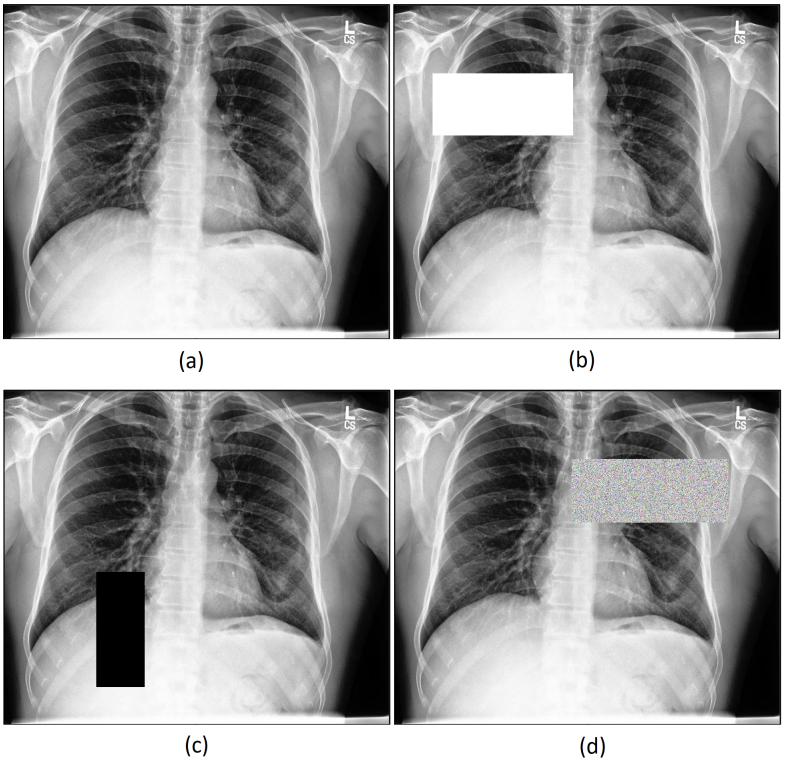


Figure 2.3. Example of random erasing transformation applied to chest x-ray image from CheXpert [53]. Respectively: (a) original image (b) white erasing value (c) black erasing value (d) random erasing value

2.2.3 Elastic transformation

Elastic transformations distort the image applying shape variations or some type of mapping. They can bring lots of noise and damage into the training set if the deformation field is seriously varied, because as opposed to affine transformations, there is no preservation of col-linearity or ratio of distances. For this reason, it is important to distort the images with appropriate values to not de-contextualize the subject.
To perform this augmentation, it has to apply a *diffeomorphism*. A diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse [83].

Mathematically, the *diffeomorphism* ϕ (also referred to as a *diffeomorphic mapping*) is given in the spatial domain Ω of a source image I, and transforms I to the target image

$$J : I \cup \phi^1(x, 1)$$

The mapping is the solution of the differential equation:

$$\frac{d\phi(x,t)}{dt} = v(\phi(x, t), t)$$

where $\phi(x,0) = x$, v is a time-dependent smooth velocity field, $v : \Omega \times t \to R^d$, $\phi(x,t)$ is a geodesic path(d denotes the dimensionality of the spatial domain Ω), and $\phi(x,t) : \Omega \times t \to \Omega$. In practice, a grid is created on an image mapped on each pixel, then this grid is distorted through various deformations, finally each pixel is re-positioned on the new grid following the old mapping in order to generate the elastically deformed image.

This type of transformation can also help to reconstruct nonoptimal situations typical of the medical domain. For example, considering the fact that when medical researchers capture medical images from patients, movements like breathing may lead to image deformation, the use of image deformation could help to resolve this problem, just like in [137] or [105].

Example of elastic transformation is shown in Figure 2.4.

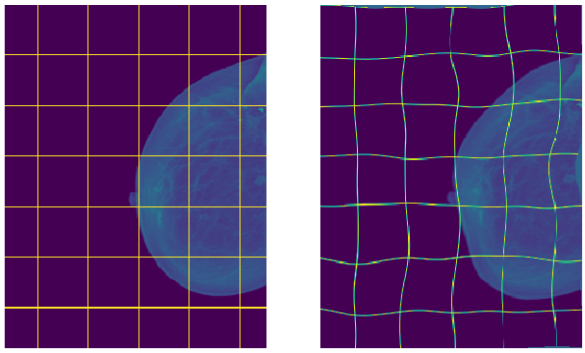


Figure 2.4. Example of elastic transformation applied to a mammography. A Gaussian filter is applied after the distortion. (Image credit: [17])

2.2.4 Pixel-level transformation

Pixel-level transformations change the pixel values in order to modify pixel-level intensity, gamma correction, blurring, etc. There are several ways of changing pixel-level to have different results, such as increasing or decreasing the brightness, contrast, or saturation of an image. In the medical domain, in most cases, we work with gray scale images, so it is quite useless to perform a color normalization. Due to this abundance of options, it is common to apply a combination of pixel-level transformation according to the final task of the model.
Example of pixel-level transformations are showed in Figure 2.5.

2.3 Generation of artificial data

The transformations belonging to this group create new synthetic images using deep learning methods or other pattern-based methods. The implementation of these methods is

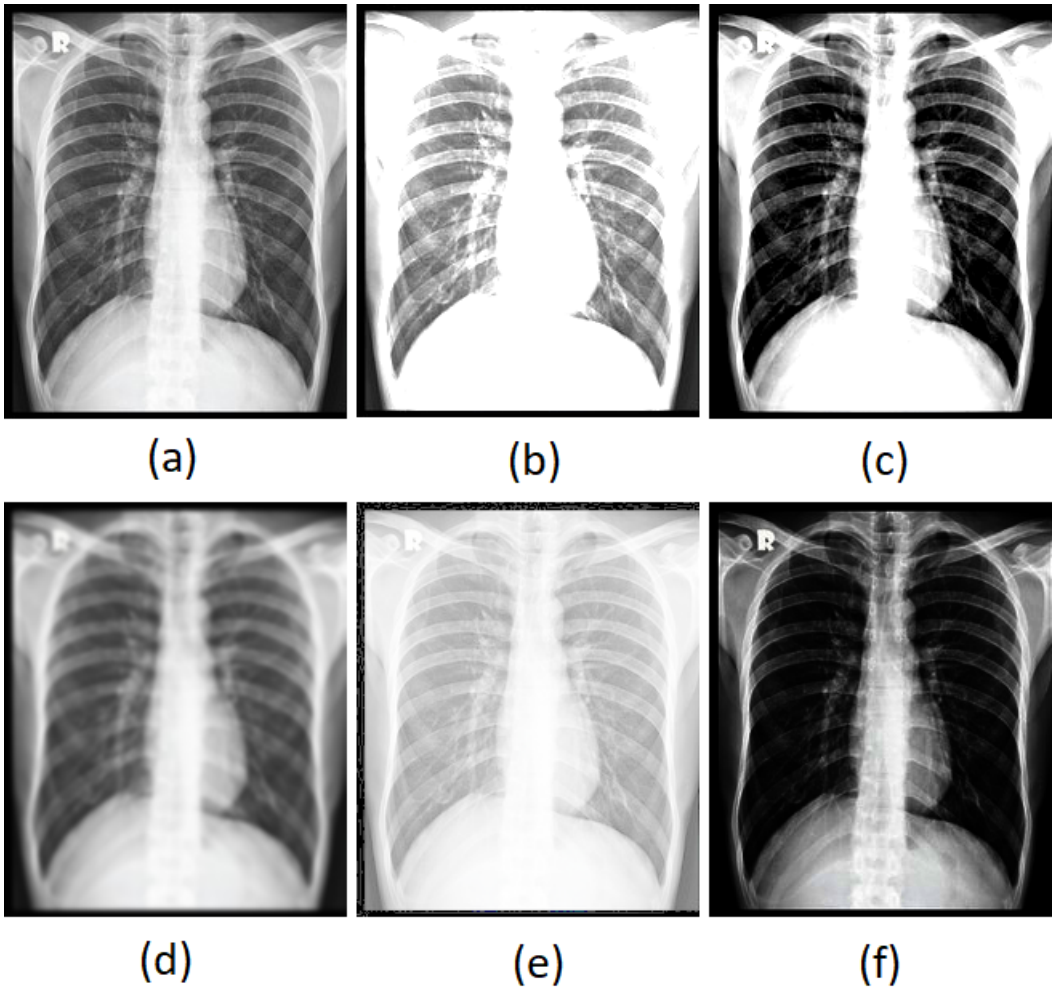


Figure 2.5. Example of pixel-level transformations applied to a chest x-ray image. Respectively: (a) original image (b) brightness (c) contrast (d) blur (e) light gamma correction (f) dark gamma correction

computationally heavier and not trivial with respect to the first group but the possibilities of approach are wider and richer in terms of developing new strategies.

2.3.1 Generative Networks

Generative Networks are a fast-growing area in deep neural networks that can be used to generate realistic images, speech, prose, and more. They can be *adversarial* or *not adversarial* with respect to presence or not of the model discriminator.
The most widespread are the Generative Adversarial Networks (GANs), introduced by [41]. The architecture of GANs, illustrated in Figure 2.6, is composed by two networks:

- **Generator**: creates a new image that tries to trick the discriminator into believing the fake data is authentic;
- **Discriminator**: evaluates samples passed from the generator and attempts to discern if the image belongs to the training dataset, meaning it's authentic, or if it was generated, meaning it's fake.

When used for image generation, the generator is typically a *deconvolutional neural network*, and the discriminator is a *convolutional neural network*.
Thus, the two neural networks contest with each other in a game (in the form of a *zero-sum game*, where one net's gain is another net's loss).

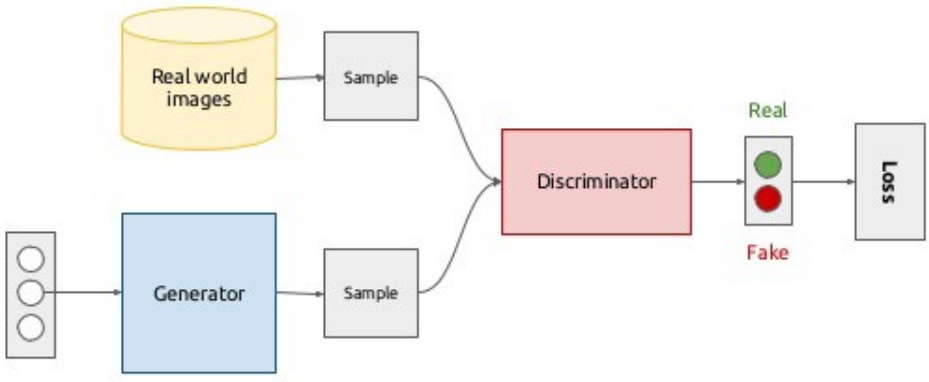


Figure 2.6. Generative Adversarial Network architecture
(Image credit: `https://becominghuman.ai/`)

The mode of operation is quite simple: the generator creates a fake sample, which is fed into the discriminator alongside a real sample taken from the real training dataset. At this point, the discriminator returns a prediction with a probability from $[0, 1]$ where 0 indicates the fake one and 1 indicates the authentic one.
Examples of synthetic images generated by GAN are shown in Figure 2.7.

There are so many improvements and changes to the GAN architecture based on own goals and tasks, starting from the the architecture of generator and discriminator to the different kind of losses. Examples of this type are:

- **Conditional GAN** [79]: it is a generative adversarial network whose generator and discriminator are conditioned during training by using some additional information. This auxiliary information could be, in theory, anything, such as a class label, a set of tags, or even a written description. In a Conditional GAN, the discriminator does not learn to identify which class is which. It learns only to accept real, matching pairs while rejecting pairs that are mismatched and pairs in which the example

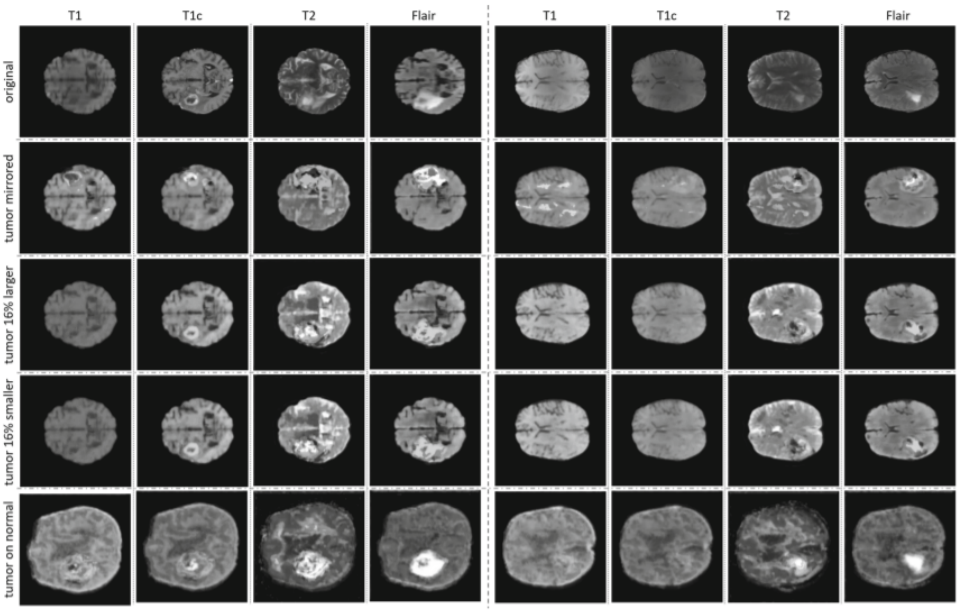


Figure 2.7. Example of GAN generated images. The first row depicts the real images on which the synthetic tumors were based. Generated images without adjustment of the segmentation label are shown in the second row. Examples of generated images with various adjustments to the tumor segmentation label 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: [116])

is fake. Thus, Conditional GAN can generate artificial samples of a specific category to improve the accuracy of the deep learning classifier when the sample size is insufficient.

- **CycleGAN** [148]: it is one of the most powerful and complex GAN architectures. It is fundamentally hallucinating part of the content it creates. Its outputs are predictions of *"what might it look like if"* and the predictions, thought plausible, may largely differ from the ground truth. CycleGAN should only be used with great care and calibration in domains where critical decisions are to be taken based on its output. This is especially true in medical applications, such as translating MRI to CT data. Indeed, CycleGAN has been widely used for cross-domain medical image synthesis tasks particularly due to its ability to deal with unpaired data. (`https://junyanz.github.io/CycleGAN/`)

A more complete in-depth analysis regarding the various GAN data augmentation techniques and architectures is given by [93] in their review.

2.3.2 Feature mixing method

Feature mixing method combines 2 or more images from the original dataset to create a new one. Mixing images together by averaging their pixel values or cut and paste some part of one image in another are very counter-intuitive approaches because the images produced by doing this will not look like a useful transformation to a human observer [117]. They are very useful for generalizing deep learning models, especially when we are not sure about selecting a set of augmentation transforms for a given dataset, medical imaging datasets, for example. Feature mixing can be extended to a variety of data modalities such as computer vision, natural-language processing, speech, and so on.

One of the most used feature mixing method is called **mix-up**. This data augmentation technique can be mathematically expressed with the following formulas:

$$\widetilde{x} = \lambda x_i + (1 - \lambda)x_j \text{ , where } x_i \text{ and } x_j \text{ are raw input vectors}$$
$$\widetilde{y} = \lambda y_i + (1 - \lambda)y_j \text{ , where } y_i \text{ and } y_j \text{ are one-hot label encodings}$$

Note that the lambda values are values with the [0,1] range and are sampled from the Beta distribution. (`https://keras.io/examples/vision/mixup/`)

Examples of mix-up data augmentation are shown in Figure 2.8.

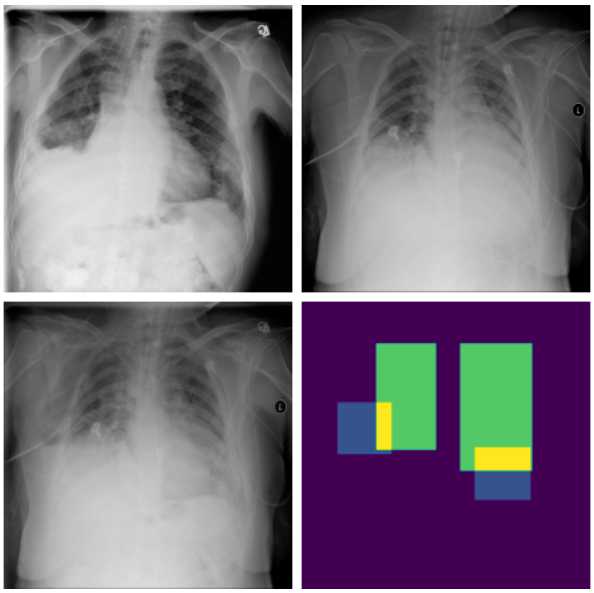


Figure 2.8. Example of mix-up generated images. In the first row there original chest x-rays, in the bottom left there is the result of mix-up augmentation and in the bottom right the corresponding mask.
(Image credit: [112])

2.3.3 Model-based methods

Under the umbrella of model-based methods, we group a variety of techniques that are based on physically or biologically inspired models to generate new images or modifying existing ones. These transformations are sometimes implemented using deep learning techniques (not necessarily generative models) or traditional image processing techniques, such as shape modelling/deformation or image blending.

Since medical datasets are often very skewed towards normal cases, the most common model-based techniques involve the synthesis and injection of artificial lesions in otherwise healthy subjects. This methods has been applied to simulate multiple sclerosis lesions in brain MR images [108] or to add cancer signs to breast mammography images [18]. An example of such model-based data augmentation is shown in Figure 2.9.

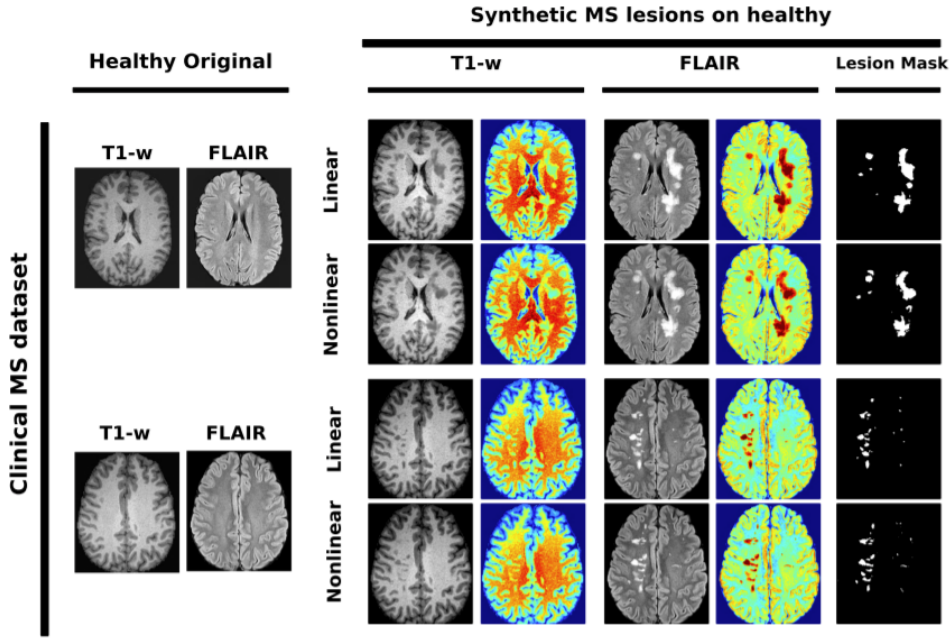


Figure 2.9. Example of model-based method: synthetic multiple sclerosis lesions generated on a healthy subject. Slices are also displayed using jet color maps to visually enhance the intensities
(Image credit: [108])

Another class of model-based techniques is concerned with the simulation of physiological time progression, such as reproducing the effect of aging on organs, or simulating disease progression and regression [73].

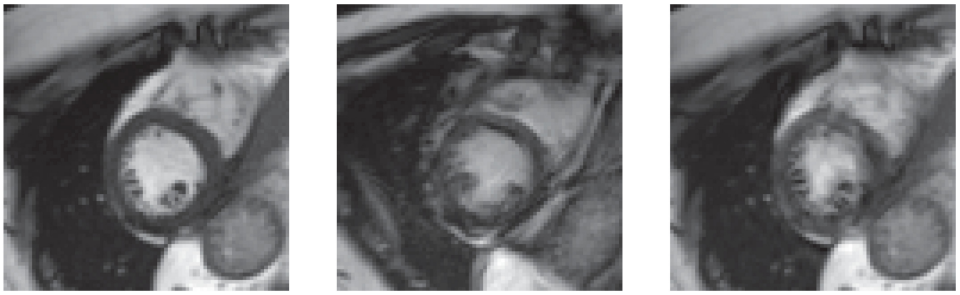
Finally, some model-based techniques exploit prior knowledge about the human anatomy and/or the image formation process to generate plausible variants of existing images, or

simulate common acquisition artifacts to increase the robustness of image analysis algorithms. For instance, in [123] statistical shape models are created for different organs (e.g., prostate and liver); then, the shape models are deformed and, after obtaining the new shape, a realistic texture is generated via image warping and interpolation. This allows to generate random variations representative of inter-subject variability. Unlike GAN-based techniques, these techniques typically require minimal or no training, and directly incorporate anatomical constraints on the generated images; however, they only be applied to specific organs for which mathematical models are available.

2.3.4 Reconstruction-based method

Three-dimensional medical imaging modalities are reconstructed using a mathematical process that transforms the raw data collected by the image scanner into a (usually three-dimensional volume) that can be viewed by the radiologist. For instance, image reconstruction in CT involves generating tomographic images from X-ray projections acquired at many different angles.

The reconstruction algorithms and its 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.



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

Figure 2.10. Example of reconstruction-based method: a good quality cine CMR image (a), an image with blurring motion artefacts (b), and a k-space corrupted image (c). The k-space corruption process is able to simulate realistic motion-related artefacts (Image credit: [89])

For instance, in [89] 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 related set of techniques involve simulating 2D images (e.g., X-rays) from 3D volumes (e.g, CT scans), in order to simulate different acquisition angles and views [15].

A visual example of model-based data augmentation is shown in Figure 2.10.

2.4 Learnable data augmentation

Learnable augmentation is a new and very active subfield of deep learning research that has the goal of reducing the human labor inherent in selecting, designing, and validating data augmentation [68].

This type of approach proposes techniques that overlap the data augmentation techniques shown in the taxonomy of Figure 2.1, but it automates the choice for the best data augmentation technique to use. They are the following:

- **Learning to Augment Data With a Neural Network**: this augmentation uses 2 types of networks: the first learns the desired task (Network B) and second learns to perform the augmentation for the network that learns the desired task (Network A, i.e. the augmenters). Network A takes two or more randomly selected images as an input and learns to utilize the information in them to generate a single third image with information from a target classifier and a random target image [69].
- **Augmentation Policy Learning**: this approach, also called *autoaugment*, take a list of known augmentation policies for a specific dataset (such as affine, pixel-level, etc.), and use a machine learning technique to automatically learn policies that result in a better model [27].
- **Evolutionary Image Augmentation**: it uses automatic construction of tree-structural image transformation is used to create augmented images that are shown to improve the results on various two-class problems [35].
- **Approaches Based on Classic Statistical Models**: in this idea, augmentations are a type of deformation. A parametric model is used for generating transformations of input images. After the model has been trained to generate images similar to those in the dataset, generated images are sampled randomly from the parametric model, allowing for a large number of realistic transformations [48].
- **Approaches Based on GANs**: GANs learn the distribution of the training set by competing against a discriminator and this competitive process alone determines the type of data augmentation that the GAN generates [58].

2.5 Training-Time & Test-Time Data Augmentation

In addition to using different data augmentation techniques, you can choose **when** to use these techniques. In particular, it is possible perform data augmentation during the *training phase* or during the test phase:

- **Training-time data augmentation**: it is the classic method, where the dataset size increases to better generalize model learning. Most of the time this approach is chosen.

- **Test-time data augmentation**: it can be seen as analogous to ensemble learning techniques in the data space. This method takes a test image and augments it in the same way as the training image. This comes at a computational cost depending on the augmentations performed, and it can restrict the speed of the model. After the augmentation, the final result is chosen performing a majority vote on the various transformations applied to the test image. This could be a very costly bottleneck in models that require real-time prediction, but it could bring more robust results [117].

Despite test-time augmentation limitations, it is a promising practice for applications such as medical image diagnosis, because its impact is a useful mechanism for measuring the robustness of a classifier, limiting the number of classification error, and in particular the number of false positives/negatives.

Chapter 3

Systematic Literature Review: Methods

3.1 Research questions

I performed this review according to the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, `prisma-statement.org`). This method guarantees that the outcome of the review is verifiable and repeatable by other researches. The protocol (which is graphically illustrated in Figure 3.1), was jointly developed by the authors.

The purpose of this systematic literature review is to identify, analyze, and summarize the studies that use data augmentation techniques in medical domain in order to compare current applications of data augmentation in the medical domain, understand its potential impact on the performance of deep learning models and identify current gaps in knowledge, also by comparing with the corresponding literature in the non-medical domain. For these reasons, we defined 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?

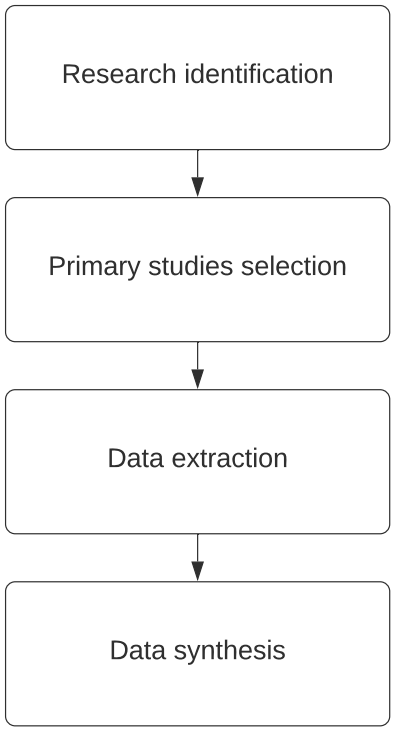


Figure 3.1. Systematic Literature Review Protocol

3.2 Search strings and Study selection

We selected four scientific digital libraries that contain primary studies related to the field of computer science, medical imaging and biomedical engineering. Other more general sources, like Google Scholar, were not included because they usually index studies already available in the primary sources. The search strings for each scientific digital library are shown in Table 3.1. Note that for each search, the studies are filtered by date: we consider only studies published between January 2018 and June 2021, which mostly represent the state of the art, since this is a rapid evolution research field.

The selection process was divided in different stages. First, I defined inclusion and exclusion criteria related to the language, time window, type of study, modality, context, and task. I select only English primary studies from 2018 to 2021, including both proceedings and journal papers, but excluded preprints. I also excluded dermatology, ophthalmology and histopathology papers (in general, I excluded any article that not uses x-ray (XR), magnetic resonance (MR), computed tomography (CT), positron emission tomography

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 3.1. Digital libraries with respective URL and search string.

(PET) and ultrasound scan (US) as dataset). I excluded studies that not use data augmentation or of which the full text is unavailable. The only downstream tasks allowed are classification, detection and segmentation. Moreover, I selected only selected papers that use deep learning models to perform their downstream tasks, excluding any other machine learning techniques that are not artificial neural network-based. This exclusion is necessary to ensure some homogeneity in the studies included in the Systematic Literature Review.
Inclusion and exclusion criteria are summarized in Table 3.2. Then, we checked the titles and the authors in order to discover possible duplicates. Furthermore, I apply the criteria only to the title and the abstract of each paper to perform a first preliminary exclusion.

Lastly, we read all the remaining papers applying the criteria to the full text. At the end of this stage, we perform the last round of exclusions obtaining the final set of paper that will be involved in the Systematic Literature Review.

	Inclusion criteria	Exclusion criteria
Language	English	Any other languages
Time window	from 2018 to 2021	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 3.2. Inclusion and exclusion criteria.

For each paper, the exclusion criterion was recorded and classified as follows:

- **No D.A.**: papers discarded because they do not use existing data augmentation methods,
- **Out of Scope**: papers discarded because they are out of modalities, tasks or type o study,

- **No Medical Domain**: papers discarded because they are out of medical domain,
- **Missing text**: papers discarded because their full text is unavailable

3.3 Data extraction

After the selection process, the last step is to extract the data that are of interest to us for the purposes of Systematic Literature Review. To give an answer to the **first research question**, it is necessary to define a classification for the analyzed papers. Thus, we defined the following categories (N.B. a paper can belong to more than one category):

- **Type 1**: papers that use existing data augmentation methods in the pipeline but do not assess their impact on the performance on the downstream task
- **Type 2**: papers that propose new data augmentation methods tailored to the medical imaging domain, but do not necessarily validate its impact on the performance of downstream tasks
- **Type 3**: papers that compare performance obtained on one or more downstream tasks with and without applying data augmentation
- **Type 4**: papers that compare performance obtained on one or more downstream tasks with different data augmentation techniques

Since a paper can belong to more than one category, we have assigned it the type with the largest number as it extends the useful features of the papers underneath it.

Subsequently, to answer to the **second research question** I only consider the type 2, type 3 and type 4 papers, instead to answer to the **third research question**, I only consider the type 3 and type 4 papers. However, to answer both research questions, I extracted and grouped the information in order to analyze and comment the values in the next chapter. In particular, the columns extracted with relative description and possible values, are summarized in Table 3.3.

The value "-" as possible value indicates that the column can assume any string value. Lastly, note that the columns marked with (*) are the ones automatically extracted using the reference management software **Zotero** (`https://www.zotero.org/`), on the contrary, the remaining columns have been extracted manually after the full text reading by me, and verified and validated by the professor L. Morra.

N.B. In the table, (1) indicates the exclusion reason before the full text review, instead, (2) indicates the exclusion reason after the full text review. Besides, *D.A. Objective* column values could be *Balance Dataset* if the data augmentation strategy is applied in a targeted way to fix the class imbalance, or could be *Random* if the only purpose of data augmentation is to increase the size of dataset randomly.

3.4 Data synthesis

The extracted data are used to generate bar plots, tables, box plots, and scatter plots to clearly analyze the distribution of the extracted values.

Bar-plots help to visualize the distribution of type of paper, type of data augmentation technique, modalities, organs and and their possible correlations. They are generated using *matplotlib* python library.

Tables help to organize data in order to calculate simple statistics, such as percentage increments obtained from the several data augmentations. They are generated in the following way:

1. The articles will be grouped into 5 groups according to the organ studied: *brain*, *lung*, *heart*, *breast* and *others*, which contains different less studied organs.
2. Subsequently, they are sub-grouped by task
3. Lastly, they are sub-grouped by pathology
4. They are compared using homogeneous performance metrics (i.e., as similar as possible) wherever possible. Where more metrics were available, more robust metrics such as *Area under the ROC Curve* and *Dice Score* were prioritized.

Box plots show the percentile distribution of the statistics calculated in the tables, following different comparisons. In particular, comparisons will be made on DA versus w/o DA results for each tabulated organ and *only affine* versus *no-only-affine* transformations. They are generated using *seaborn* python library.

Scatter plot shows, using Cartesian coordinates, the relative performance increase values (on the ordinate axis) as a function of the original training set size(on the abscissa axis). In addition, each different data augmentation is indicated with a different marker and color. It is generated using *seaborn* python library.

Lastly, analyzing the data extracted or calculated in the previous sections and qualitative comparing them with the distributions of the most used of data augmentation techniques in the literature, I will be able to answer to the **fourth research question**.

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 3.3. Description and possible values of data extracted from the papers

Chapter 4

Systematic Literature Review: Results

4.1 Study selection result

The selection process was performed during July 2021. The preliminary set initially contained 484 papers. We checked their titles and authors in order to discover possible duplicates: after having removed 30 duplicated results, the preliminary set was reduced to 454 papers.

We first applied the inclusion and exclusion criteria [Table 3.2] only to title and abstract, obtaining a list of 266 papers. In particular, we discarded 32 studies that not belong to the medical domain, 22 studies that are reviews or surveys, 42 studies that do not use data augmentation, and 92 studies classified as *Out of Scope*.
This last grouping includes studies with different tasks, modalities or context with respect to the inclusion criteria.

We also added 7 papers from records identified through other sources (e.g., studies that were already known to the authors or cited by the retrieved papers). Then, we analyzed all the 273 papers in detail. At the end of the process, we discarded 67 papers and we finally selected 206 papers as part of this literature review.

The control flow diagram of the selection process is shown in Figure 4.1.

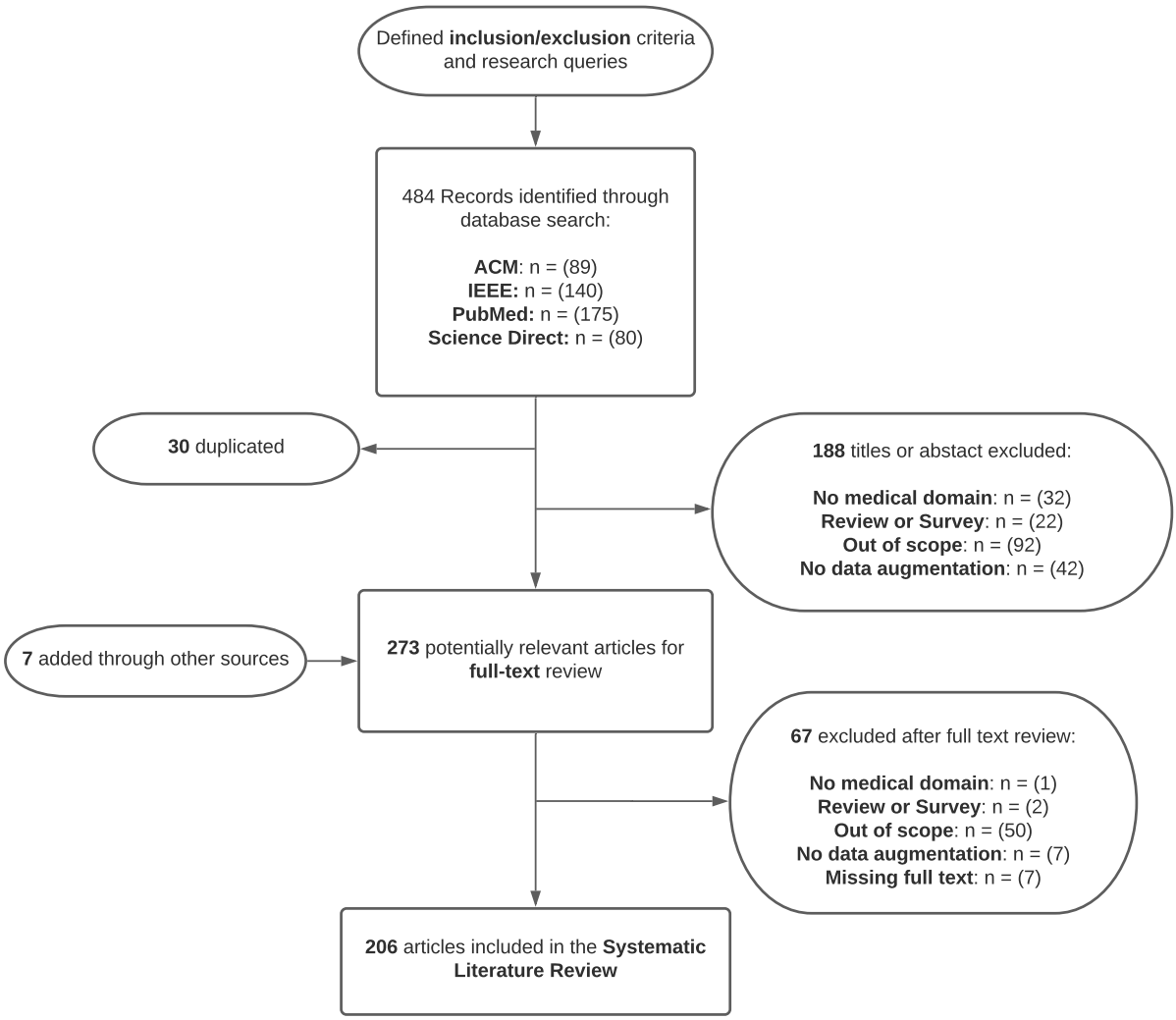


Figure 4.1. Systematic Literature Review Control Flow Diagram

4.2 RQ1 & RQ2: Types of Data Augmentation

Each paper has been assigned to a category according to the methods described in the Section 3.3. The distribution is shown in Figure 4.2.
It is immediate to note that more than half of the papers, about 62% (128/206), do a comparison among different methods, instead only 10% (22/206) propose a new method without evaluating their impact on one or more downstream tasks.

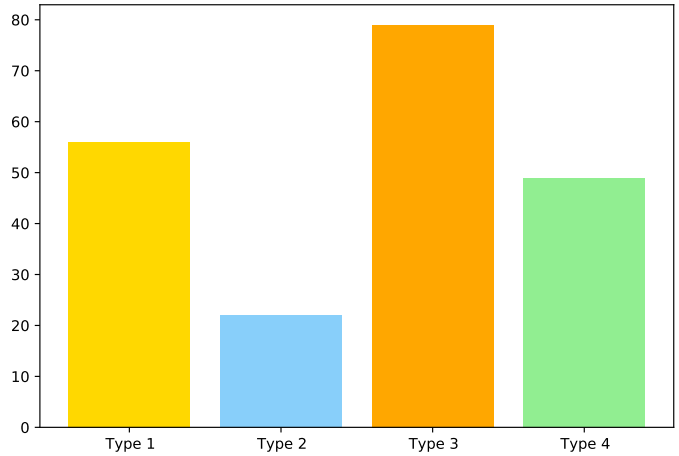


Figure 4.2. Number of papers for each type

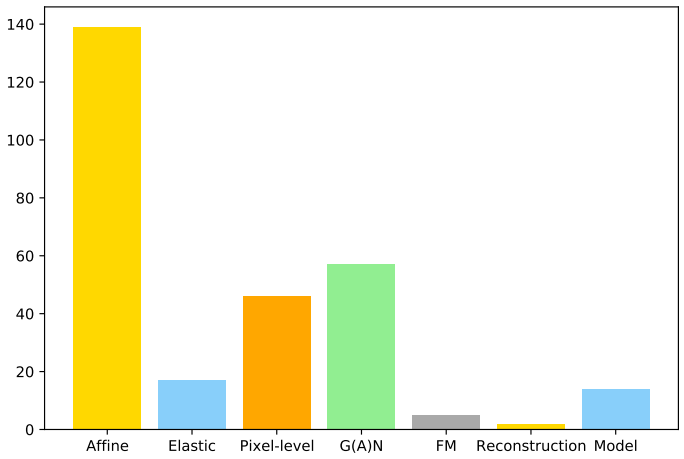


Figure 4.3. Number of papers for each class of D.A. transformation used.
Note that a single paper can uses more than one D.A. technique. In this case, every D.A. technique implemented is considered.

Figure 4.3 represent the distribution of papers for each D.A. transformation, Figure 4.4 represent the distribution of papers for each modality and Figure 4.5 represent the distribution of papers for each organ.

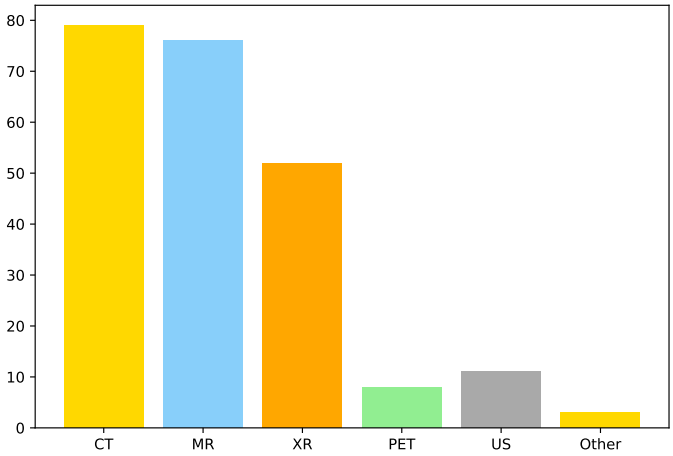


Figure 4.4. Number of papers for each modality

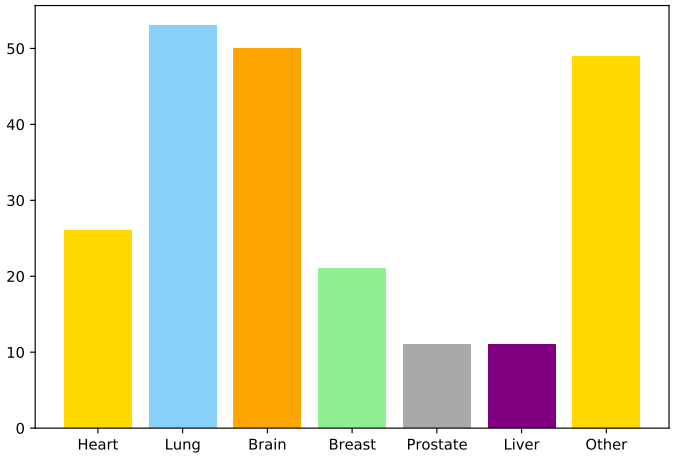


Figure 4.5. Number of papers for each organ

Looking at the histogram in Figure 4.3 it is easy to answer to the second research question. The most used type of data augmentation are affine transformations, used in 49.6% (139/280) of papers. This result was really expected, because the type of data augmentation is very easy to implement.

To follow, we have GAN-based and pixel-level transformations with both above 20.3%

(57/280) and 16.4% (46/280), respectively. I think that probably the latter technique is particularly quite widespread because it allows to be more robust to different vendors and acquisition protocols.

Model-based methods and elastic transformation are used in 11% of the papers (14/280 and 17/280 respectively) and the remaining methods are around 2.5% (7/280), which indicates that *reconstruction-based* method, used in [89] which implements Gaussian blurring corrupted data augmentation and [15] which creates X-ray images from the testing CT data set, and *feature mixing* transformation, used in [11] [74] [87] [120] [147] which use different mix-up techniques, still needs to be further investigated in the medical domain image analysis.

In Figure 4.4, we can see that CT, MR, and XR are the most used (and probably the most common) modalities. Particularly, under the *Other* category, there are 3 different modalities: *Coronary Angioscopy (CAS)* [80], *Myocardial perfusion imaging (MPI)* [10], *Digital breast tomosynthesis (DBT)* [145]. Besides, there are 3 papers that use a cross-modality approach: MR as the source and CT as the target [110]; CT as source and MR as target [20]; modality manipulation from CT to MR and the other way around [32]. Due to data privacy concerns and unavailability of medical annotators and due to difficult to obtain a lot of labeled medical images, cross-modality data augmentation could mitigate the data deficiency issue in the medical imaging domain.

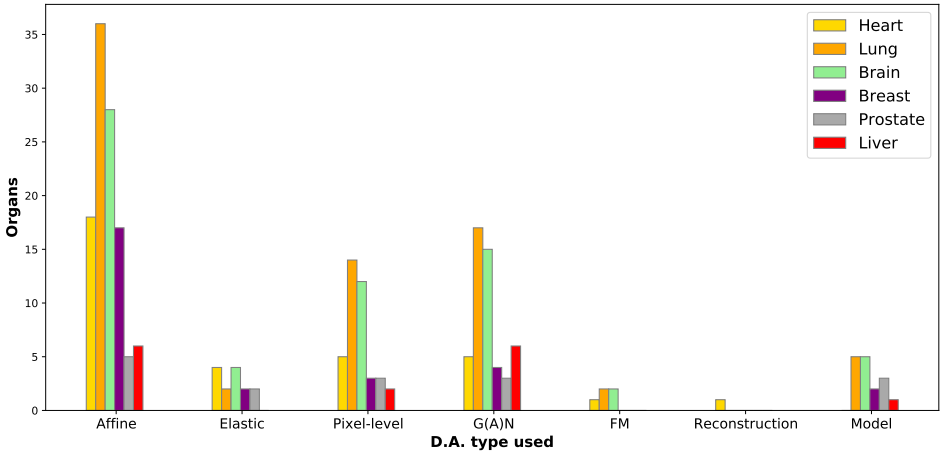


Figure 4.6. Number of different data augmentation techniques for each organ

In Figure 4.5, it is immediate to see that the most studied organs are lung and brain with 53 and 50 papers respectively. Note that *other* indicates different organs studied in less than 10 papers, which include the musculoskeletal system (arm, calcaneus, knee, hand, humerus, hip, maxillary, neck, pelvis, skull, shoulder, spine), the abdomen (colon, kidney, lymph node, pancreas, stomach) and espohagus, pituitary membrane, rectum,

teeth, thyroid and urinary conduct.

Moreover, we plot multiple bar plots in order to understand if there exists a type of correlation between D.A. technique used and (i) organ, (ii) modality, (iii) type of paper, (iv) downstream task.
Taking into consideration the Figure 4.6, it is immediate to note that affine transformation in the technique most used for all the different organs. Then, looking at Figure 4.7, we notice that CT and MR are the most used modalities predictable as they are the most widespread.

Lastly, in Figure 4.9, except for affine transformations, the majority of papers do comparison between no data augmentation results and different techniques of data augmentation.

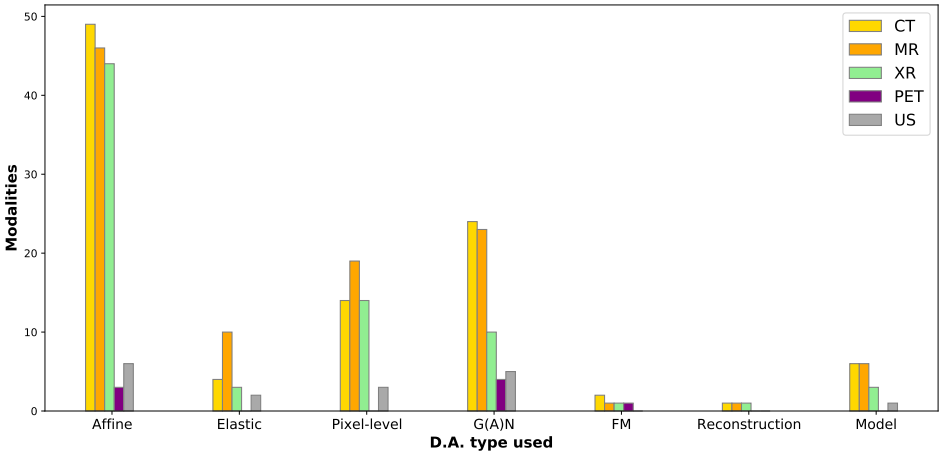


Figure 4.7. Number of different modality for each data augmentation technique

The type 1 papers were not included in the following research questions, because they neither experiment new data augmentation methods nor do any comparison with other results. In most cases, the type 1 papers include very standard and simple affine data augmentation techniques, like flip and rotation, and thus they provide limited information.

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, however, are consistent with a previous study by Nalepa and colleagues, who analyzed data augmentation strategies employed in the BraTS challenge [84].

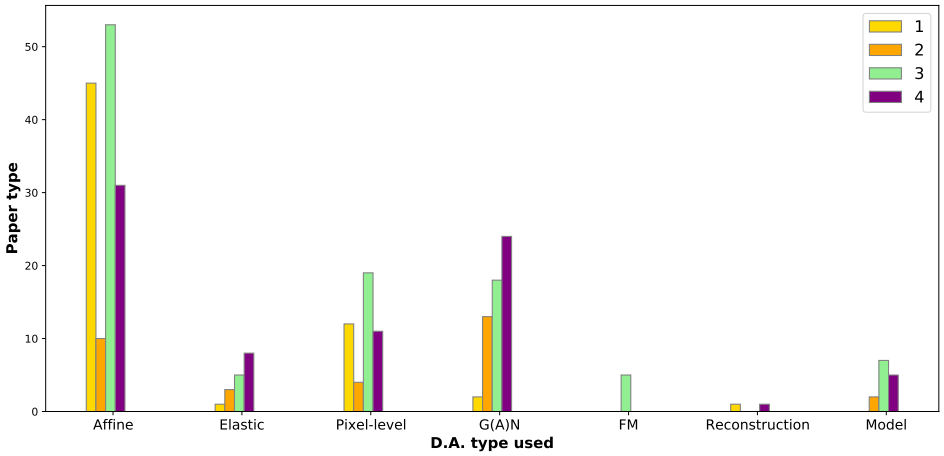


Figure 4.8. Number of different data augmentation techniques for each type of paper

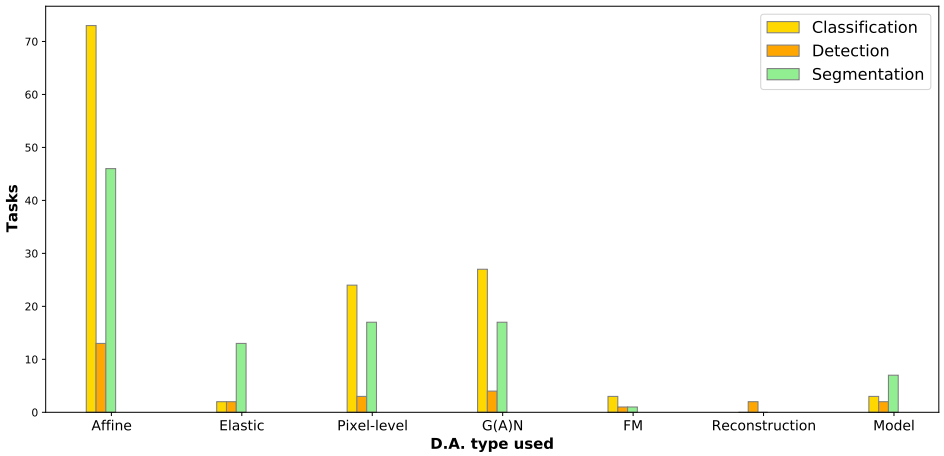


Figure 4.9. Number of different data augmentation techniques for each task

4.3 RQ3: Effect of Data Augmentation on Performance

To evaluate the effect of the various data augmentation strategies we have to select only papers that make comparison with/without other data augmentation strategies. For this reason, in this section we will only consider type 3 and type 4 papers.
A total of **128** papers was included in this analysis. The approach to the table analysis is described in **Section 3.4**.

The comparisons between the different studies are organized in tables, which report the paper modality, pathology, the type of D.A. studied (and the baseline D.A. method for the type 4 papers), the metrics used for the evaluation and the results w/w.o. data augmentation. Notice that results w/w.o. data augmentation are not available for all papers.

Moreover, the last 2 columns indicates the difference (in percentage with respect to the baseline) between the proposed D.A. and the relative NO D.A. result or other D.A. result. The use of relative differences allows us to compare studies based on different metrics.
The papers in which it was not possible to extract these data were excluded from the quantitative analysis. A total of 9 papers were excluded from the tables for this reason.

We report the results separately for papers that include experiments on different organs or tasks, hence the total number of records is **131**, out of 119 papers.

The following abbreviations have been used and the results are discussed in the following section:

- **S** - Segmentation
- **C** - Classification
- **D** - Detection
- **OSD** - One Stage Detector
- **H** - Healthy
- **CAN** - Cancer
- **L** - Lesion
- **MS** - Multiple Sclerosis
- **ALZ** - Alzheimer
- **HAND** - Hand Injuries and Disorders
- **PAR** - Parkinson
- **ACS** - Acute Coronary Sindrome
- **BIF** - Bifurcation
- **PNE** - Pneumothorax
- **NOD** - Nodule
- **COV** - COVID-19
- **TUB** - Tubercolosis
- **POL** - Polyp
- **NMD** - Neuromuscolar Diseases
- **OS** - Osteoporosis
- **FRC** - Fractures
- **URS** - Urinary Stones
- **CY** - Cyst
- **AF** - Affine
- **M-AF** - Mix Affine

- **S-AF** - Single Affine
- **GAN** - Generative Adversial Network
- **pGAN** - custom GAN
- **GN** - Generative Network
- **wGAN** - Wasserstein GAN
- **iGAN** - Info GAN
- **EL** - Elastic
- **PX** - Pixel-level
- **FM** - Feature Mixing
- **MOD** - Model-based
- **REC** - Reconstruction-based
- **ER** - Erasing
- **(*)** - Test Time Augmentation
- **MR** - Magnetic Resonance
- **CT** - Computed Tomography
- **XR** - X-Ray scan
- **MG** - Mammography
- **US** - Ultrasound scan
- **CAS** - Coronary Angioscopy
- **DICE** - DICE Score
- **CPM** - Competition Performing Metric
- **F1** - F1 Score
- **ACC** - Accuracy
- **SNS** - Sensitivity
- **PSNR** - Peak Signal-to-Noise Ratio
- **AUC** - Area under the ROC Curve
- **SP** - Specificity
- **FPI** - False Per Image
- **TPR** - True Positive Rate
- **PACC** - Pixel Accuracy
- **FROC** - Free-Response Receiver Operating Characteristic
- **MD** - Mean Distance (mm)

4.4 Brain

Neuroimaging is used to directly or indirectly assess 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 various studies are Alzheimer's disease, Multiple Sclerosis, Parkinson's disease and brain tumor.
In Table 4.1 there are 31 records, of which 10 studies perform classification tasks, 1 performs a detection task and 20 perform segmentation task. Classification is useful in determining the stage of progress in brain disease, like for example [149] for *glioma grading*, instead the segmentation helps to determine a healthy brain model with respect to a sick brain model, like in [121]. Lastly, detection task is not very used in this research field.
Every data augmentation technique used in these studies improves performance with respect to the w/o DA baseline and the other DA baseline, either by using data augmentation for balance imbalanced dataset or by using simple random data augmentation. The best percentage increase from the result w/o DA is given by [116] with 26.56%, and the best percentage increase from the baseline DA result is given by [105] and [57] with 11.76%. However, the most impressive increase in w/o DA results is given by [100] with an increase in accuracy from 0.976 to 1 using only *affine techniques*.

Furthermore, *G(A)N* and *model-based* techniques are used more frequently in neuroimaging than that in other organs: this data is very interesting because it suggests that these techniques probably perform better for the brain than other organs because the brain has an anatomical structure that is more 'stable' and less variable from one subject to another.

In particular, [108] and [59] are very interesting model-based techniques. The first proposes to generate synthetic MS lesions on patient or healthy images using different intensity level masks, and the second proposes the generation of brain tumor MR images and tumor masks used for supervised learning in brain tumor segmentation.

The study [13] is interesting because it offers a comparison with *SMOTE* (Synthetic Minority Over-sampling Technique as presented in [21]) proposing a framework based on generative adversarial network to create synthetic structural brain networks in multiple sclerosis.

In [57] the proposed data augmentation simulates a type of artifact that has been tried to correct for years. Now we try to do the opposite (make the network robust to this type of artifact). The ablation studies they propose are also interesting.
Lastly, the paper [146] deserves mention, because it proposes a model-based data augmentation that learns spatial and appearance transformations by transforming a reference atlas to match the images of a real subject. The purpose of this DA strategy is to perform a *Few-shot learning*: since the segmentation is known for the atlas, the images generated through the proposed methodology will be automatically labelled.

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	Proposed vs. no DA	Proposed vs. Baseline
[104]	C	CNN	MR	CAN	M-AF	S-AF	ACC	0.64	0.77	0.79	23.44%	2.60%
[86]	C	CNN	CT	H	AF	-	ACC	0.73	-	0.923	26.44%	-
[11]	C	CNN	PET	PAR	MOD	-	ACC	0.876	-	0.882	0.68%	-
[100]	C	CNN	MR	CAN	AF	-	ACC	0.976	-	1	2.46%	-
[129]	C	CNN	MR	CAN	AF	-	ACC	0.923	-	0.983	6.50%	-
[51]	C	CNN	PET-MR	ALZ	GAN	-	ACC	0.67	-	0.74	10.45%	-
[107]	C	CNN	MR	CAN	AF - PX	-	ACC	0.874	-	0.907	3.78%	-
[135]	C	CNN	MR	CAN	pGAN	GAN	SNS	0.689	0.74	0.789	14.51%	6.62%
[61]	C	CNN	MR	ALZ	AF - PX - ER	-	ACC	0.834	-	0.888	6.47%	-
[13]	C	AU + RF	MR	MS	GAN	-	F1	0.656	-	0.81	23.48%	-
[45]	D	Yolo-3	MR	CAN	pGAN	GAN	SNS	0.83	-	0.91	9.64%	-
[62]	S	U-Net	MR	MS	AF	-	DICE	0.803	-	0.866	7.85%	-
[3]	S	U-Net	MR	MS	MOD	AF	CPM	-	0.77	0.79	-	2.60%
[120]	S	AMRUNet	MR	CAN	FM	-	DICE	0.771	-	0.787	2.08%	-
[37]	S	U-Net	MR	CAN	GAN	AF - PX	DICE	-	0.841	0.851	-	1.19%
[42]	S	U-Net	MR	CAN	AF	-	DICE	0.636	-	0.701	10.22%	-
[59]	S	U-Net	MR	CAN	GAN - AF	AF	DICE	0.623	0.644	0.653	4.82%	1.40%
[71]	S	U-Net	MR	CAN	GAN	-	DICE	0.71	-	0.736	3.66%	-
[116]	S	U-Net	MR	CAN - ALZ	GAN	-	DICE	0.64	-	0.81	26.56%	-
[134]	S	U-Net	MR	L	GAN	-	DICE	0.438	-	0.55	25.57%	-
[42]	S	U-Net	MR	L	AF	-	DICE	0.427	-	0.46	7.73%	-
[57]	S	U-Net	MR	H	MOD-AF	AF	DICE	-	0.876	0.892	-	1.83%
[24]	S	Custom	MR-CT	H	pGAN	GAN	DICE	-	0.783	0.81	-	3.45%
[125]	S	U-Net	MR	H	EL	AF	DICE	-	0.884	0.906	-	2.49%
[121]	S	U-Net	PET	H	GAN	-	DICE	0.872	-	0.889	1.95%	-
[28]	S	Custom	MR	H	PX	-	DICE	0.805	-	0.82	1.86%	-
[95]	S	U-Net	MR	H	GN	-	DICE	0.83	-	0.88	6.02%	-
[73]	S	U-Net	CT	CAN	MOD	AF	DICE	0.852	0.859	0.875	2.70%	1.86%
[146]	S	U-Net	MR	H	MOD - AF - PX	AF - PX	DICE	0.76	0.775	0.815	7.24%	5.16%
[85]	S	U-Net	MR	CAN	AF - EL	-	DICE	0.763	-	0.8	4.85%	-
[108]	S	U-Net	MR	MS	MOD	-	DICE	0.57	-	0.63	10.53%	-

Table 4.1. Comparison between **31** results that studied data augmentation techniques in brain imaging.See the legend for the acronyms.

4.5 Heart

Cardiac imaging refers to non-invasive imaging of the heart using primarily ultrasound (US), magnetic resonance imaging (MR) and computed tomography (CT). The heart is the most important organ in our body, for this reason many diseases involving the circulatory system are also analyzed in these studies.

In Table 4.2, there are 11 records, of which 3 perform classification tasks, 2 perform detection task and 6 perform segmentation task. Note that the segmentation is practically performed only on healthy subjects.

Every data augmentation technique used in these studies improves performance with respect to the w/o DA baseline and the other DA baseline, either by using data augmentation for balance imbalanced dataset or by using simple random data augmentation.

In interpreting the results, attention must be paid to the experimental setting of each paper, which could inflate the comparison. For example, [144] starts from a very poor and easily improved result, therefore the 352.13% increase is not an amazing result as it might seem. With this reading key, the best results are achieved by [70] (DICE: 0.911 to 0.927) from the result w/o DA and [12] (DICE: 0.833 to 0.846) e from the baseline DA.

The study [89] deserves a mention because it is the only one paper that performs data augmentation using a *reconstruction-based* method. They presented a study on automatic cardiac motion artefact detection using spatio-temporal deep learning techniques, in particular by using a k-space based corruption strategy to increase the robustness of the classification: this strategy is able to simulate different levels of realistic motion-related artefacts (e.g., breathing and mistrigering artefacts).

Lastly, different variants of GAN are used, such as *SpeckleGAN* [12], *ScreenGAN* [40] or *CycleGAN* [20].

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	Proposed vs. no DA	Proposed vs. Baseline
[39]	C	CNN	CT	Many	AF - ER	-	ACC	0.74	-	0.903	22.03%	-
[40]	C	CNN	US	H	GAN	-	ACC	0.7	-	0.85	21.43%	-
[80]	C	CNN	CAS	Many	GAN	-	AUC	0.77	-	0.81	5.19%	-
[74]	D	DCNN	CT	ACS	AF - FM	-	AUC	0.808	-	0.867	7.30%	-
[89]	D	3D CNN	MR	Many	AF - PX - REC	AF	AUC	0.581	0.674	0.735	26.51%	9.05%
[19]	S	U-Net	MR - CT	H	MOD	FM	DICE	0.332	0.652	0.747	125.00%	14.57%
[20]	S	U-Net	MR - CT	H	GAN	AF	DICE	0.613	0.676	0.712	16.15%	5.33%
[144]	S	U-Net	MR - US	H	AF - PX - EL	GAN	DICE	0.188	0.575	0.85	352.13%	47.83%
[119]	S	U-Net	MR	H	PX	AF - EL	DICE	-	0.652	0.894	-	37.12%
[12]	S	U-Net	US	H	pGAN	GAN	DICE	0.824	0.833	0.846	2.67%	1.56%
[70]	S	U-Net	CT	Many	AF	-	DICE	0.911	-	0.927	1.76%	-

Table 4.2. Comparison between **11** results that studied data augmentation techniques for heart imaging.See the legend for the acronyms.

4.6 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 the leading cause of cancer deaths worldwide and due to the COVID-19 epidemic of the last 2 years, the lung is probably the most studied and analyzed organ in the entire SLR.
In Table 4.3, there are 34 records, of which 26 perform classification tasks, 4 perform detection task and 4 perform segmentation task. Every data augmentation technique used in these studies improves performance with respect to the w/o DA baseline and the other DA baseline, either by using data augmentation for balance imbalanced dataset or by using simple random data augmentation, **except for** [97] which observes a decrease of 14.45% in COVID-19 classification using simple affine transformation. There might be different reasons why the performance get worse, one of these could be that after the affine transformation (i.e., flip, rotation, etc) the image is not a valid image in the domain. However, this remains an isolated case.
Moreover, there are 2 papers that use particular strategies to deepen. The first is [55] which implements a tumor-aware unsupervised cross-domain adaptation (preserving the anatomical details between CT and MRI), followed by a semi-supervised tumor segmentation using U-Net trained with synthesized artefacts created from a limited number of original magnetic resonances.
The second one is [92] which implements a particular *rotation invariant* CNN: three image slices of cross sections normal to the axial, coronal and sagittal axes of the pulmonary nodule were captured using the extracted VOI. Then, data augmentation was performed to increase the training data. To facilitate consideration of information from adjacent slices, multiplanar images of the pulmonary nodule were generated, with slice angles varying between $[-40°, 40°]$ in steps of $5°$ with reference to the three slice images. This increased the amount of data by 32 times. Furthermore, it was increased by eight times by rotating and reversing the multiplanar images. Therefore, the training data were expanded 256 times and served as input to the networks.

Lastly, paper [8] and paper [2] implements *test time data augmentation*, according to the methodology described in the Section 2.5., obtaining good results (DICE: 0.941 to 0.979 and 0.901 to 0.902, respectively). These results may confirm that the models can benefit from the increased robustness given by test-time data augmentation.

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	Proposed vs. no DA	Proposed vs. Baseline
[97]	C	CNN	XR	COV	AF	-	ACC	0.868	-	0.745	-14.17%	-
[72]	C	CNN	CT - XR	PNE	PX	-	ACC	0.928	-	0.988	6.47%	-
[63]	C	CNN	CT	CAN	AF	-	ACC	0.9	-	1	11.11%	-
[7]	C	CNN	XR	TUB	AF	-	ACC	0.78	-	0.8	2.56%	-
[87]	C	CNN	XR	COV - PNE	AF - ER - FM	-	ACC	0.787	-	0.837	6.35%	-
[14]	C	CNN	XR	PNE	GAN - PX - AF	-	ACC	0.795	-	0.91	14.47%	-
[111]	C	CNN - RNN	CT - XR	COV	AF - GAN	-	AUC	0.91	-	0.99	8.79%	-
[8]	C	CNN	XR	COV	AF (*)	-	ACC	0.941	-	0.979	4.04%	-
[132]	C	CNN	XR	PNE	AF - PX	AF - PX	ACC	0.84	0.86	0.945	12.50%	9.88%
[98]	C	CNN	CT	CAN	MOD	-	ACC	0.5	-	0.786	57.20%	-
[91]	C	CNN	CT	CAN	AF - PX - GAN	AF - PX	ACC	0.5	0.761	0.825	65.00%	8.41%
[9]	C	CNN	CT	CAN	AF	-	ACC	0.842	-	0.942	11.88%	-
[92]	C	CNN	CT	CAN	AF - GAN	-	AUC	0.635	-	0.812	27.87%	-
[78]	C	CNN	CT	NOD	MOD	-	CPM	0.77	-	0.79	2.60%	-
[124]	C	CNN	CT	NOD	pGAN	wGAN	ACC	-	0.923	0.95	-	2.93%
[38]	C	CNN	XR	PNE	AF	AF - PX	ACC	0.74	0.79	0.85	14.86%	7.59%
[131]	C	PSSPNN	CT	COV	AF - PX	-	F1	0.939	-	0.958	2.02%	-
[133]	C	CNN	CT	NOD	AF	AF (*)	ACC	-	0.848	0.88	-	3.77%
[126]	C	CNN	CT	CAN	pGAN	wGAN	ACC	0.342	0.547	0.577	68.71%	5.48%
[90]	C	CNN	CT	H	MOD	AF	SNS	-	0.839	0.858	-	2.26%
[147]	C	CNN	CT	NOD	AF - FM	-	ACC	0.694	-	0.74	6.63%	-
[6]	C	CNN	CT	CAN	AF	-	AUC	0.598	-	0.655	9.53%	-
[136]	C	RNN	XR	TUB	AF - PX	-	ACC	0.471	-	0.928	97.03%	-
[130]	C	CNN	CT	NOD	AF - EL	wGAN	ACC	-	0.57	0.65	-	14.04%
[52]	C	CNN	XR	Many	GAN	-	ACC	0.662	-	0.957	44.56%	-
[99]	C	CNN	CT	CAN	MOD	AF	ACC	-	0.83	0.92	-	10.84%
[96]	D	3D CNN	CT	NOD	AF	-	SNS	0.75	-	0.9	20.00%	-
[49]	D	-	CT	NOD	AF	-	ACC	0.697	-	0.788	13.06%	-
[122]	D	3D CNN	CT	NOD	AF	-	SNS	0.923	-	0.99	7.26%	-
[46]	D	R-CNN	CT	NOD	GAN	-	CPM	0.518	-	0.55	6.18%	-
[2]	S	U-Net	XR	PNE	AF - PX - EL (*)	-	DICE	0.9011	-	0.9023	0.13%	-
[54]	S	U-Net	MR - CT	CAN	GAN	AF	DICE	-	0.63	0.7	-	11.11%
[55]	S	U-Net	CT	CAN	pGAN	GAN	DICE	0.55	0.63	0.8	45.45%	26.98%
[73]	S	U-Net	XR	NOD	MOD	AF	DICE	0.725	0.794	0.809	11.59%	1.89%

Table 4.3. Comparison between **34** results that studied data augmentation techniques over lung and relative pathologies. See the previous legend for full comprehension of the acronyms.

4.7 Breast

Breast imaging is a sub-speciality of diagnostic radiology that involves imaging of the breast for screening or diagnostic purposes using principally mammography (MG). Although the death rate is decreasing, breast cancer remains one of the most common cancers, especially among women. Indeed, breast cancer is the only pathology analyzed in this table.

In Table 4.4, there are 14 records, of which 9 perform classification tasks, 3 perform detection task and 2 perform segmentation task.

Every data augmentation technique used in these studies improves performance with respect to the w/o DA baseline and the other DA baseline, either by using data augmentation for balance imbalanced dataset or by using simple random data augmentation. In particular, using the Sigmoid correction, the best increase from the w/o DA result is given by [143] using only *affine* transformation and, the best increase from the baseline DA is given by [106] using a GAN with respect to affine techniques.

The studies [18] and [73] deserve a mention because they use model-based methods: in the first, virtual three-dimensional anthropomorphic phantoms were produced using a procedural analytic model in which the major anatomical structures (including fat and glandular tissues, ductal tree, vasculature, and ligaments) are stochastically generated within a predefined breast volume; instead in the second they propose a method called stochastic evolution (SE): the irregular deterioration and healing processes of the diseased tissue is simulated according to the direction of the local distortion, thereby producing a natural sample that is indistinguishable by humans.

Lastly, there is another interesting study conducted by [17] where *only elastic* technique is compared to *affine* technique. The proposed data augmentation decreases the False per Image (FPI) of about 25.66% in classification task: this is probably favored by the shape of the breast that rewards this type of transformation.

Paper	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	Proposed vs. no DA	Proposed vs. Baseline
[30]	C	CNN	MG	CAN	AF	-	ACC	0.77	-	0.82	6.49%	-
[29]	C	CNN	MG	CAN	GAN	-	ACC	0.782	-	0.87	11.25%	-
[143]	C	CNN	US	CAN	AF	-	AUC	0.94	-	0.96	2.13%	-
[16]	C	OSD	MG	CAN	AF - EL - PX	-	TPR	0.861	-	0.913	6.04%	-
[1]	C	U-Net	MG	CAN	AF	-	DICE	0.922	-	0.951	3.15%	-
[115]	C	CNN	MG	CAN	AF - PX	-	ACC	0.782	-	0.836	6.91%	-
[118]	C	CNN	MG	CAN	AF	-	ACC	0.71	-	0.859	20.99%	-
[17]	C	CNN	MG	CAN	EL	AF	FPI	-	2934	2181	-	-25.66%
[106]	C	CNN	MG	CAN	AF - GAN	AF	ACC	0.699	0.88	0.94	34.48%	6.82%
[82]	D	CNN	CT	CAN	GAN	AF	AUC	-	0.741	0.869	-	17.27%
[113]	D	R-CNN	MG	CAN	GAN	AF	AUC	0.151	0.159	0.172	13.91%	8.18%
[94]	D	CNN	US	CAN	pGAN	iGAN	ACC	-	0.887	0.904	-	1.92%
[18]	S	R-CNN	CT	CAN	MOD	-	SNS	0.802	-	0.833	3.87%	-
[73]	S	U-Net	CT	CAN	MOD	AF	DICE	0.547	0.643	0.666	21.76%	3.58%

Table 4.4. Comparison between **14** results that studied data augmentation techniques for breast imaging.See the legend for the acronyms.

4.8 Other Organs

In Table 4.5 there are 42 records representing studies conducted on minor organs. The organs most present in the table are *liver* and *prostate* with 10 and 7 instances, respectively. These studies analyze more particular and less common cases and/or conditions than in the previous tables, such as detection of urinary stones in [101], classification of cyst and tumors of both jaws in [64] or classification of calcaneal fractures in [4].

Even if not included in the table because they do not compare different DA techniques, the studies [4] and [50] deserve a mention because they use a curious approach: same data augmentation techniques, but they change the number of data generated for each experiment. In particular, the first study compares 10000 vs 2000 calcaneous images, and the second compares 3000 vs 500 pelvic images, noting that generating more images bring to better results.

The study [60] uses train-time data augmentation to randomly increase the size of the training set and, it uses test-time data augmentation to estimate the uncertainty of the predictions. For both strategies there is an improvement with respect w/o DA baseline.

The studies [139] and [101] deserve a mention because they use model-based methods to generate spine and urinary tract images, respectively.

In [36] we noticed a positive effect of augmentation for the *"severe" data*, instead for the *"moderate" data* there are no significant improvements (note that "moderate" and "severe" referring to the level of generated fatty infiltration in human thighs).

The study [102] is very interesting because it uses an adaptive data augmentation in which the optimal policy is learned through reinforcement learning, being the only study in this Systematic Literature Review that implements a learnable augmentation strategy.

Also the study [26] is very interesting because, in addition to studying the application of their data augmentation technique on various organs, it performs 3-dimensional transformations to augment the datasets to train the same model 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 up performance.

Lastly, in this table there are another study in which the data augmentation used does not improve the w/o DA performance: in [65] we have a decrease in performance from 0.89 to 0.88 using affine and pixel-level transformations in teeth classification.

Paper	Organ	Task	Network	Modality	Pathology	Proposed DA	Baseline DA	Metric	w/o DA	w/Baseline	w/Proposed	Proposed vs. no DA	Proposed vs. Baseline
[44]	Abdomen	D	YOLO	CT	H	GAN	-	MD	8.66	-	7.95	-8.20%	-
[26]	Abdomen	S	U-Net	MR	H	AF	-	DICE	0.897	-	0.9	0.33%	-
[88]	Arm	S	U-Net	US	H	EL	AF	PACC	0.98	0.998	0.999	1.94%	0.10%
[25]	Colon	D	3D FCN	CT	POL	AF - EL	AF	FROC	-	0.9	0.97	-	7.78%
[128]	Colon	C	3D CNN	CT	POL	GN	GAN	AUC	0.811	0.846	0.881	8.63%	4.14%
[142]	Humerus	S	U-Net	US	H	GAN - AF - PX	AF - PX	DICE	0.58	0.605	0.66	13.79%	9.09%
[64]	Teeth	C	CNN	XR	CAN - CY	AF - PX	-	AUC	0.86	-	0.94	9.30%	-
[102]	Kidney	S	U-Net	CT	CAN	AF - ER - PX	AF	DICE	0.749	0.832	0.84	12.15%	0.96%
[138]	Kidney	S	U-Net	US	H	MOD	-	DICE	0.93	-	0.942	1.29%	-
[109]	Kidney	S	U-Net	CT	H	GAN - AF	AF	DICE	0.92	0.94	0.944	2.61%	0.43%
[26]	Kidney	S	U-Net	MR	H	AF	-	DICE	0.935	-	0.954	2.03%	-
[123]	Liver	S	3D U-Net	CT	H	MOD	-	ACC	0.861	-	0.883	2.56%	-
[22]	Liver	C	CNN	CT	CAN	GAN	GAN + AF	ACC	0.802	-	0.855	6.61%	-
[33]	Liver	C	CNN	CT	L	GAN	AF	ACC	-	0.786	0.857	-	9.03%
[47]	Liver	S	U-Net	CT	H	AF - PX	-	DICE	0.896	-	0.915	2.12%	-
[47]	Liver	S	U-Net	CT	CAN	AF - PX	-	DICE	0.659	-	0.702	6.53%	-
[127]	Liver	S	U-Net	CT	CAN	AF - FM	-	DICE	0.915	-	0.945	3.28%	-
[34]	Liver	C	CNN	CT	CAN	GAN	AF	SNS	-	0.797	0.865	-	8.53%
[109]	Liver	S	U-Net	CT	H	GAN - AF	AF	DICE	0.944	0.941	0.947	0.32%	0.64%
[26]	Liver	S	U-Net	MR	H	AF	-	DICE	0.961	-	0.962	0.10%	-
[36]	Muscle	S	U-Net	MR	NMD	GAN - EL	EL	DICE	0.84	0.87	0.88	4.76%	1.15%
[75]	Muscle	S	U-Net	CT	H	AF - PX	-	DICE	0.833	-	0.915	9.84%	-
[26]	Pancreas	S	U-Net	MR	H	AF	-	DICE	0.86	-	0.864	0.47%	-
[137]	Pelvis	S	U-Net	MR	Many	EL	-	PACC	0.724	-	0.841	16.16%	-
[73]	Prostate	S	U-Net	XR	H	MOD	AF	DICE	0.785	0.81	0.848	8.03%	4.69%
[123]	Prostate	S	3D U-Net	MR	H	MOD	-	ACC	0.742	-	0.868	16.98%	-
[31]	Prostate	S	U-Net	MR	CAN	AF - GAN	-	DICE	0.678	-	0.738	8.85%	-
[56]	Prostate	S	U-Net	MR	H	MOD	-	DICE	0.813	-	0.863	6.15%	-
[77]	Prostate	S	AMS	MR	CAN	GAN	-	DICE	0.858	-	0.882	2.80%	-
[140]	Prostate	C	Custom	MR	CAN	GAN	AF	ACC	-	0.815	0.892	-	9.45%
[43]	Prostate	S	U-Net	MR - PET - CT	CAN	AF - FM	-	DICE	0.776	-	0.819	5.54%	-
[67]	Rectal	S	MRSN	MR	H	AF - PX	-	DICE	0.938	-	0.943	0.53%	-
[67]	Rectal	S	MRSN	MR	CAN	AF - PX	-	DICE	0.732	-	0.742	1.37%	-
[139]	Spine	S	U-Net	CT	H	MOD	-	PACC	0.743	-	0.948	27.59%	-
[109]	Spleen	S	U-Net	CT	H	GAN - AF	AF	DICE	0.884	0.89	0.919	3.96%	3.26%
[26]	Spleen	S	U-Net	MR	H	AF	-	DICE	0.944	-	0.944	0.00%	-
[60]	Teeth	S	U-Net	XR	H	AF	AF (*)	ACC	0.9474	-	0.9476	0.02%	-
[76]	Teeth	C	CNN	XR	CAN	AF	-	ACC	0.699	-	0.904	29.33%	-
[65]	Teeth	C	CNN	XR	H	AF - PX	-	AUC	0.89	-	0.88	-1.12%	-
[66]	Teeth	C	CNN	XR	OS	AF	-	ACC	0.925	-	0.98	5.95%	-
[114]	Thyroid	C	CNN	US	CAN	GAN	AF	ACC	0.701	0.756	0.915	30.53%	21.03%
[101]	Urinary Tract	D	U-Net	XR	URS	MOD	-	F1	0.561	-	0.603	7.49%	-

Table 4.5. Comparison between **42** results that studied data augmentation techniques for different organs related pathologies. See the legend for the acronyms.

4.9 Visual Data Analysis

The box plot in Figure 4.10 represents the distribution of the percentage increase in performance with respect to each organ and with respect to the totality of the data. Only papers comparing w/o DA results and proposed DA result are taken into consideration. The detail of the number of papers included is summoned in Table 4.6.

Organ	# papers included
Heart	8
Brain	26
Lung	28
Breast	12
Others	38
All	112

Table 4.6. Number of papers involved for each organ in the box plot of Figure 4.10

The average increase compared to all organs is about 10%, however, heart, lung and breast appear to have larger increases. It could be due to various factors: the data augmentation techniques are more effective, the datasets are on average smaller, or the variability of the organ is greater, and therefore the data augmentation helps more to generalize. Instead, brain has the lowest average increment.

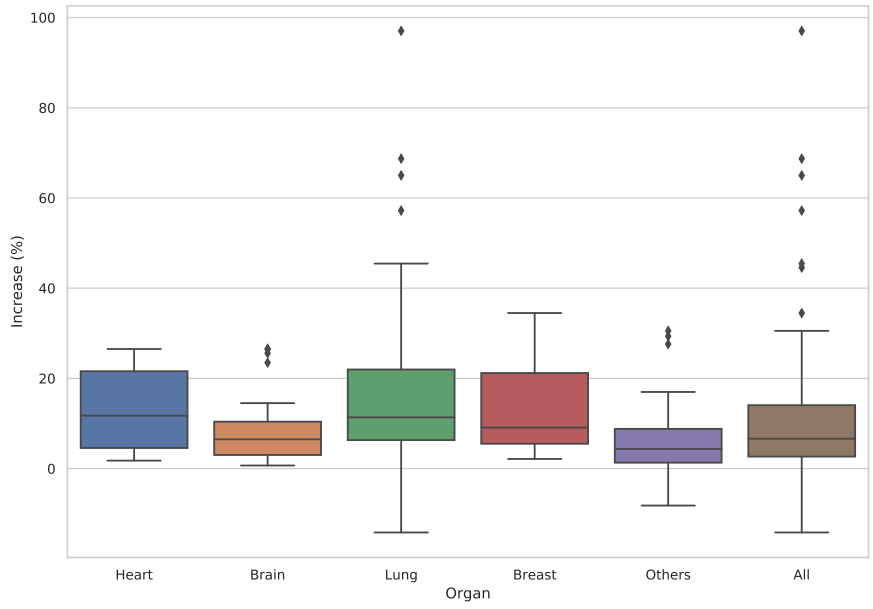


Figure 4.10. Box plot of the percentage increases from w/o DA for each organ.

The box plot in Figure 4.11 represents the distribution of the percentage increase in performance with respect to **affine** and no-affine data augmentation. Moreover, the increases of *no-affine* techniques when the baseline uses *affine* transformation are reported. The detail of the number of papers included is summoned in Table 4.7.

Comparison	**# papers included**
Only Affine vs w/o DA	29
Other vs w/o DA	91
Other vs Affine	29

Table 4.7. Number of papers involved for each DA techniques comparison in the box plot in Figure 4.11

A total of 149 papers are involved in this box plot.
Excluding the few outliers, the affine transformations reach on average results comparable to the other techniques, and, as they are very simple to implement, they can be the first way to go to perform data augmentation, before attempting *"more complex"* techniques that will still give similar results.

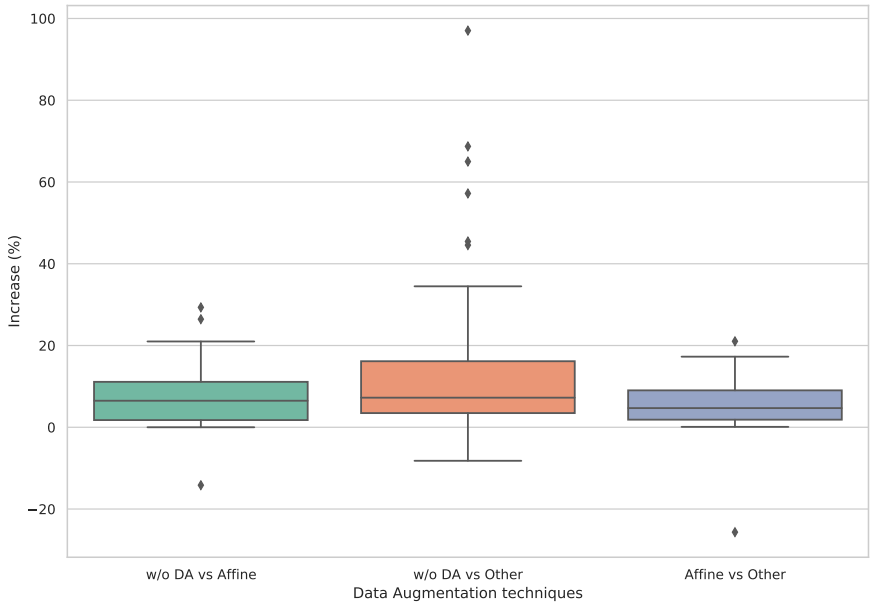


Figure 4.11. Box plot of the percentage increases from w/o DA and Baseline results over affine and no-affine transformations.

Scatter plot in Figure 4.12 shows the distribution of every percentage increase from the w/o DA result with respect to its original data set size. Only papers comparing w/o DA results and proposed DA result are taken into consideration.

There are 111 points in the plot. Methods that combine multiple transformations were assigned to the most complex transformation according to the following hierarchy:

1. GAN
2. Model-based
3. Reconstruction-based
4. Feature Mixing
5. Erasing
6. Affine

Moreover, the x-axis is represented in logarithmic scale in order to facilitate the reading of the plot. The scatter plot is shown in Figure 4.12.

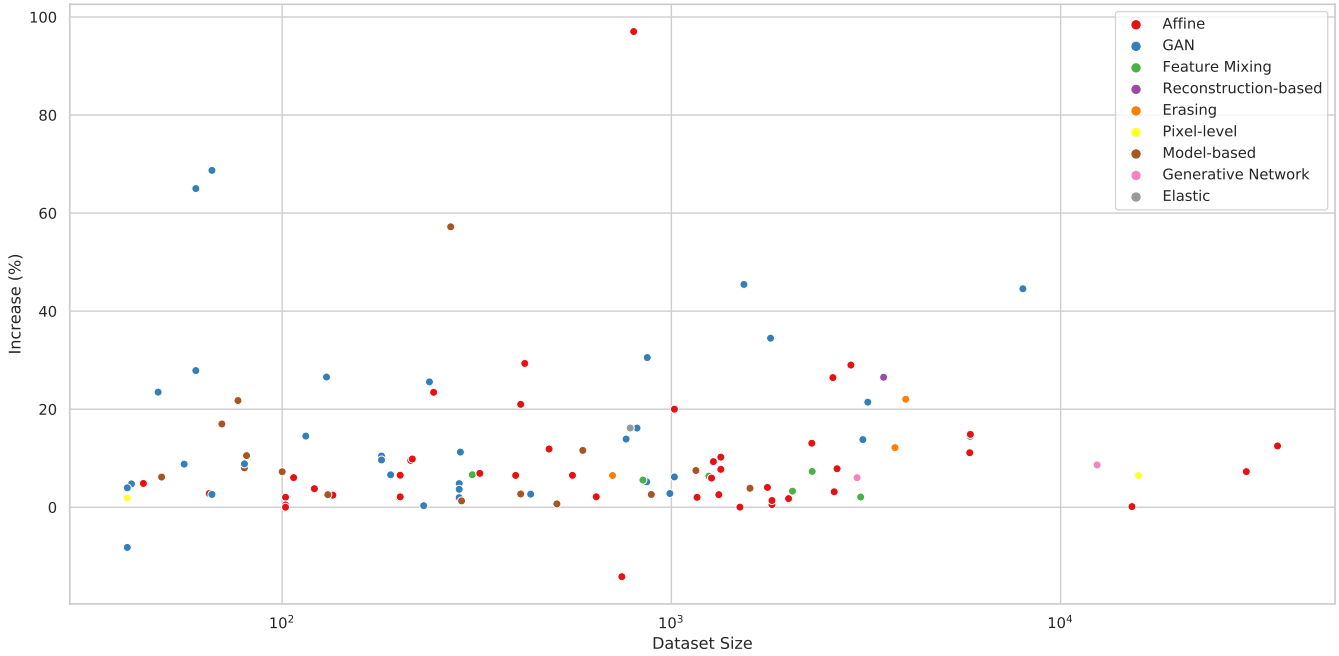


Figure 4.12. Scatter plot of percentage increase from w/o DA result over the original training set size.

More than half of the datasets have less than 1000 records, confirming the initial hypothesis that it is difficult to collect this type of data and data augmentation must be done to better generalize the model.

There are only 2 negative increments. Furthermore, it can be seen that the GANs seem to have a higher trend. In any case, almost all values and finally almost all are in the 0-40% range.

It is interesting to draw a regression line of all these values, as shown in the Figure 4.13.

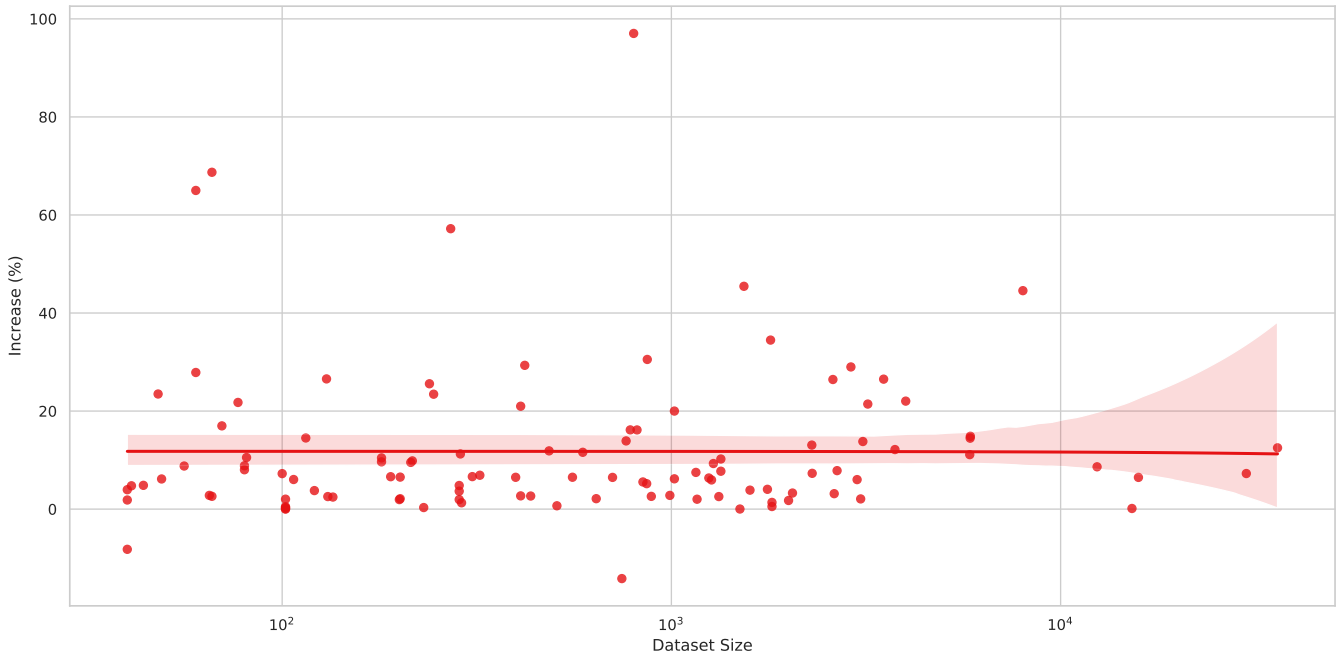


Figure 4.13. Regression line between percentage increase from w/o DA result and the original training set size.

The regression line has a slope of $-1.46356e-05$, confirming that there is practically **no correlation** between the size of the training set and the percentage increase of the data augmentation.

4.10 RQ4: Differences between Computer Vision and Medical Data Augmentation

In this section we will answer to what are the types of data augmentation that are not adopted/less used in the medical domain. Considering as a point of reference the taxonomy defined in [117], and looking to the previous results of this Systematic Literature Review, the following consideration emerge:

- **Geometric transformations**: different augmentations based on geometric transformations and many other image processing functions. It is characterized by their ease of implementation. They are called *affine transformations* in our taxonomy.
- **Random Erasing**: this technique was specifically designed to combat image recognition challenges due to occlusion, but in the medical domain does not exist the concept of occlusion. Rarely used in the medical domain literature.
- **Color Space Transformations**: these techniques modify the pixel values of the image. In the medical domain, it is usual to modify the brightness, the contrast, or the saturation, rather than modifying the color space, since most reports are in grayscale. All these transformations are called pixel-level transformations in our taxonomy.
- **Mixing images**: these techniques are mostly used in object detection tasks. They are called *feature mixing* in our taxonomy. Rarely used in the medical domain literature.
- **Adversarial training**: it is used as a search algorithm for the best data augmentation policy. Very widespread in the learnable data augmentation field, but less in medical domain.
- **Neural style transfer**: the algorithm works by manipulating the sequential representations across a CNN such that the style of one image can be transferred to another while preserving its original content. We have considered these techniques as *GAN-based* methods.
- **GAN Data Augmentation**: classic approach to generate synthetic images. It is called *GAN-based* methods in our taxonomy.

Taking into account the results of this literature, the two main data augmentation techniques that are rarely used in the medical literature: *random erasing* and *feature-mixing techniques*.

The reasons and motivations behind this kind of exclusion are never expressly exhibited, so we can only do a hypothesis about it. Probably, because erasing or mixing a part of an image can lead to lose the important feature of that. To discover if can be future studies about these techniques in medical domain, in the next chapter, we conduct some experiments using random erasing and mix-up data augmentation on ChestX-ray14 dataset.

Chapter 5

Experiments

5.1 Dataset

The dataset chosen for the experiments is ChestX-ray14 from [53]. It contains 112,000 images from 32,717 patients. The dataset is labeled for the presence/absence of 14 independent observations starting from radiological reports using two different semiautomatic labellers. The five classes chosen for the exeriments are: *Atelectasis*, *Cardiomegaly*, *Consolidation*, *Edema* and *Pleural Effusion*, for which prior results are available in literature.

CheXpert ground truth takes into account uncertainty in the reporting and labelling process by assigning each observation one of three values: positive, *negative* and *uncertain*. To handle uncertain labels during training, we used following policies [81]:

- **U-Ones**: it converts an uncertain label to a positive label.
  It was used on Atelectasis, Edema, and Pleural Effusion labels.

- **U-Zeroes**: it converts an uncertain label to a negative label.
  It was used on Cardiomegaly and Consolidation labels.

5.2 Experimental setup

We performed the experiments using the baseline described in [81] based on ResNet18 architecture (pretrained on ImageNet). The output layer is the combination of five binary heads, one for each observation. During training, classical weak data augmentation is performed: random rotation ($\pm$ 10°), random zooming (0% to 10%), random cropping and rescaling to network input size. Every transformation is applied with probability $p = 0.75$.

Moreover, all the images were normalized with the mean and standard deviation calculated over CheXpert. The network was trained using the SGD optimizer with the *One Cycle Policy* for 12 epochs. The starting Learning Rate (LR) is equal to $3e-3$. Weights

were saved every 1024 iterations and the checkpoint with the highest validation performance was used.

The metric chosen for the evaluation is the Area under the ROC curve (AUC). It was calculated for each observation, and then its mean value was used to summarize the performance. For each experiment configuration, the training was repeated three times and the mean and standard deviation were reported.

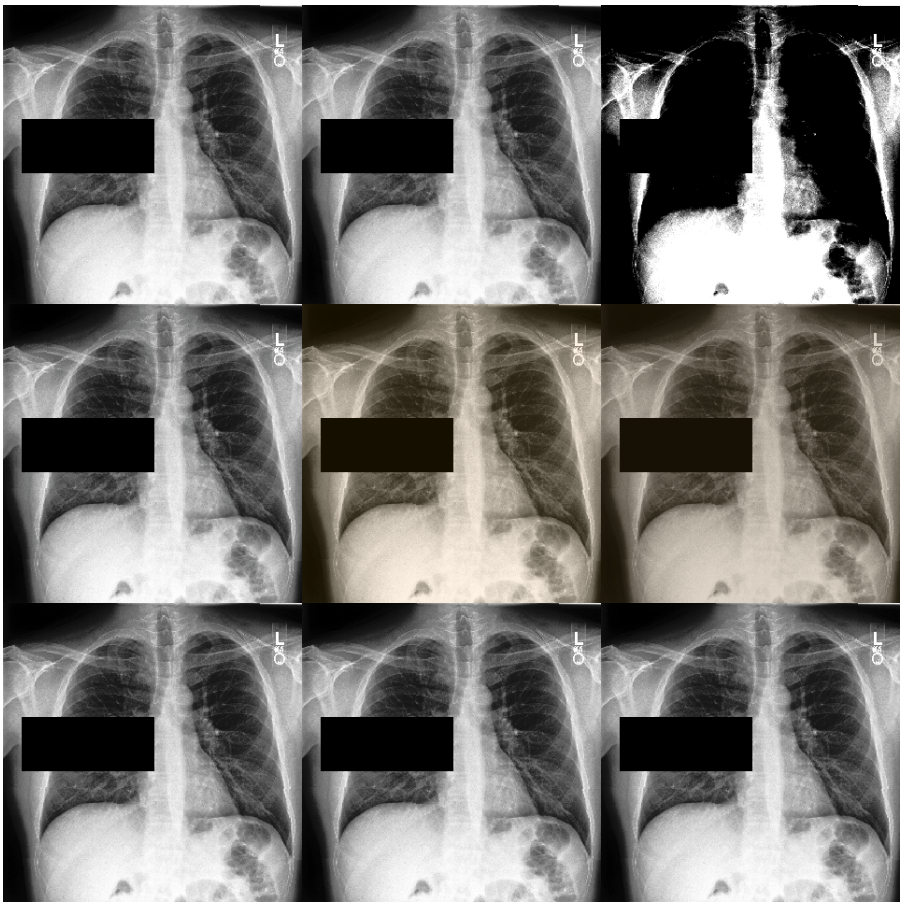


Figure 5.1. Example of images generated using *random erasing* technique over the baseline weak data augmentation.

5.3 Tested Data Augmentation Techniques

For the first experiment, **random erasing** data augmentation is applied to the baseline with probability $p = 0.5$, scale from 0.02 to 1 (i.e. range of the proportion of erased area against the input image), ratio from 0.3 to 1 (i.e. the range of aspect ratio of erased area)

and pixel erasing value equal to 0 (i.e white color in the grayscale).
Examples of images generated using this technique are shown in Figure 5.1.

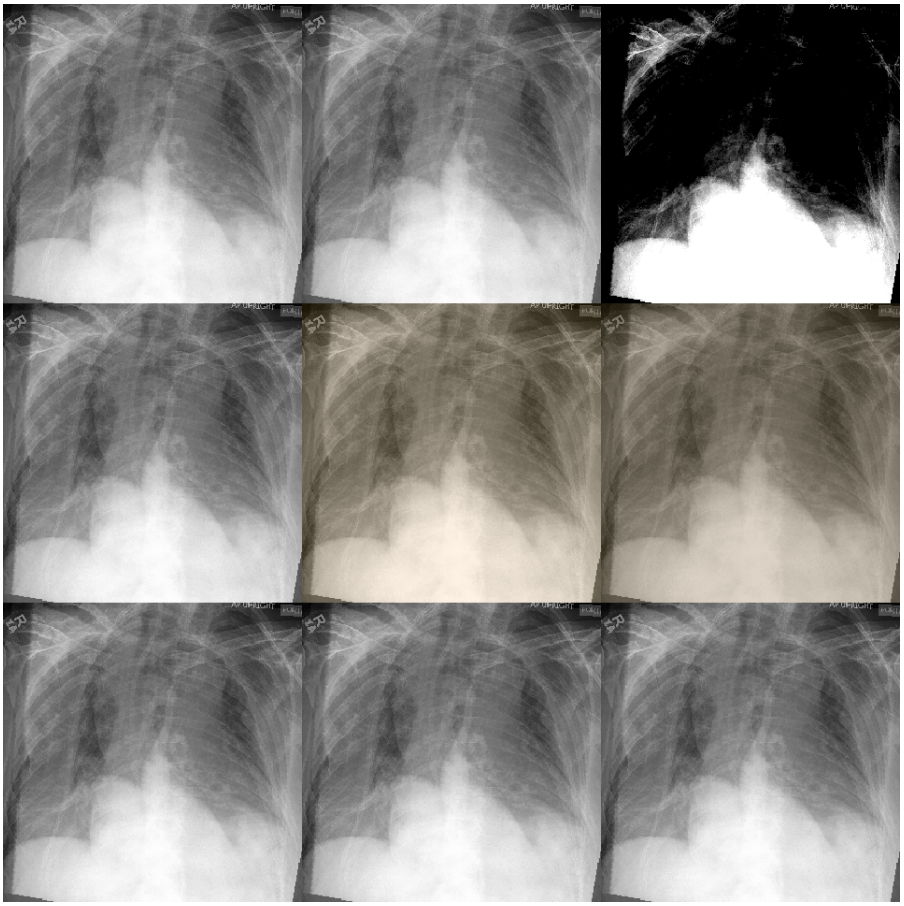


Figure 5.2. Example of images generated using *mix-up* technique over the baseline weak data augmentation.

For the second experiment, **mix-up** data augmentation is applied to the baseline. Inspired by [103], we used this formula to generate the mixed images:

$$\widetilde{x} = \lambda x_i + (1 - \lambda)x_j \text{ , where } x_i \text{ and } x_j \text{ are image input vectors}$$

and this formula to generate the new labels:

$$\widetilde{y} = \lambda y_i + (1 - \lambda)y_j \text{ , where } y_i \text{ and } y_j \text{ are label vectors}$$

and where lambda values is randomly sampled from Beta distribution (α, α).
Every transformation is applied with probability $p = 0.5$ and parameter $\alpha = 0.3$.
Examples of images generated using this technique are shown in Figure 5.2.

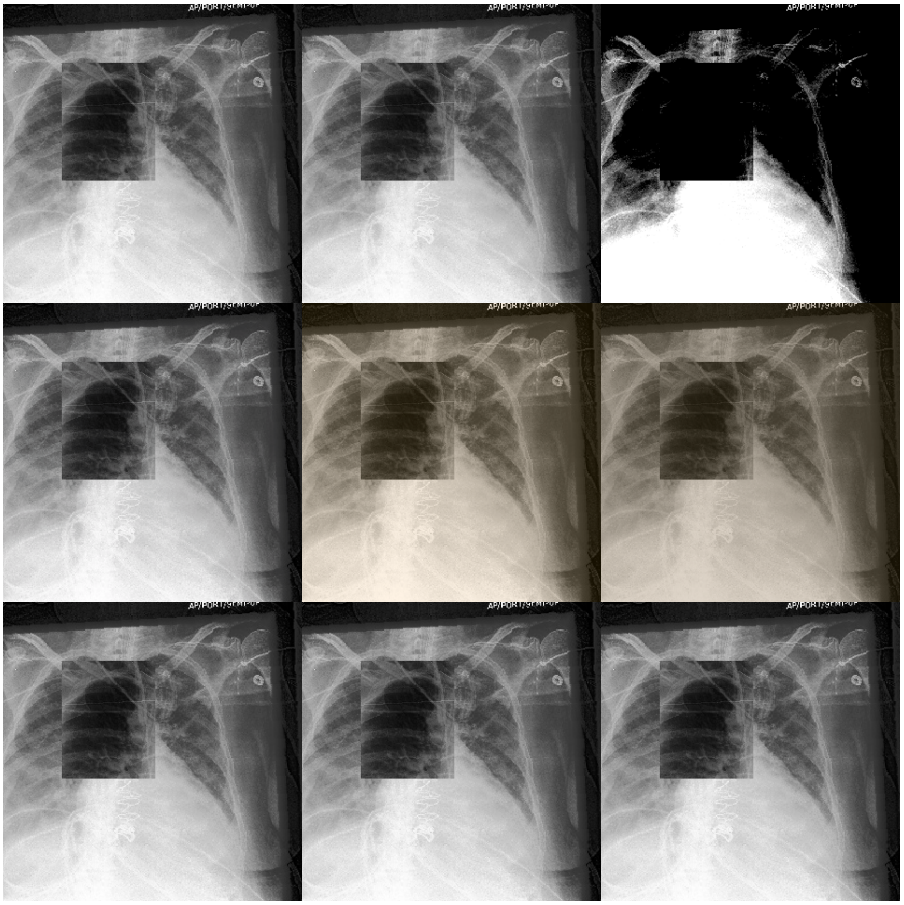


Figure 5.3. Example of images generated using *cut-mix* technique over the baseline weak data augmentation.

In the last experiment, we decided to *merge* the techniques used in the previous experiments, implementing **cut-mix** data augmentation. Instead of removing pixels and filling them with black or grey pixels (like random erasing), you replace the removed regions with a patch from another image (like feature mixing with $\alpha = 1$), while the ground truth labels are mixed proportionally to the number of pixels of combined images [141].
In particular, we used this formula to generate the new labels:

$$\widetilde{y} = (1 - A)y_i + Ay_j \text{ , where } y_i \text{ and } y_j \text{ are label vectors}$$

and where A is the ratio between area of patch applied and the total area of the image (320x320). Every transformation is applied with probability $p = 0.5$ and its width w and heigth h both variable in the range $[50, 160]$ so that the largest possible patch modifies the label by a maximum of 25%.
Examples of images generated using this technique are shown in Figure 5.3.

5.4 Results

In this section, we compare the different data augmentation strategies illustrated above. All the experiments are trained by fine-tuning all layers. The average AUC values are shown in Table 5.1.

Setup	Data Augmentation Strategy	Mean AUC
Baseline	Weak DA	89.6 ± 0.2
Exp1	Weak DA + Random Erasing	89.1 ± 0.2
Exp2	Weak DA + Mix-up	89.0 ± 0.2
Exp3	Weak DA + Cut-Mix	89.3 ± 0.3

Table 5.1. Average AUC for different data augmentation strategies on CheXpert.

Based on the observed standard deviations, differences between the different data augmentation schemes appear to be compatible with the variability due to random initialization.

Moreover, in validation losses and AUC curves showed in Figure 5.4 for random erasing, Figure 5.5 for mix-up and Figure 5.6 for cut-mix, it is observed that the random erasing experiment is more subject to overfitting than the other two experiments.

However, all the experiments performed slightly worse than the baseline result. These results confirm that random erasing and feature mixing strategies are less used in the medical literature because they are not so effective with respect to other data augmentation techniques.

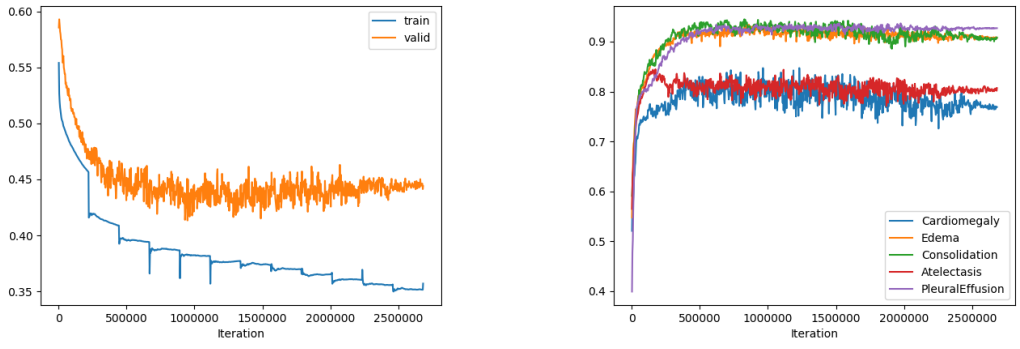


Figure 5.4. Random erasing: losses and AUC curves

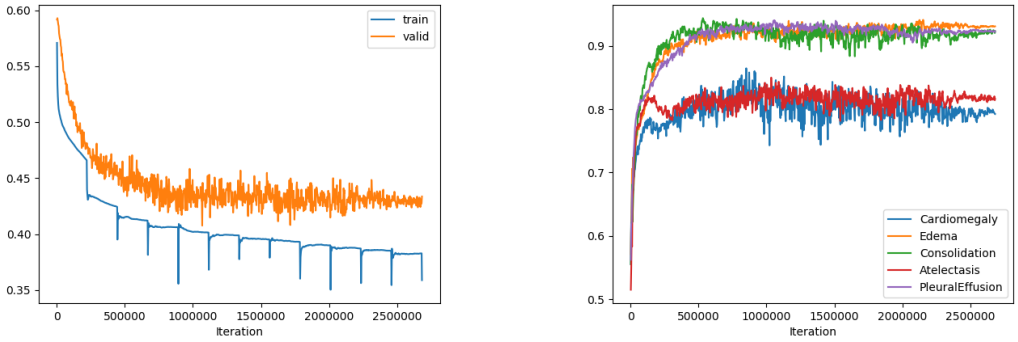


Figure 5.5. Mix-up: losses and AUC curves

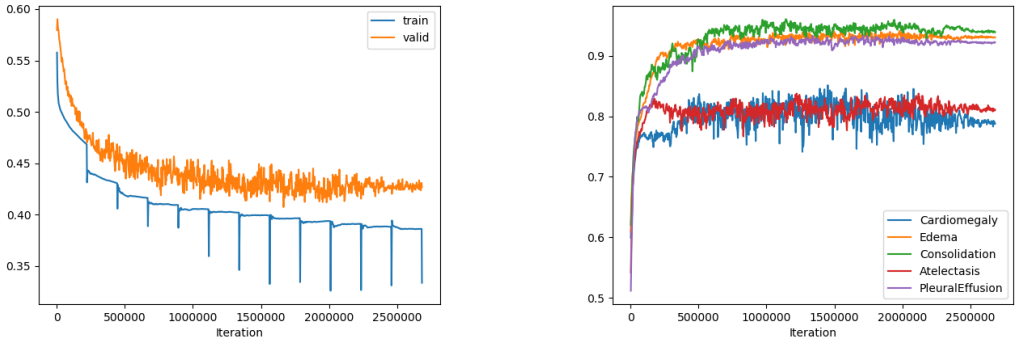


Figure 5.6. Cut-mix erasing: losses and AUC curves

Chapter 6

Conclusion

6.1 Systematic Literature Review conclusion

This Systematic Literature Review presents an analysis of the main data augmentation techniques and strategies used in the medical literature, to solve the problem of lack of high-quality ground-truth data, due to time, money and human resource limitations.

The most common study designs to present and assess data augmentation methods compare the effect of different data augmentation models on downstream tasks.

Affine transformation and **GAN-based** models are the most widespread and used data augmentation techniques to fight the phenomenon of overfitting, with significant increases in the performance of downstream tasks.

Anyway, in most cases, using any data augmentation technique brings advantages in terms of performance of downstream tasks.

Instead, **feature mixing** techniques and **random erasing** are the less used data augmentation strategies in the medical literature. These transformations were initially designed to make computer vision models more robust to occlusions, which do not occur in medical images; in the case of feature mixing, it is also difficult to generate reliable labels for the resulting images. These techniques need further study to solve these problems and perhaps be used in the world of medical data augmentation.

6.2 Experimental conclusion

The experiments performed on *ChestXpert* dataset confirm that feature mixing and random erasing strategies are the less used and less effective data augmentation techniques. They do not have a positive impact on the performance of the model, obtaining lower results than the baseline.

Bibliography

[1] 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. doi: 10.1109/ICCABS.2018.8542071.

[2] 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. doi: 10.1109/IJCNN48605.2020.9207268.

[3] 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. doi: 10.1109/DICTA47822.2019.8946022.

[4] 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. ISSN 1879-0267 0020-1383. doi: 10.1016/j.injury.2020.09.010. Place: Netherlands.

[5] Jamil Ahmad, Khan Muhammad, and Sung Baik. Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search. *PLOS ONE*, 12: e0183838, 08 2017. doi: 10.1371/journal.pone.0183838.

[6] 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. ISSN 1558-254X. doi: 10.1109/TMI.2020.2971258.

[7] 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. doi: 10.1109/EIT.2019.8833768.

[8] Saleh Albahli and Waleed Albattah. Detection of coronavirus disease from x-ray images using deep learning and transfer learning algorithms. 28(5):841–850. ISSN 1095-9114 0895-3996. doi: 10.3233/XST-200720.

[9] Vincent Andrearczyk, Julien Fageot, Valentin Oreiller, Xavier Montet, and Adrien Depeursinge. Local rotation invariance in 3d CNNs. 65:101756. ISSN 1361-8423 1361-8415. doi: 10.1016/j.media.2020.101756. Place: Netherlands.

[10] 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. ISSN 1790-5427. doi: 10.1967/s002449912101. Place: Greece.

[11] 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. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.115271. URL `https://www.sciencedirect.com/science/article/pii/S0957417421007028`.

[12] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-020-02203-1.

[13] 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. 206:106113. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2021.106113. Place: Ireland.

[14] 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. doi: 10.1109/ICIIP47207.2019.8985892.

[15] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-019-01975-5.

[16] Haichao Cao, Shiliang Pu, Wenming Tan, and Junyan Tong. Breast mass detection in digital mammography based on anchor-free architecture. 205:106033. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2021.106033. Place: Ireland.

[17] 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. doi: 10.1109/BHI.2018.8333411.

[18] 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. doi: 10.1117/1.JMI.7.1.012703. URL https://doi.org/10.1117/1.JMI.7.1.012703.

[19] 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. ISBN 978-3-030-20351-1.

[20] 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. ISBN 978-3-319-68127-6.

[21] 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. ISSN 1076-9757.

[22] 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, . ISBN 978-1-4503-6841-4. doi: 10.1145/3338533.3366577. URL https://doi.org/10.1145/3338533.3366577.

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

[24] 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. 71: 102060, . ISSN 1361-8423 1361-8415. doi: 10.1016/j.media.2021.102060.

[25] 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. doi: 10.1109/EMBC.2018.8512305.

[26] 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, . ISSN 2473-4209 0094-2405. doi: 10.1002/mp.14429.

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

[28] Quentin Delannoy, Chi-Hieu Pham, Clément Cazorla, Carlos Tor-Díez, Guillaume Dollé, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. SegSRGAN: Super-resolution and segmentation using generative adversarial networks - application to neonatal brain MRI. 120:103755. ISSN 1879-0534 0010-4825. doi: 10.1016/j.compbiomed.2020.103755. Place: United States.

[29] 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. doi: 10.1109/CICN49253.2020.9242551.

[30] 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. doi: 10.1109/ICIP.2018.8451510.

[31] 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. doi: 10.1109/ISBI48211.2021.9433793.

[32] 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. ISSN 0939-3889. doi: https://doi.org/10.1016/j.zemedi.2020.05.001. URL `https://www.sciencedirect.com/science/article/pii/S0939388920300544`.

[33] 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. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.09.013. URL `https://www.sciencedirect.com/science/article/pii/S0925231218310749`.

[34] 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. doi: 10.1109/ISBI.2018.8363576.

[35] 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. doi: 10.1109/DICTA.2018.8615799.

[36] 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. ISSN 1522-2586 1053-1807. doi: 10.1002/jmri.26544. Place: United States.

[37] 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. doi: 10.1109/WCNCW48565.2020.9124813.

[38] 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. doi: 10.23919/SPA50552.2020.9241305.

[39] 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. ISSN 1558-254X 0278-0062. doi: 10.1109/TMI.2018.2865659. Place: United States.

[40] 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. ISSN 1558-254X 0278-0062. doi: 10.1109/TMI.2019.2946059. Place: United States.

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

[42] Charley Gros, Andreanne Lemay, and Julien Cohen-Adad. Softseg: Advantages of soft versus binary training for image segmentation. *Medical Image Analysis*, 71:102038, 2021. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2021.102038. URL `https://www.sciencedirect.com/science/article/pii/S1361841521000840`.

[43] 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. ISSN 1932-6203. doi: 10.1371/journal.pone.0212550.

[44] 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. doi: 10.1109/ICIP40778.2020.9191127.

[45] 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. ISBN 978-1-4503-6976-3. doi: 10.1145/3357384.3357890. URL `https://doi.org/10.1145/3357384.3357890`.

[46] 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. doi: 10.1109/3DV.2019.00085.

[47] 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. ISSN 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2021.107118. URL `https://www.sciencedirect.com/science/article/pii/S0045790621001221`.

[48] 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.

[49] Zhehao He, Wang Lv, and Jian Hu. A simple method to train the AI diagnosis model of pulmonary nodules. 2020:2812874. ISSN 1748-6718 1748-670X. doi: 10.1155/2020/2812874.

[50] Robert Hemke, Colleen G. Buckless, Andrew Tsao, Benjamin Wang, and Martin Torriani. Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. 49(3):387–395. ISSN 1432-2161 0364-2348. doi: 10.1007/s00256-019-03289-8.

[51] 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. doi: 10.1109/ICCC51575.2020.9344912.

[52] 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. doi: 10.1109/INDIACom51348.2021.00018.

[53] 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. doi: 10.1609/aaai.v33i01.3301590.

[54] 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. ISSN 2473-4209 0094-2405. doi: 10.1002/mp.13695.

[55] 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. ISBN 978-3-030-00934-2.

[56] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-018-1785-8. Place: Germany.

[57] 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. ISSN 1873-5894 0730-725X. doi: 10.1016/j.mri.2019.05.020. Place: Netherlands.

[58] Seung-Taek Kim and Hyo Jong Lee. Lightweight stacked hourglass network for human pose estimation. *Applied Sciences*, 10(18), 2020. ISSN 2076-3417. doi: 10.3390/app10186497. URL `https://www.mdpi.com/2076-3417/10/18/6497`.

[59] Sunho Kim, Byungjai Kim, and HyunWook Park. Synthesis of brain tumor multicontrast MR images for improved data augmentation. 48(5):2185–2198. ISSN 2473-4209 0094-2405. doi: 10.1002/mp.14701. Place: United States.

[60] 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. doi: 10.1109/ISBI.2019.8759563.

[61] 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. doi: 10.1109/IWSSIP.2019.8787315.

[62] 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. doi: 10.1109/TENCON.2019.8929615.

[63] 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. ISSN 0895-6111. doi: https://doi.org/10.1016/j.compmedimag.2020.101812. URL `https://www.sciencedirect.com/science/article/pii/S0895611120301075`.

[64] 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. ISSN 0250-832X 1476-542X. doi: 10.1259/dmfr.20200185.

[65] Nikolaos Kyventidis and Christos Angelopoulos. Intraoral radiograph anatomical region classification using neural networks. 16(3):447–455. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-021-02321-4. Place: Germany.

[66] 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. ISSN 0250-832X 1476-542X. doi: 10.1259/dmfr.20170344.

[67] 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. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2960371.

[68] 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. doi: 10.1109/MCE.2019.2959075.

[69] Joseph Lemley, Shabab Bazrafkan, and Peter Corcoran. Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, 5:5858–5869, 2017. doi: 10.1109/ACCESS.2017.2696121.

[70] 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, . ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2020.105876. Place: Ireland.

[71] Qingyun Li, Zhibin Yu, Yubo Wang, and Haiyong Zheng. TumorGAN: A multi-modal data augmentation framework for brain tumor segmentation. 20(15), . ISSN 1424-8220. doi: 10.3390/s20154203.

[72] 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. doi: 10.1109/ITNEC48623.2020.9084734.

[73] 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, . ISSN 2473-4209 0094-2405. doi: 10.1002/mp.14651. Place: United States.

[74] 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, . ISSN 1869-4098 1869-408X. doi: 10.1007/s13239-019-00425-2. Place: United States.

[75] 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. doi: 10.1109/IPTA.2019.8936106.

[76] 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, . ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-021-02309-0.

[77] 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. ISBN 978-1-4503-7262-6. doi: 10.1145/3364836.3364906. URL https://doi.org/10.1145/3364836.3364906.

[78] 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. doi: 10.1109/ISBI48211.2021.9433893.

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

[80] Toru Miyoshi, Akinori Higaki, Hideo Kawakami, and Osamu Yamaguchi. Automated interpretation of the coronary angioscopy with deep convolutional neural networks. 7(1). ISSN 2053-3624. doi: 10.1136/openhrt-2019-001177.

[81] Lia Morra, Luca Piano, F. Lamberti, and Tatiana Tommasi. Bridging the gap between natural and medical images through deep colorization, 05 2020.

[82] 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. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2020.103698. URL https://www.sciencedirect.com/science/article/pii/S001048252030086X.

[83] Jakub Nalepa, Michal Marcinkiewicz, and Michal Kawulok. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13: 83, 2019. ISSN 1662-5188. doi: 10.3389/fncom.2019.00083. URL https://www.frontiersin.org/article/10.3389/fncom.2019.00083.

[84] Jakub Nalepa, Michal Marcinkiewicz, and Michal Kawulok. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13: 83, 2019. ISSN 1662-5188. doi: 10.3389/fncom.2019.00083. URL `https://www.frontiersin.org/article/10.3389/fncom.2019.00083`.

[85] 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. doi: 10.1109/ICIP.2019.8803423.

[86] 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. doi: 10.1109/ICT-ROBOT.2018.8549917.

[87] 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. ISSN 2045-2322. doi: 10.1038/s41598-020-74539-2.

[88] 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. doi: 10.1109/EMBC.2019.8857669.

[89] 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.

[90] 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. ISSN 2473-4209 0094-2405. doi: 10.1002/mp.13755. Place: United States.

[91] 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, . ISSN 1865-0341 1865-0333. doi: 10.1007/s12194-020-00564-5. Place: Japan.

[92] 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, . ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-019-02092-z. Place: Germany.

[93] 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.

[94] 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. 203:106018. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2021.106018. Place: Ireland.

[95] 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. ISSN 1558-254X. doi: 10.1109/TMI.2019.2914656.

[96] 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. ISSN 2168-2208. doi: 10.1109/JBHI.2018.2879449.

[97] Tuan Pham. A comprehensive study on classification of covid-19 on computed tomography with pretrained convolutional neural networks, 05 2020.

[98] Tuan D. Pham. Geostatistical simulation of medical images for data augmentation in deep learning. *IEEE Access*, 7:68752–68763, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2919678.

[99] 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. doi: 10.1109/BIBE50027.2020.00124.

[100] 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. doi: 10.1109/WiSPNET45539.2019.9032811.

[101] 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. doi: 10.1109/ICSIP49896.2020.9339452.

[102] 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. doi: 10.1109/ICASSP40776.2020.9053403.

[103] Deepta Rajan, Jayaraman J. Thiagarajan, Alexandros Karargyris, and Satyananda Kashyap. Self-training with improved regularization for sample-efficient chest x-ray classification. page 61, 02 2021. doi: 10.1117/12.2582290.

[104] 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. doi: 10.1109/CSDE50874.2020.9411553.

[105] 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. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2019.05.006. Place: Ireland.

[106] 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. doi: 10.1109/ICCCSP49186.2020.9315189.

[107] 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. ISSN 1877-7503. doi: https://doi.org/10.1016/j.jocs.2018.12.003. URL `https://www.sciencedirect.com/science/article/pii/S1877750318307385`.

[108] 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. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2900198.

[109] 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. ISSN 2045-2322. doi: 10.1038/s41598-019-52737-x.

[110] 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. ISSN 1879-176X 0300-5712. doi: 10.1016/j.jdent.2019.103260. Place: England.

[111] 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). ISSN 1999-4915. doi: 10.3390/v12070769.

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

[113] 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. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2021.106019. URL `https://www.sciencedirect.com/science/article/pii/S0169260721000948`.

[114] 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. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2020.105611. Place: Ireland.

[115] 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. doi: 10.1109/ITME.2019.00034.

[116] 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. ISBN 978-3-030-00536-8.

[117] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, Jul 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0197-0. URL `https://doi.org/10.1186/s40537-019-0197-0`.

[118] 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. ISBN 978-1-4503-8910-5. doi: 10.1145/3447587.3447620. URL `https://doi.org/10.1145/3447587.3447620`.

[119] 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. ISSN 1941-0484. doi: 10.1109/JSTSP.2020.3013351.

[120] 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. doi: 10.1109/ICCC51575.2020.9344915.

[121] 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. doi: 10.1109/ICBK50248.2020.00041.

[122] 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. ISSN 2473-4209 0094-2405. doi: 10.1002/mp.14161. Place: United States.

[123] 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. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2941154.

[124] Hitesh Tekchandani, Shrish Verma, and Narendra Londhe. Performance improvement of mediastinal lymph node severity detection using GAN and inception network. 194:105478. ISSN 1872-7565 0169-2607. doi: 10.1016/j.cmpb.2020.105478. Place: Ireland.

[125] Lin Teng, Hang Li, and Shahid Karim. DMCNN: A deep multiscale convolutional neural network model for medical image segmentation. 2019:8597606. ISSN 2040-2309 2040-2295. doi: 10.1155/2019/8597606.

[126] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-021-02308-1. Place: Germany.

[127] 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. doi: 10.1109/CRC.2018.00010.

[128] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-020-02275-z.

[129] Ahammed Muneer K. V, V. R. Rajendran, and Paul Joseph K. Glioma tumor grade identification using artificial intelligent techniques. 43(5):113. ISSN 1573-689X 0148-5598. doi: 10.1007/s10916-019-1228-2. Place: United States.

[130] 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. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2896409.

[131] 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. ISSN 1748-670X. doi: 10.1155/2021/6633755. URL `https://doi.org/10.1155/2021/6633755`.

[132] 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. ISSN 1557-9964. doi: 10.1109/TCBB.2019.2911947.

[133] Marysia Winkels and Taco S. Cohen. Pulmonary nodule detection in CT scans with equivariant CNNs. 55:15–26. ISSN 1361-8423 1361-8415. doi: 10.1016/j.media.2019.03.010. Place: Netherlands.

[134] 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. doi: 10.1109/EMBC44109.2020.9176537.

[135] 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. doi: 10.1109/BIBM47256.2019.8982932.

[136] Ojasvi Yadav, Kalpdrum Passi, and Chakresh Kumar Jain. Using deep learning to classify x-ray images of potential tuberculosis patients. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2368–2375, Dec 2018. doi: 10.1109/BIBM.2018.8621525.

[137] 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. doi: 10.1109/BIBM47256.2019.8983319.

[138] 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. ISSN 1361-8423 1361-8415. doi: 10.1016/j.media.2019.101602.

[139] 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. doi: 10.1109/CISP-BMEI48845.2019.8965817.

[140] Houqiang Yu and Xuming Zhang. Synthesis of prostate MR images for classification using capsule network-based GAN model. 20(20). ISSN 1424-8220. doi: 10.3390/s20205736.

[141] Sangdoo Yun, Dongyoon Han, Sanghyuk Chun, Seong Joon Oh, Youngjoon Yoo, and Junsuk Choe. Cutmix: Regularization strategy to train strong classifiers with localizable features. pages 6022–6031, 10 2019. doi: 10.1109/ICCV.2019.00612.

[142] 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. ISSN 1861-6429 1861-6410. doi: 10.1007/s11548-020-02192-1. Place: Germany.

[143] 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. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.3010863.

[144] 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. ISSN 1558-254X. doi: 10.1109/TMI.2020.2973595.

[145] 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. ISSN 1558-2639 1536-1241. doi: 10.1109/TNB.2018.2845103. Place: United States.

[146] 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. doi: 10.1109/CVPR.2019.00874.

[147] 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. ISSN 2045-7634. doi: 10.1002/cam4.2233.

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

[149] 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. ISSN 2473-4209 0094-2405. doi: 10.1002/mp.14168. Place: United States.