Medical Image Analysis 84 (2023) 102704

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Survey paper

Data synthesis and adversarial networks: A review and meta-analysis in cancer imaging

Richard Osuala [a,*], Kaisar Kushibar [a], Lidia Garrucho [a], Akis Linardos [a], Zuzanna Szafranowska [a], Stefan Klein [b], Ben Glocker [c], Oliver Diaz [a,1], Karim Lekadir [a,1]

[a] Artificial Intelligence in Medicine Lab (BCN-AIM), Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
[b] Biomedical Imaging Group Rotterdam, Department of Radiology & Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
[c] Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK

ARTICLE INFO

Keywords:
Generative Adversarial Network
Adversarial training
Synthetic data
Trustworthiness

ABSTRACT

Despite technological and medical advances, the detection, interpretation, and treatment of cancer based on imaging data continue to pose significant challenges. These include inter-observer variability, class imbalance, dataset shifts, inter- and intra-tumour heterogeneity, malignancy determination, and treatment effect uncertainty. Given the recent advancements in image synthesis, Generative Adversarial Networks (GANs), and adversarial training, we assess the potential of these technologies to address a number of key challenges of cancer imaging. We categorise these challenges into (a) data scarcity and imbalance, (b) data access and privacy, (c) data annotation and segmentation, (d) cancer detection and diagnosis, and (e) tumour profiling, treatment planning and monitoring. Based on our analysis of 164 publications that apply adversarial training techniques in the context of cancer imaging, we highlight multiple underexplored solutions with research potential. We further contribute the Synthesis Study Trustworthiness Test (*SynTRUST*), a meta-analysis framework for assessing the validation rigour of medical image synthesis studies. *SynTRUST* is based on 26 concrete measures of thoroughness, reproducibility, usefulness, scalability, and tenability. Based on *SynTRUST*, we analyse 16 of the most promising cancer imaging challenge solutions and observe a high validation rigour in general, but also several desirable improvements. With this work, we strive to bridge the gap between the needs of the clinical cancer imaging community and the current and prospective research on data synthesis and adversarial networks in the artificial intelligence community.

1. Introduction

1.1. The burden of cancer and early detection

The evident improvement in global cancer survival in the last decades is arguably attributable not only to health care reforms, but also to advances in clinical research (e.g., targeted therapy based on molecular markers) and diagnostic imaging technology e.g whole-body magnetic resonance imaging (MRI) (Messiou et al., 2019), and positron emission tomography–computed tomography (PET-CT) (Arnold et al., 2019). Nonetheless, cancers still figure among the leading causes of morbidity and mortality worldwide (Ferlay et al., 2015), with an approximated 9.6 million cancer related deaths in 2018 (World Health Organization, 2018). The most frequent cases of cancer death worldwide in 2018 are lung (1.76 million), colorectal (0.86 million), stomach (0.78 million), liver (0.78 million), and breast (0.63 million) (World Health Organization, 2018). These figures are prone to continue to increase in consequence of the ageing and growth of the world population (Jemal et al., 2011).

A large proportion of the global burden of cancer could be prevented due to treatment and early detection (Jemal et al., 2011). For example, an early detection can provide the possibility to treat a tumour before it acquires critical combinations of genetic alterations (e.g., metastasis with evasion of apoptosis Hanahan and Weinberg, 2000). Solid tumours become detectable by medical imaging modalities only at an approximate size of 10^9 cells ($\approx 1\ cm^3$) after evolving from a single neoplastic cell typically following a Gompertzian (Norton et al., 1976) growth pattern (Frangioni, 2008).[2] To detect and diagnose tumours, radiologists

* Corresponding author.
E-mail address: richard.osuala@ub.edu (R. Osuala).
[1] Authors contributed equally
[2] In vitro studies reported a *theoretical* detection limit around 10^5 to 10^6 for human cancer cell lines using PET. In clinical settings, the *theoretical* detection limit is larger and depends, among others, on background radiation, cancer cell line, and cancer type (Fischer et al., 2006).

https://doi.org/10.1016/j.media.2022.102704
Received 23 July 2021; Received in revised form 2 November 2022; Accepted 21 November 2022
Available online 24 November 2022

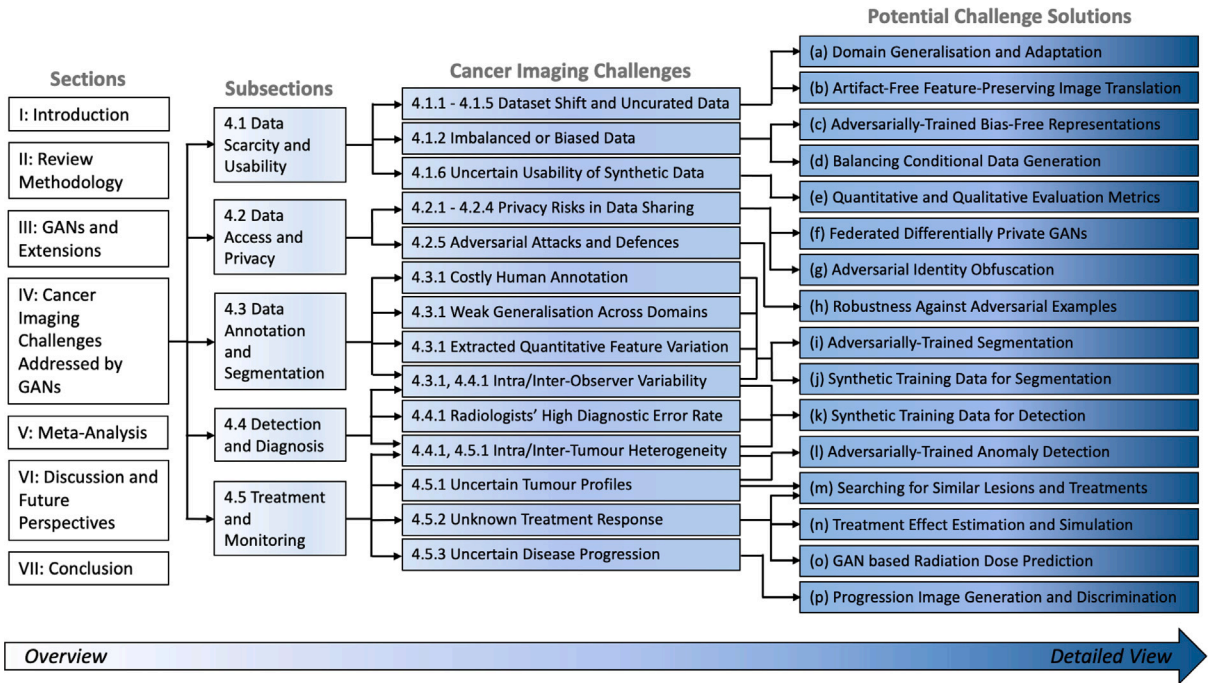

Fig. 1. Section organisation: Illustration of the structure of the present paper starting with paper sections on the left and going into detail towards the right culminating in a selection of cancer imaging challenges and solutions. These solutions (a)–(p) are part of the solutions found in the surveyed GAN literature or are proposed extensions thereof, as is further discussed in Section 4.

inspect, normally by visual assessment, medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), ultrasound (US), X-ray mammography (MMG), PET (Frangioni, 2008; Itri et al., 2018; McCreadie and Oliver, 2009).

Medical imaging data evaluation is time demanding and therefore costly in nature. In addition, volumes of new technologies (e.g., digital breast tomosynthesis Swiecicki et al., 2021) become available and studies generally show an extensive increase in analysable imaging volumes (McDonald et al., 2015). Also, the diagnostic quality in radiology varies and is very much dependent on the personal experience, skills and invested time of the data examiner (Itri et al., 2018; Elmore et al., 1994; Woo et al., 2020). Hence, to decrease cost and increase quality, automated or semi-automated diagnostic tools can be used to assist radiologists in the decision-making process. Such diagnostic tools comprise traditional machine learning, but also recent deep learning methods, which promise an immense potential for detection performance improvement in radiology.

1.2. The promise of deep learning and the need for data

The rapid increase in graphics processing unit (GPU) processing power has allowed training deep learning algorithms such as convolutional neural networks (CNNs) (Fukushima, 1980; LeCun et al., 1989, 1998) on large image datasets achieving impressive results in Computer Vision (CireAan et al., 2012; Krizhevsky et al., 2012), and Cancer Imaging (Cireşan et al., 2013). In particular, the success of AlexNet in the 2012 ImageNet challenge (Krizhevsky et al., 2012) triggered an increased adoption of deep neural networks to a multitude of problems in numerous fields and domains including medical imaging, as reviewed in Shen et al. (2017), Zhou et al. (2021) and Litjens et al. (2017). Despite the increased use of medical imaging in clinical practice, the public availability of medical imaging data remains limited (McDonald et al., 2015). This represents a key impediment for the training, research, and use of deep learning algorithms in radiology and oncology. Clinical centres refrain from sharing such data for ethical, legal, technical, and financial (e.g., costly annotation) reasons (Bi et al., 2019).

Such cancer imaging data not only is necessary to train deep learning models, but also to provide them with sufficient learning possibility to acquire robustness and generalisation capabilities. We define robustness as the property of a predictive model to remain accurate despite of variations in the input data (e.g., noise levels, resolution, contrast, etc.). We refer to a model's generalisation capability as its property of preserving predictive accuracy on new data from unseen sites, hospitals, scanners, etc. Both of these properties are in particular desirable in cancer imaging considering the frequent presence of biased or unbalanced data with sparse or noisy labels.[3] Both robustness and generalisation are essential to demonstrate the trustworthiness of a deep learning model for usage in a clinical setting, where every edge-case needs to be detected and a false negative can potentially cost the life of a patient.

1.3. Synthetic cancer imaging data

We hypothesise that the variety of data needed to train robust and well-generalising deep learning models for cancer images can be largely synthetically generated using Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). The adversarial learning scheme in GANs is based on a generator that generates synthetic (alias 'fake') samples of a target distribution trying to fool a discriminator, which classifies these samples as either real or fake. Various papers have provided reviews of GANs in the medical imaging domain (Yi et al., 2019; Kazeminia et al., 2020; Tschuchnig et al., 2020; Sorin et al., 2020; Lan et al., 2020; Singh and Raza, 2020), but they focused on general presentation of the main methods and possible applications. In cancer imaging, however, there are specificities and challenges that call for specific implementations and solutions based on GANs and the adversarial learning scheme at large, including:

(i) the small size and complexity of cancerous lesions

[3] Alongside tumour manifestation heterogeneity, and multi-centre, multi-organ, multi-modality, multi-scanner, and multi-vendor data.

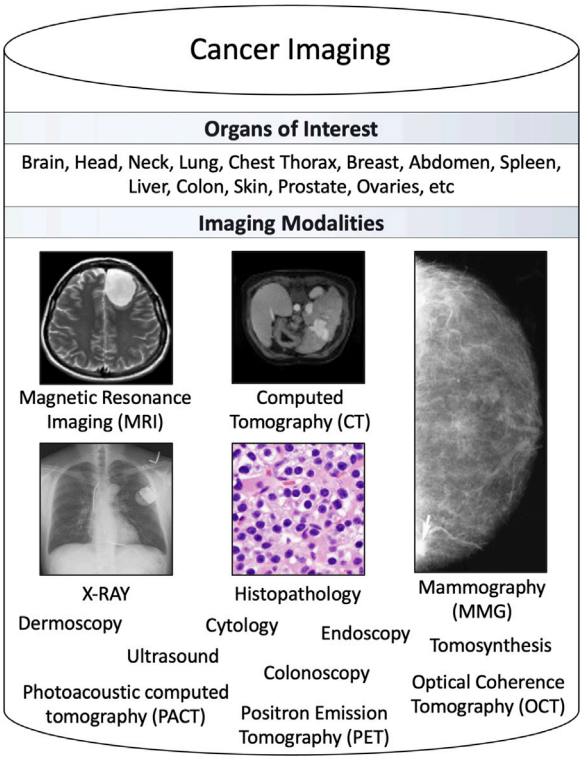

Fig. 2. Overview of the most common organs and modalities targeted by the surveyed cancer imaging publications. A respective histogram that shows the number of papers per modality and per organ can be found in Fig. 15.

(ii) the high heterogeneity between tumours within as well as between patients and cancer types
(iii) the difficulty to annotate, delineate and label cancer imaging studies at large scale
(iv) the high data imbalance in particular between healthy and pathological subjects or between benign and malignant cases
(v) the difficulty to gather large consented datasets from highly vulnerable patients undergoing demanding care plans

Hence, the present paper contributes a unique perspective and comprehensive analysis of adversarial networks attempting to address the specific challenges in the cancer imaging domain. To the authors' best knowledge, this is the first survey that exclusively focuses on GANs and adversarial training in cancer imaging. In this context, we define cancer imaging as the entirety of approaches for research, diagnosis, and treatment of cancer based on medical images. Our survey comprehensively analyses cancer imaging GAN and adversarial training applications focusing on radiology modalities. As presented in Fig. 2, we recognise that non-radiology modalities are also widely used in cancer imaging. For this reason, we do not restrict the scope of our survey to radiology, but rather also analyse relevant publications in these other modalities including histopathology and cytopathology (e.g., in Section 4.5), and dermatology (e.g., in Sections 4.3 and 4.4).

Further, our survey uncovers and highlights promising research directions for adversarial networks and image synthesis that can facilitate the sustainable adoption of AI in clinical oncology and radiology.

1.4. Section organisation

The remainder of this paper is organised as follows. In Section 2, we introduce the methodology of this review. Section 3 provides an overview of GANs and highlights extensions of the adversarial learning framework relevant to cancer imaging. Section 4 contains the main contribution that encompasses the systematic review of challenges of cancer imaging and potential solutions based on adversarial networks. This organisation is depicted in more detail in Fig. 1.

The different challenges are categorised into groups in the Sections 4.1, 4.2, 4.3, 4.4, and 4.5. Each of the challenges categories contains several specific cancer imaging challenges, which we introduce and discuss in 4.1.1–4.5.3. The sections are organised in an independent way allowing the reader to directly jump to a particular cancer imaging category (4.1–4.5) of interest without requiring context from previous sections. For each of the specific challenges, we survey and discuss potential solutions, as depicted in Fig. 1(a)–(p).

The subsequent Section 5 contains our second core contribution, which consists of the *SynTRUST* framework for systematic analysis of trustworthiness criteria of image synthesis and adversarial training publications in medical imaging. Based on this framework, we meta-analyse a set of studies selected based on their strong performance and promising methodology for solving a specific cancer imaging challenge.

After learning how and to what extent image synthesis and adversarial training solutions have addressed cancer imaging challenges in the past, we highlight and discuss prospective avenues of future research in the Discussion Section 6 and point out unexploited potential of image synthesis and adversarial networks in cancer imaging.

2. Review methodology

Our review comprises two comprehensive literature screening processes. The first screening process surveyed the current challenges in the field of cancer imaging with a focus on radiology imaging modalities. After screening and gaining a deepened understanding of AI-specific and general cancer imaging challenges, we grouped these challenges for further analysis into the following five categories.

- *Data scarcity and usability challenges* (Section 4.1); discussing dataset shifts, class imbalance, fairness, generalisation, domain adaptation and the evaluation of synthetic data.
- *Data access and privacy challenges* (Section 4.2); comprising patient data sharing under privacy constraints, security risks, and adversarial attacks.
- *Data annotation and segmentation challenges* (Section 4.3); discussing costly human annotation, high inter and intra-observer variability, and the consistency of extracted quantitative features.
- *Detection and diagnosis challenges* (Section 4.4); analysing the challenges of high diagnostic error rates among radiologists, early detection, and detection model robustness.
- *Treatment and monitoring challenges* (Section 4.5); examining challenges of high inter and intra-tumour heterogeneity, phenotype to genotype mapping, treatment effect estimation and disease progression.

The second screening process comprised first of a generic and second a specific literature search to find all papers that apply adversarial learning (i.e. GANs) to cancer imaging. In the generic literature search, generic search queries such as 'Cancer Imaging GAN', 'Tumour GANs' or 'Nodule Generative Adversarial Networks' were used to recall a high number of papers. The specific search focused on answering key questions of interest to the aforesaid challenges such as 'Carcinoma Domain Adaptation Adversarial', 'Skin Melanoma Detection GAN', 'Brain Glioma Segmentation GAN', or 'Cancer Treatment Planning GAN'.

In Section 4, we map the papers that propose adversarial training and GAN applications applied to cancer imaging (second screening) to the surveyed cancer imaging challenges (first screening). The mapping of these GAN-related papers to challenge categories facilitates analysing the extent to which existing solutions solve the current cancer imaging challenges and helps to identify gaps and further potential for adversarial networks in this field. The mapping is based on the evaluation criteria used in the GAN-related papers and on the relevance

of the reported results to the corresponding section. For example, if a GAN generates synthetic data that is used to train and improve a tumour detection model, then this paper is assigned to the detection and diagnosis challenge Section 4.4. If a papers describes a GAN that improves a segmentation model, then this paper is assigned to the segmentation and annotation challenge Section 4.3, and so forth.

To gather the literature (e.g., first papers describing cancer imaging challenges, second papers proposing GAN solutions), we have searched in medical imaging, computer science and clinical conference proceedings and journals, but also freely on the web using the search engines Google, Google Scholar, and PubMed. After retrieving all papers with a title related to the subject, their abstract was read to filter out non-relevant papers. A full-text analysis was done for the remaining papers to determine whether they were to be included into our manuscript. We analysed the reference sections of the included papers to find additional relevant literature, which also underwent filtering and full-text screening. Applying this screening process, we reviewed and included a total of 164 GAN and adversarial training cancer imaging publications comprising both peer-reviewed articles and conference papers, but also relevant preprints from arXiv and bioRxiv.

Details about these 164 cancer imaging applications can be found in Tables 2–6. The distribution of these publications across challenge category, year, modality, and anatomy is outlined in Fig. 15.

The methodology for deriving and applying the *SynTRUST* meta-analysis framework, which assesses the validity and trustworthiness of medical image synthesis studies, is provided in Section 5.

3. GANs and extensions

3.1. Introducing the theoretical underpinnings of GANs

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are a type of generative model with a differentiable generator network (Goodfellow et al., 2016). GANs are formalised as a minimax two-player game, where the generator network (G) competes against an adversary network called discriminator (D). As visualised in Fig. 3, given a random noise distribution z, G generates samples $x = G(z; \theta(g))$ that D classifies as either real (drawn from training data, i.e. $x \sim p_{data}$) or fake (drawn from G, i.e. $x \sim p_g$). x is either sampled from p_{data} or from p_g with a probability of 50%. D outputs a value $p = D(x; \theta(d))$ indicating the probability that x is a real training example rather than one of G's fake samples (Goodfellow et al., 2016). As defined by Goodfellow et al. (2014), the task of the discriminator can be characterised as binary classification (CLF) of samples x. Hence, the discriminator can be trained using binary-cross entropy resulting in the following loss function L_D:

$$L_D = -\mathbb{E}_{x\sim p_{data}}[log\, D(x)] + \mathbb{E}_{z\sim p_z}[log(1 - D(G(z)))] \tag{1}$$

D's training objective is to minimise L_D (or maximise $-L_D$) while the goal of the generator is the opposite (i.e. minimise $-L_D$) resulting in the value function $V(G, D)$ of a two-player zero-sum game between D and G:

$$\min_G \max_D V(D,G) = \min_G \max_D [\mathbb{E}_{x\sim p_{data}}[log\, D(x)] + \mathbb{E}_{z\sim p_z}[log(1 - D(G(z)))]] \tag{2}$$

In theory, in convergence, the generator's samples become indistinguishable from the real training data ($x \sim p_{data} = p_g$) and the discriminator outputs $p = 1/2$ for any given sample x (Goodfellow et al., 2016). As this is a state where both D and G cannot improve further on their objective by changing only their own strategy, it represents a Nash equilibrium (Farnia and Ozdaglar, 2020; Nash et al., 1950). In practice, achieving convergence for this or related adversarial training schemes is an open research problem (Kodali et al., 2017; Mescheder et al., 2018; Farnia and Ozdaglar, 2020).

3.2. Extensions of the Vanilla GAN methodology

As indicated by Fig. 4, numerous extensions of GANs have shown to generate synthetic images with high realism (Karras et al., 2017, 2019, 2020; Chan et al., 2020) and under flexible conditions (Mirza and Osindero, 2014; Odena et al., 2017; Park et al., 2018). GANs have been successfully applied to generate high-dimensional data such as images and, more recently, have also been proposed to generate discrete data (Hjelm et al., 2017). Apart from image generation, GANs have also widely been proposed and applied for paired and unpaired image-to-image translation, domain-adaptation, data augmentation, image inpainting, image perturbation, super-resolution, and image registration and reconstruction (Yi et al., 2019; Kazeminia et al., 2020; Wang et al., 2019b).

Table 1 introduces a selection of common GAN extensions found to be frequently applied to cancer imaging. For each GAN methodology in this and the Tables 1–6, we define the 'Task' describing the application of the respective adversarial network. For instance, in 'noise-to-image synthesis' the input into the generator G consists of a noise vector that G translates into an image. A further input into G can be a class label as in 'class-conditional-image-synthesis' based on which an output is generated that corresponds to this class. Paired and unpaired translation refer to the task where the input into G is a sample (e.g. an image in the source domain) based on which G generates another sample (e.g. an image in the target domain). This translation is paired if the training data consists of target and source domain sample pairs. The key characteristics of each of the GAN extensions of Table 1 are described in the following paragraphs.

3.2.1. Noise-to-image GAN extensions

As depicted in blue in Fig. 3, cGAN adds a discrete label as conditional information to the original GAN architecture that is provided as input to both generator and discriminator to generate class conditional samples (Mirza and Osindero, 2014).

AC-GAN feeds the class label only to the generator while the discriminator is tasked with correctly classifying both the class label and whether the supplied image is real or fake (Odena et al., 2017).

WGAN is motivated by mathematical rationale and based on the Wasserstein-1 distance (alias 'earth mover distance' or 'Kantorovich distance') between two distributions. WGAN extends on the theoretic formalisation and optimisation objective of the vanilla GAN to better approximate the distribution of the real data. By applying an alternative loss function (i.e. Wasserstein loss), the discriminator (alias 'critic' or 'f_w') maximises – and the generator minimises – the difference between the critic's scores for generated and real samples. A important benefit of WGAN is the empirically observed correlation of the loss with sample quality, which helps to interpret WGAN training progress and convergence (Arjovsky et al., 2017).

In WGAN, the weights of the critic are clipped, which means they have to lie within a compact space $[-c, c]$. This is needed to fulfil that the critic is constraint to be in the space of 1-Lipschitz functions. With clipped weights, however, the critic is biased towards learning simpler functions and prone to have exploding or vanishing gradients if the clipping threshold c is not tuned with care (Gulrajani et al., 2017; Arjovsky et al., 2017).

In WGAN-GP, the weight clipping constraint is replaced with a gradient penalty. Gradient penalty of the critic is a tractable and soft version of the following notion: By constraining that the norm of the gradients of a differentiable function is at most 1 everywhere, the function (i.e. the critic) would fulfil the 1-Lipschitz criterion without the need of weight clipping. Compared, among others, to WGAN, WGAN-GP was shown to have improved training stability (i.e. across many different GAN architectures), training speed, and sample quality (Gulrajani et al., 2017).

DCGAN generates realistic samples using a convolutional network architecture with batch normalisation (Ioffe and Szegedy, 2015) for

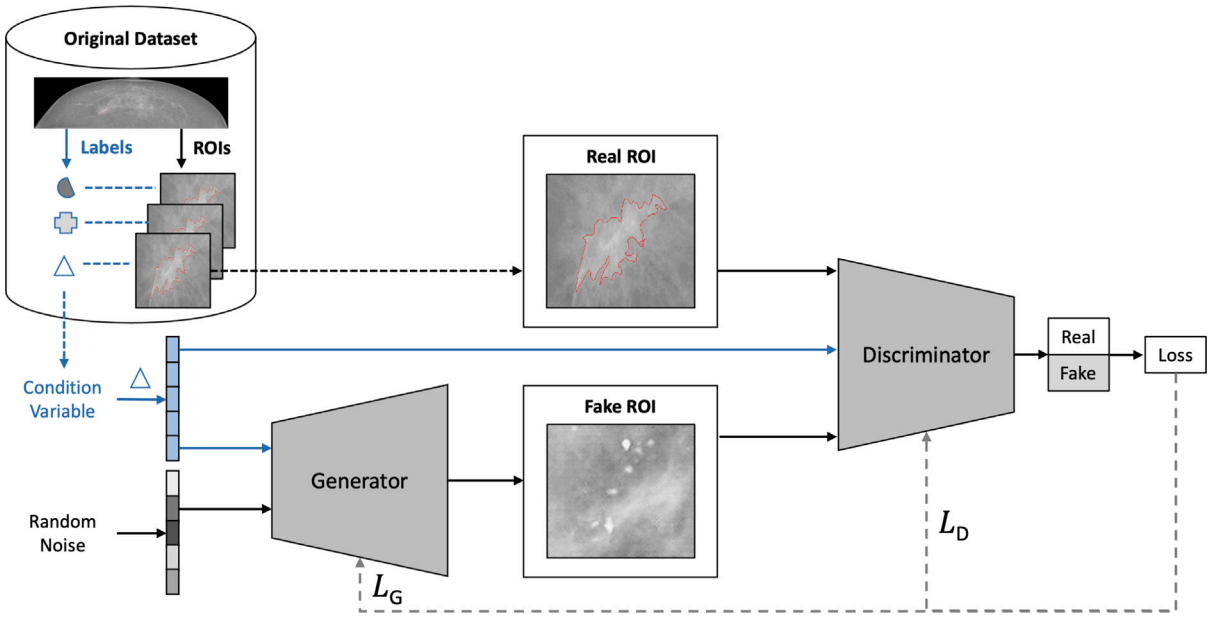

Fig. 3. An example of a generic GAN architecture applied to generation of synthetic mammography region of interest (ROI) images based on the INbreast dataset (Moreira et al., 2012). Note that including the 'Condition' depicted in blue colour extends the vanilla GAN architecture (Goodfellow et al., 2014) to the cGAN architecture (Mirza and Osindero, 2014).

both generator and discriminator and progressively increases the spatial dimension in the layers of the generator using transposed convolution (alias 'fractionally-strided convolution') (Radford et al., 2015).

PGGAN is tested with loss and configurations introduced in WGAN GP. It starts by generating low pixel resolution images, but progressively adds new layers to the generator and discriminator during training resulting in increased pixel resolution and finer image details. It is suggested that after early convergence of initial low-resolution layers, the introduced additional layers enforce the network to only refine the learned representations by increasingly smaller-scale effects and features (Karras et al., 2017).

In SRGAN, the generator transforms a low-resolution (LR) to a high-resolution (HR, alias 'super-resolution') image, while the discriminator learns to distinguish between real high-resolution images and fake super-resolution images. Apart from an adversarial loss, a perceptual loss called 'content loss' measures how well the generator represents higher level image features. This content loss is computed as the euclidean distance between feature representations of the reconstructed image and the reference image based on feature maps of a pretrained 19 layer VGG (Simonyan and Zisserman, 2014) network (Ledig et al., 2017).

3.2.2. Image-to-image GAN extensions

In image-to-image translation, a mapping is learned from one image distribution to another. For example, images from one domain can be transformed to resemble images from another domain via a mapping function implemented by a GAN generator.

CycleGAN achieves realistic unpaired image-to-image translation using two generators (G, F) with one traditional adversarial loss each and an additional cycle-consistency loss. Unpaired image-to-image translation transforms images from domain X to another domain Y in the absence of paired training data i.e. corresponding image pairs for both domains. In CycleGAN, the input image x from domain X is translated by generator $G(x)$ to resemble a sample from domain Y. Next, the sample is translated back from domain Y to domain X by generator $F(G(x))$. The cycle consistency loss enforces that $F(G(x)) \approx x$ (forward cycle consistency) and that $G(F(y)) \approx y$ (backward cycle consistency) (Zhu et al., 2017).

Both pix2pix and SPADE are used in paired image-to-image translation where corresponding image pairs for both domains X and Y are available. pix2pix (alias 'condGAN') is a conditional adversarial network that adapts the U-Net architecture[4] (Ronneberger et al., 2015) for the generator to facilitate encoding an conditional input image into a latent representation before decoding it back into an output image. pix2pix uses L1 loss to enforce low level (alias 'low frequency') image reconstruction and a patch-based discriminator ('PatchGAN') to enforce high level (alias 'high frequency') image reconstruction that the authors suggest to interpret as texture/style loss. Note that the input into the PatchGAN discriminator is a concatenation[5] of the original image (i.e. the generator's input image; e.g. this can be a segmentation map) and the real/generated image (i.e. the generator's output image) (Isola et al., 2017).

In SPADE, the generator architecture does not rely on an encoder for downsampling, but uses a conditional normalisation method during upsampling instead: A segmentation mask as conditional input into the SPADE generator is provided to each of its upsampling layers via spatially-adaptive residual blocks. These blocks embed the masks and apply two two-layer convolutions to the embedded mask to get two tensors with spatial dimensions. These two tensors are multiplied/added to each upsampling layer prior to its activation function. The authors demonstrate that this type of normalisation achieves better fidelity and preservation of semantic information in comparison to other normalisation methods that are commonly applied in neural networks (e.g., Batch Normalisation). The multi-scale discriminators and the loss functions from pix2pixHD (Wang et al., 2018a) are adapted in SPADE, which contains a hinge loss (i.e. as substitute of the adversarial loss), a perceptual loss, and a feature matching loss (Park et al., 2019).

3.2.3. GAN network architectures and adversarial loss

For further methodological detail on the aforementioned GAN methods, loss functions, and architectures, we point the interested reader to the GAN methods review by Wang et al. (2019b). Due to the

[4] To reduce information loss in latent space compression, U-Net uses skip connections between corresponding layers (e.g., first to last) in the encoder and decoder.

[5] Note the concatenation of real_A and fake_B before computing the loss in the discriminator backward pass (L93) in the authors' pix2piximplementation.

Table 1
A selection of the GAN architectures that we found to be the ones most frequently applied in cancer imaging.

Publication	Input G	Input D	Losses	Task
Noise to image				
GAN (Goodfellow et al., 2014)	Noise	Image	Binary cross-entropy based adversarial loss (L_{ADV})	Noise-to-image synthesis
Conditional GAN (cGAN) (Mirza and Osindero, 2014)	Noise & label	Image & label	L_{ADV}	Class-conditional image synthesis
Auxiliary classifier GAN (AC-GAN) (Odena et al., 2017)	Noise & label	Image	L_{ADV} & cross-entropy loss (label classification)	Class-conditional image synthesis
Deep convolutional GAN (DCGAN) (Radford et al., 2015)	Noise	Image	L_{ADV}	Noise-to-image synthesis
Wasserstein GAN (WGAN) (Arjovsky et al., 2017)	Noise	Image	Wasserstein loss (L_{WGAN})	Noise-to-image synthesis
WGAN gradient penalty (WGAN GP) (Gulrajani et al., 2017)	Noise	Image	L_{WGAN} with GP ($L_{WGAN-GP}$)	Noise-to-image synthesis
Progressively growing GAN (PGGAN) (Karras et al., 2017)	Noise	Image	$L_{WGAN-GP}$	Noise-to-image synthesis
Image to Image				
Super-Resolution GAN (SRGAN) (Ledig et al., 2017)	Image (LR)	Image (HR)	L_{ADV} & content loss (based on VGG features)	Super-resolution
CycleGAN (Zhu et al., 2017)	Source image	Target image	L_{ADV} & cycle consistency loss & identity loss	Unpaired image-to-image translation
pix2pix (Isola et al., 2017)	Source image	Concatenated source and target images	L_{ADV} & reconstruction loss (i.e. L1)	Paired image-to-image translation
SPatially-adaptive (DE)normalization (SPADE) (Park et al., 2019)	Noise or encoded source image & segmentation map	Concatenated target image and segmentation map	Hinge & perceptual & feature matching losses (from Wang et al., 2018a)	Paired image-to-image translation

image processing capabilities of CNNs (LeCun et al., 1989), the above-mentioned GAN architectures generally rely on CNN layers internally. Recently, TransGAN (Jiang et al., 2021) and VQGAN (Esser et al., 2021) were proposed, which diverges from the CNN design pattern to using Transformer Neural Networks (Vaswani et al., 2017). Due to the promising performances of these approaches in computer vision tasks, we encourage future studies to investigate the potential of transformer-based GANs for applications in medical and cancer imaging.

Multiple deep learning architectures apply the adversarial loss proposed in Goodfellow et al. (2014) together with other loss functions (e.g., segmentation loss functions) for other tasks than image generation (e.g., image segmentation). This adversarial loss is useful for unsupervised learning of features and representations that are invariant to some part of the training data. For instance, adversarial learning can be useful to discriminate a domain to learn domain-invariant representations (Ganin and Lempitsky, 2015), as has been successfully demonstrated for medical images (Kamnitsas et al., 2017). Such methods that apply the adversarial loss internally are referred to as 'adversarial training' methods and are included in the scope of our survey. That is, we include and consider all relevant cancer imaging papers that apply or build upon the adversarial learning scheme defined in Goodfellow et al. (2014), which comprises GANs as well as adversarial training methods.

4. Cancer imaging challenges addressed by data synthesis and adversarial networks

In this section we follow the structure presented in Fig. 1, where we categorise cancer imaging challenges into five categories consisting of *data scarcity and usability* (4.1), *data access and privacy* (4.2), *data annotation and segmentation* (4.3), *detection and diagnosis* (4.4), and *treatment and monitoring* (4.5). In each subsection, we group and analyse respective cancer imaging challenges and discuss the potential and the limitations of corresponding GAN-based data synthesis and adversarial training solutions. In this regard, we also identify and highlight key needs to be addressed by researchers in the field of cancer imaging GANs towards solving the surveyed cancer imaging challenges. We provide respective Tables 2–6 for each Sections 4.1–4.5 containing relevant information (publication, method, dataset, modality, task, highlights) for all of the reviewed cancer imaging GAN solutions.

Chronology of key innovations. The most commonly applied adversarial network methodologies in cancer imaging are summarised chronologically in Fig. 4. Next to each network (a)–(m), the number of occurrence per cancer imaging challenge category 4.1–4.5 is highlighted.

Following Vanilla GANs 4(a), four main lines of innovations have been widely adopted in cancer imaging. These are methods that condition the synthetic data generation e.g. cGAN 4(b), methods that improve upon the network architecture e.g. DCGAN 4(c), methods that improve upon the adversarial loss function e.g. WGAN 4(g), and methods that backpropagate the adversarial loss for representation learning, e.g. domain-invariant representations 4(d).

As to conditional methods, further key innovations have been AC-GAN's 4(f) discriminator classifying the input condition, and methods that conditioning the generation based on an input image using additional reconstruction (e.g., pix2pix 4(e), cycleGAN 4(h)) or perceptual (e.g., SRGAN 4(i)) losses. Recent approaches (e.g., SPADE 4(l)) innovate regarding how the input image is provided to the generator network, e.g., via spatially-adaptive residual blocks in upsampling layers.

WGAN's 4(g) loss based on the discriminator estimating the Wasserstein-1 distance between real and synthetic image distributions is a widely used and extended (e.g., WGAN-GP 4(j)) alternative to the vanilla binary-cross entropy adversarial loss in cancer imaging.

The architectural innovation of progressive network growing 4(k) unlocked high-resolution cancer image generation and is adopted by recent approaches such as StyleGAN 4(m), which introduced adaptive instance normalisation and pioneered noise (and style condition) input via intermediate activation maps.

4.1. Data scarcity and usability challenges

4.1.1. Challenging dataset sizes and shifts

Although data repositories such as The Cancer Imaging Archive (TCIA) (Clark et al., 2013) have made a wealth of cancer imaging

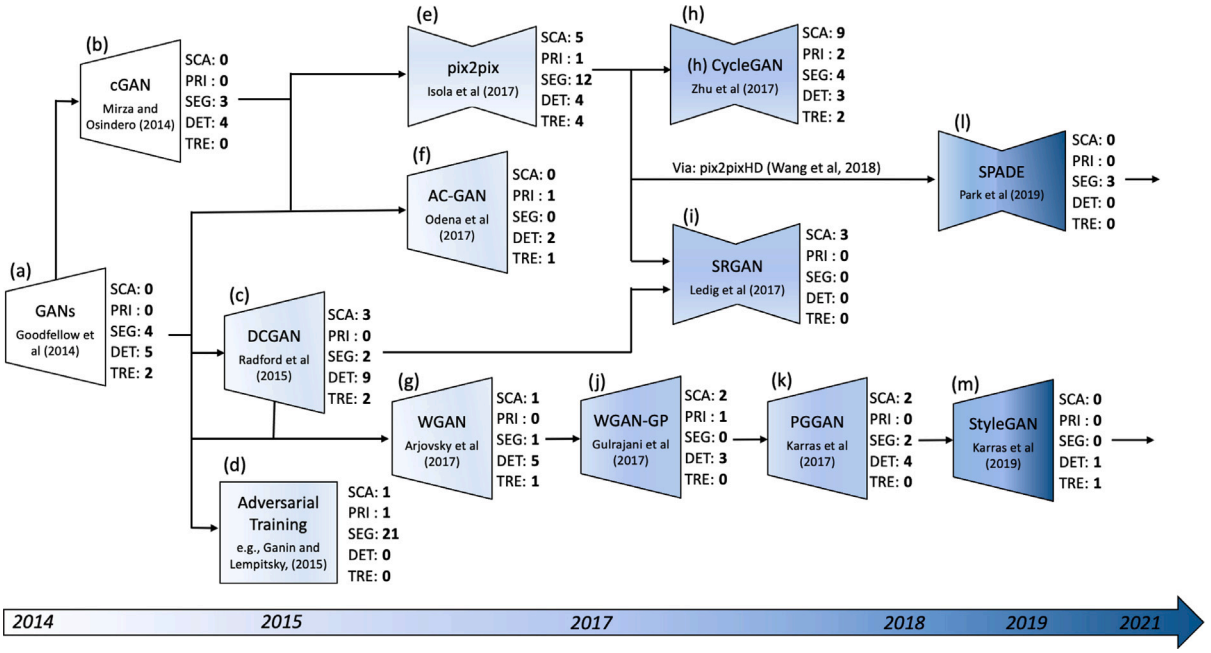

Fig. 4. The chronological evolution of adversarial networks in cancer imaging based on the commonly applied GAN types (a)–(m). The shape of the respective GAN indicates its mapping function, e.g., a trapezium represents a noise-to-image mapping function. For each cancer imaging challenge category surveyed in 4, the number of occurrences of each GAN type (a)–(m) is counted across reviewed publications. The illustration abbreviates challenge categories as SCA: 4.1 Data scarcity and usability challenges, PRI: 4.2 Data access and privacy challenges, SEG: 4.3 Data annotation and segmentation challenges, DET: 4.4 Detection and diagnosis challenges, TRE: 4.5 Treatment and monitoring challenges.

data available for research, the demand is still far from satisfied. As a result, data augmentation techniques are widely used to artificially enlarge the existing datasets, traditionally including simple spatial (e.g., flipping, rotation) or intensity transformations (e.g., noise insertion) of the true data. GANs have shown promise as a more advanced augmentation technique and have already seen use in medical and cancer imaging (Han et al., 2018; Yi et al., 2019).

Aside from the issue of lacking sizeable data, data scarcity often forces studies to be constrained on small-scale single-centre datasets. The resulting findings and models are likely to not generalise well due to diverging distributions between the (synthetic) datasets seen in training and those seen in testing or after deployment, a phenomenon known as dataset shift (Quionero-Candela et al., 2009).[6] An example of this in clinical practice are cases where training data is preselected from specific patient sub-populations (e.g., only high-risk patients) resulting in bias and limited generalisability to the broad patient population (Troyanskaya et al., 2020; Bi et al., 2019).

From a causality perspective, dataset shift can be split into several distinct scenarios (Castro et al., 2020):

- *Population shift*, caused by differences in age, sex, ethnicities etc.
- *Acquisition shift*, caused by differences in scanners, resolution, contrast etc.
- *Annotation shift*, caused by differences in annotation policy, annotator experience, segmentation protocols etc.
- *Prevalence shift*, caused by differences in the disease prevalence in the population, often resulting from artificial sampling of data
- *Manifestation shift*, caused by differences in how the disease is manifested

GANs may inadvertently introduce such types of dataset shifts (e.g., due to mode collapse Goodfellow et al., 2014), but it has been shown that this shift can be studied, measured and avoided (Santurkar et al., 2018; Arora et al., 2018). GANs can be a sophisticated tool for data augmentation or curation (Diaz et al., 2021) and by calibrating the type of shift introduced, they have the potential to turn it into an advantage, generating diverse training data that can help models generalise better to unseen target domains. The research line studying this problem is called *domain generalisation*, and it has presented promising results for harnessing adversarial models towards learning of domain-invariant features (Zhou et al., 2021). GANs and adversarial training have been used in various ways in this context, using multi-source data to generalise to unseen targets (Rahman et al., 2019; Li et al., 2018) or in unsupervised domain generalisation using adaptive data augmentation to append adversarial examples iteratively (Volpi et al., 2018). As indicated in Fig. 1(a), the domain generalisation research line has recently been further extended to cancer imaging (Lafarge et al., 2019; Chen et al., 2021).

In the following, further cancer imaging challenges in the realm of data scarcity and usability are described and related GAN solutions are referenced. Given these challenges and solutions, we derive a workflow for clinical adoption of (synthetic) cancer imaging data, which is illustrated in Fig. 5.

4.1.2. Imbalanced data and fairness

Apart from the rise of data-hungry deep learning solutions and the need to cover the different organs and data acquisition modalities, a major problem that arises from data scarcity is that of imbalance—i.e. the overrepresentation of a certain type of data over others (Bi et al., 2019). In its more common form, imbalance of diagnostic labels can hurt a model's specificity or sensitivity, as a prior bias from the data distribution may be learned. The Lung Screening Study (LSS) Feasibility Phase exemplifies the common class imbalance in cancer imaging data: 325 (20.5%) suspicious lung nodules were detected in the 1586 first low-dose CT screening, of which only 30 (1.89%) were lung cancers (Gohagan et al., 2004, 2005; NLST Research Team, 2011). This problem directly translates to multi-task classification (CLF), with imbalance between different types of cancer leading to worse sensitivity on the underrepresented categories (Yu et al., 2013). It is important to note that by solving the imbalance with augmentation techniques,

[6] More concretely, this describes a case of covariate shift (Quionero-Candela et al., 2009; Shimodaira, 2000) defined by a change of distribution within the independent variables between two datasets.

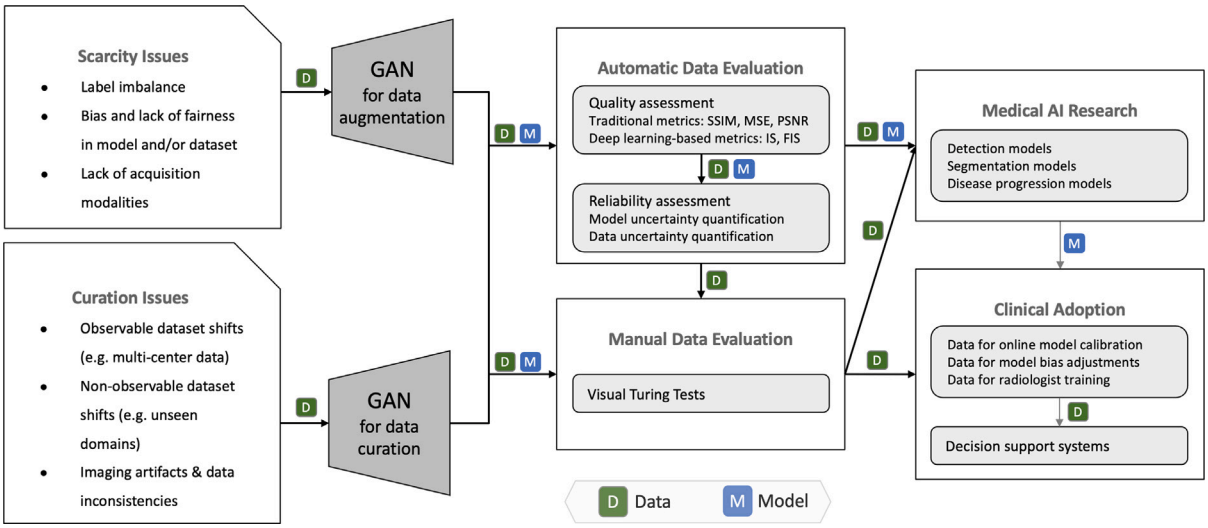

Fig. 5. Illustration of a workflow that applies GANs to the challenges of data scarcity and data curation. After the GAN generates synthetic data specific to the issue at hand, the data is automatically and manually evaluated before further used in medical AI research. Ultimately, both synthetic data and medical AI models are integrated as decision support tools into clinical practice.

bias is introduced as the prior distribution is manipulated, causing prevalence shift. As such, the test set should preserve the population statistics. Aside from imbalance of labels, more insidious forms of imbalance such as that of race/ethnicity (Adamson and Smith, 2018) or gender (Larrazabal et al., 2020) of patients are easily omitted in studies. This leads to fairness problems in real world applications as underrepresenting such categories in the training set will hurt performance on these categories in the real world (population shift) (Li et al., 2021a). Because of their potential to generate synthetic data, GANs are a promising solution to the aforementioned problems and have already been thoroughly explored in this regard in Computer Vision (Sampath et al., 2021; Mullick et al., 2019). Concretely, the discriminator and generator can be conditioned on underrepresented labels, forcing the generator to create images for a specific class,[7] as indicated in Fig. 1(d). Many lesions classifiable by complex scoring systems such as RADS reporting are rare and, hence, effective conditional data augmentation is needed to improve the recognition of such lesions by ML detection models (Kazuhiro et al., 2018). GANs have already been used to adjust label distributions in imbalanced cancer imaging datasets, e.g. by generating underrepresented grades in a risk assessment scoring system (Hu et al., 2018b) for prostate cancer. A further promising applicable method is to enrich the data using a related domain as proxy input (Addepalli et al., 2020). Towards the goal of a more diverse distribution of data with respect to gender and ethnicity, similar principles can be applied. For instance, Li et al. (2021a) proposed an adversarial training scheme to improve fairness in classification of skin lesions for underrepresented groups (age, sex, skin tone) by learning a neutral representation using an adversarial bias discrimination loss. Fairness imposing GANs can also generate synthetic data with a preference for underrepresented groups, so that models may ingest a more balanced dataset, improving demographic parity without excluding data from the training pipeline. Such models have been trained in computer vision tasks (Sattigeri et al., 2018; Wang et al., 2019a; Zhang et al., 2018a; Xu et al., 2018; Beutel et al., 2017), but corresponding research on medical and cancer imaging denoted by Fig. 1(c) has been limited (Li et al., 2021a; Ghorbani et al., 2020).

[7] The class can be something as simple as 'malignant' or 'benign', or a more complex score for risk assessment of a tumour such as the BiRADs scoring system for breast tumours (Liberman and Menell, 2002).

4.1.3. Cross-modal data generation

In cancer, multiple acquisition modalities are enlisted in clinical practice (Kim et al., 2016; Chen et al., 2017; Barbaro et al., 2017; Chang et al., 2020b,a); thus automated diagnostic models should ideally learn to interpret various modalities as well or learn a shared representation of these modalities. Conditional GANs offer the possibility to generate one or multiple (Yurt et al., 2019; Li et al., 2019a; Zhou et al., 2020) modalities from another, alleviating the need to actually perform the potentially more harmful screenings—i.e. high-dose CT, PET—that expose patients to radiation, or require invasive contrast agents such as intravenous iodine-based contrast media (ICM) in CT (Haubold et al., 2021), gadolinium-based contrast agents in MRI (Zhao et al., 2020a) (in Table 5) or radioactive tracers in PET (Wang et al., 2018b; Zhao et al., 2020b). Furthermore, extending the acquisition modalities used in a given task would also enhance the performance and generalisability of AI models, allowing them to learn shared representations among these imaging modalities (Bi et al., 2019; Hosny et al., 2018). Towards this goal, multiple GAN domain-adaptation solutions have been proposed to generate CT using MRI (Wolterink et al., 2017; Kearney et al., 2020b; Tanner et al., 2018; Kaiser and Albarqouni, 2019; Nie et al., 2017; Kazemifar et al., 2020; Prokopenko et al., 2019), PET from MRI (Wang et al., 2018b), PET from CT (Ben-Cohen et al., 2017; Bi et al., 2017) (in Table 5), and CT from PET as in Armanious et al. (2020), where also GAN-based PET denoising and MR motion correction are demonstrated. If not indicated otherwise, these image-to-image translation studies are outlined in Table 2. Because of its complexity, clinical cancer diagnosis is based not only on imaging but also non-imaging data (genomic, molecular, clinical, radiological, demographic, etc.). In cases where this data is readily available, it can serve as conditional input to GANs towards the generation of images with the corresponding phenotype–genotype mapping, as is also elaborated in regard to tumour profiling for treatment in Section 4.5.1. A multimodal cGAN was recently developed, conditioned on both images and gene expression code (Xu et al., 2020); however, research along this line is otherwise limited.

4.1.4. Feature hallucinations in synthetic data

As displayed in Fig. 6 and denoted in Fig. 1(b), conditional GANs can unintentionally[8] hallucinate non-existent artifacts into a patient image. This is particularly likely to occur in cross-modal data augmentation, especially but not exclusively if the underlying dataset is

[8] Intentional feature injection or removal is discussed in 4.2.5.

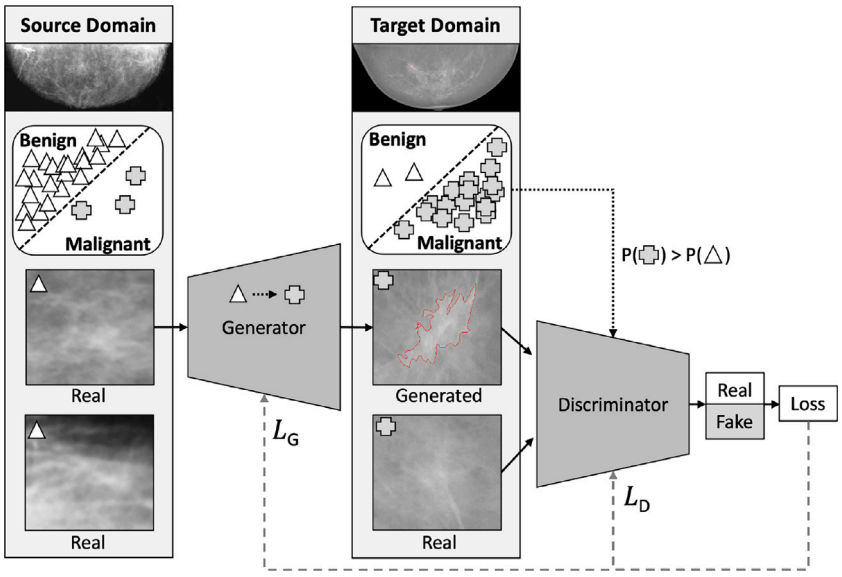

Fig. 6. Example of a GAN that translates Film Scanned MMG (source) to Full-Field Digital MMG (target). The generator transforms 'benign' source images (triangles) into 'malignant' target images (plus symbols). As opposed to source, the target domain contains more malignant MMGs than benign ones. If the discriminator thus learns to associate malignancy with realness, this incentivises the generator to inject malignant features (depicted by dotted arrows). For simplicity additional losses (e.g., reconstruction losses) are omitted.

imbalanced. For instance, Cohen et al. (2018a) describe GAN image feature hallucinations embodied by added and removed brain tumours in cranial MRI. The authors tested the relationship between the ratio of tumour images in the GAN target distribution and the ratio of images diagnosed with tumours by a classifier. The classifier was trained on the GAN generated target dataset, but tested on a balanced holdout test set. It was thereby shown that the generator of CycleGAN effectively learned to hide source domain image features in target domain images, which arguably helped it to fool its discriminator. Paired image-to-image translation with pix2pix (Isola et al., 2017) was more stable, but still some hallucinations were shown to likely have occurred. A cause for this can be a biased discriminator that has learned to discriminating specific image features (e.g., tumours) that are more present in one domain. Cohen et al. (2018a,b) and Wolterink et al. (2018) warn that models that map source to target images, have an incentive to add/remove features during translation if the feature distribution in the target domain is distinct from the feature distribution in the source domain.[9]

Domain-adaptation with unpaired image-to-image translation GANs such as CyleGAN has become increasingly popular in cancer imaging (Wolterink et al., 2017; Tanner et al., 2018; Modanwal et al., 2019; Fossen-Romsaas et al., 2020; Zhao et al., 2020b; Hognon et al., 2019; Mathew et al., 2020; Kearney et al., 2020b; Peng et al., 2020; Jiang et al., 2018; Sandfort et al., 2019). As described, these methods are hallucination-prone and, thus, can put patients at risk when used in clinical settings. More research is needed on how to robustly avoid or detect and eliminate hallucinations in generated data. To this end, we highlight the potential of investigating feature preserving image translation techniques and methods for evaluating whether features have been accurately translated. For instance, in the presence of feature masks or annotations, an additional local reconstruction loss can be introduced in GANs that enforces feature translation in specific image areas.

4.1.5. Data curation and harmonisation

Aside from the limited availability of cancer imaging datasets, a major problem is that the ones available are often not readily useable and require further curation (Hosny et al., 2018). Curation includes dataset formatting, normalising, structuring, de-identification, quality assessment and other methods to facilitate subsequent data processing steps, one of which is the ingestion of the data into AI models (Diaz et al., 2021). In the past, GANs have been proposed for curation of data labelling, segmentation and annotation of images (details in Section 4.3) and de-identification of facial features, EHRs, etc (details in Section 4.2). Particular to cancer imaging datasets and of significant importance is the correction of artifacts, such as patient motion, metallic objects, chemical shifts and others caused by the image processing pipeline (Pusey et al., 1986; Nehmeh et al., 2002), which run the risk of confusing models with spurious information. Towards the principled removal of artifacts, several GAN solutions have been proposed (Vu et al., 2020b; Koike et al., 2020; Armanious et al., 2020). As for the task of reconstruction of compressed data (e.g., compressed sensing MRI Mardani et al., 2017), markedly, Yang et al. (2018a) proposed DAGAN, which is based on U-Net (Ronneberger et al., 2015), reduces aliasing artifacts, and faithfully preserves texture, boundaries and edges (of brain tumours) in the reconstructed images. Kim et al. (2018a) feed down-sampled high-resolution brain tumour MRI into a GAN framework similar to pix2pix to reconstruct high-resolution images with different contrast. The authors highlight the possible acceleration of MR imagery collection while retaining high-resolution images in multiple contrasts, necessary for further clinical decision-making. As relevant to the context of data quality curation, GANs have also been proposed for image super-resolution in cancer imaging (e.g., for lung nodule detection Gu et al., 2020, abdominal CT You et al., 2019, and breast histopathology Shahidi, 2021).

Beyond the lack of curation, a problem particular to multi-centre studies is that of inconsistent curation between data derived in different centres. These discontinuities arise from different scanners, segmentation protocols, demographics, etc, and can cause significant problems to subsequent ML algorithms that may overfit or bias towards one configuration over another (i.e. acquisition and annotation shifts). GANs have the potential to contribute in this domain as well by bringing the distributions of images across different centres closer together. In this context recent work by Li et al. (2021b) and Wei et al. (2020) used GAN-based volumetric normalisation to reduce the variability of heterogeneous 3D chest CT scans of different slice thickness and dose levels. The authors showed that features in subsequent radiomics analysis exhibit increased alignment. Other works in this

[9] For example, if one domain contains mainly healthy images, while the other domain contains mainly pathological images.

domain include a framework that could standardise heterogeneous datasets with a single reference image and obtained promising results on an MRI dataset (Hognon et al., 2019), and GANs that learn bidirectional mappings between different vendors to normalise dynamic contrast enhanced (DCE) breast MRI (Modanwal et al., 2019). An interesting research direction to be explored in the future is synthetic multi-centre data generation using GANs, simulating the distribution of various scanners/centres.

4.1.6. Synthetic data assessment

As indicated in Fig. 1(e), a condition of paramount importance is proper evaluation of GAN-generated or GAN-curated data. This evaluation is to verify that synthetic data is useable for a desired downstream task (e.g., segmentation, classification) and/or indistinguishable from real data while ensuring that no private information is leaked. GANs are commonly evaluated based on *fidelity* (realism of generated samples) and *diversity* (variation of generated samples compared to real samples) (Borji, 2021). Different quantitative measures exist to assess GANs based on the fidelity and diversity of its generated synthetic medical images (Yi et al., 2019; Borji, 2021).

Visual Turing tests (otherwise referred to as Visual Assessment, Mean Opinion Score (MOS) Test, and sometimes used interchangeably with In-Silico Clinical Trials) are arguably the most reliable approach, where clinical experts are presented with samples from real and generated data and are tasked to identify which one is generated. Korkinof et al. (2020) showed that their PGGAN-generated (Karras et al., 2017) 1280×1024 mammograms were inseparable by the majority of participants, including trained breast radiologists. A similar visual Turing test was successfully done in the case of skin disease (Ghorbani et al., 2020), super-resolution of CT (You et al., 2019), brain MRI (Kazuhiro et al., 2018; Han et al., 2018), lung cancer CT scans (Chuquicusma et al., 2018), and histopathology images (Levine et al., 2020). For instance, Chuquicusma et al. (2018) trained a DCGAN (Radford et al., 2015) on the LIDC-IDRI dataset (Armato III et al., 2011) to generate 2D (56×56 pixel) pulmonary lung nodule scans that were realistic enough to deceive 2 radiologists with 11 and 4 years of experience. In contrast to computer vision techniques where synthetic data can often be easily evaluated by any non-expert, the requirement of clinical experts makes Visual Turing Tests in this domain much more costly. Furthermore, a lack of scalability and consistency in medical judgement needs to be taken into account as well Brennan and Silman (1992) and visual Turing tests should in the ideal case engage a range of experts to address inter-observer variation in the assessments. Also, iterating over the same observer addresses intra-observer variation—i.e. repeating the process within a certain amount of intervals that could be days or weeks. These problems are further magnified by the shortage of radiology experts (Mahajan and Venugopal, 2020; Rimmer, 2017) which brings up the necessity for supplementary metrics that can automate the evaluation of generative models. Such metrics allow for preliminary evaluation and can enable research to progress without the logistical hurdle of enlisting experts.

Furthermore, in cases where the sole purpose of the generated data is to improve a downstream task—i.e. classification or segmentation—then the prediction success of the downstream task would be the metric of interest. The latter can reasonably be prioritised over other metrics given that the underlying reasons why the synthetic data alters downstream task performance are examined and clarified.[10]

Image quality assessment metrics. Wang et al. (2004) have thoroughly investigated image quality assessment metrics. The most commonly applied metrics include structural similarity index measure (SSIM)[11] between generated image and reference image (Wang et al., 2004), mean squared error (MSE)[12] and peak signal-to-noise ratio (PSNR).[13] In a recent example that followed this framework of evaluation, synthetic brain MRI with tumours generated by edge-aware EA-GAN (Yu et al., 2019) was assessed using three such metrics: PSNR, SSIM, and normalised mean squared error (NMSE). The authors integrated an end-to-end sobel edge detector to create edge maps from real/synthetic images that are input into the discriminator in the dEa-GAN variant to enforce improved textural structure and object boundaries. Interestingly, aside from evaluating on the whole image, the authors demonstrated evaluation results focused on the tumour regions, which were overall significantly lower than the whole image. Other works that have evaluated their synthetic images in an automatic manner have focused primarily on the SSIM and PSNR metrics and include generation of CT (Kearney et al., 2020b; Mathew et al., 2020) and PET scans (Zhao et al., 2020b). While indicative of image quality, these similarity-based metrics might not generalise well to human judgement of image similarity, the latter depending on high-order image structure and context (Zhang et al., 2018c). Finding evaluation metrics that are strong correlates of human judgement of perceptual image similarity is a promising line of research. In the context of cancer and medical imaging, we highlight the need for evaluation metrics for synthetic images that correlate with the perceptual image similarity judged by medical experts. Apart from perceptual image similarity, further evaluation metrics in cancer and medical imaging are to be investigated that are able to estimate the diagnostic value of (synthetic) images and, in the presence of reference images, the diagnostic value proportion between target and reference image.

Deep generative model-specific assessment metrics. In recent years, the Inception score (IS) (Salimans et al., 2016) and Fréchet Inception distance (FID) (Heusel et al., 2017) have emerged, offering a more sophisticated alternative for the assessment of synthetic data. The IS uses a classifier to generate a probability distribution of labels given a synthetic image. If the probability distribution is highly skewed, it is indicative that a specific object is present in the image (resulting in a higher IS), while in the case where it is uniform, the image contains a jumble of objects and that is more likely to be non-sense (resulting in a lower IS).[14] The FID metric compares the distance between the synthetic image distribution to that of the real image distribution by comparing extracted high-level features from one of the layers of a classifier (e.g., Inception v3 as in IS). Both metrics have shown promise in the evaluation of GAN-generated data; however, they come with several bias issues that need to be taken into account during evaluation (Chong and Forsyth, 2020; DeVries et al., 2019; Borji, 2019). As these metrics have not been widely used in cancer imaging yet, their applicability on GAN-synthesised cancer images remains to be investigated. In contrast to computer vision datasets containing diverse objects, medical imaging datasets commonly only contain images of one specific organ. In this regard, we promote further research as to how object diversity based methods such as IS can be applied to medical and cancer imaging,

[10] For example, synthetic data may balance imbalanced datasets, reduce overfitting on limited training data, or improve model robustness to better capture domain shifts in the test dataset.

[11] SSIM predicts perceived quality and considers image statistics to assess structural information based on luminance, contrast, and structure.

[12] MSE is computed by averaging the squared intensity differences between corresponding pixels of the generated image and the reference image.

[13] PSNR is an adjustment to the MSE score, commonly used to measure reconstruction quality in lossy compression.

[14] Not only a low label entropy within an image is desired, but also a high label entropy across images: IS also assesses the variety of peaks in the probability distributions generated from the synthetic images, so that a higher variety is indicative of more diverse objects being generated by the GAN (resulting in a higher IS).

which requires, among others, meaningful adjustments of the dataset-specific pretrained classifications models (i.e. Inception v3) that IS and FID rely upon.

Uncertainty quantification as GAN evaluation metric?. A general problem facing the adoption of deep learning methods in clinical tasks is their inherent unreliability exemplified by high prediction variation caused by minimal input variation (e.g., one pixel attack Korpihalkola et al., 2020). This is further exacerbated by the nontransparent decision making process inside deep neural networks thus often described as 'black box models' (Bi et al., 2019). Also, the performance of deep learning methods in out-of-domain datasets has been assessed as unreliable (Lim et al., 2019). To eventually achieve beneficial clinical adoption and trust, examining and reporting the inherent uncertainty of these models on each prediction becomes a necessity. Besides classification, segmentation (Hu et al., 2020; Alshehhi and Alshehhi, 2021), etc, uncertainty estimation is applicable to models in the context of data generation as well (Lim et al., 2019; Abdar et al., 2020; Hu et al., 2020). Edupuganti et al. (2019) studied a GAN architecture based on variational autoencoders (VAE) (Kingma and Welling, 2013) on the task of MRI reconstruction, with emphasis on uncertainty studies. Due to their probabilistic nature, VAEs allowed for a Monte Carlo sampling approach which enables quantification of pixel-variance and the generation of uncertainty maps. Furthermore, they used Stein's Unbiased Risk Estimator (SURE) (Stein, 1981) as a measure of uncertainty that serves as surrogate of MSE even in the absence of ground truth. Their results indicated that adversarial losses introduce more uncertainty. Parallel to image reconstruction, uncertainty has also been studied in the context of brain tumours (glioma) in MRI enhancement (Tanno et al., 2021). In this study, a probabilistic deep learning framework for model uncertainty quantification was proposed, decomposing the problem into two uncertainty types: *intrinsic uncertainty* (particular to image enhancement and pertaining to the one-to-many nature of the super-resolution mapping) and *parameter uncertainty* (a general challenge, it pertains to the choice of the optimal model parameters). The overall model uncertainty in this case is a combination of the two and was evaluated for image super-resolution. Through a series of systematic studies the utility of this approach was highlighted, as it resulted in improved overall prediction performance of the evaluated models even for out-of-distribution data. It was further shown that predictive uncertainty highly correlated with reconstruction error, which not only enabled spotting unrealistic synthetic images, but also highlights the potential in further exploring uncertainty as an evaluation metric for GAN-generated data. A further use-case of interest for GAN evaluation via uncertainty estimation is the 'adherence' to provided conditional inputs. As elaborated in 4.1.4 for image-to-image translation, conditional GANs are likely to introduce features that do not correspond to the conditional class label or source image. After training a classification model on image features of interest (say, tumour vs non-tumour features), we can examine the classifier's prediction and estimated uncertainty[15] for the generated images. Given the expected features in the generated images are known beforehand, the classifier's uncertainty of the presence of these features can be used to estimate not only image fidelity (e.g., image features are not generated realistic enough), but also 'condition adherence' (e.g., expected image features are altered during generation).

Outlook on clinical adoption. Alongside GAN-specific and standard image assessment metrics, uncertainty-based evaluation schemes can further automate the analysis of generative models. To this end, the challenge of clinical validation for predictive uncertainty as a reliability metric for synthetic data assessment remains (Tanno et al., 2021). In practice, building clinical trust in AI models is a non-trivial endeavour and will require rigorous performance monitoring and calibration especially in the early stages (Kelly et al., 2019; Durán and Jongsma, 2021). This is particularly the case when CADe and CADx models are trained on entirely (or partially) synthetic data given that the data itself was not first assessed by clinicians. Until a certain level of trust is built in these pipelines, automatic metrics will be a preliminary evaluation step that is inevitably followed by diligent clinical evaluation for deployment. A research direction of interest in this context would be 'gatekeeper' GANs—i.e. GANs that simulate common data (and/or difficult edge cases) of the target hospital, on which deployment-ready candidate models (e.g., segmentation, classification, etc.) are then tested to ensure they are sufficiently generalisable. If the candidate model performance on such test data satisfies a predefined threshold, it has passed this quality gate for clinical deployment.

4.2. Data access and privacy challenges

Access to sufficiently large and labelled data resources is the main constraint for the development of deep learning models for medical imaging tasks (Esteva et al., 2019). In cancer imaging, the practice of sharing validated data to aid the development of AI algorithms is restricted due to technical, ethical, and legal concerns (Bi et al., 2019). The latter is exemplified by regulations such as the Health Insurance Portability and Accountability Act (HIPAA, 1996) in the United States of America (USA) or the European Union's General Data Protection Regulation (GDPR, 2016) with which respective clinical centres must comply with. Alongside the need and numerous benefits of patient privacy preservation, it can also limit data sharing initiatives and restrict the availability, size and usability of public cancer imaging datasets. Bi et al. (2019) assess the absence of such datasets as a noteworthy challenge for AI in cancer imaging.

The published GANs and adversarial training methods that are suggested for or applied to cancer imaging challenges within this Section 4.2 are summarise below in Table 3.

4.2.1. Decentralised data generation

As AI systems are often developed and trained outside of medical institutions, prior approval to transfer data out of their respective data silos is required, adding significant hurdles to the logistics of setting up a training pipeline or rendering it entirely impossible. In addition, medical institutions can often not guarantee a secured connection to systems deployed outside their centres (Hosny et al., 2018), which further limits their options to share valuable training data.

One privacy preserving approach solving this problem is federated learning (McMahan et al., 2017), where copies of an AI model are trained in a distributed fashion inside each clinical centre in parallel and are aggregated to a global model in a central server. This eliminates the need for sensitive patient data to leave any of the clinical centres (Kaissis et al., 2020; Sheller et al., 2020). However, it is to be noted that federated learning cannot guarantee full patient privacy. Hitaj et al. (2017) demonstrated that any malicious user can train a GAN to violate the privacy of the other users in a federated learning system. While difficult to avoid, the risk of such GAN-based attacks can be minimised, e.g., by using a combination of selective parameter updates (Shokri and Shmatikov, 2015) (sharing only a small selected part of the model parameters across centres) and the sparse vector technique[16] as shown by Li et al. (2019b).

To solve the challenge of privacy assurance of clinical medical imaging data, a distributed GAN (Hardy et al., 2019; Xin et al., 2020; Guerraoui et al., 2020; Rasouli et al., 2020; Zhang et al., 2021) can

[15] The uncertainty can be estimated using methods such as Bayesian Neural Networks (MacKay, 1992; Neal, 2012), Monte-Carlo Dropout (Gal and Ghahramani, 2016) or Deep Ensembles (Lakshminarayanan et al., 2016).

[16] Sparse Vector Technique (SVT) (Lyu et al., 2016) is a Differential Privacy (DP) (Dwork, 2006) method that introduces noise into a deep learning model's gradients.

Table 2
Overview of adversarially-trained models applied to cancer imaging **data scarcity and usability** challenges. Publications are clustered by section and ordered by year in ascending order.

Publication	Method	Dataset	Modality	Task	Highlights
Imbalanced data & fairness					
Hu et al. (2018b)	ProstateGAN	Private	Prostate MRI	Class-conditional synthesis	Gleason score (cancer grade) class conditions.
Ghorbani et al. (2020)	DermGAN	Private	Dermoscopy	Paired translation	Adapted pix2pix evaluated via Turing Tests and rare skin condition CLF.
Li et al. (2021a)	Encoder	ISIC 2018 (Codella et al., 2018)	Dermoscopy	Adversarial training, Representation learning	Fair Encoder with bias discriminator and skin lesion CLF.
Cross-modal data generation					
Wolterink et al. (2017)	CycleGAN	Private	Cranial MRI/CT	Unpaired translation	First CNN for unpaired MR-to-CT translation. Evaluated via PSNR and MAE.
Ben-Cohen et al. (2017)	pix2pix	Private	Liver PET/CT	Paired translation	Paired CT-to-PET translation with focus on hepatic malignant tumours.
Nie et al. (2017)	context-aware GAN	ADNI (Wyman et al., 2013; Weiner et al., 2017)	Cranial/pelvic MRI/CT	Paired translation	Supervised 3D GAN for MR-to-CT translation with 'Auto-Context Model' (ACM).
Wang et al. (2018b)	Locality Adaptive GAN (LA-GAN)	BrainWeb phantom (Cocosco et al., 1997)	Cranial MRI, PET phantom	Paired translation	3D auto-context, synthesising PET from low-dose PET and multimodal MRI.
Tanner et al. (2018)	CycleGAN	VISCERAL (Jimenez-del Toro et al., 2016)	Lung/abdominal MRI/CT	Image registration	MR-CT CycleGAN for registration.
Kaiser and Albarqouni (2019)	pix2pix, context-aware GAN (Nie et al., 2017)	RIRE (Fitzpatrick, 1998)	Cranial MRI/CT	Paired translation	Detailed preprocessing description, MR-to-CT translation, comparison with U-Net.
Prokopenko et al. (2019)	DualGAN, SRGAN	CPTAC3 (National Cancer Institute, 2018) & Head-Neck-PET-CT (Vallières et al., 2017)	Cranial MRI/CT	Unpaired translation	DualGAN for unpaired MR-to-CT, visual Turing tests.
Yurt et al. (2019) mrirecon	mustGAN	IXI (IXI Dataset, 2007) & ISLES (Maier et al., 2017)	Cranial MRI	Paired translation	FLAIR, T1, T2 synthesis via feature fusion of one-to-one and many-to-one pix2pix networks.
Li et al. (2019a)	diamondGAN	Private & MICCAI-WMH (Kuijf et al., 2019)	Cranial MRI	Unpaired translation	Target modality synthesis from flexible set of non-aligned source modalities.
Zhou et al. (2020)	hi-Net	BRATS 2018 (Menze et al., 2014; Bakas et al., 2018)	Cranial MRI	Paired translation	Domain-specific encoder network features fused into layers of pix2pix-based network.
Zhao et al. (2020b)	S-CycleGAN	Private	Cranial low/full dose PET	Paired translation	Low (LDPET) to full dose (FDPET) translation with supervised loss for paired images.
Kearney et al. (2020b)	VAE-enhanced A-CycleGAN	Private	Cranial MRI/CT	Unpaired translation	MR-to-CT evaluated via PSNR, SSIM, MAE. Superior to paired alternatives.
Kazemifar et al. (2020)	context-aware GAN	Private	Cranial MRI/CT	Paired translation	Feasibility of generated CT from MRI for dose calculation for radiation treatment.
Armanious et al. (2020)	MedGAN	Private	Liver PET/CT	Paired translation	CasNet architecture, PET-to-CT, MRI motion artifact correction, PET denoising.
Xu et al. (2020)	multi-conditional GAN	NSCLC (Zhou et al., 2018)	Lung CT, gene expression	Multi-input conditional synthesis	Image-gene data fusion, nodule generator input: background, segmentation, gene code.
Haubold et al. (2021)	Pix2PixHD (Wang et al., 2018a)	Private	Arterial phase CT	Paired translation	Low-to-full ICM CT (thorax, liver, abdomen), 50% reduction in intravenous ICM dose.

(*continued on next page*)

Table 2 (*continued*).

Publication	Method	Dataset	Modality	Task	Highlights
Feature hallucinations					
Cohen et al. (2018a,b) dist-bias	CycleGAN, pix2pix	BRATS2013 (Menze et al., 2014)	Cranial MRI	Paired/unpaired translation	Removed/added tumours during image translation can lead to misdiagnosis.
Data curation					
Yang et al. (2018a)	DAGAN	MICCAI 2013 grand challenge dataset	Cranial MRI	Image reconstruction	Fast GAN compressed sensing MRI reconstruction outperformed conventional methods.
Kim et al. (2018a)	pix2pix-based	BRATS (Menze et al., 2014)	Cranial MRI	Reconstruction/ super-resolution	Information transfer between different contrast MRI, effective pretraining/fine-tuning.
Hognon et al. (2019)	CycleGAN, pix2pix	BRATS (Menze et al., 2014), BrainWeb phantom (Cocosco et al., 1997)	Cranial MRI	Paired/unpaired translation, normalisation	CycleGAN translation to BrainWeb reference image, pix2pix back-translation to source.
Modanwal et al. (2019)	CycleGAN	Private	Breast MRI	Unpaired translation	Standardising DCE-MRI across scanners, anatomy preserving mutual information loss.
You et al. (2019)	CycleGAN-based	Mayo Low Dose CT (AAPM, 2016)	Abdominal CT	Super-resolution	Joint constraints to Wasserstein loss for structural preservation. Evaluated by 3 radiologists.
Gu et al. (2020)	MedSRGAN	LUNA16 (Setio et al., 2017)	MRI/thoracic CT	Super-resolution	Residual Whole Map Attention Network (RWMAN) in G. Evaluated by 5 radiologists.
Vu et al. (2020b)	WGAN-GP-based	k-Wave toolbox (Treeby and Cox, 2010)	Photoacoustic CT (PACT)	Paired translation	U-Net & WGAN-GP based generator for artifact removal. Evaluated via SSIM, PSNR.
Koike et al. (2020)	CycleGAN	Private	Head/neck CT	Unpaired translation	Metal artifact reduction via CT-to-CT translation, evaluated via radiotherapy dose accuracy.
Wei et al. (2020)	WGAN-GP-inspired	Private	Chest CT	Paired translation	CT normalisation of dose/slice thickness. Evaluated via Radiomics Feature Variability.
Shahidi (2021)	WA-SRGAN	BreakHis (Benhammou et al., 2020), Camelyon (Litjens et al., 2018)	Breast/lymph node histopathology	Super-resolution	Wide residual blocks, self-attention SRGAN for improved robustness & resolution
Li et al. (2021b)	SingleGAN-based (Yu et al., 2018b)	Private	Spleen/colorectal CT	Unpaired translation	Multi-centre (4) CT normalisation. Evaluated via cross-centre radiomics features variation. Short/long-term survivor CLF improvement.
Synthetic data assessment					
Kazuhiro et al. (2018)	DCGAN	Private	Cranial MRI	Noise-to-image synthesis	Feasibility study for brain MRI synthesis evaluated by 7 radiologists.
Han et al. (2018)	DCGAN, WGAN	BRATS 2016 (Menze et al., 2014)	Cranial MRI	Noise-to-image synthesis	128 × 128 brain MRI synthesis evaluated by one expert physician.
Chuquicusma et al. (2018)	DCGAN	LIDC-IDRI (Armato III et al., 2011)	Thoracic CT	Noise-to-image synthesis	Malignant/benign lung nodule ROI generation evaluated by two radiologists.
Yu et al. (2019)	Ea-GAN	BRATS 2015 (Menze et al., 2014), IXI (IXI Dataset, 2007)	Cranial MRI	Paired image-to-image translation	Loss based on edge maps of synthetic images. Evaluated via PSNR, NMSE, SSIM.
Korkinof et al. (2020)	PGGAN	Private	Full-field digital MMG	Noise-to-image synthesis	1280 × 1024 MMG synthesis from > 10^6 image dataset. Evaluated by 55 radiologists.
Levine et al. (2020)	PGGAN, VAE, ESRGAN	TCGA (Grossman et al., 2016), OVCARE archive	Ovarian Histopathology	Noise-to-image synthesis	1024 × 1024 whole slide synthesis. Evaluated via FID and by 15 pathologists (9 certified)

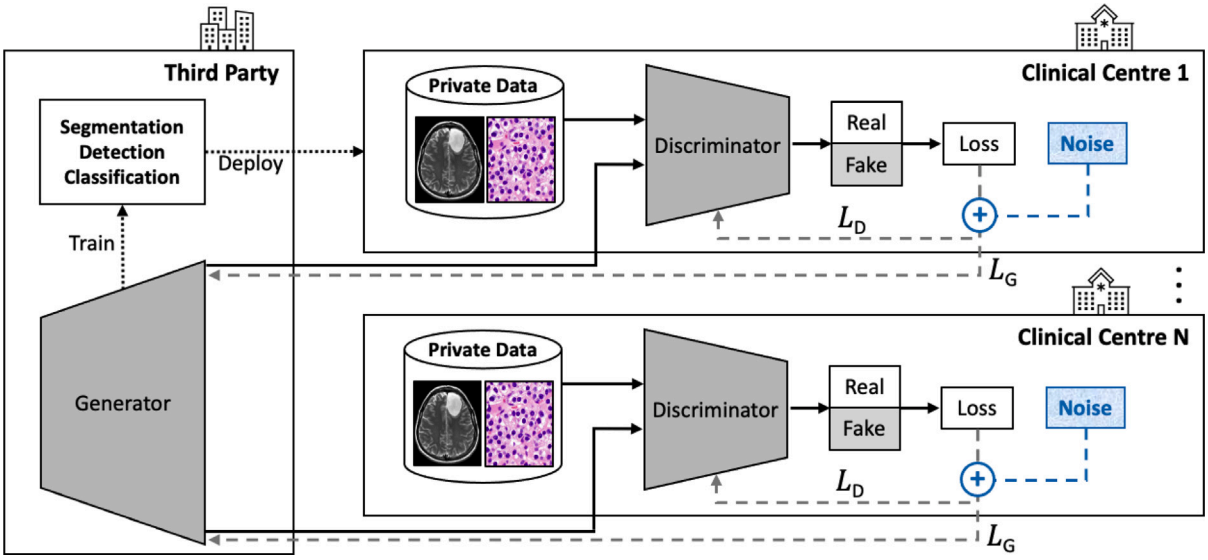

Fig. 7. Visual example of a GAN in a federated learning setup with a central generator trying to generate realistic samples that fool all of the discriminators, which are distributed across clinical centres as in Chang et al. (2020b,a). Once trained, the generator can produce training data for a downstream task model (e.g., segmentation, detection, classification). As depicted in blue colour, we suggest to extend the federated learning setup by adding 'Noise' to the gradients providing a differential privacy guarantee. This reduces the possibility of reconstruction of specific records of the training data (i.e. images of a specific patient) by someone with access to the trained GAN model (i.e. to the generator) or by someone intercepting the synthetic images while they are transferred from the central generator to the centres during training.

be trained on sensitive patient data to generate synthetic training data. The technical, legal, and ethical constraints for sharing de-identified synthetic data are typically less restrictive than for real patient data. Such generated data can be used instead of the real patient data to train models on disease detection, segmentation, or prognosis.

For instance, Chang et al. (2020b,a) proposed the Distributed Asynchronized Discriminator GAN (AsynDGAN), which consists of multiple discriminators deployed inside various medical centres and one central generator deployed outside the medical centres. The generator never needs to see the private patient data, as it learns by receiving the gradient updates of each of the discriminators. The discriminators are trained to differentiate images of their medical centre from synthetic images received from the central generator. After training AsynDGAN, its generator is used and evaluated based on its ability to provide a rich training set of images to successfully train a segmentation model. AsynDGAN is evaluated on MRI brain tumour segmentation and cell nuclei segmentation. The segmentation models trained only on AsynDGAN-generated data achieves a competitive performance when compared to segmentation models trained on the entire dataset of real data. Notably, models trained on AsynDGAN-generated data outperform models trained on local data from only one of the medical centres. To our best knowledge, AsynDGAN is the only distributed GAN applied to cancer imaging to date. Therefore, we promote further research in this line to fully exploit the potential of privacy-preservation using distributed GANs. As demonstrated in Fig. 7 and suggested in Fig. 1(f), for maximal privacy preservation we recommend exploring methods that combine privacy during training (e.g., federated GANs) with privacy after training (e.g., differentially-private GANs), the latter being described in the following section.

4.2.2. Differentially-private data generation

Shin et al. (2018a) train a GAN to generate brain tumour images and highlight the usefulness of their method for anonymisation, as their synthetic data cannot be attributed to a single patient but rather only to an instantiation of the training population. However, it is to be scrutinised whether such synthetic samples are indeed fully private, as, given a careful analysis of the GAN model and/or its generated samples, a risk of possible reconstruction of part of the GAN training data exists (Papernot et al., 2016). For example, Chen et al. (2020a) propose a GAN for model inversion (MI) attacks, which aim at reconstructing the training data from a target model's parameters. A potential solution to avoid training data reconstruction is highlighted by Xie et al. (2018), who propose the Differentially Private Generative Adversarial Network (DPGAN). In Differential Privacy (DP) (Dwork, 2006) the parameters (ε, δ) denote the privacy budget (Torfi et al., 2020), where ε measures the privacy loss and δ represents the probability that a range of outputs with a privacy loss $> \varepsilon$ exists.[17] Hence, the smaller the parameters (ε, δ) for a given model, the less effect a single sample in the training data has on model output. The less effect of such a single sample, the stronger is the confidence in the privacy of the model to not reveal samples of the training data.

Examples of GANs with differential privacy guarantees. In DPGAN noise is added to the model's gradients during training to ensure training data privacy. Extending on the concept of DPGAN, Jordon et al. (2018) train a GAN coined PATE-GAN based on the Private Aggregation of Teacher Ensembles (PATE) framework (Papernot et al., 2016, 2018). In the PATE framework, a student model learns from various unpublished teacher models each trained on data subsets. The student model cannot access an individual teacher model nor its training data. PATE-GAN consists of k discriminator teachers, $T_1, \ldots, T_k$, and a student discriminator S that backpropagates its loss back into the generator. This limits the effect of any individual sample in PATE-GAN's training. In a $(\varepsilon = 1, \delta = 10^{-5})$-DP setting, classification models trained on PATE-GAN's synthetic data achieves competitive performances e.g. on a *non-imaging* cervical cancer dataset (Fernandes et al., 2017) compared to an upper bound vanilla GAN baseline without DP.

On the same dataset, Torfi et al. (2020) achieve competitive results using a Rényi Differential Privacy and Convolutional Generative Adversarial Networks (RDP-CGAN) under an equally strong $(\varepsilon = 1, \delta = 10^{-5})$-DP setting.

For the generation of biomedical participant data in clinical trials, Beaulieu-Jones et al. (2019) apply an AC-GAN under a $(\varepsilon = 3.5, \delta =$

[17] For example, if an identical model M is trained two times, once with training data D resulting in M_D and once with marginally different training data D' resulting in $M_{D'}$, it is (ϵ)-DP if the following holds true: For any possible output x, the output probability for x of model M_D differs no more than $exp(\epsilon)$ from the output probability for x of $M_{D'}$.

10^{-5})-DP setting based on Gaussian noise added to AC-GAN's gradients during training.

Bae et al. (2020) propose AnomiGAN to anonymise private medical data via some degree of output randomness during inference. This randomness of the generator is achieved by randomly adding, for each layer, one of its separately stored training variances. AnomiGAN achieves competitive results on a *non-imaging* breast cancer dataset and a *non-imaging* prostate cancer for any of the reported privacy parameter values $\varepsilon \in [0.0, 0.5]$ compared to DP, where Laplacian noise is added to samples.

Outlook on synthetic cancer image privacy. Despite the above efforts, DP in GANs has only been applied to non-imaging cancer data which indicates research potential for extending these methods reliably to cancer imaging data. According to Stadler et al. (2021), using synthetic data generated under DP can protect outliers in the original data from linkage attacks, but likely also reduces the statistical signal of these original data points, which can result in lower utility of the synthetic data. Apart from this privacy-utility tradeoff, it may not be readily controllable/predictable which original data features are preserved and which omitted in the synthetic datasets (Stadler et al., 2021). In fields such as cancer imaging where patient privacy is critical, desirable privacy-utility tradeoffs need to be defined and thoroughly evaluated to enable trust, shareability, and usefulness of synthetic data. Consensus is yet to be found as to how privacy preservation in GAN-generated data can be evaluated and verified in the research community and in clinical practice. Promising approaches include methods that define a privacy gain/loss for synthetic samples (Stadler et al., 2021; Yoon et al., 2020). Yoon et al. (2020), for instance, define and backpropagate an identifiability loss to the generator to synthesis anonymised electronic health records (EHRs). The identifiability loss is based on the notion that the minimum weighted euclidean distance between two patient records from two different patients can serve as a desirable anonymisation target for synthetic data. Designing or extending reliable methods and metrics for standardised quantitative evaluation of patient privacy preservation in synthetic medical images is a line of research that we call attention to.

4.2.3. Obfuscation of identifying patient features in images

If the removal of all sensitive patient information within a cancer imaging dataset allows for sharing such datasets, then GANs can be used to obfuscate such sensitive data. As indicated by Fig. 1(g), GANs can learn to remove the features from the imaging data that could reveal a patient's identity, e.g. by learning to apply image inpainting to pixel or voxel data of burned in image annotations or of identifying body parts. Such identifying body parts could be the facial features of a patient, as was shown by Schwarz et al. (2019) on the example of cranial MRI. Numerous studies exist where GANs accomplish facial feature de-identification on non-medical imaging modalities (Wu et al., 2018b; Hukkelås et al., 2019; Li and Lin, 2019; Maximov et al., 2020). For medical imaging modalities, GANs have yet to prove themselves as tool of choice for anatomical and facial feature de-identification against common standards (Ségonne et al., 2004; Bischoff-Grethe et al., 2007; Schimke et al., 2011; Milchenko and Marcus, 2013) with solid baselines. These standards, however, have shown to be susceptible to reconstruction achieved by unpaired image-to-image GANs on MRI volumes with high reversibility for blurred faces and partial reversibility for removed facial features (Abramian and Eklund, 2019). Van der Goten et al. (2021) provide a first proof-of-concept for GAN-based facial feature de-identification in 3D (128^3 voxel) cranial MRI. The generator of their conditional de-identification GAN (C-DeID-GAN) receives brain mask, brain intensities and a convex hull of the brain MRI as input and generates de-identified MRI slices. C-DeID-GAN generates the entire de-identified brain MRI scan and, hence, may not be able to guarantee that the generation process does not alter any of the original brain features. A solution to this can be to only generate and replace the 2D MRI slices or parts thereof that do contain non-pathological facial features while retaining all other original 2D MRI slices. Presuming preservation of medically relevant features and robustness of de-identification, GAN-based approaches can allow for subsequent medical analysis, privacy preserving data sharing and provision of de-identified training data. Hence, we highlight the research potential of GANs for robust medical image de-identification e.g. via image inpainting GANs that have already been successful applied to other tasks in cancer imaging such as synthetic lesion inpainting into mammograms (Wu et al., 2018a; Becker et al., 2019) and lung CT scans (Mirsky et al., 2019). Also, GAN-based patient feature de-identification methods that are adjustable and trainable to remain quantifiably robust against adversarial image reconstruction are a research line of interest.

4.2.4. Identifying patient features in latent representations

In line with Fig. 1(g), a further example for privacy preserving methods are autoencoders[18] that learn patient identity-specific features and obfuscate such features when encoding input images into latent space representation. Such an identity-obfuscated representation can be used as input into further models (classification, segmentation, etc.) or decoded back into a de-identified image. Adversarial training has been shown to be effective for learning a privacy-preserving encoding function, where a discriminator tries to succeed at classifying the private attribute from the encoded data (Raval et al., 2017; Wu et al., 2018c; Yang et al., 2018c; Pittaluga et al., 2019). Apart from being trained via the backpropagated adversarial loss, the encoder needs at least one further utility training objective to learn to generate useful representations, such as denoising (Vincent et al., 2008) or classification of a second attribute (e.g., facial expressions Chen et al., 2018a; Oleszkiewicz et al., 2018). Siamese Neural Networks (Bromley et al., 1994) such as the Siamese Generative Adversarial Privatizer (SGAP) (Oleszkiewicz et al., 2018) have been used effectively for adversarial training of an identity-obfuscated representation encoder. In SGAP, two weight-sharing Siamese Discriminators are trained using a distance based loss function to learn to classify whether a pair of images belongs to the same person. As visualised in Fig. 8, Kim et al. (2019a) follow a similar approach with the goal of de-identifying and segmenting brain MRI data. Two feature maps are encoded from a pair of MRI scans and fed into a Siamese Discriminator that evaluates via binary classification whether the two feature maps are from the same patient. The generated feature maps are also fed into a segmentation model that backpropagates a Dice loss (Sudre et al., 2017) to train the encoder. Fig. 8 illustrates the scenario where the encoder is deployed in a trusted setting after training, e.g. in a clinical centre, and the segmentation model is deployed in an untrusted setting, e.g. outside the clinical centre at a third party. The encoder shares the identity-obfuscated feature maps with the external segmentation model without the need of transferring the sensitive patient data outside the clinical centre. This motivates further research into adversarial identity-obfuscated encoding methods e.g., to allow sharing and usage of cancer imaging data representations and models across clinical centres.

4.2.5. Adversarial attacks putting patients at risk

Examples of GAN-based tampering with cancer imaging data. For instance, Mirsky et al. (2019) added and removed evidence of cancer in lung CT scans. Of two identical deep 3D convolutional cGANs (based on pix2pix), one was used to inject (diameter $\geq$ 10 mm) and the other to remove (diameter $<$ 3 mm) multiple solitary pulmonary nodules indicating lung cancer. The GANs were trained on 888 CT scans from the Lung Image Database Consortium image collection (LIDC-IDRI) dataset (Armato III et al., 2011) and inpainted on an extracted

[18] For example adversarial autoencoders (Makhzani et al., 2015; Creswell et al., 2018), which adversarially learn latent space representations that match a chosen prior distribution.

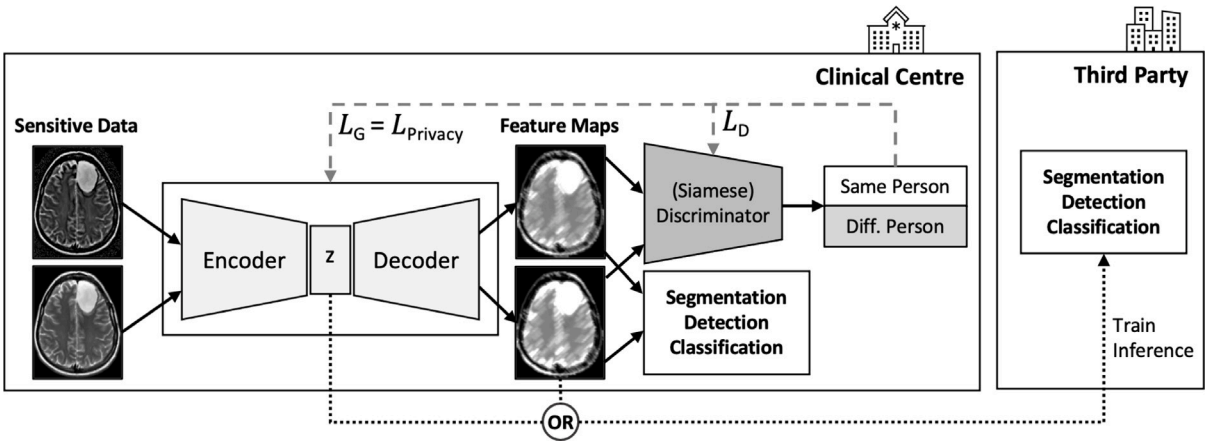

Fig. 8. Example of an autoencoder architecture trained via adversarial loss to learn privacy-preserving feature maps as in Kim et al. (2019a) and/or a privacy-preserving latent representation z. Once trained and after thorough periodic manual verification of its ability to preserve privacy, the representation z and/or the feature maps can be sent to third parties outside the clinical centre for model training or inference requests.

region of interest of 32^3 voxel cuboid shape. The trained GANs can be autonomously executed by malware and are capable of ingesting nodules into standard CT scans that are realistic enough to deceive both radiologists and AI disease detection systems. Three radiologists with 2, 5 and 7 years of experience analysed 70 tampered and 30 authentic CT scans. Spending on average 10 min on each scan, the radiologists diagnosed 99% of the scans with added nodules as malignant and 94% of the scans with removed nodules as healthy. After disclosing the presence of the attack to the radiologists, the percentages dropped to 60% and 87%, respectively (Mirsky et al., 2019).

Becker et al. (2019) trained a CycleGAN (Zhu et al., 2017) on 680 down-scaled mammograms from the Breast Cancer Digital Repository (BCDR) (Lopez et al., 2012) and the INbreast (Moreira et al., 2012) datasets to generate suspicious features and was able to remove or inject them into existing mammograms. They showed that their approach can fool radiologists at lower pixel dimensions (i.e. 256 × 256) demonstrating that alterations in patient images by a malicious attacker can remain undetected by clinicians, influence the diagnosis, and potentially harm the patient (Becker et al., 2019).

Defending adversarial attacks. In regard to fooling diagnostic models, one measure to circumvent adversarial attacks is to increase model robustness against adversarial examples (Madry et al., 2017), as suggested by Fig. 1(h). Augmenting the robustness has been shown to be effective for medical imaging segmentation models (He et al., 2019; Park et al., 2020), lung nodule detection models (Liu et al., 2020b; Paul et al., 2020), skin cancer recognition (Huq and Pervin, 2020; Hirano et al., 2021), and classification of histopathology images of lymph node sections with metastatic tissue (Wetstein et al., 2020). Liu et al. (2020b) provide model robustness by adding adversarial chest CT examples to the training data. These adversarial examples are composed of synthetic nodules that are generated by a 3D convolutional variational encoder trained in conjunction with a WGAN-GP (Gulrajani et al., 2017) discriminator. To further enhance robustness, Projected Gradient Descent (PGD) (Madry et al., 2017) is applied to find and protect against noise patterns for which the detector network is prone to produce over-confident false predictions (Liu et al., 2020b).

Apart from being the adversary, GANs can also detect adversarial attacks and thus are applicable as security counter-measure enabling attack anticipation, early warning, monitoring and mitigation. Defence-GAN, for example, learns the distribution of non-tampered images and can generate a close output to an inference input image that does not contain adversarial modifications (Samangouei et al., 2018).

We highlight the research potential in adversarial attacks and examples, alongside prospective GAN detection and defence mechanisms that can, as elaborated, highly impact the field of cancer imaging. Apart from the image injection of entire tumours and the generation of adversarial radiomics examples, a further attack vector to consider in future studies is the perturbation of the specific imaging features within an image that are used to compute radiomics features.

4.3. Data annotation and segmentation challenges

4.3.1. Annotation-specific issues in cancer imaging

Missing annotations in datasets. In cancer imaging, not only the availability of large datasets is rare, but also the availability of labels, annotations, and segmentation masks within such datasets. The generation and evaluation of such labels, annotations, and segmentation masks is a task for which trained health professionals (radiologists, pathologists) are needed to ensure validity and credibility (Hosny et al., 2018; Bi et al., 2019). Nonetheless, radiologist annotations of large datasets can take years to generate (Bi et al., 2019). The tasks of labelling and annotating (e.g., bounding boxes, segmentation masks, textual comments) cancer imaging data is, hence, expensive both in time and cost, especially considering the large amount of data needed to train deep learning models.

Intra/inter-observer annotation variability. This cancer imaging challenge is further exacerbated by the high intra- and inter-observer variability between both pathologists (Gilles et al., 2008; Dimitriou et al., 2018; Martin et al., 2018; Klaver et al., 2020) and radiologists (Elmore et al., 1994; Hopper et al., 1996; Hadjiiski et al., 2012; Teh et al., 2017; Wilson et al., 2018; Woo et al., 2020; Brady, 2017) in interpreting cancer images across imaging modalities, affected organs, and cancer types. Automated annotation processes based on deep learning models allow to produce reproducible and standardised results in each image analysis. In one of most common case where the annotations consist of a segmentation mask, reliably segmenting both tumour and non-tumour tissues is crucial for disease analysis, biopsy, and subsequent intervention and treatment (Hosny et al., 2018; Huynh et al., 2020), the latter being further discussed in Section 4.5. For example, automatic tumour segmentation models are useful in the context of radiotherapy treatment planning (Cuocolo et al., 2020).

Human biases in cancer image annotation. During routine tasks, such as medical image analysis, humans are prone to account for only a few of many relevant qualitative image features. On the contrary, the strength of GANs and deep learning models is the evaluation of large numbers of multi-dimensional image features alongside their (non-linear) inter-relationships and combined importance (Hosny et al., 2018). Deep learning models are likely to react to unexpected and subtle patterns in the imaging data (e.g., anomalies, hidden comorbidities, etc.) that

Table 3
Overview of adversarially-trained models applied/applicable to **data access and privacy** cancer imaging challenges. Publications are clustered by section and ordered by year in ascending order.

Publication	Method	Dataset	Modality	Task	Highlights
Decentralised GANs					
Chang et al. (2020b,a) AsynDGAN	AsynDGAN, PatchGAN (Isola et al., 2017)	BRATS 2018 (Bakas et al., 2018), Multi-Organ (Kumar et al., 2017)	Cranial MRI, nuclei images	Paired translation	Mask-to-image, central G gets distributed Ds' gradients, synthetic only-trained segmentation.
Differential-privacy GANs					
Xie et al. (2018)	DPGAN	MNIST (LeCun et al., 1998), MIMIC-III (Johnson et al., 2016)	MNIST images, EHRs	Noise-to-image synthesis	Noisy gradients during training ensure DP guarantee.
Jordon et al. (2018)	PATE-GAN	Cervical cancer (Fernandes et al., 2017)	[non-imaging] Medical records	Data synthesis	DP via PATE framework. G gradient from student D that learns from teacher Ds.
Beaulieu-Jones et al. (2019)	AC-GAN	MIMIC-III (Johnson et al., 2016)	[non-imaging] EHRs, clinical trial data	Class-conditional synthesis	DP via Gaussian noise added to AC-GAN gradient. Treatment arm (standard/intensive) as condition.
Bae et al. (2020)	AnomiGAN	UCI b & Prostate (Blake, 1998)	[non-imaging] Cell nuclei tabular data	Multi-class-conditional synthesis, classification	DP via training variances added to G's layers in inference. Real data row as G's condition.
Torfi et al. (2020)	RDP-CGAN	Cervical cancer (Fernandes et al., 2017), MIMIC-III (Johnson et al., 2016)	[non-imaging] Medical records, EHRs	Data synthesis	DP GAN based on Rényi divergence. Allows to track a DP loss.
Patient de-identification					
Abramian and Eklund (2019)	CycleGAN	IXI (IXI Dataset, 2007)	Cranial MRI	Unpaired translation	Reconstruction of blurring/removed faces in MRI shows privacy vulnerability.
Kim et al. (2019a)	PrivacyNet	PPMI (Marek et al., 2011)	Cranial MRI	Adversarial training, segmentation	Segmenting de-identified representations learned via same-person CLF by Siamese Ds.
Van der Goten et al. (2021)	C-DeID-GAN	ADNI (Wyman et al., 2013; Weiner et al., 2017), OASIS-3 (LaMontagne et al., 2019)	Cranial MRI	Paired translation	Face de-id. Concatenated convex hull, brain mask & brain volumes as G & D inputs.
Adversarial data tampering					
Mirsky et al. (2019)	pix2pix-based CT-GAN	LIDC-IDRI (Armato III et al., 2011)	Lung CT	Image inpainting	Injected/removed lung nodules in CT fool radiologists and AI models.
Becker et al. (2019)	CycleGAN	BCDR (Lopez et al., 2012), INbreast (Moreira et al., 2012)	Digital/Film MMG	Unpaired image-to-image translation	Suspicious features can be learned and injected/removed from MMG.
Liu et al. (2020b)	Variational Encoder, WGAN-GP	LUNA (Setio et al., 2017), NLST (NLST Research Team, 2011)	Lung CT	Noise-to-image synthesis	Robustness via adversarial data augmentation, reduce false positives in nodule detection.

medical practitioners are prone to overlook for instance due to any of multiple existing cognitive biases (e.g., anchoring bias, framing bias, availability bias) (Brady, 2017) or inattentional blindness (Drew et al., 2013). Inattentional blindness occurs when radiologists (or pathologist) have a substantial amount of their attention drawn to a specific task, such as finding an expected pattern (e.g., a lung nodule) in the imaging data, that they become blind to other patterns in that data.

Implications of low segmentation model robustness. As for the common annotation task of segmentation mask delineation, automated segmentation models can minimise the risk of the aforesaid human biases. However, to date, segmentation models have difficulties when confronted with intricate segmentation problems including domain shifts, rare diseases with limited sample size, or small lesion and metastasis segmentation. In this sense, the performance of many automated and semi-automated clinical segmentation models has been suboptimal (Sharma and Aggarwal, 2010). This emphasises the need for expensive manual verification of segmentation model results by human experts (Hosny et al., 2018). The challenge of training automated models for difficult segmentation problems can be approached by applying methods for learning discriminative features without explicit labels. Such methods include GANs and variational autoencoders (Kingma and Welling, 2013) capable of automating robust segmentation (Hosny et al., 2018).

In addition, segmented regions of interest (ROI) are commonly used to extract quantitative imaging features with diagnostic value such as radiomics features. The latter are used to detect and monitor tumours

Table 4
Overview of adversarial training and GAN-based approaches applied to **segmentation** in cancer imaging tasks. Publications are clustered by organ type and ordered by year in ascending order. '*' indicates that the metrics are only available in figures and the baseline numbers are lower than using GANs in the corresponding paper. 'n.a.' indicates that there was no comparison with a specific baseline with the reason for this being indicated in the 'Highlights' column.

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Head/Brain/Neck							
Kamnitsas et al. (2017) deepmedic	deepmedic	Private	MRI	Adversarial training	Dice: 0.60 Recall: 0.56 Precision: 0.70	0.63, 0.59, 0.72	Multi-domain (MPRAGE, FLAIR, T2, Proton Density, and Gradient-Echo) segmentation with domain discriminator loss.
Rezaei et al. (2017)	pix2pix, MarkovianGAN (Li and Wand, 2016)	BRATS 2017 (Menze et al., 2014; Bakas et al., 2018, 2017)	MRI	Adversarial training	Dice: n.a.	0.80	D detects G-generated masks for high/low grade glioma segmentation. Benchmarked in a challenge, thus missing metric w/o GAN.
Mok and Chung (2018)	CB-GAN	BRATS (Menze et al., 2014)	MRI	Data augmentation	Dice: 0.79	0.84	Coarse-to-fine G captures training set manifold, generates generic samples in HGG & LGG segmentation.
Yu et al. (2018a)	pix2pix-based	BRATS (Menze et al., 2014)	MRI	Data augmentation	Dice: 0.67	0.68	Cross-modal paired FLAIR to T1 translation, training tumour segmentation with T1+real/synthetic FLAIR. Baseline: only-T1 training
Shin et al. (2018a)	pix2pix	BRATS (Menze et al., 2014)	MRI	Data augmentation	Dice: 0.81	0.81	Training on synthetic, before fine-tuning on 10% of the real data.
Xue et al. (2018)	SegAN	BRATS (Menze et al., 2014)	MRI	Adversarial training	Dice: 0.80	0.85	Paired image-to-mask. New multi-scale loss: L1 of D representations between GT- & prediction-masked input MRI. U-Net baseline.
Giacomello et al. (2020)	SegAN-CAT	BRATS (Menze et al., 2014)	MRI	Adversarial training	Dice: 0.71	0.86	Paired image-to-mask. Combined dice loss & multi-scale loss using concatenation on channel axis instead of masking. SegAN baseline.
Kim et al. (2020) BrainTumor	GAN	BRATS (Bakas et al., 2018; Menze et al., 2014)	MRI	Image inpainting, data augmentation	Dice: 0.57	0.59	Simplifying tumour features into concentric circle & grade mask to inpaint.
Hu et al. (2020)	UNet-based GAN segmenter	Private	CT, PET	Adversarial training	Dice: 0.69	0.71	Spatial context information & hierarchical features. Nasal-type lymphoma segmentation with uncertainty estimation.
Qasim et al. (2020)	SPADE-based	BRATS (Bakas et al., 2018), ISIC (Codella et al., 2019)	MRI, dermoscopy	Cross-domain translation	Dice: *	B:0.66 S:0.62	Brain and skin segmentation. Frozen segmenter as 3rd player in GAN to condition on local apart from global information.
Foroozandeh and Eklund (2020)	PGGAN, SPADE	BRATS (Menze et al., 2014)	MRI	Data augmentation	Av. dice error(%): 16.80	16.18	Sequential noise-to-mask and paired mask-to-image translation to synthesise labelled tumour images.
Cirillo et al. (2020)	vox2vox: 3D pix2pix	BRATS (Menze et al., 2014)	MRI	Adversarial training	Dice: 0.87	0.93	3D adversarial training to enforce segmentation results to look realistic.
Han et al. (2021)	Symmetric adaptation network	BRATS (Menze et al., 2014)	MRI	Cross-domain translation	Dice: 0.48	0.67	Simultaneous source/target (T2 to other sequences) translation and segmentation. Compared to CycleGAN baseline.
Alshehhi and Alshehhi (2021)	U-Net based GAN segmenter	BRATS (Menze et al., 2014)	MRI	Adversarial training	Dice: n.a.	n.a.	Quantitative comparison of 7 active learning acquisition functions using existing adversarial networks (including SegAN, SegAN-CAT)
Breast							
Singh et al. (2018)	pix2pix	DDSM (Heath et al., 2001)	Film MMG	Adversarial training	Dice: 0.86	0.94	Adversarial loss to make automated segmentation close to manual masks for breast mass segmentation.
Caballo et al. (2020)	DCGAN (Radford et al., 2015)	Private	CT	Data augmentation	Dice: 0.70	0.93	GAN based data augmentation; Validated by matching extracted radiomics features.
Negi et al. (2020)	GAN, WGAN-RDA-UNET	Private	Ultrasound	Adversarial training	Dice: 0.85	0.88	Outperforms state-of-the-art using Residual-Dilated-Attention-Gate-UNet and WGAN for lesion segmentation.
Abdominal							
Huo et al. (2018)	Conditional PatchGAN	Private	MRI	Adversarial training	Dice: 0.93	0.94	Adversarial loss as segmentation post-processing for spleen segmentation. ResNet baseline.
Chen et al. (2019)	DC-FCN-based GAN segmenter	LiTS (Bilic et al., 2019)	CT	Adversarial training	Dice: 0.62	0.68	Cascaded framework with densely connected adversarial training. DC-Fully connected network baseline.

(*continued on next page*)

Table 4 (*continued*).

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Sandfort et al. (2019)	CycleGAN	NIH (Prior et al., 2017), Decathlon (Simpson et al., 2019), DeepLesion (Yan et al., 2018)	CT	Cross-domain translation	Dice (od): 0.92, 0.10	0.93, 0.75	Contrast enhanced to non-enhanced translation to improve out-of-distribution (od) segmentation.
Xiao et al. (2019)	Radiomics-guided GAN	Private	MRI	Adversarial training	Dice: 0.72	0.92	Radiomics-guided adversarial mechanism to map relationship between contrast and non-contrast images. U-Net baseline comparison.
Oliveira (2020b)	pix2pix, SPADE	LiTS (Bilic et al., 2019)	CT	Image inpainting	Dice: 0.58	0.61	Realistic lesion inpainting in CT slices to provide controllable set of training samples.
Chest and lungs							
Jiang et al. (2018)	CycleGAN-based	NSCLC (Prior et al., 2017)	CT, MRI	Cross-domain translation	Dice: 0.66	0.80	Tumour-aware loss for unsupervised cross-domain adaptation compared with standard cycleGAN benchmark.
Jin et al. (2018)	cGAN	LIDC (Armato III et al., 2011)	CT	Image inpainting	Dice: 0.96	0.99	Generated lung nodules to improve segmentation robustness; A novel multi-mask reconstruction loss.
Dong et al. (2019)	UNet-based GAN segmenter	AAPM Challenge (Yang et al., 2018b)	CT	Adversarial training	Dice: 0.97 (l), 0.83 (sc), 0.71 (o), 0.85 (h)	0.97, 0.90, 0.75, 0.87	Adversarial training to discriminate manual and automated segmentation of lungs, spinal cord, oesophagus, heart.
Tang et al. (2019) XLSor	MUNIT (Huang et al., 2018)	JSRT (Shiraishi et al., 2000), Montgomery (Jaeger et al., 2013)	Chest X-ray	Lung segmentation	Dice: 0.97	0.98	Unpaired normal-to-abnormal (pathological) translation. Synthetic data augmentation for lung segmentation.
Shi et al. (2020)	AUGAN	LIDC-IDRI (Armato III et al., 2011)	CT	Adversarial training	Dice: 0.82	0.85	A deep layer aggregation based on U-Net++.
Munawar et al. (2020)	Unet-based GAN segmenter	JSRT (Shiraishi et al., 2000), Montgomery & Shenzhen (Jaeger et al., 2013)	Chest X-ray	Adversarial training	Dice: 0.96	0.97	Adversarial training to discriminate manual and automated segmentation.
Prostate							
Kohl et al. (2017)	UNet-based GAN segmenter	Private	MRI	Adversarial training	Dice: 0.35	0.41	Adversarial loss to discriminate manual and automated segmentation.
Grall et al. (2019) ProstatecGAN	pix2pix	Private	MRI	Adversarial training	Dice: 0.67 (ADC), 0.74 (DWI), 0.67 (T2)	0.73, 0.79, 0.74	Paired prostate image-to-mask translation. Investigated the robustness of the pix2pix against noise.
Nie and Shen (2020)	GAN	PROMISE12 (Litjens et al., 2014)	MRI	Adversarial confidence learning	Dice: 88.25 (b), 90.11 (m), 86.67 (a)	89.52, 90.97, 88.20	Difficulty-aware mechanism to alleviate the effect of easy samples during training. b = base, m = middle, a = apex.
Zhang et al. (2020)	PGGAN	Private	CT	Data augmentation	Dice: 0.85	0.90	Semi-supervised training using both annotated and un-annotated data. Synthetic data augmentation using PGGAN.
Cem Birbiri et al. (2020)	pix2pix, CycleGAN	PROMISE12 (Litjens et al., 2014)	MRI	Data augmentation	Dice: 0.72	0.76	Compared the performance of pix2pix, U-Net, and CycleGAN.
Chaitanya et al. (2021)	cGAN, DCGAN-based D	ACDC (Bernard et al., 2018), Decathlon (Simpson et al., 2019)	MRI, CT	Data augmentation	Dice: 0.40	0.53	GAN data augmentation for intensity and shape variations. Tested on cardiac, prostate, and pancreas datasets.
Colorectal							
Liu et al. (2019a)	GAN, LAGAN	Private	CT	Adversarial training	Dice: 0.87	0.92	Automatic post-processing to refine the segmentation of deep networks.
Poorneshwaran et al. (2019)	pix2pix	CVC-clinic (Vázquez et al., 2017)	Endoscopy	Adversarial training	Dice: n.a.	0.88	Adversarial learning to make automatic segmentation close to manual. Ablations were compared instead of baselines.
Xie et al. (2020)	MI^{2}GAN, CycleGAN	CVC-clinic (Vázquez et al., 2017), ETIS-Larib (Silva et al., 2014)	Endoscopy	Cross-domain translation	Dice: 0.66	0.73	Content features and domain information decoupling and maximising the mutual information.

(*continued on next page*)

Table 4 (*continued*).

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Pathology							
Pandey et al. (2020)	GAN & cGAN	Kaggle (Ljosa et al., 2012a)	Histopathology	Data augmentation	Dice: 0.79	0.83	Two-stage GAN to generate masks and conditioned synthetic images.
Other							
Chi et al. (2018)	pix2pix	ISBI ISIC (Codella et al., 2018)	Dermoscopy	Data augmentation	Dice: 0.85	0.84	Similar performance replacing half of real with synthetic data. Colour labels as lesion specific characteristics.
Abhishek and Hamarneh (2019)	pix2pix	ISBI ISIC (Codella et al., 2018)	Dermoscopy	Data augmentation	Dice: 0.77	0.81	Generate new lesion images given any arbitrary mask.
Sarker et al. (2019)	MobileGAN	ISBI ISIC (Codella et al., 2018)	Dermoscopy	Adversarial training	Dice: 0.76	0.91	Lightweight and efficient GAN model with position and channel attention. Dice was compared to a U-Net baseline.
Zaman et al. (2020)	pix2pix	Private	Ultrasound	Data augmentation	Dice: *	≈ Δ+0.05	Recommendations on standard data augmentation approaches. Bone surface segmentation. pix2pix based discriminator.

(e.g., lymphoma Kang et al., 2018), biomarkers, and tumour-specific phenotypic attributes (Lambin et al., 2012; Parmar et al., 2015). The accuracy and success of such commonly applied diagnostic image feature quantification methods, hence, depends on accurate and robust ROI segmentations. Segmentation models need to be able to provide reproducibility of extracted quantitative features and biomarkers (Bi et al., 2019) with reliably-low variation, among others, across different scanners, CT slice thicknesses, and reconstruction kernels (Balagurunathan et al., 2014; Zhao et al., 2016). To this end, we promote lines of research that use adversarial training schemes to target the robustification of segmentation models. Progress in this open research challenge can beneficially unlock trust, usability, and clinical adoption of biomarker quantification methods in clinical practice.

4.3.2. GAN applications for cancer image segmentation

Table 4 summarises the collection of segmentation publications that utilise such adversarial training approaches and GAN-based data synthesis for cancer imaging.

In Table 4, we further report the baseline performance alongside the performance increase attributable to applying GANs or adversarial training for each surveyed publication. For the common Dice Score segmentation performance metric, Fig. 10 visualises these differences. Comparing the figure's black identity line and the red trend line over publications, we observe a general improvement of approximately 5 percentage points of adversarial learning methods compared to their baselines. Fig. 11 further displays the variation in performance between baselines and adversarial network methods for the years 2017 to 2021. Based on visual analysis, performance gains seem to be both anatomy-invariant and invariant to the strength of the baseline, where similar gains are achieved for initially low (e.g., < 0.7) and high (e.g., $>= 0.7$) baseline Dice scores. While Figs. 10 and 11 offer interesting quantitative insights, we recommend taking potential publication bias[19] into account when drawing conclusions from these figures. Trends in the presented data in these plots can be analysed holistically, however, they are not intended to benchmark and compare individual publications against each other. This is due to multiple limiting factors of such comparisons including the differences in (a) the used baselines, (b) organs, (c) modalities, (d) the segmentation task and its associated difficulty, (e) the amount of training and testing data, (f) data and annotation quality, (g) pre- and post-processing methods, or (h) the study's objectives. In regard to (g), some studies may focus on other benefits of adversarial learning methods instead of or apart from Dice Score improvement, such as, reducing the needed training dataset size, domain adaptation in general, protecting patient privacy with synthetic data, or simply improving other metrics (e.g. Hausdorff distance, FID).

In the following sections, we provide a summary of the commonly used techniques and trends in the GAN literature that address the challenges in cancer image segmentation.

Robust quantitative imaging feature extraction. For example, Xiao et al. (2019) addressed the challenge of robustification of segmentation models and reliable biomarker quantification. Xiao et al. (2019) provide radiomics features as conditional input to the discriminator of their adversarially trained liver tumour segmentation model. Their learning procedure strives to inform the generator to create segmentations that are specifically suitable for subsequent radiomics feature computation. Apart from adversarially training segmentation models, we also highlight the research potential of adversarially training quantitative imaging feature extraction models (e.g., deep learning radiomics) for reliable application in multi-centre and multi-domain settings.

Synthetic segmentation model training data. By augmenting and varying the training data of segmentation models, it is possible to substantially decrease the amount of manually annotated images during training while maintaining the performance (Foroozandeh and Eklund, 2020). A general pipeline of such usage of GAN based generative models is demonstrated in Fig. 9(a) and mentioned in Fig. 1(j).

Over the past few years, CycleGAN (Zhu et al., 2017) based approaches have been widely used for synthetic data generation due to the possibility of using unpaired image sets in training, as compared to paired image translation methods like pix2pix (Isola et al., 2017) or SPADE (Park et al., 2019). CycleGAN based data augmentation has been shown to be useful for segmentation model training, in particular, for generating images with different acquisition characteristics such as contrast enhanced MRI from non-contrast MRI (Wang et al., 2021), cross-modality image translation between different modalities such as CT and MRI images (Huo et al., 2018), and domain adaptation tasks (Jiang et al., 2018). The popularity of the CycleGAN based methods lies not only in image synthesis or domain adaptation, but also in the inclusion of simultaneous image segmentation in its pipeline (Lee et al., 2020).

Although pix2pix methods require paired samples, it is also a widely used type of GAN in data augmentation for medical image segmentation (see Table 4). Several works on segmentation have demonstrated its effectiveness in generating synthetic medical images. By manipulating its input, the variability of the training dataset for image segmentation could be remarkably increased in a controlled manner (Abhishek and

[19] Publication bias likely influences the trends observable in Figs. 10 and 11: Only papers that show an improvement attributable to adversarial networks were published and therefore only such studies could be included.

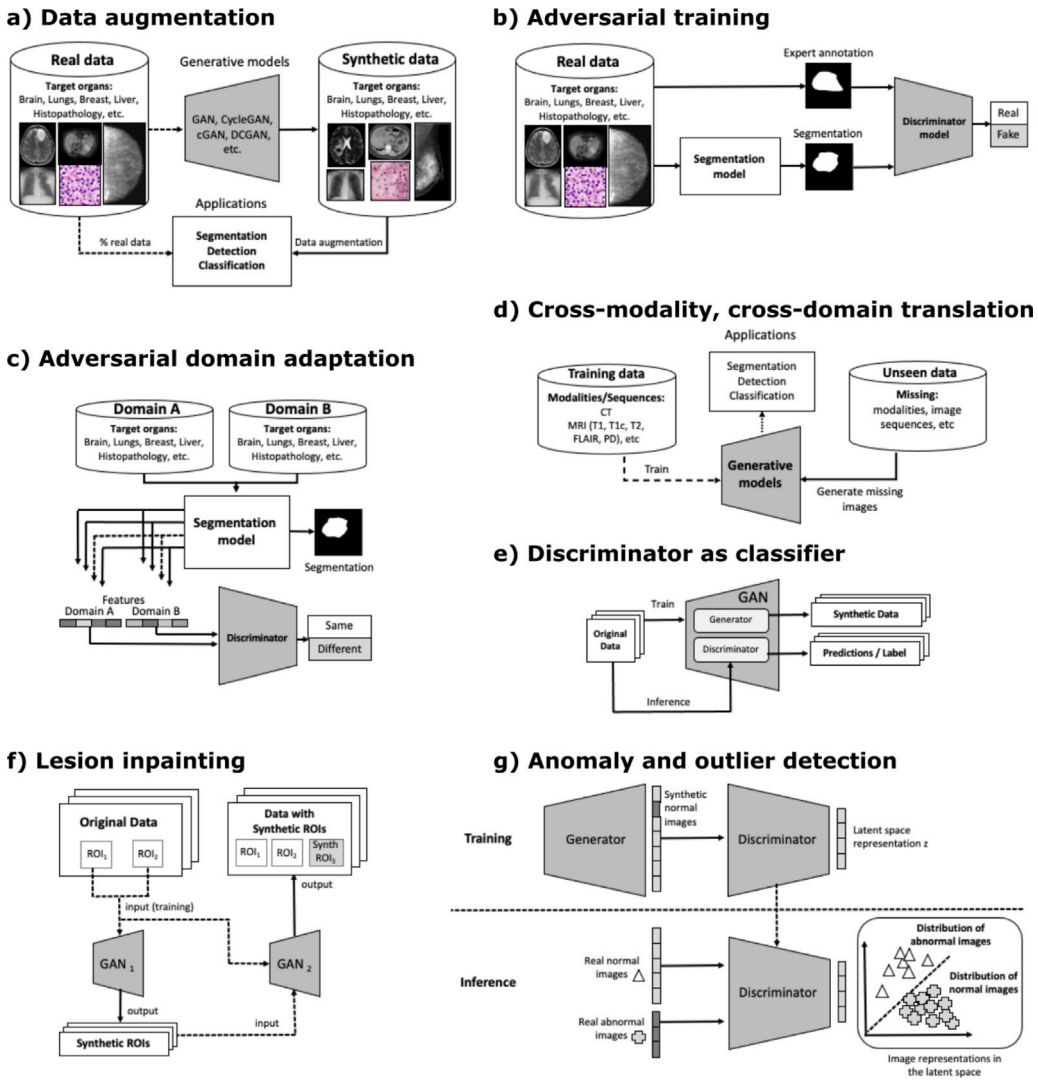

Fig. 9. Overview of cancer imaging GAN applications for detection and segmentation. (a) describes training data augmentation of downstream task models (e.g., segmentation, detection, classification, etc.). In (b) a discriminator scrutinises the segmentations created by a segmentation model, while in (c) the discriminator enforces the model to create domain-agnostic latent representations. (d) illustrates domain-adaptation, where the translated target domain images are used for downstream model training. In (e), the AC-GAN (Odena et al., 2017) discriminator classifies original data. In (f), one GAN generates ROIs while another inpaints them into full-sized images. (g) uses the discriminator's latent space to find abnormal/outlier representations.

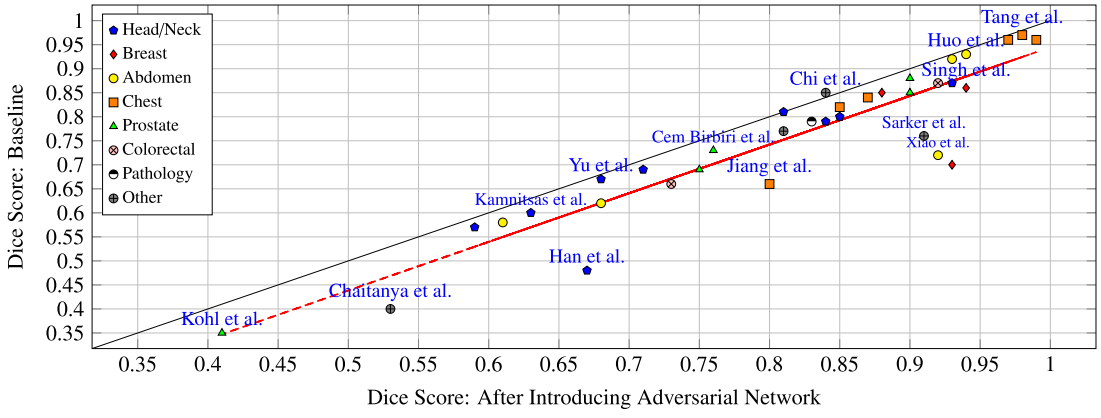

Fig. 10. Scatter plot illustrating the segmentation performance improvement attributable to adversarial networks for the surveyed publications. Each publication is represented by a marker with a colour and shape encoding depicting the publication's anatomical category. Only the publications are included that measure performance via Dice Score and compare against a baseline, as reported in Table 4. For publications reporting multiple Dice Scores, their mean was computed and included herein. The black identity line indicates no change between baseline and adversarial network intervention, while dots below this line represent an improvement. The red regression line depicts the trend of improvement across publications. The author names of a few publications have been manually selected for highlighting based on the distance to the trend line.

Table 5
Overview of adversarially-trained models applied to **detection and diagnosis** tasks in cancer imaging. Publications are clustered by organ type and ordered by year in ascending order.

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Brain							
Chen et al. (2018b)	VAE GAN	Cam-CAN (Shafto et al., 2014) BRATS (Menze et al., 2014) ATLAS (Liew et al., 2018)	MRI	Anomaly detection	AUC(%): 80.0	70.0	Comparison of unsupervised outlier detection methods.
Han et al. (2020)	PGGAN	BRATS (Menze et al., 2014)	MRI	Data augmentation	Accuracy(%): 90.06	91.08	PGGAN-based augmentation method of whole brain MRI.
Han et al. (2019c)	PGGAN, SimGAN	BRATS (Menze et al., 2014)	MRI	Data augmentation	Sensitivity(%): 93.67	97.48	Two-step GAN for noise-to-image and image-to-image data augmentation.
Benson and Beets-Tan (2020)	GAN	TCIA (Schmainda and Prah, 2018)	MRI	Anomaly detection	Accuracy(%): 73.48	74.96	Multi-modal MRI images as input to the GAN.
Han et al. (2019b)	GAN	Private	MRI	Data augmentation	Sensitivity(%): 67.0	77.0	Synthesis and detection of brain metastases in MRI with bounding boxes.
Siddiquee et al. (2019)	Fixed-Point GAN	BRATS (Menze et al., 2014)	MRI	Anomaly detection	AUC(%): 95, IoU:0.261	92, 0.348	Brain lesion detection and localisation with fixed-point image-to-image translation concept.
Sun et al. (2020b)	ANT-GAN	BRATS 2013 (Menze et al., 2014)	MRI	Data augmentation, Anomaly detection	F1-Score(%): 89.6	91.7	Abnormal to normal image translation in cranial MRI, VGG lesion classification.
Breast							
Wu et al. (2018a) Mammo-ciGAN	ciGAN	DDSM (Heath et al., 2001)	Film MMG	Data augmentation	ROC AUC(%): 88.7	89.6	Synthetic lesion in-filling in healthy mammograms.
Guan and Loew (2019)	GAN	DDSM (Heath et al., 2001)	Film MMG	Data augmentation	Accuracy(%): 73.48	74.96	Generate patches containing benign and malignant tumours.
Jendele et al. (2019) BreastGAN	CycleGAN	BCDR (Lopez et al., 2012), INbreast (Moreira et al., 2012), CBIS–DDSM (Lee et al., 2016)	Digital/Film MMG	Data augmentation	ROC AUC(%): 83.50, F1(%): 62.53	Δ–1.46, Δ+1.28	Scanned & digital mammograms evaluated together for lesion detection.
Lee and Nishikawa (2020)	CGAN	Private	Digital MMG	Data augmentation	ROC AUC(%): 57	67	Synthesising contralateral mammograms.
Wu et al. (2020)	U-Net based GAN	OPTIMAM (Halling-Brown et al., 2020)	Digital MMG	Data augmentation	ROC AUC(%): 82.9	84.6	Removed/added MMG lesions for malignant/benign classification.
Alyafi et al. (2020)	DCGAN	OPTIMAM (Halling-Brown et al., 2020)	Digital MMG	Data augmentation	F1-Score(%): $\approx$90	$\approx$96	Synthesise breast mass patches with high diversity.
Desai et al. (2020)	DCGAN	DDSM (Heath et al., 2001)	Film MMG	Data augmentation	Accuracy(%): 78.23	87.0	Synthesise full view MMGs and used them in visual Turing test.
Muramatsu et al. (2020)	CycleGAN	DDSM (Heath et al., 2001)	CT, Film MMG	Data augmentation	Accuracy(%): 79.2	81.4	CT lung nodule to MMG mass translation and vice versa.
Swiderski et al. (2021)	AGAN	DDSM (Heath et al., 2001)	Film MMG	Data augmentation	ROC AUC(%): 92.50	94.10	AutoencoderGAN (AGAN) augments data in normal abnormal classification.
Kansal et al. (2020)	DCGAN	Private	OCT	Data augmentation	Accuracy(%): 92.0	93.7	Synthetic Optical Coherence Tomography (OCT) images.
Shen et al. (2021)	ciGAN	DDSM (Heath et al., 2001)	Film MMG	Data augmentation	FROC-AUC(%): 15.1	17.2	Generate labelled breast mass images for precise detection.
Pang et al. (2021)	TripleGAN-based	Private	Ultrasound	Data augmentation	Sensitivity(%): 86.60	87.94	Semi-supervised GAN-based Radiomics model for mass CLF.
Liver							
Frid-Adar et al. (2018)	DCGAN, ACGAN	Private	CT	Data augmentation	Sensitivity(%): 78.6	85.7	Synthesis of high quality focal liver lesions from CT for lesion CLF.
Zhao et al. (2020a)	Tripartite GAN	Private	MRI	Data augmentation	Accuracy(%): 79.2	89.4	Synthetic contrast-enhanced MRI → tumour detection without contrast agents.

(*continued on next page*)

Table 5 (*continued*).

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Doman et al. (2020)	DCGAN	JAMIT (JAMIT Japanese Society of Medical Imaging Technology), 3Dircadb (Soler et al., 2010)	CT	Data augmentation	Detection rate(%): 65	95	Generate metastatic liver lesions in abdominal CT for improved cancer detection.
Stomach/Colon/Prostate							
Kanayama et al. (2019)	DCGAN	Private	Endoscopy	Data augmentation	AP(%): 59.6	63.2	Synthesise lesion images for gastric cancer detection.
Shin et al. (2018b)	cGAN	CVC-CLINIC (Bernal et al., 2015), CVC-ClinicVideoDB (Angermann et al., 2017)	Colonoscopy	Data augmentation	Precision(%): 81.9	85.3	Synthesise polyp images from normal colonoscopy images for polyp detection.
Rau et al. (2019) ColonoscopyDepth	pix2pix-based	Private	Colonoscopy	Data augmentation	Mean L1(%): 33.9	24.7	Transform monocular endoscopic images from two domains to depth maps.
Yu and Zhang (2020)	CapGAN	BrainWeb phantom (Cocosco et al., 1997), Prostate MRI (Choyke et al., 2016)	MRI	Data augmentation	ROC AUC(%): 85.1	88.5	Synthesise prostate MRI using Capsule Network-Based DCGAN instead of CNN.
Krause et al. (2021)	CGAN	TCGA, NLCS (van den Brandt et al., 1990)	Histopathology	Data augmentation	ROC AUC(%): 75.7	77.7	GANs to enhance genetic alteration detection in colorectal cancer histology.
Skin							
Bissoto et al. (2018) gan-skin-lesion	PGAN, DCGAN, pix2pix	Dermofit (Ballerini et al., 2013), ISIC 2017 (Codella et al., 2018), IAD (Argenziano et al., 2002)	Dermoscopy	Data augmentation	ROC AUC(%): 83.4	84.7	Comparative study of GANs for skin lesions generation
Creswell et al. (2018)	ssDAAE	ISIC 2017 (Codella et al., 2018)	Dermoscopy	Representation learning, classification	ROC AUC(%): 89.0	89.0	Adversarial autoencoder fine-tuned on few labelled lesion classification samples.
Baur et al. (2018)	DCGAN, LAPGAN	ISIC 2017 (Codella et al., 2018)	Dermoscopy	Data augmentation	Accuracy(%) : 71.6	74.0	Comparative study, 256 × 256 px skin lesions synthesis. LAPGAN acc = 74.0%
Rashid et al. (2019)	GAN	ISIC 2017 (Codella et al., 2018)	Dermoscopy	Data augmentation	Accuracy(%): 81.5	86.1	Boost CLF performance with GAN-based skin lesions augmentation.
Fossen-Romsaas et al. (2020)	AC-GAN, CycleGAN	HAM10000 & BCN20000 (Tschandl et al., 2018; Combalia et al., 2019), ISIC 2017 (Codella et al., 2018)	Dermoscopy	Data augmentation	Recall(%): 72.1	76.3	Realistic-looking, class-specific synthetic skin lesions.
Qin et al. (2020)	Style-based GAN	ISIC 2017 (Codella et al., 2018)	Dermoscopy	Data augmentation	Precision(%): 71.8	76.9	Style control & noise input tuning in G to synthesise high quality lesions for CLF.
Lung							
Bi et al. (2017)	M-GAN	Private	PET	Data augmentation	F1-Score(%): 66.38	63.84	Synthesise PET data via multi-channel GAN for tumour detection.
Salehinejad et al. (2018)	DCGAN	Private	Chest X-rays	Data augmentation	Accuracy(%): 70.87	92.10	Chest pathology CLF using synthetic data.
Zhao et al. (2018)	F(&)BGAN	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	Accuracy(%): 92.86	95.24	Forward GAN generates diverse images, Backward GAN improves their quality.
Onishi et al. (2019)	WGAN	Private	CT	Data augmentation	Accuracy(%): 63 (Benign), 82 (Malign)	67, 94	Synthesise pulmonary nodules on CT images for nodule CLF.

(*continued on next page*)

Table 5 (*continued*).

Publication	Method	Dataset	Modality	Task	Metric w/o GAN (Baseline)	Metric with GAN (Baseline+Δ)	Highlights
Gao et al. (2019) 3DGANLungNodules	WGAN-GP	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	Sensitivity(%): 84.8	95.0	Synthesise lung nodule 3D data for nodule detection.
Han et al. (2019a)	3DMCGAN	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	CPM(%): 51.8	55.0	3D multi-conditional GAN (2 Ds) for misdiagnosis prevention in nodule detection.
Yang et al. (2019)	GAN	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	ROC AUC(%): 87.56	88.12	Class-aware 3D lung nodule synthesis for nodule CLF.
Wang et al. (2020b) CA-MW-AS	CGAN	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	F1-Score(%): 85.88	89.03	Nodule synthesis conditioned on semantic features.
Kuang et al. (2020)	Multi-D GAN	LIDC-IDRI (Armato III et al., 2011)	CT	Anomaly detection	Accuracy(%): 91.6	95.32	High anomaly scores on malignant images, low scores on benign.
Ghosal et al. (2020)	WGAN-GP	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	ROC AUC(%): 95.0	97.0	Unsupervised AE & clustering augmented learning method for nodule CLF.
Sun et al. (2020a)	DCGAN	LIDC-IDRI (Armato III et al., 2011)	CT	Data augmentation	Accuracy(%): 93.8	94.5	Nodule CLF: Pre-training AlexNet (Krizhevsky et al., 2012) on synthetic, fine-tuning on real.
Wang et al. (2020c)	pix2pix, PGWGAN, WGAN-GP	Private	CT	Data augmentation	Accuracy(%): 53.2	60.5	Augmented CNN for subcentimeter pulmonary adenocarcinoma CLF.
Bu et al. (2020)	3D CGAN	LUNA16 (Setio et al., 2017)	CT	Data augmentation	Sensitivity(%): 97.81	98.57	Squeeze-and-excitation mechanism and residual learning for nodule detection.
Nishio et al. (2020)	3D pix2pix	LUNA16 (Setio et al., 2017)	CT	Data augmentation	Accuracy(%): 85	85	Nodule size CLF. Masked image + mask + nodule size conditioned paired translation.
Onishi et al. (2020)	WGAN	Private	CT	Data augmentation	Specificity(%): 66.7	77.8	AlexNet pretrained on synthetic, fine-tuned on real nodules for malign/benign CLF.
Teramoto et al. (2020)	PGGAN	Private	Cytopathology	Data augmentation	Accuracy(%): 81.0	85.3	Malignancy CLF: CNN pretrained on synthetic lung cytology images, fine-tuned on real.
Others							
Schlegl et al. (2017)	AnoGAN	Private	OCT	Anomaly detection	ROC AUC(%): 73	89	D representations trained on healthy retinal image patches to score abnormal patches.
Zhang et al. (2018b)	DCGANs, WGAN, BEGANs	Private	OCT	Data augmentation	Accuracy(%): 95.67	98.83	CLF of thyroid/non-thyroid tissue. Comparative study for GAN data augmentation.
Chaudhari et al. (2019)	MG-GAN	NCBI (Edgar et al., 2002)	Expression microarray data	Data augmentation	Accuracy(%) P:71.43 L:68.21 B:69.8 C:67.59	93.6, 88.1, 90.3, 91.7	Prostate, Lung, Breast, Colon. Interesting for fusion with imaging data.
Liu et al. (2019b)	WGAN-based	Private	Serum sample staging data	Data augmentation	Accuracy(%): 64.52	70.97	Synthetic training data for CLF of stages of Hepatocellular carcinoma.
Rubin et al. (2019)	TOP-GAN	Private	Holographic microscopy	Data augmentation	AUC(%): 89.2	94.7	Pretrained D adapted to CLF of optical path delay maps of cancer cells (colon, skin).

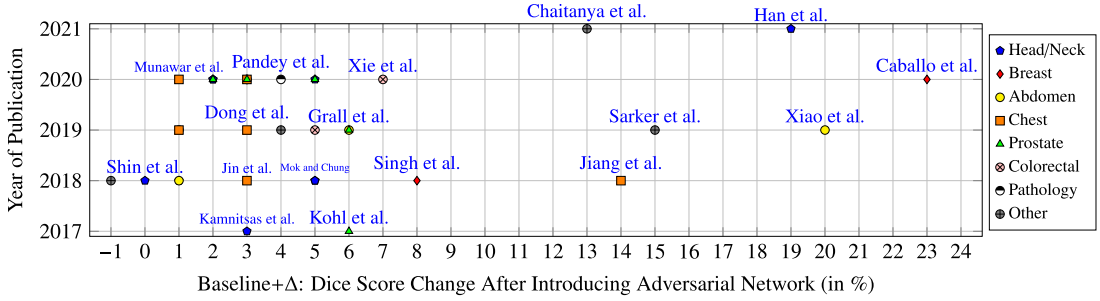

Fig. 11. Scatter plot displaying year of publication and Dice Score improvement (in %). Each marker represents a publication and its colour and shape encoding represents its corresponding anatomical category. Only the publications are included that report Dice Score alongside a baseline comparison. For publications reporting multiple Dice Scores, their mean was computed and included herein. Author names have been manually selected at random for highlighting.

Hamarneh, 2019; Oliveira, 2020b). Similarly, the conditional GAN methods have also been used for controllable data augmentation for improving lesion segmentation (Oliveira, 2020a). Providing a condition as an input to generate a mask is particularly useful to specify the location, size, shape, and heterogeneity of the synthetic lesions. One of the recent examples, proposed by Kim et al. (2020), demonstrates

this in brain MRI tumour synthesis by conditioning an input with simplified controllable concentric circles to specify lesion location and characteristics. A further method for data augmentation is the inpainting of generated lesions into healthy real images or into other synthetic images, as depicted by Fig. 9(f). Overall, the described data augmentation techniques have shown to improve generalisability and performance of segmentation models by increasing both the number and the variability of training samples (Qasim et al., 2020; Foroozandeh and Eklund, 2020; Lee et al., 2020).

Segmentation models with integrated adversarial loss. As stated in Fig. 1(i), GANs can also be used as the algorithm that generates robust segmentation masks, where the generator is used as a segmenter and the discriminator scrutinises the segmentation masks given an input image. One intuition behind this approach is the detection and correction of higher-order inconsistencies between the ground truth segmentation maps and the ones created by the segmenter via adversarial learning (Luc et al., 2016; Hu et al., 2020; Cirillo et al., 2020). This approach is demonstrated in Fig. 9(b). With the additional adversarial loss when training a segmentation model, this approach has been shown to improve semantic segmentation accuracy (Hung et al., 2019; Sarker et al., 2019; Shi et al., 2020). Using adversarial training, similarity of a generated mask to manual segmentation given an input image is taken under consideration by the discriminator allowing a global assessment of the segmentation quality. This approach further offers a practical solution towards handling intra- and inter-observer annotation variability, as the mask discriminator learns an average over observers, which is backpropagated to the segmenter via adversarial loss.

A unique way of incorporating the adversarial loss from the discriminator has been recently proposed in Nie and Shen (2020). In their work, the authors utilise a fully-convolutional network as a discriminator, unlike its counterparts that use binary, single neuron output networks. In doing so, a dense confidence map is produced by the discriminator, which is further used to train the segmenter with an attention mechanism.

Overall, using an adversarial loss as an additional global segmentation assessment is likely to be a helpful further signal for segmentation models, in particular, for heterogeneously structured datasets of limited size (Kohl et al., 2017), as is common for cancer imaging datasets. We highlight potential further research in GAN-based segmentation models to learn to segment increasingly fine radiologic distinctions. These models can help to solve further cancer imaging challenges, for example, accurate differentiation between neoplasms and tissue response to injury in the regions surrounding a tumour after treatment (Bi et al., 2019).

Segmentation models with integrated adversarial domain discrimination. Moreover, a similar adversarial loss can also be performed internally on the segmentation model features as illustrated in Fig. 9(c). Such an approach can benefit unsupervised domain adaptation and domain generalisation by enforcing the segmentation model to learn to base its prediction on domain-invariant feature representations (Kamnitsas et al., 2017).

4.3.3. Limitations and future prospects for cancer imaging segmentation

As shown in Table 4, the applications of GANs in cancer image segmentation cover a variety of clinical requirements. Remarkable steps have been taken to advance this field of research over the past few years. However, the following limitations and future prospects can be considered for further investigation:

- Although the data augmentation using GANs could increase the number of training samples for segmentation, the variability of the synthetic data is limited to the training data. Hence, it may limit the potential of improving the performance in terms of segmentation accuracy. Moreover, training a GAN that produce high sample variability requires a large dataset also with a high variability, and, in most of the cases, with corresponding annotations. Considering the data scarcity challenge in the cancer imaging domain, this can be difficult to achieve.
- In some cases, using GANs could be excessive, considering the difficulties related to convergence of competing generator and discriminator parts of the GAN architectures. For example, the recently proposed SynthSeg model (Billot et al., 2020) is based on Gaussian Mixture Models to generate images and train a contrast agnostic segmentation model. Such approaches can be considered as an alternative to avoid common pitfalls of the GAN training process (e.g., mode collapse). However, this approach needs to be further investigated for cancer imaging tasks where the heterogeneity of tumours is challenging.
- A great potential for using synthetic cancer images is to generate common shareable datasets as benchmarks for automated segmentation methods (Bi et al., 2019). Although this benchmark dataset needs its own validation, it can be beneficial in testing the limits of automated methods with systematically controlled test cases. Such benchmark datasets can be generated by controlling the shape, location, size, intensities of tumours, and can simulating diverse images of different domains that reflect the distributions from real institutions. To avoid learning patterns that are only available in synthetic datasets (e.g., checkerboard artifacts), it is a prospect to investigate further metrics that measure the distance of such synthetic datasets to real-world datasets and the generalisation and extrapolation capabilities of models trained on synthetic benchmarks to real-world data.

4.4. Detection and diagnosis challenges

4.4.1. Common issues in diagnosing malignancies

Clinicians' high diagnostic error rates. Studies of radiological error report high ranges of diagnostic error rates (e.g., discordant interpretations in 31%–37% in Oncologic CT, 13% major discrepancies in Neuro CT and MRI) (Brady, 2017). After (McCreadie and Oliver, 2009) critically reviewed the radiology cases of the last 30 months in their clinical centre, they found that from 256 detected errors (62% CT, 12% Ultrasound, 11% MRI, 9% Radiography, 5% Fluoroscopy) in 222 patients, 225 errors (88%) were attributable to poor image interpretation (14 false positive, 155 false negative, 56 misclassifications). A recent literature review on diagnostic errors by Newman-Toker et al. (2021) estimated a false negative rate,[20] of 22.5% for lung cancer, 8.9% for breast cancer, 9.6% for colorectal cancer, 2.4% for prostate and 13.6% for melanoma. These findings exemplify the uncomfortably high diagnostic and image interpretation error rates that persist in the field of radiology despite decades of interventions and research (Itri et al., 2018).

The challenge of reducing clinicians' high workload. In some settings, radiologists must interpret one CT or MRI image every 3–4 s in an average 8-h workday (McDonald et al., 2015). Automated CADe and CADx systems can provide a more balanced quality-focused workload for radiologists, where radiologists focus on scrutinising the automated detected lesions (false positive reduction) and areas/patches with high predictive uncertainty (false negative reduction). A benefit of CADe/CADx deep learning models are their real-time inference and strong pattern recognition capabilities that are not readily susceptible to cognitive bias (discussed in 4.3.1), environmental factors (Itri et al., 2018), or inter-observer variability (discussed in 4.3.1).

[20] In Newman-Toker et al. (2021) the false negative rates includes both missed (patient encounters at which the diagnosis might have been made but was not) and delayed diagnosis (diagnostic delay relative to urgency of illness detection).

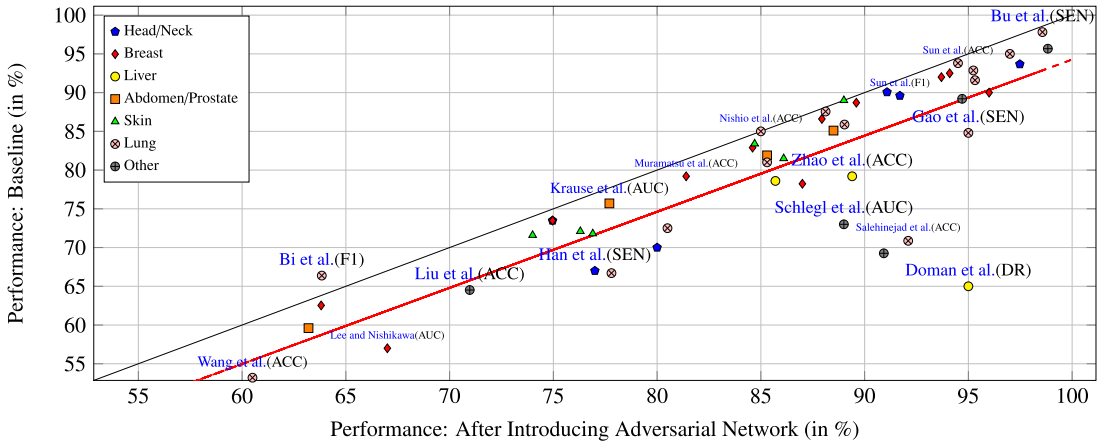

Fig. 12. Scatter plot illustrating the performance improvement attributable to adversarial networks for the surveyed disease diagnosis and detection publications. The shown performance is based on the respective publication's metric reported in Table 5 and include f1-score (F1), sensitivity (SEN), accuracy (ACC), area under the receiver operating characteristic curve (AUC), and detection rate (DR). Each publication is represented by a marker with a colour and shape encoding depicting the publication's anatomical category. The black identity line indicates no change between baseline and adversarial network intervention, while dots below this line represent an improvement. The red regression line depicts the trend of improvement across publications. The author names of a few publications have been manually selected for highlighting based on the distance to the trend line.

Detection model performance on critical edge cases. Challenging cancer imaging problems are the high intra- and inter-tumour heterogeneity (Bi et al., 2019), the detection of small lesions and metastasis across the body (e.g., lymph node involvement and distant metastasis Hosny et al., 2018) and the accurate distinction between malignant and benign tumours (e.g., for detected lung nodules that seem similar on CT scans Hosny et al., 2018). Methods are needed to extend on and further increase the current performance of deep learning detection models (Bi et al., 2019).

4.4.2. GAN applications for cancer detection and diagnosis

As we detail in the following, the capability of adversarial learning to improve malignancy detection has been demonstrated for multiple tumour types and imaging modalities. To this end, Table 5 summarises the collection of recent publications that utilise GANs and adversarial training for cancer detection, classification, and diagnosis.

Figs. 12 and 13 visualise the publications' performance metric values reported in Table 5. Fig. 12 provides visual estimate of the effectiveness of GANs and adversarial training in increasing downstream task performance. A performance increase of approximately 5 percentage points can be observed by comparing the figure's black identity line with the red trend line over publications. We note that no visual pattern seems to be observable that indicates a difference in performance gain between anatomical categories. Across all publications, the performance gains seem not to be a function of the strength of the baseline, as they remain approximately constant with increasing baseline performance. This is indicated by the minimal change of distance between the black identity line and the red trend line throughout the graph. Fig. 13 shows the GAN-induced variation in performance for the years 2017 to 2021 with multiple adversarial models achieving a performance increase of over 10% and most models over 3% on their respective diagnostic downstream task. As emphasised in Section 4.3.2, conclusion drawn from Figs. 12 and 13 have to take publication bias into account. Further, benchmarking and comparison of individual publications based on the presented data in these figures is not part of their intended use due to the differences in baselines, modalities, organs, train and test datasets, and publication objectives.

Adversarial anomaly and outlier detection examples. Schlegl et al. (2017) captured imaging markers relevant for disease prediction using a deep convolutional GAN named AnoGAN. AnoGAN learnt a manifold of normal anatomical variability, accompanying a novel anomaly scoring scheme based on the mapping from image space to a latent space. While Schlegl et al. validated their model on retina optical coherence tomography images, their unsupervised anomaly detection approach is applicable to other domains including cancer detection, as indicated in Fig. 1(l). Chen et al. (2018b) used a Variational Autoencoder GAN for unsupervised outlier detection using T1 and T2 weighted brain MRI images. The scans from healthy subjects were used to train the auto-encoder model to learn the distribution of healthy images and detect pathological images as outliers. Creswell et al. (2018) proposed a semi-supervised Denoising Adversarial Autoencoder (ssDAAE) to learn a representation based on unlabelled skin lesion images. The semi-supervised part of their CNN-based architecture corresponds to malignancy classification of labelled skin lesions based on the encoded representations of the pretrained DAAE. As the amount of labelled data is smaller than the unlabelled data, the labelled data is used to fine-tune classifier and encoder. In ssDAAE, not only the adversarial autoencoder's chosen prior distribution (Makhzani et al., 2015), but also the class label distribution is discriminated by a discriminator, the latter distinguishing between predicted continuous labels and real binary (malignant/benign) labels. Kuang et al. (2020) applied unsupervised learning to distinguish between benign and malignant lung nodules. In their multi-discriminator GAN (MDGAN) various discriminators scrutinise the realness of generated lung nodule images. After GAN pretraining, an encoder is added in front of the generator to the end-to-end architecture to learn the feature distribution of benign pulmonary nodule images and to map these features into latent space. The benign and malignant lung nodules were scored similarly as in the f-AnoGAN framework (Schlegl et al., 2019), computing and combining an image reconstruction loss and a feature matching loss, the latter comparing the discriminators' feature representations between real and encoded-generated images from intermediate discriminator layers. As exemplified in Fig. 9(g), the model yielded high anomaly scores on malignant images and low anomaly scores on benign images despite limited dataset size. Benson and Beets-Tan (2020) used GANs trained from multi-modal MRI images as a 3-channel input (T1-T2 weighted, FLAIR, ADC MRI) in brain anomaly detection. The training of the generative network was performed using only healthy images together with pseudo-random irregular masks. Despite the training dataset consisting of only 20 subjects, the resulting model increased the anomaly detection rate.

Synthetic detection model training data. Among the GAN publications trying to improve classification and detection performance, data augmentation is the most recurrent approach to balance, vary, and increase the detection model's training set size, as suggested in Fig. 1(k).

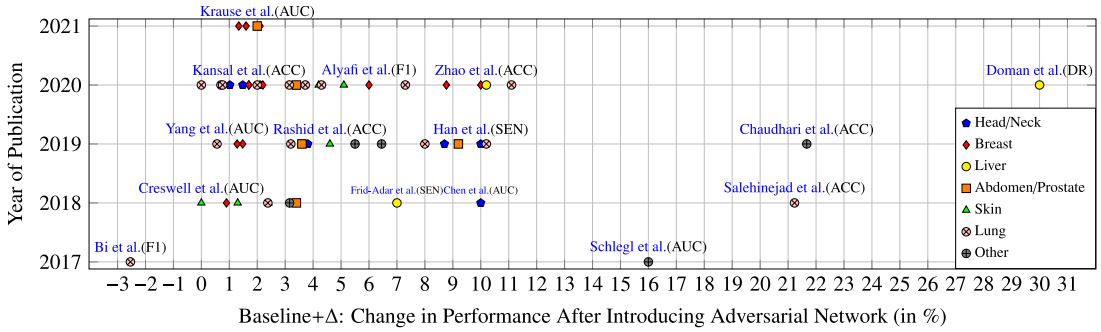

Fig. 13. Scatter plot displaying year of publication and change in performance between baseline and adversarial network method (in %). As in Table 5, the underlying performance metrics include f1-score (F1), sensitivity (SEN), accuracy (ACC), area under the receiver operating characteristic curve (AUC), and detection rate (DR). Each publication is represented by a marker with a colour and shape encoding depicting the publication's anatomical category. Author names have been manually selected at random for highlighting.

For instance in breast imaging, Wu et al. (2018a) trained a class-conditional GAN to perform contextual in-filling to synthesise lesions in healthy scanned mammograms. Guan and Loew (2019) trained a GAN on the same dataset (Heath et al., 2001) to generate synthetic patches with benign and malignant tumours. The synthetic generated patches had clear artifacts and did not match the original dataset distribution. Jendele et al. (2019) used a CycleGAN (Zhu et al., 2017) and both film scanned and digital mammograms to improve binary (malignant/benign) lesion detection using data augmentation. Detecting mammographically-occult breast cancers is another challenging topic addressed by GANs. For instance, Lee and Nishikawa (2020) exploit asymmetries between mammograms of the left and right breasts as signals for finding mammography-occult cancer. They trained a image-conditioned GAN (pix2pix) to generate a healthy synthetic mammogram image of the contralateral breast (e.g., left breast) given the corresponding single-sided mammogram (e.g., right breast) as input. The authors showed that there is a higher similarity (MSE, 2D-correlation) between simulated–real (SR) mammogram pairs than real–real (RR) mammogram pairs in the presence of mammography-occult cancer. Consequently, distinguishing between healthy and mammography-occult mammograms, their classifier yielded a higher performance when trained with both RR and SR similarity as input (AUC = 0.67) than when trained only with RR pair similarity as input (AUC = 0.57). 3-dimensional conditional image synthesis with GANs has been shown, for instance, by Han et al. (2019a), who proposed a 3D Multi-Conditional GAN (3DMCGAN) to generate realistic and diverse nodules placed naturally on lung CT images to boost sensitivity in 3D object detection. Bu et al. (2020) built a 3D image-conditioned GAN based on pix2pix, where the input is a 3D volume of interest (VOI) that is cropped from a lung CT scan and contains a missing region in its centre. Both generator and discriminator contain squeeze-and-excitation (Hu et al., 2018a) residual (He et al., 2016) neural network (SE-ResNet) modules to improve the quality of the synthesised lung nodules. Another example based on lung CT images is the method by Nishio et al. (2020), where the proposed GAN model used masked 3D CT images and nodule size information to generate images.

As to multi-modal training data synthesis, Van Tulder and de Bruijne (2015) replaced missing sequences of a multi-sequence MRI with synthetic data. The authors illustrated that if the synthetic data generation model is more flexible than the classification model, the synthetic data can provide features that the classifier has not extracted from the original data, which can improve the performance. During colonoscopy, depth maps can enable navigation alongside aiding detection and size measurements of polyps. For this reason, Rau et al. (2019) demonstrated the synthesis of depth maps using a image-conditioned GAN (pix2pix) with monocular endoscopic images as input, reporting promising results on synthetic, phantom and real datasets. In breast cancer detection, Muramatsu et al. (2020) translated lesions from lung CT to breast MMG using cycleGAN yielding a performance improvement in breast mass classification when training a classifier with the domain-translated generated samples.

4.4.3. Future prospects for cancer detection and diagnosis

Granular class distinctions for synthetic tumour images. Further research opportunity exists in exploring a more fine-grained classification of tumours that characterises different subtypes and disease grades instead of binary malignant–benign classification. Being able to robustly distinguish between different disease subtypes with similar imaging phenotypes (e.g., glioblastoma versus primary central nervous system lymphoma Kang et al., 2018) addresses the challenge of reducing diagnostic ambiguity (Bi et al., 2019). GANs can be explored to augment training data with samples of specific tumour subtypes to improve the distinction capabilities of disease detection models. This can be achieved by training a detection model on training data generated by various GANs, where each GAN is trained on a different tumour subtype distribution. Another option we estimate worth exploring is to use the tumour subtype or the disease grade (e.g., the Gleason Score for prostate cancer Hu et al., 2018b) as a conditional input into the GAN to generate additional labelled synthetic training data.

Cancer image interpretation and risk estimation. Besides the detection of prospectively cancerous characteristics in medical scans, ensuring a high accuracy in the subsequent interpretation of these findings are a further challenge in cancer imaging. Improving the interpretation accuracy can reduce the number of unnecessary biopsies and harmful treatments (e.g., mastectomy, radiation therapy, chemotherapy) of indolent tumours (Bi et al., 2019). For instance, the rate of overdiagnosis of non-clinically significant prostate cancer ranges widely between 1.7% up to a noteworthy 67% (Loeb et al., 2014). To address this, detection models can be extended to provide risk and clinical significance estimations. For example, given both an input image, and an array of risk factors (e.g., BRCA1/BRCA2 status for breast cancer Li et al., 2017, comorbidity risks), a deep learning model can weight and evaluate a patient's risk based on learned associations between risk factors and input image features. The GAN framework is an example of this, where clinical, non-clinical and imaging data can be combined, either as conditional input for image generation or as prediction targets. For instance, given an input image, an AC-GAN (Odena et al., 2017; Kapil et al., 2018) can classify the risk as continuous label (see Fig. 9(e)) or, alternatively, a discriminator can be used to assess whether a risk estimate provided by a generator is realistic. Also, a generator can learn a function for transforming and normalising an input image given one or several conditional input target risk factors or tumour characteristics (e.g., a specific mutation status, a present comorbidity, etc.) to generate labelled synthetic training data.

4.5. Treatment and monitoring challenges

After a tumour is detected and properly described, new challenges arise related to planning and execution of medical intervention. In this section we examine these challenges, in particular: tumour profiling and prognosis; challenges related to choice, response and discovery of treatments; as well as further disease monitoring. Table 6 provides an overview of the cancer imaging GANs that are applied to treatment and monitoring challenges, which are discussed in the following.

4.5.1. Disease prognosis and tumour profiling

Challenges for disease prognosis. An accurate prognosis is crucial to plan suitable treatments for cancer patients. However, in specific cases, it could be more beneficial to actively monitor the tumours instead of treating them Bi et al. (2019). Challenges in cancer prognosis include the differentiation between long-term and short term survivors (Bi et al., 2019), patient risk estimation considering the complex intra-tumour heterogeneity of the tumour microenvironment (TME) (Nearchou et al., 2021), or the estimation of the probability of disease stages and tumour growth patterns, which can strongly affect outcome probabilities (Bi et al., 2019). In this sense, GANs (Li et al., 2021b; Kim et al., 2018b) and AI models in general (Cuocolo et al., 2020; Dimitriou et al., 2018) have shown potential in prognosis and survival prediction for oncology patients.

GAN disease prognosis examples. Li et al. (2021b) (in Table 2) show that their GAN-based CT normalisation framework for overcoming the domain shift between images from different centres significantly improves accuracy of classification between short-term and long-term survivors. Ahmed et al. (2021) trained omicsGAN to translate between microRNA and mRNA expression data pairs, but could be readily enhanced to also translate between cancer imaging features and genetic information. The authors evaluate omicsGAN on breast and ovarian cancer datasets and report improved prediction signals fo synthetic data tested via cancer outcome classification. Another non-imaging approach is provided by Kim et al. (2018b), who apply a GAN for patient cancer prognosis prediction based on identification of prognostic biomarker genes. They train their GAN on reconstructed human biology pathways data, which allows for highlighting genes relevant to cancer development, resulting in improvement of the prognosis prediction accuracy. In regard to these works on non-imaging approaches, we promote future extensions combining prognostic biomarker genes and -omics data with the phenotypic information present in cancer images into multi-modal prognosis models.

GAN tumour profiling examples. Related to Fig. 1(l), Vu et al. (2020a) propose that image-conditioned GANs (pix2pix) can learn latent characteristics of tissues of tumours that correlate with specific tumour grade. The authors show that when inferring their proposed BenignGAN on malignant tumour tissue images after training it exclusively on benign ones, it generates less realistic results. This allows for quantitative measurement of the differences between the original and the generated image, whereby these differences can be interpreted as tumour grade.

Kapil et al. (2018) explore AC-GAN (Odena et al., 2017) on digital pathology imagery for semi-supervised quantification of the Non-Small-Cell-Lung-Cancer biomarker *programmed death ligand 1 (PD-L1)*. Their class-conditional generator receives a one-hot encoded PD-L1 label as input to generate a respective biopsy tissue image, while their discriminator receives the image and predicts both PD-L1 label and whether the image is fake or real. The AC-GAN method compares favourably to other supervised and non-generative semi-supervised approaches, and also systematically yields high agreement with visual[21] tumour proportional scoring (TPS).

As for the analysis of the TME, Quiros et al. (2019) propose PathologyGAN, which they train on breast and colorectal cancer tissue imagery. This allows for learning the most important tissue phenotype descriptions, and provides a continuous latent representation space, enabling quantification and profiling of differences and similarities between different tumours' tissues. Quiros et al. (2019) show that lesions encoded in an GAN's latent space enable using vector distance measures to find similar lesions that are close in the latent space within large patient cohorts. We highlight the research potential in lesion latent space representations to assess inter-tumour heterogeneity. Also, the treatment strategies and successes of patients with a similar lesion can inform the decision-making process of selecting treatments for a lesion at hand, as denoted by Fig. 1(m).

Outlook on genotypic tumour profiling with phenotypic data. A further challenge is that targeted oncological therapies require genomic and immunological tumour profiling (Cuocolo et al., 2020) and effective linking of tumour genotype and phenotype. Biopsies only allow to analyse the biopsied portion of the tumour's genotype, while also increasing patient risk due to the possibility of dislodging and seeding of neoplastic altered cells (Shyamala et al., 2014; Parmar et al., 2015). Therefore, a trade-off[22] exists between minimising the number of biopsies and maximising the biopsy-based information about a tumour's genotype. These reasons and the fact that current methods are invasive, expensive, and time-consuming (Cuocolo et al., 2020) make genotypic tumour profiling an important issue to be addressed by AI cancer imaging methods. In particular adversarial deep learning models are promising to generate the non-biopsied portion of a tumour's genotype after being trained on paired genotype and radiology imaging data.[23] We recommend future studies to explore this line of research, which is regarded as a key challenge for AI in cancer imaging (Bi et al., 2019; Parmar et al., 2015).

4.5.2. Treatment planning and response prediction

Challenges for cancer treatment predictions. A considerable number of malignancies and tumour stages have various possible treatment options and almost no head-to-head evidence to compare them to. Due to that, oncologists need to subjectively select an approved therapy based on their individual experience and exposure (Troyanskaya et al., 2020).

Furthermore, despite existing treatment response assessment frameworks in oncology, inter- and intra-observer variability regarding choice and measurement of target lesions exists among oncologists and radiologists (Levy and Rubin, 2008). To achieve consistency and accuracy in standardised treatment response reporting frameworks (Levy and Rubin, 2008), AI and GAN methods can identify quantitative biomarkers[24] from medical images in a reproducible manner useful for risk and treatment response predicts (Hosny et al., 2018).

Apart from the treatment response assessment, treatment response prediction is also challenging, particularly for cancer treatments such as immunotherapy (Bi et al., 2019). In cancer immunogenomics, for instance, unsolved challenges comprise the integration of multi-modal data (e.g., radiomic and genomic biomarkers Bi et al., 2019), immunogenicity prediction for neoantigens, and the longitudinal non-invasive monitoring of the therapy response (Troyanskaya et al., 2020). In regard to the sustainability of a therapy, the inter- and intra-tumour heterogeneity (e.g., in size, shape, morphology, kinetics, texture, etiology) and potential sub-clone treatment survival complicates individual treatment prediction, selection, and response interpretation (Bi et al., 2019).

[21] A visual estimation of pathologists of the tumour cell percentage showing PD-L1 staining.

[22] Due to this and due to the high intra-tumour heterogeneity, available biopsy data likely only describes a subset of tumour's clonal cell population.

[23] Imaging data on which the entire lesion is visible to allow learning correlations between phenotypic tumour manifestations and genotype signatures.

[24] For example, characteristics and density variations of the parenchyma patterns on breast images (Bi et al., 2019).

Table 6
Overview of adversarially-trained models applied to **treatment and monitoring** challenges. Publications are clustered by section and ordered by year in ascending order.

Publication	Method	Dataset	Modality	Task	Highlights
Disease prognosis					
Kim et al. (2018b)	GAN-based	TCGA (Tomczak et al., 2015), Reactome (Croft et al., 2014; Fabregat et al., 2017)	[non-imaging] multi-omics cancer data	Data synthesis	Biomarker gene identification for pancreas, breast, kidney, brain, and stomach cancer with GANs and PageRank.
Ahmed et al. (2021)	omicsGAN	TCGA (Network et al., 2011; Ciriello et al., 2015)	[non-imaging] ovarian/ breast gene expression	Paired translation	microRNA to mRNA translation and vice versa. Synthetic data improves cancer outcome classification.
Tumour profiling					
Kapil et al. (2018)	AC-GAN	Private	Lung histopathology	Classification	AC-GAN CLF of PD-L1 levels for lung tumour tissue images obtained via needle biopsies.
Quiros et al. (2019) PathologyGAN	PathologyGAN	VGH/NKI (Beck et al., 2011), NCT (Kather et al., 2018)	Breast/colorectal histopathology	Representation learning	Learning tissue phenotype descriptions & tumour representations. Combines BigGAN (Brock et al., 2018), StyleGAN (Karras et al., 2019) & RAD (Jolicoeur-Martineau, 2018)
Vu et al. (2020a)	BenignGAN	Private	Colorectal histopathology	Paired translation	Edge map-to-image. As trained on only benign, malignant images quantifiable via lower realism.
Treatment response prediction					
Kadurin et al. (2017a) and Kadurin et al. (2017b)	AAE-based druGAN	Pubchem BioAssay (Wang et al., 2014)	[non-imaging] Molecular fingerprint data	Representation learning	AAE for anti-cancer agent drug discovery. AAE input/output: molecular fingerprints & log concentration vectors.
Goldsborough et al. (2017)	CytoGAN	BBBC021 (Ljosa et al., 2012b)	Cytopathology	Representation learning	Grouping cells with similar treatment response via cell image representations. Based on DCGAN, LSGAN, WGAN.
Yoon et al. (2018)	GANITE	USA 89-91 Twins (Almond et al., 2004)	[non-imaging] individualised treatment effects	Multi-class-conditional synthesis	cGANs for individual treatment effect prediction, including unseen counterfactual outcomes and confidence intervals.
Ge et al. (2020)	MGANITE	AML clinical trial (Kornblau et al., 2009)	[non-imaging] individualised treatment effects	Multi-class-conditional synthesis	GANITE extension introducing dosage quantification and continuous and categorical treatment effect estimation.
Bica et al. (2020)	SCIGAN	CGA (Weinstein et al., 2013), MIMIC III (Johnson et al., 2016)	[non-imaging] individualised treatment effects	Multi-class-conditional synthesis	GANITE extension introducing continuous interventions and theoretical explanation for GAN counterfactuals.
Radiation dose planning					
Mahmood et al. (2018)	pix2pix-based	Private	Oropharyngeal CT	Paired translation	Translating CT to 3D dose distributions without requiring hand-crafted features.
Maspero et al. (2018)	pix2pix	Private	Prostate/rectal/cervical CT/MRI	Paired translation	MR-to-CT translation for MR-based radiation dose planning without CT acquisition.
Murakami et al. (2020)	pix2pix	Private	Prostate CT	Paired translation	CT-to-radiation dose distribution image translation without time-consuming contour/organs at risk (OARs) data.
Peng et al. (2020)	pix2pix, CycleGAN	Private	Nasopharyngeal CT/MRI	Unpaired/Paired translation	Comparison of pix2pix & CycleGAN-based generation of CT from MR for radiation dose planning.
Kearney et al. (2020a)	DoseGAN	Private	Prostate CT/PTV/OARs	Paired translation	Synthesis of volumetric dosimetry from CT+PTV+OARs even in the presence of diverse patient anatomy.
Disease tracking & monitoring					
Kim et al. (2019b)	CycleGAN	Private	Liver MRI/CT/dose	Unpaired translation	Pre-treatment MR+CT+dose translation to post-treatment MRI → predicting hepatocellular carcinoma progression.
Elazab et al. (2020)	GP-GAN	BRATS 2014 (Menze et al., 2014)	Cranial MRI	Paired translation	3D U-Net G generating progression image from longitudinal MRI to predict glioma growth between time-step.
Li et al. (2020a)	DC-AL GAN, DCGAN	Private	Cranial MRI	Noise-to-image synthesis	CLF uses D representations to distinguish pseudo- and true glioblastoma progression.

GAN treatment effect estimation examples. In line with Fig. 1(n), Yoon et al. (2018) propose the conditional GAN framework 'GANITE', where individual treatment effect prediction allows for accounting for unseen, counterfactual outcomes of treatment. GANITE consists of two GANs: first, a counterfactual GAN is trained on feature and treatment vectors along with the factual outcome data. Then, the trained generator's output is used for creating a dataset, on which the other GAN, called ITE (Individual Treatment Response) GAN, is being trained. GANITE provides confidence intervals along with the prediction, while being readily scalable for any number of treatments. However, it does not allow for taking time, dosage or other treatment parameters into account. MGANITE, proposed by Ge et al. (2020), extends GANITE by introducing dosage quantification, and thus enables continuous and categorical treatment effect estimations. SCIGAN (Bica et al., 2020) also extends upon GANITE and predicts outcomes of continuous rather than one-time interventions and the authors further provide theoretical justification for GANs' success in learning counterfactual outcomes. As to the problem of individual treatment response prediction, we suggest that quantitative comparisons of GAN-generated expected post-treatment images with real post-treatment images can yield interesting insight for tumour interpretation. We encourage future work to explore

generating such post-treatment tumour images given a treatment parameter and a pre-treatment tumour image as conditional inputs. With varying treatment parameters as input, it is to be investigated whether GANs can inform treatment selection by simulating various treatment scenarios prior to treatment allocation or whether GANs can help to understand and evaluate treatment effects by generating counterfactual outcome images after treatment application.

Goldsborough et al. (2017) present an approach called CytoGAN, where they synthesise fluorescence microscopy cell images using DCGAN, LSGAN, or WGAN. The discriminator's latent representations learnt during synthesis enable grouping encoded cell images together that have similar cellular reactions to treatment by chemicals of known classes (morphological profiling).[25] Even though the authors reported that CytoGAN obtained inferior result[26] compared to classical, widely applied methods such as CellProfiler (Singh et al., 2014), using GANs to group tumour cells representations to inform chemical cancer treatment allocation decisions is an interesting approach in the realm of treatment selection, development (Kadurin et al., 2017a,b) and response prediction.

GAN radiation dose planning examples. As radiation therapy planning is labour-intensive and time-consuming, researchers have been spurred to pursue automated planning processes (Sharpe et al., 2014). As outlined in the following and suggested by Fig. 1(o), the challenge of automated radiation therapy planning can be approached using GANs.

By framing radiation dose planning as an image colourisation problem, Mahmood et al. (2018) introduced an end-to-end GAN-based solution, which predicts 3D radiation dose distributions from CT without the requirement of hand-crafted features. They trained their model on Oropharyngeal cancer data along with three traditional ML models and a standard CNN as baselines. The authors trained a pix2pix (Isola et al., 2017) GAN on 2D CT imagery, and then fed the generated dose distributions to an inverse optimisation (IO) model (Babier et al., 2018), in order to generate optimised plans. Their evaluation showed that their GAN plans outperformed the baseline methods in all clinical metrics.

Kazemifar et al. (2020) (in Table 2) proposed a cGAN with U-Net generator for paired MRI to CT translation. Using conventional dose calculation algorithms, the authors compared the dose computed for real CT and generated CT, where the latter showed high dosimetric accuracy. The study, hence, demonstrates the feasibility of synthetic CT for intensity-modulated proton therapy planning for brain tumour cases, where only MRI scans are available.

Maspero et al. (2018) proposed a GAN-assisted approach to quicken the process of MR-based radiation dose planning, by using a pix2pix for generating synthetic CTs (sCTs) required for this task. They show that a conditional GAN trained on prostate cancer patient data can successfully generate sCTs of the entire pelvis.

A similar task has also been addressed by Peng et al. (2020). Their work compares two GAN approaches: one is based on pix2pix and the other on a CycleGAN (Zhu et al., 2017). The main difference between these two approaches was that pix2pix was trained using registered MR-CT pairs of images, whereas CycleGAN was trained on unregistered pairs. Ultimately, the authors report pix2pix to achieve results (i.e. mean absolute error) superior to CycleGAN, and highlight difficulties in generating high-density bony tissues using CycleGAN.

The recently introduced attention-aware DoseGAN (Kearney et al., 2020a) overcomes the challenges of volumetric dose prediction in the presence of diverse patient anatomy. As illustrated in Fig. 14, DoseGAN is based on a variation of the pix2pix architecture with a 3D encoder–decoder generator (L1 loss) and a patch-based patchGAN discriminator (adversarial loss). The generator was trained on concatenated CT, planning target volume (PTV) and organs at risk (OARs) data of prostate cancer patients, and the discriminator's objective was to distinguish the real dose volumes from the generated ones. Both qualitatively and quantitatively, DoseGAN was able to synthesise more realistic volumetric doses compared to current alternative state-of-the-art methods.

Murakami et al. (2020) published another GAN-based fully automated approach to dose distribution of Intensity-Modulated Radiation Therapy (IMRT) for prostate cancer. The novelty of their solution is that it does not require the tumour contour information, which is time-consuming to create, to successfully predict the dose based on the given CT dataset. Their approach consists of two pix2pix-based architectures, one trained on paired CT and radiation dose distribution images, and the other trained on paired structure images and radiation dose distribution images. From the generated radiation dose distribution images the dosimetric parameters for the PTV and OARs are computed. The generated dosimetric parameters differed on average only between 1%–3% with respect to the original ground truth dosimetric parameters.

Koike et al. (2020) proposed a CycleGAN for dose estimation for head and neck CT images with metal artifact removal in CT-to-CT image translation as described in Table 2. Providing consistent dose calculation against metal artifacts for head and neck IMRT, their approach achieves dose calculation performance similar to commercial metal artifact removal methods.

4.5.3. Disease tracking and monitoring

Challenges in tracking and modelling tumour progression. Tumour progression is challenging to model (Huang et al., 2020) and commonly requires rich, multi-modal longitudinal data sets. As cancerous cells acquire growth advantages through genetic mutation in a process arguably analogous to Darwinian evolution (Hanahan and Weinberg, 2000), it is difficult to predict which of the many sub-clones in the TME will outgrow the other clones. A tumour lesion is, hence, constantly evolving in phenotype and genotype (Bi et al., 2019) and might acquire dangerous further mutations over time, anytime. The TME's respective impact is exemplified by the stage II colorectal cancer outcome classification performance gain in Dimitriou et al. (2018), which is likely attributable to the high prognostic value of the TME information in their training data.

In addition, concurrent conditions and alterations in the organ system surrounding a tumour, but also in distant organs may not only remain undetected, but could also influence patient health and progression (Bi et al., 2019). GANs can generate hypothetical comorbidity data[27] to aid awareness, testing, finding, and analysis of complex disease and comorbidity patterns. A further difficulty for tumour progression modelling is the a priori unknown effect of treatment. Treatment effects may even remain partly unknown after treatment for example in the case of radiation therapy[28] (Verma et al., 2013) or after surgery[29] (Bi et al., 2019).

GAN tumour progression modelling examples. Relating to Fig. 1(p), GANs can not only diversify the training data, but can also be applied to simulate and explore disease progression scenarios (Elazab et al., 2020). For instance, Elazab et al. (2020) propose GP-GAN, which uses stacked 3D conditional GANs for growth prediction of glioma based on longitudinal MR images. The generator is based on the U-Net architecture (Ronneberger et al., 2015) and the segmented feature maps

[25] CytoGAN uses an approach comparable to the one shown in Fig. 9(g).

[26] i.e. mechanism-of-action classification accuracy.

[27] For example from EHR (Hwang et al., 2017; Dashtban and Li, 2020), imaging data, or a combination thereof.

[28] Radiation therapy can result in destruction of the normal tissue (e.g., radionecrosis) surrounding the tumour. Such heterogeneous normal tissue can become difficult to characterise and distinguish from the cancerous tissue (Verma et al., 2013).

[29] It is challenging to quantify the volume of remaining tumour residuals after surgical removal (Bi et al., 2019).

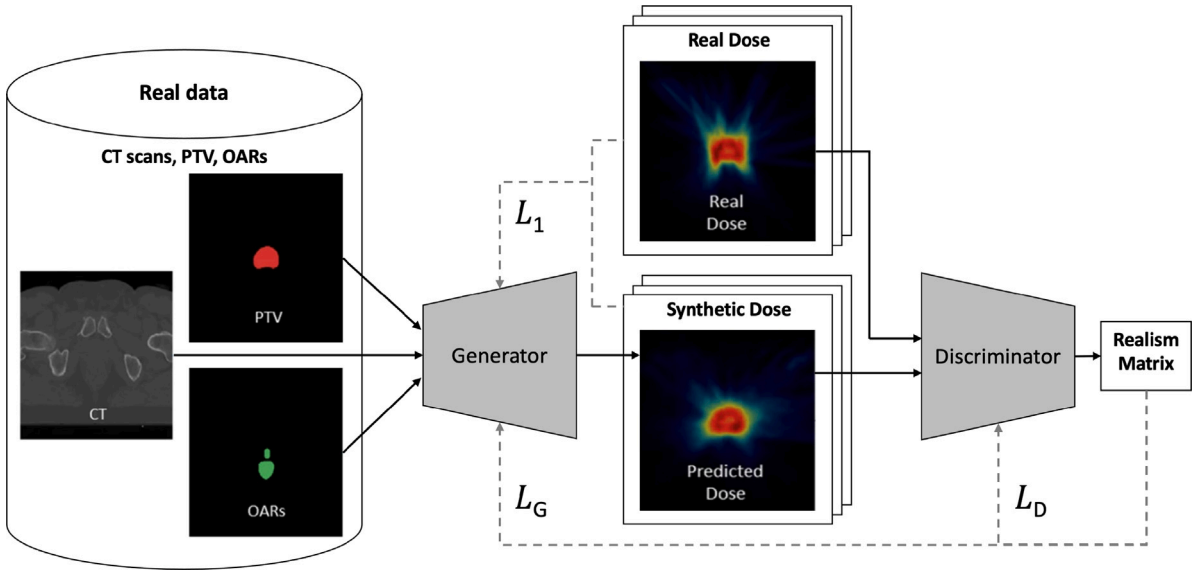

Fig. 14. GAN architecture of DoseGAN adapted from Kearney et al. (2020a) and based on pix2pix (Isola et al., 2017). Given concatenated CT scans, planning target volume (PTV) and organs at risk (OARs), the generator of DoseGAN addresses the challenge of volumetric dose prediction for prostate cancer patients.

are used in the training process. Kim et al. (2019b) trained a CycleGAN on concatenated pre-treatment MR, CT and dose images (i.e. resulting in one 3-channel image) of patients with hepatocellular carcinoma to generate follow-up enhanced MR images. This enables tumour image progression prediction after radiation treatment, whereby CycleGAN outperformed a vanilla GAN baseline.

The deep convolutional (DC) (Radford et al., 2015) - AlexNet (AL) (Krizhevsky et al., 2012) GAN (DC-AL GAN) proposed by Li et al. (2020a) is trained on longitudinal diffusion tensor imaging (DTI) data of pseudoprogression (PsP) and true tumour progression (TTP) in glioblastoma multiforme (GBM) patients. Both of these progression types can occur after standard treatment[30] and they are often difficult to differentiate due to similarities in shape and intensity. In DC-AL GAN, representations are extracted from various layers of its AlexNet discriminator that is trained on discriminating between real and generated DTI images. These representations are then used to train a support vector machine (SVM) classifier to distinguish between PsP and TTP samples achieving promising performance.

We recommend further studies to extend on these first adversarial learning disease progression modelling approaches. One potential research direction are GANs that simulate environment and tumour dependent progression patterns based on conditional input data such as the tumour's gene expression data (Xu et al., 2020) or the progressed time between original image and generated progression image (e.g., time passed between image acquisitions or since treatment exposure). To this end, unexpected changes of a tumour may be uncovered between time points or deviations from a tumour's biopsy proven genotypic growth expectations.[31]

5. Meta-analysis: Towards a framework for the assessment of trustworthiness and validation

5.1. Trustworthiness of medical image synthesis studies

Section 4 presented an extensive analysis of the challenges, existing publications, and state-of-the-art data synthesis and adversarial network methods in cancer imaging. While the methodologies, experiments, and results of these studies were elaborated, their validity and trustworthiness was not specifically addressed. The validity and trustworthiness varies between studies and depends on the breadth and depth of the methodological evaluation and the analysis of potential limitations. In the absence of a rigorous evaluation indicating otherwise, the methodology and experimental results of a study cannot be readily assumed to be transferable across domains, settings, tasks, datasets and modalities. Hence, while a study reports promising results for a particular task and seemingly solves the task's underlying (cancer imaging) challenge, modest changes in the dataset, evaluation method or evaluation metrics can lead to different results and conclusions. This points to the need of a principled assessment of trustworthiness and validity of studies in the cancer and medical imaging domains, in particular, the ones contributing and evaluating synthetic data and data generation methodology.

Some frameworks have proposed guidelines and best practices for the development of trustworthy artificial intelligence solutions in medical imaging (Lekadir et al., 2021; Hasani et al., 2022). However, to the best of our knowledge, no framework has been proposed for trustworthiness assessment of studies focused on medical image synthesis solutions. Building upon the FUTURE-AI consensus guidelines (Lekadir et al., 2021) and the lesson's learned from the extensive analysis of the 164 publications presented in Section 4, we propose the Synthesis Study Trustworthiness Test (*SynTRUST*) as a principled framework to evaluate medical image synthesis studies.

5.2. Proposing the SynTRUST framework

The Synthesis Study Trustworthiness Test (*SynTRUST*) framework consists of a principled set of measures to assess the trustworthiness and validity of studies proposing generative models, synthetic data, or adversarial training methods in medical and cancer imaging. It is based on five core principles, namely,

(i) ***T**horoughness* of experimental design and validation.
(ii) ***R**eproducibility* and transparency of results, data, models, and implementation.
(iii) ***U**sefulness* and versatility of synthesis method, model, and generated data.
(iv) ***S**calability* and transferability of the methodology and the results across clinical domains.

[30] Pseudoprogression occurs in 20%–30% of GBM patients (Li et al., 2020a).
[31] For example, by comparing the original patient image after progression with the GAN-generated predicted image (or its latent representation) after progression for time spans of interest.

Table 7
Illustration of the ***SynTRUST*** *Framework* for evaluation of trustworthiness and validation vigour of studies that propose generative models, synthetic data, or adversarial training in medical and cancer imaging. ***SynTRUST*** is based on the 5 core principles ***T****horoughness*, ***R****eproducibility*, ***U****sefulness*, ***S****calability*, and ***T****enability*. For each overarching principle, a set of concrete corresponding measures is defined. Each of the 26 ***SynTRUST*** measures is associated to an ID for reference and is assigned an importance rating, where 1 stands for *Essential*, 2 for *Desirable*, and 3 for *Recommended*.

Principle	ID	Rating	Definition
Thoroughness: Validity of experiments			
Minimum test set size	Th1	1	Representative test set size should allow confident conclusions (e.g., > 30 cases, > 100 images, ≥ 20% of training data).
Multi-metric reporting	Th2	1	Multiple standardised metrics (e.g., FID, SSIM, downstream task metric) evaluate the synthesis method (e.g., ≥ 2).
Multiple result validation runs	Th3	2	Mean & variance over multiple runs to be reported for all metrics (e.g., ≥ 3 random seeds, or ≥ 5-fold cross-validation).
Fair baseline comparison	Th4	2	Fairest-possible comparison with closest-possible generative/adversarial/downstream task model baseline.
Principled benchmark definition	Th5	2	Systematic construction of heterogeneous (patients, pathologies, acquisitions) test set(s), without data leaking.
Statistical significance testing	Th6	3	Validation that reported performance variation & improvements are statistically significant.
Ablation study of ket components	Th7	3	Testing removal of both downstream & generative model parts for insight on impact.
Effect of varying training set sizes	Th8	3	Testing data scarcity impact by systematically reducing downstream and generative model training data.
Reproducibility: Transparency of study			
Detailed reporting of design decisions	R1	1	Study design & experiments are defined & reported with rationales and attention to detail.
Public availability of dataset	R2	2	At least one reported evaluation dataset is publicly accessible allowing repetition of experiments.
Public availability of software code	R3	2	Source code is shared in publicly accessible repository providing method implementation, ideally with documentation.
Public availability of model weights	R4	3	Sharing of model weights for reusing the trained model for faster and sustainable reproducing of experiments.
Usefulness: Versatility of synthesis method			
Synthesis method usability testing	U1	1	Solid generative/adversarial model usefulness evaluation on at least one community-defined (downstream) clinical task.
Quantitative quality measurement	U2	2	Assessment of synthetic data quality (e.g., via FID) or adversarial loss and their correlation with downstream task metrics.
Qualitative quality measurement	U3	3	Observer study with clinicians to assess synthesis model output on realism, utility, quality, and diversity.
Mode collapse analysis	U4	3	Diversity of generative model modes is analysed (e.g., via visual inspection and t-sne of synthetic & real distributions).
Scalability: Transferability of methodology			
Real-world representing data	S1	1	Evaluation on cases & samples highly representative of medically-relevant real-world clinical data.
Multi-dataset evaluation	S2	2	Evaluation of generative model on multiple datasets/modalities demonstrating scalability, ideally for different organs.
Multi-centre evaluation	S3	3	Evaluation of generative model per centre showing generalisability across centre-specific variations.
Multi-downstream task evaluation	S4	3	Evaluate generative model versatility via test with multiple downstream models & tasks (e.g. segmentation, classification).
Downstream task robustness evaluation	S5	3	Performance variation test for simulated train & test acquisition, manifestation, population, annotation, prevalence shifts.
Tenability: Acceptability of trained model			
Condition adherence testing	Te1	1	Test of preciseness & reliability of presence of (input) conditions in the (synthetic) data (e.g., via classification).
Bias awareness analysis	Te2	2	Discussion & analysis how bias from dataset (e.g., age, gender, ethnicity, in/exclusion criteria) transfers into model.
Model hallucination tendency analysis	Te3	3	Analysis of undesired removal/addition of features such as artifacts or tumours (e.g., via inspection or classification).
Fairness variation testing	Te4	3	Change in fairness is measured for generative/adversarial model intervention (e.g., via equalised odds in downstream task).
Privacy preservation testing	Te5	3	Investigation of patient-identifying feature leakage and training data reconstruction risk given generative model (output).

(v) ***T****enability*, acceptability, and reliability of the properties of the model and respective synthetic data.

The methodology applied to derive the SynTRUST framework is composed of several consecutive steps, outlined as follows.

1. Observation of experimental evaluation methods in the surveyed cancer imaging papers.
2. Questioning to which extent an observed study concludes with a generally-applicable, scientifically-sound finding.
3. Definition of causes as to why the results of the study are limited in general-applicability and trustworthiness.
4. Suggestion of additional validation methods that can increase the study's general-applicability.
5. Grouping and formalisation of suggestions into 26 concrete validation measures.
6. Definition of an overarching principle for each group of measures resulting in the 5 core principles: ***T****horoughness*, ***R****eproducibility*, ***U****sefulness*, ***S****calability*, and ***T****enability*.
7. Refinement of the measures to complement with and extend on expert consensus on best practices for the application of artificial intelligence in medical imaging (Lekadir et al., 2021).
8. Importance rating of each measure from 1 to 3 based on their estimated impact on trustworthiness. A rating of 1 indicates *essential* measures with the highest importance, a rating of 2 characterises *desirable* measures, and a rating of 3 depicts measures that are *recommended* additions to a study.

The resulting *SynTRUST* framework is illustrated in Table 7. Table 7 contains the title, the definition, the importance rating, and an ID for reference for each of the 26 measures, grouped by the 5 *SynTRUST* principles.

5.3. Analysis of cancer imaging challenges using SynTRUST

5.3.1. SynTRUST study curation

Towards the objective of evaluating the trustworthiness of cancer imaging solutions, we demonstrate in the following how the *SynTRUST* framework can be used to analyse medical imaging publications. This not only shows the practicability of the *SynTRUST* framework, but also estimates the trustworthiness of current results in the field. The latter allows to corroborate concrete quality-controlled conclusions about the progress and state-of-the-art in adversarial networks in cancer imaging.

Table 8
Selection of studies that employ data synthesis and adversarial networks methodology curated based on their promising potential towards solving the cancer imaging challenges surveyed in Sections 4.1–4.5. Each of the *studies* represents one concrete *proposed solution* to one of the *challenges*.

Cancer imaging challenge	Proposed solution	Representative study
Imbalanced/biased data (4.1.2)	Adversarially-trained bias-free representations	Li et al. (2021a)
Dataset shifts (4.1.3)	Multi-modal image translation	Yurt et al. (2019)
Uncertain synthetic data usability (4.1.4)	Feature hallucination evaluation metric	Cohen et al. (2018b,a)
Uncurated data (4.1.5)	Generative image correction & denoising model	Armanious et al. (2020)
Privacy risks in data sharing (4.2.1)	Federated (differentially-private) image synthesis	Chang et al. (2020b,a)
Adversarial attacks and defences (4.2.5)	Adversarial example-based augmentation	Liu et al. (2020b)
Costly human annotation (4.3.1)	Uncertainty-aware annotation generation	Hu et al. (2020)
Weak domain generalisation (4.3.2)	Adversarially-trained cross-domain segmentation	Kamnitsas et al. (2017)
Extracted feature variation (4.3.2)	Discriminator learning radiomics correlations	Xiao et al. (2019)
Intra/inter-observer variability (4.3.2)	Observer averaging via mask discriminator	Sarker et al. (2019)
Radiologists' high error rate (4.4.1)	Detection improving synthetic data augmentation	Zhao et al. (2020a)
Intra/inter-tumour heterogeneity (4.4.2)	Adversarially-trained anomaly detection	Kuang et al. (2020)
Uncertain tumour profiles (4.5.1)	Adversarially-trained representation comparison	Quiros et al. (2019)
Unknown treatment response (4.5.2)	Semi-supervised treatment biomarker quantification	Kapil et al. (2018)
Unknown treatment dose (4.5.2)	Synthesis of volumetric dosimetry images	Kearney et al. (2020a)
Uncertain disease progression (4.5.3)	Tumour progression image generation	Elazab et al. (2020)

Table 9
Results of the in-depth analysis of all *essential* and *desirable* measures of the *SynTRUST* framework for studies proposing adversarial network methodology. The analysed studies are selected in Table 8 and represent solutions to key cancer imaging challenges. The *SynTRUST* measures are referenced by ID from Table 7. The blue check mark symbol indicates a positive evaluation, while the red and orange cross symbols respectively indicate a negative evaluation of an essential or desirable measure. The evaluated *essential* measures are *minimum test set size* (Th1), *multi-metric reporting* (Th2), *detailed reporting of design decisions* (R1), *synthesis method usability testing* (U1), *real-world representing data* (S1), *condition adherence testing* (Te1). The evaluated *desirable* measures are *multiple result validation runs* (Th3), *fair baseline comparison* (Th4), *principled benchmark definition* (Th5), *public availability of dataset* (R2), *public availability of software code* (R3), *quantitative quality measurement* (U2), *multi-dataset evaluation* (S2), *bias awareness analysis* (Te2).

Representative study	SynTRUST framework													
	1: Essential measures						2: Desirable measures							
	Th1	Th2	R1	U1	S1	Te1	Th3	Th4	Th5	R2	R3	U2	S2	Te2
Li et al. (2021a)	✔	✔	✔	✔	✔	✔	✔	✔	✘	✔	✘	✔	✘	✔
Yurt et al. (2019)	✔	✔	✔	✔	✔	✔	✘	✔	✔	✔	✔	✘	✔	✘
Cohen et al. (2018b,a)	✔	✔	✔	✔	✔	✔	✘	✔	✘	✔	✔	✔	✘	✘
Armanious et al. (2020)	✘	✔	✔	✔	✔	✔	✘	✔	✘	✘	✘	✔	✔	✘
Chang et al. (2020b,a)	✔	✔	✔	✔	✔	✔	✘	✔	✘	✔	✔	✘	✔	✘
Liu et al. (2020b)	✔	✘	✔	✔	✔	✔	✘	✔	✔	✔	✘	✘	✔	✘
Hu et al. (2020)	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✘	✘	✔	✘
Kamnitsas et al. (2017)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✔	✔	✔	✘
Xiao et al. (2019)	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✘	✘	✘	✘
Sarker et al. (2019)	✔	✔	✔	✔	✔	✔	✘	✔	✘	✔	✘	✘	✔	✘
Zhao et al. (2020a)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✘	✘	✘
Kuang et al. (2020)	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✘	✘
Quiros et al. (2019)	✔	✔	✔	✘	✔	✔	✔	✘	✔	✔	✔	✘	✔	✘
Kapil et al. (2018)	✔	✔	✔	✔	✔	✔	✘	✔	✘	✘	✘	✘	✘	✘
Kearney et al. (2020a)	✘	✔	✔	✔	✔	✔	✘	✔	✘	✘	✘	✔	✘	✘
Elazab et al. (2020)	✘	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✘	✔	✘

In our analysis we first sample the present-day challenges in cancer imaging that were surveyed in Section 4 and summarised in Fig. 1. Next, we carefully select representative adversarial network publications to represent a particular challenge and its solution. This selection is based on the criteria that the publication (a) proposes a particularly promising solution to its respective challenge, (b) contributes a methodology that is generally-applicable across domains and (c) report promising results. Most of the sampled publications further (d) have shown more impact and were referenced in other relevant studies. The selected studies are displayed in Table 8 together with their representative solution and associated cancer imaging challenge.

5.3.2. SynTRUST study assessment

Next, we analyse each of the selected publications independently based on the *SynTRUST* framework. We choose to base our analysis on the most important measures of the *SynTRUST* framework that, as shown in Table 7, have received either a rating of *1* as *essential* or a rating of 2 as *desirable*. For the sake of conciseness, we leave the analysis of less critical measures rated as 3 (*recommended*) to further studies. The results of our analysis of each of the selected publications are summarised in Table 9.

Essential SynTRUST measures. We observe that the analysed studies overall show strong trustworthiness and validity considering the *essential* measures: 11 out of 16 studies fulfil all of the essential criteria, while the remaining 5 studies fulfil all but one essential measures. For 3 out of these 5 studies, the only essential measure that is not fulfilled is Th1 (*minimum test set size*). For instance, studies that pioneer methodologies on promising new clinical applications, such as generative tumour progression modelling (Elazab et al., 2020), it is particularly challenging to encounter datasets suitable for the clinical task at hand. Even though the number of test images exceeds the defined minimum of 100 in these studies, the number of different patients (cases) is lower than 30. 30 was defined as the indicative minimum of cases to allow for conclusions for the larger patient population.[32] All 16 studies have a detailed reporting of design decisions (R1), train and test on real-world representing clinical data (S1), and test the conditions of their adversarial network (Te1). Also, 15 out of 16 studies report multiple standardised performance metrics to evaluate the adversarial network (Th2) and demonstrate their method's usefulness on a clinically relevant downstream task (U1). In sum, the result for the essential

[32] Based on the central limit theorem, 30 is a popular choice and rule-of-thumb for the minimum sample size of a population.

measures demonstrates that the reported performance and progress of the analysed studies are considerably reliable and trustworthy.

Desirable SynTRUST measures. While the 6 *essential* basic trustworthiness requirements are mostly fulfilled, the result for the 8 *desirable* measures is more varied. This highlights that the studies have a general high level of trustworthiness, but a lower level of trustworthiness for the more specific and nuanced aspects of their reported results and validations. For instance, while 15 out of 16 studies included a comparison with a suitable baseline (Th4), multiple studies did not accomplish a positive evaluation of Th3 (8), Th5 (9), R2 (7), R3 (11), U2 (11), S2 (7), and Te2 (15).

- Regarding Th3, often studies defined a static train and test set without running experiments multiple times. For example, multiple different random seed network weight initialisations or k-fold cross-validation are options to corroborate results by demonstrating stable performance with reported mean and standard deviation across runs/folds.
- Regarding Th5, in general the train-test split ensured no data leaking between training and testing sets, e.g., with images from the same patient not being in both sets. However, the benchmark test sets were often not defined systematically to ensure validating the methods on a varied distribution of, e.g., cases, patients, pathologies, and acquisition parameters.
- For R2, we observe that often the studies' datasets are not public available, which limits the reproducibility of the results. Often, this is due to the collection and usage of private patient data from hospitals. Further limiting factors are the high effort to repeat the study on public datasets or the specificity of the clinical task rendering its evaluation non-viable on the available public datasets.
- Analysing R3 shows that the software implementing the studies' methods and experiments is often not shared publicly in code repositories, which reduces reproducibility and impedes rerunning experiments with exactly the same code base used in the respective study.
- Regarding U2, often the correlation between (a) the downstream tasks and (b) either the synthetic data quality (e.g., in the case of generative models) or the adversarial loss (e.g. in the case of adversarial training) is not analysed. Such an analysis informs on the usefulness of the quality of the respective model and on its contribution to the results on the clinical task.
- As to S2, often the method is validated on, both, (a) a single dataset and (b) a single modality, while a desirable evaluation would use multiple datasets, modalities, ideally further demonstrating the method's transferability across organs, clinical domains and acquisition protocols.
- For Te2, we note the general absence of an analysis of the bias that is transferred from the training dataset into the models. For instance, a model trained on a homogeneous patient population sample, e.g., in terms of gender, sex, ethnicity, geography, likely is biased towards this subset of the overall population and can result in unequal treatment of patients from other subsets. Model biases can be detected by reviewing (a) the dataset statistics, (b) the model performance shifts on carefully subset patient samples, and (c) the exclusion and inclusion criteria applied in the data acquisition and curation processes. This enables to report and potentially mitigate otherwise unknown model biases, which increases the knowledge and reliability of a model's properties.

In concluding our meta-analysis, we highlight the high general level of trustworthiness of the selected adversarial network publications based on our assessment of the *essential SynTRUST* measures. This demonstrates technical maturity of adversarial training and image synthesis methods in cancer imaging. As described in the Sections 4.1–4.5, many approaches towards solving the challenges in cancer imaging are not yet fully explored. Nonetheless, the solutions that have been pioneered and validated are shown to be relatively trustworthy and solid.

However, our meta-analysis also revealed that specific *desirable* trustworthiness criteria that go beyond basic *essential* validation are often not fulfilled, even by the most promising and in-depth studies in the field. For instance, a wider practice of data and code sharing is desirable. Closing this gap will not only increase reproducibility, but also accelerate adoption of existing methods and further innovation. part from that, the validation of biases and fairness criteria in datasets and models is largely overlooked despite its importance to ensure a model's acceptability and trust in the clinical setting.

We motivate further studies to address and build upon the gaps our analysis has revealed regarding the trustworthiness of existing cancer imaging studies. In this regard, we highlight the *SynTRUST* framework not only as a means for study evaluation, but also as a guideline guiding the design of future image synthesis studies.

6. Discussion and future perspectives

6.1. Adversarial methods in cancer imaging over the years

As presented in Fig. 15(c), we have included 164 of the surveyed GAN-based data synthesis and adversarial training publications in the timeframe from 2017 until March 7th 2021. We observe that the numbers of these cancer imaging GAN publications has been increasing from 2017 to 2020 from 10 to 63 with a surprising slight drop between 2018 to 2019 (41 to 38). The final number of respective publications for 2021 is still pending. The trend towards publications that propose GANs and adversarial training to solve cancer imaging challenges demonstrates the considerable research attention that the adversarial learning scheme has been receiving in this field. Following our literature review in Section 4, the need for further research in adversarial networks seems not yet to be met. We were able to highlight various lines of research for GANs and adversarial training in oncology, radiology, and pathology that have received limited research attention or are untapped research potentials. These potentials indicate a continuation of the trend towards more data synthesis and adversarial training applications and standardised integration of GAN-generated synthetic data into medical image analysis pipelines and software solutions.

6.2. Modality biases

In regard to imaging modalities, we analyse in Fig. 15(b) how much research attention each modality has received in terms of the number of corresponding publications. By far, MRI and CT are the most dominant modalities with 61, and 53 publications, respectively, followed by MMG (13), dermoscopy (12) and PET (6). The wide spread between MRI and CT and less investigated domains such as endoscopy (3), ultrasound (3), and digital tomosynthesis (0) is to be critically remarked. Due to variations in the imaging data between these modalities (e.g., spatial resolutions, pixel dimensions, domain shifts), it cannot be readily assumed that a GAN application with desirable results in one modality will produce equally desirable results in another. Due to that and with awareness of the clinical importance of MRI and CT, we suggest a more balanced application of GANs and adversarial training across modalities including experiments on rare modalities to demonstrate the clinical versatility and applicability of GAN-based solutions. Alongside the open-access datasets described by Diaz et al. (2021), we highlight the following additional recent open datasets to facilitate experiments on some of the cancer imaging modalities that we found to be less explored:

- Breast tomosynthesis: BCS-DBT (Buda et al., 2021)
- PET-CT: Lung-PET-CT-Dx (Li et al., 2020b)
- Endoscopy: HyperKvasir (Borgli et al., 2020)
- Dermatology: HAM10000 (Tschandl et al., 2018)
- Cytology: CERVIX93 (Phoulady and Mouton, 2018)
- Thoracic X-ray: Node21 (Sogancioglu et al., 2021)

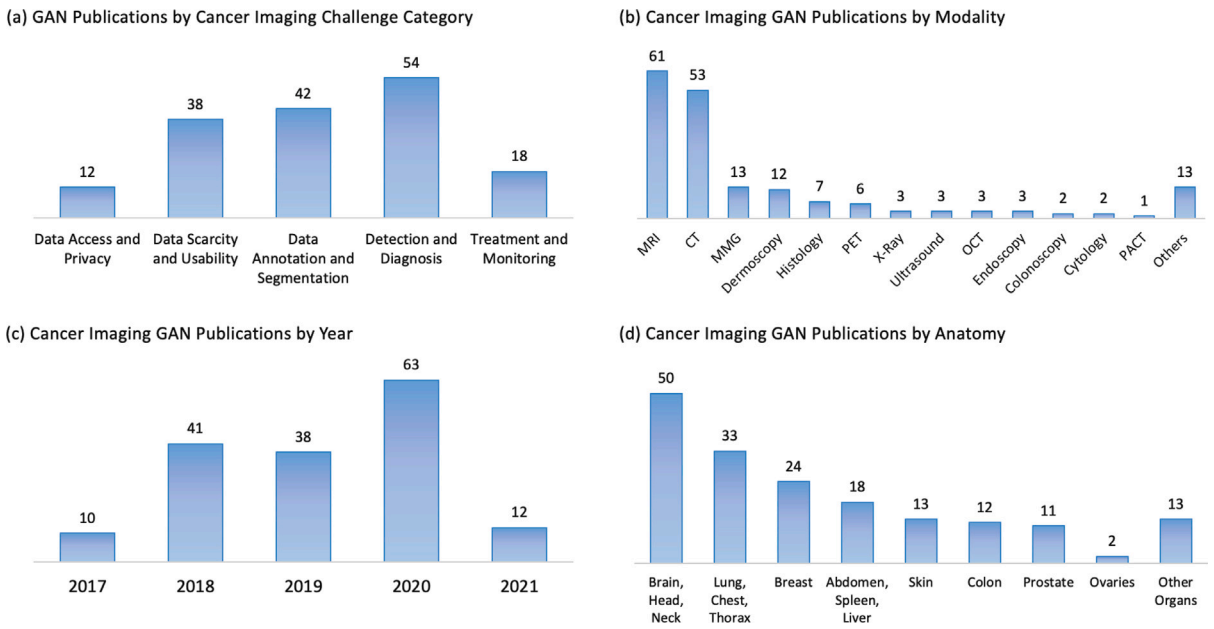

Fig. 15. Histograms showing the distribution of the 164 analysed GAN publications in this paper by (a) cancer imaging challenge category, (b) imaging modality, (c) year of publication, and (d) anatomy/organ. These numbers are retrieved exclusively from the information in Tables 2–6 of the respective Sections 4.1–4.5. Note that (b) and (d) contain more publications in total than (a) and (c), which is caused by GAN publications that evaluate on (and are assigned to) more than one modality (b) and/or anatomy (d) due to multiple experiments or cross-domain translation. In (c), the count for 2021 is not final, as the GAN papers herein analysed have been published on or before 7th March 2021.

6.3. Anatomy biases

In comparison, the GAN-based solutions per anatomy are more evenly spread, but still show a clear trend towards brain, head, neck (50), lung, chest, thorax (33) and breast (24). We suspect these spreads are due to the availability of few well-known widely-used curated benchmark datasets (Menze et al., 2014; Armato III et al., 2011; Heath et al., 2001; Moreira et al., 2012) resulting in underexposure of organs and modalities with less publicly available data resources. Where possible, we recommend evaluating GAN-based data synthesis and adversarial training on a range of different tasks and organs. This can avoid iterating towards non-transferable solutions tuned for specific datasets with limited generalisation capabilities. Said generalisation capabilities are critical for beneficial usage in clinical environments where dynamic data processing requirements and dataset shifts (e.g., multi-vendor, multi-scanner, multi-modal, multi-organ, multi-centre) commonly exist.

6.4. Cancer imaging challenge category biases

Fig. 15(a) displays the distribution of GAN publications across cancer imaging challenge categories that correspond to the subsections of Section 4. While the Sections 4.4 detection and diagnosis (54) and 4.4 data annotation and segmentation (42), and 4.1 data scarcity and usability (38) have received much research attention, Sections 4.5 treatment and monitoring (18) and 4.2 data access and privacy (12) contain substantially less GAN-related publications. This spread can be anticipated considering that classification and segmentation are popular computer vision problems and common objectives in publicly available medical imaging benchmark datasets. Early detected cancerous cells likely have had less time to acquire malignant genetic mutations (Hanahan and Weinberg, 2000, 2011) than their latter detected counterparts, which, by then, might have acquired more treatment-resistant alterations and subclone cell populations. Hence, automated early detection, location and diagnosis can provide high clinical impact via improved cancer treatment prospects, which likely influences the trend towards detection and segmentation-related GAN publications.

6.5. Well-validated adversarial network solutions

Our survey uncovers in Sections 4.1, 4.3, and 4.4 that a vast amount of cancer imaging literature exists around a few common adversarial network solutions.

The most common application of GANs is data augmentation, where synthetic data is added to the training dataset to yield an improved downstream task performance. Such data augmentation can be further used to balance imbalanced datasets, which, for instance, often include much more benign tumour images than malignant ones.

A further well-explored application of GANs is domain adaptation via adversarial training, where a domain-adversarial loss is backpropagated into a downstream task model. Domain mapping is a related application, where images are translated from one domain to another. In general, GANs learn to translate between one source and one target domain. However, promising work has extended this technique to cross-modal synthesis between multiple domains (Yurt et al., 2019; Li et al., 2019a; Zhou et al., 2020), which remains an area with much clinically-relevant research potential. Similarly, GANs for super resolution and data curation including artifact removal and image denoising achieve desirable performance and real-world applicability.

Image-to-image translating GANs can remove or hallucinate features such as tumours (Cohen et al., 2018b,a) into generated images. While this can be a major concern for clinical adoption, it also opens an avenue for future research into automated detection and assessment of removed or hallucinated features and sheds light on the need for additional metrics for GAN condition-adherence and synthetic data evaluation.

Furthermore, we observe that the discriminator and its associated adversarial loss can be flexibly used to classify any type of model output without necessarily following the purpose of data generation. For example, discriminator can predict whether a segmentation mask is real or created by a segmentation model, which enables the model to learn to output more globally coherent segmentation masks.

6.6. New solutions for unexploited areas

Patient privacy. We promote future work on the less researched open challenges in Section 4.2, where we describe the promising research

potential of adversarial networks in patient data privacy and security. We note that secure patient data is required for legal and ethical patient data sharing and usage, which, on the other hand, is required for successful training of state-of-the-art downstream task models. For instance, sharing GANs instead of private patient data can reduce data sharing constraints, while maintaining data utility (Szafranowska et al., 2022). Furthermore, GANs can be trained both in a federated learning setup as well as in a differential-privacy setup. Both of these techniques can be combined to further reduce privacy risks such as the risk of generating synthetic imaging data attributable to a specific patient. Further unexploited research potential lies in adversarial identity obfuscation both on image level, as well as on latent feature representation level. In particular, devising privacy preservation testing methods to evaluate the success of adversarial identity obfuscation and related methods is a needed and not fully addressed research problem in cancer imaging and AI in healthcare at large.

Patient security. With the projected increase in clinical AI applications, adversarial learning based cybersecurity methodology becomes increasingly important to protect patients against the vulnerabilities inherent in clinically deployed deep learning solutions. Attacks can alter diagnostic markers on cancer imaging data, which can potentially result in diagnostic errors with dangerous consequences for the targeted patients. For instance, defences against adversarial examples (Liu et al., 2020b; Samangouei et al., 2018) or detection of imaging data that has been tampered with Mirsky et al. (2019) are areas where solutions based on adversarial methods will increasingly gain practical importance.

Model debiasing. The versatile ability of adversarial training to curate a model's latent space is likely to continue to increase in popularity due to the need to remove certain features in clinical AI models. For example, it is desirable to minimise a model's learned biases to increase the fairness of clinical models across patient populations (Lekadir et al., 2021). Such bias removal has been shown to be achievable via adversarial loss backpropagation (Zhang et al., 2018a; Li et al., 2021a). As Elazar and Goldberg (2018) point out, some residual biases may remain in a model's latent space after converged adversarial bias removal training. Therefore, research potential lies in automated test and evaluation methodology to assess the quantity of residual bias remaining in an adversarial networks after debiasing, particularly if applied to data unseen during training.

Generative model evaluation. A key aspect this survey observes is the absence of interpretable, standardised and exact evaluation methodology for synthetic data and generative models in the medical and cancer imaging domains. This is particularly noticeable for models without a narrow downstream task performance objective that can be used as surrogate evaluation metric nor a reconstruction objective that informs the evaluation technique. Generative models that generate a synthetic image with a clear reference value (i.e., a real image) can be evaluated based on the difference between reference and generated sample, e.g., via perceptual and reconstruction losses and metrics such as SSIM, PSNR, MSE, as discussed in Section 4.1. In the absence of such reference images, remaining methods at hand are image inspection techniques and real versus synthetic distribution comparisons, the latter including the Fréchet Inception Distance (FID) score (Heusel et al., 2017). The popularity of the FID metric for fidelity and diversity evaluation of synthetic data has largely translated from computer vision into medical imaging. The applicability of FID in the medical domain, nonetheless, is questionable, as it internally relies on an inception classifier pretrained on the ImageNet dataset consisting of 3-channel natural images as opposed to, for instance, grayscale images from radiological domains. This demonstrates a clear need for research on further evaluation methodologies of synthetic medical images. FID extensions that pretrain the internal classifier on medical imaging datasets are potential directions, but limited by the acquisition techniques, scope, modalities, and, importantly, the size of these medical imaging datasets. Recent promising work proposed the automated generation of segmentation mask from GANs based on latent space exploration (Melas-Kyriazi et al., 2021). Such latent space inspection approaches can offer further potential for generative model evaluation, e.g., by helping to measure the number and difference between modes or by providing quality and diversity estimates of the segmentation masks (or other extractable pieces of information) that the model produces.

Patient treatment. Sections 4.3 and 4.4 have shown that adversarial models for cancer detection, classification, and localisation are, at least for particular organs and modalities, well explored research areas. These applications are mostly relevant in diagnostic activities, which comprise only one part of the clinical workflow. We encourage more research on GAN-based solutions in less explored subsequent clinical workflow steps such as oncological treatment planning and disease monitoring as elaborated in Section 4.5. For example, adversarial learning offers research potential in tumour profiling and intra- and inter-tumour heterogeneity assessment via anomaly detection within the latent space of adversarial models (Schlegl et al., 2019; Quiros et al., 2019). The high intra- and inter-tumour heterogeneity increases the difficulty of assessing and selecting targeted treatment options. Research potential exists in precisely encoding a tumour based on imaging and/or non-imaging patient and tumour data in an adversarial model's multi-dimensional latent space. For example, this can unlock vector search applications to find similarly encoded tumours in databases to inform on therapy selection, success probabilities, and progression patterns. Tumour progression modelling on image-level based on generative models such as GANs remains largely unexplored. Even though not strictly necessary (Xia et al., 2021), longitudinal and time-series cancer imaging datasets will likely trigger increased exploration of this research area once such data becomes available. For instance, given a tumour image at timepoint *t1*, a GAN can learn to simulate the tumour image at timepoint *t2*. To this end, generation of image-level counterfactuals (Pawlowski et al., 2020) as a clinically impactful solution for probing interventions. For instance, GANs can generate a tumour at *t2* given the tumour image at *t1* alongside multiple input conditions such as tumour growth rate, tumour type, and applied treatments.

6.7. Future perspectives and technology trends

6.7.1. Towards state-of-the-art GAN innovations in cancer imaging

In recent years, multiple novel adversarial networks have been introduced in the field of computer vision. A lesson learned from our survey is that many of these techniques are yet to be applied thoroughly to cancer imaging. These innovations open avenues in cancer imaging that extend upon the currently used methods shown in Fig. 4, for instance, enabling improved high-resolution image generation and input-conditioned image synthesis.

Overcoming dataset and computation limitations. For instance, the recent VQGAN (Esser et al., 2021) combines the efficiency of convolutional networks with the expressiveness of transformers, which model the composition of a reusable codebook of context-rich visual parts. This approach is particularly relevant to medical and cancer imaging, as it allows high-resolution image synthesis despite limited computing resources. Apart from containing high resolution images, cancer imaging datasets are often limited in the number of image, which may not suffice to train a GAN. In these cases, the potential issues are that during training there is convergence-failure, where the synthetic image quality is low and does not improve any further during training. While the adversarial loss often is non-interpretable not corresponding to synthetic image fidelity, diversity or condition adherence, also mode collapse may occur, where the generator has learned a particular mode to fool the discriminator instead of generating a high diversity of samples. These issues are not only a function of the GAN architecture and loss

function, but also of the size of the training dataset. FastGAN (Liu et al., 2020a) and SinGAN (Shaham et al., 2019) shows great promise to overcome this data scarcity cancer imaging problem. FastGAN (Liu et al., 2020a) uses self-supervised training of discriminator as encoder for regularisation and generates high-resolution images despite limited computing resources and dataset size. SinGAN (Shaham et al., 2019) generates multiple synthetic images based on only a single training image. This has wide applicability and can substantially increase the usefulness of even very small cancer imaging datasets via SinGAN-based data augmentation. A first successful applications of SinGAN and FastGAN to cancer imaging for polyp segmentation by Thambawita et al. (2022) shows the potential of using these models to generate not only a synthetic images, but also a corresponding segmentation mask by outputting an additional channel. This type of methodology enables training data generation for tumour detection, localisation and segmentation models without the need of conditioning the GAN on input segmentation masks.

Best practice combining GAN frameworks. As a vast amount of novel additions to the GANs framework has been suggested, some work (Brock et al., 2018) has focused on collecting the best working practices and combining them into novel architectures, which are promising and not yet widely applied to challenges in cancer imaging. For example, BigGAN (Brock et al., 2018) (a) scales model parameters by increasing the size of the feature maps, (b) applies large batch sizes, (c) uses self-attention based on SAGAN (Zhang et al., 2019a), (d) provides information about the class via class-conditional batch normalisation, and (e) uses hinge-loss. BigGAN and extensions thereof (e.g., Zhang et al. (2019b), Casanova et al. (2021) and Schonfeld et al. (2020)) achieve state-of-the-art performance on class-conditional image generation.

Extending on PGGAN (Karras et al., 2017) as shown in Fig. 4(m), another such example is StyleGAN (Karras et al., 2019) and its variants (Karras et al., 2021, 2020; Sauer et al., 2022), which accomplish state-of-the-art performance in conditional and unconditional computer vision image generation benchmarks. Yielding strong results, multiple architectural innovations have been introduced by the StyleGAN family, such as a style vector generating fully connected mapping network, adaptive instance normalisation, and, instead of sampling from a noise vector, moving the noise input to intermediate activation maps. These innovations can inform cancer image generation models and improve their latent space exploration capabilities e.g. allowing to compare different tumour types and manifestations.

Image-to-image translation. Image-to-image translation problems in cancer imaging are widely approached using commonly pix2pix (Isola et al., 2017) (paired) and cycleGAN (Zhu et al., 2017) (unpaired). Nonetheless, more recent models such as OASIS (Sushko et al., 2020), ResVit (Dalmaz et al., 2021), and StarGAN V2 (Choi et al., 2020) have been proposed, which are not only applicable to cancer imagery, but also have shown superior performance on computer vision benchmarks. ResVit (Dalmaz et al., 2021), for instance, diverges away from common CNN architectures with inductive biases by using a vision transformer architecture (Dosovitskiy et al., 2020) alongside an adversarial loss (Goodfellow et al., 2014) and the common L1 losses between source and target (Isola et al., 2017) and between source and reconstructed source (Zhu et al., 2017). StarGAN V2 (Choi et al., 2020) employs besides the adversarial and cycle consistency losses also a style reconstruction loss and a style diversification loss, while OASIS (Sushko et al., 2020) shows that a perceptual loss is not necessary given an adversarial loss and a segmentation-based discriminator.

6.7.2. GAN alternatives and complementary methods

Diffusion models. In image inpainting (Saharia et al., 2021a) and super resolution (Saharia et al., 2021b), the recently proposed and increasingly popular diffusion models (Sohl-Dickstein et al., 2015; Song and Ermon, 2019; Ho et al., 2020) have been shown to achieve state-of-the-art and competitive performances for computer vision benchmarks and, thus, are an alternative to GANs. Diffusion models iteratively add noise to an image in a Markov chain of diffusion steps. Reversing this process, a noise vector z is gradually denoised and transformed into an image. While achieving promising generative modelling capabilities, it still takes longer to sample from diffusion models than from GANs due to multiple denoising steps, while also further work is needed to explore the interpretability of latent representations of diffusion models (Dhariwal and Nichol, 2021). A promising line of research suggests the combination of GANs with diffusion models to increase the stability and data efficiency of GAN training (Wang et al., 2022).

Variational autoencoders. GANs are commonly considered to achieve higher quality outputs than variational autoencoders (VAEs) (Kingma and Welling, 2013) at the cost of a training process more prone towards requiring manual intervention and tuning. A promising line of research improves upon vanilla VAE by exploring combinations of GANs and VAEs (Larsen et al., 2016; Makhzani et al., 2015). Extending on VAEs, Van Den Oord et al. (2017) proposed Vector Quantised Variational AutoEncoder (VQ-VAE), which learns discrete instead of continuous latent representations to avoid the issue of 'posterior collapse' that is common in VAEs. VQ-VAE has been shown to be an effective method for diverse high-quality synthetic image generation (Razavi et al., 2019). A promising extension combines VQ-VAE with transformers (Vaswani et al., 2017) for unsupervised anomaly detection and segmentation and demonstrates its potential for tumour segmentation in brain MRI (Pinaya et al., 2021).

Normalizing flows. The recently proposed Normalizing Flows (Rezende and Mohamed, 2015; Dinh et al., 2014, 2016) are an alternative deep generative model gaining increasing popularity for synthetic data generation tasks. As opposed to GANs and VAEs (implicit), Normalizing Flows explicitly learn the probability density function $p(x)$ and are trained via maximum likelihood estimation. Knowing $p(x)$, unobserved but realistic new data points can be sampled with exact likelihood estimates. Normalizing Flows have been shown to be combinable with GANs and the adversarial loss function, e.g., by being the building block of the generator network (Grover et al., 2018), and for image-to-image translation (Grover et al., 2020). To date, Normalizing Flows have seen less adoption in medical and cancer imaging than GANs, but promising initial applications exist. For example, Normalizing Flows have been proposed for uncertainty estimation of lung lesion segmentation (Selvan et al., 2020), counterfactual inference on brain MRI (Pawlowski et al., 2020), and low-dose CT image reconstruction (Denker et al., 2020).

Unsupervised domain adaptation. In unsupervised domain adaptation, self-training approaches are described as an alternative to domain adversarial losses. For example, state-of-the-art methods like HRDA (Hoyer et al., 2022b) and DaFormer (Hoyer et al., 2022a) show the effectiveness of self-training in domain-adaptive semantic segmentation. DaFormer uses a transformer encoder (Vaswani et al., 2017; Dosovitskiy et al., 2020) and transfers knowledge from source to target domain via a teacher network that generates pseudo-labels for the data from the target domain. A promising avenue of research combines self-training approaches and adversarial losses (Li et al., 2019c; Kim and Byun, 2020; Wang et al., 2020a).

Self-supervised learning. Given successes in learning useful representations from unlabelled data, self-supervised learning (SSL) approaches, such as BYOL (Grill et al., 2020), have become a common technique in the toolkit of deep learning researchers. Particularly when working with datasets limited in size or annotations, additional GAN-generated data can improve the learning of representations, upon which a downstream task model produces its predictions. SSL can provide an alternative, often computationally less expensive, means towards representation learning given a training task with objective function, where labels y and inputs x are extracted from an unlabelled dataset. A popular and powerful SSL method is contrastive learning, where a

model's latent space is learned by minimising the distance of similar samples and maximising the distance between dissimilar ones. Effective model pretraining methods such as SimCLR (Chen et al., 2020b) rely on such contrastive loss functions, which, e.g., maximise agreement between differently augmented views of the same image. Multiple recent studies propose the combination of GANs and self-supervised (Patel et al., 2021) and contrastive learning with promising results reporting improved performance and sample diversity, as well as reduced discriminator overfitting (Jeong and Shin, 2021; Kang and Park, 2020; Liu et al., 2021). In cancer imaging, for instance, this combination has been applied to address the problem of mode collapse while retaining phenotypic tumour features for the task of colour normalisation in histopathology images (Ke et al., 2021).

7. Conclusion

In closing, we emphasise the versatility and the resulting modality-independent wide applicability of the adversarial learning scheme of GANs. In this survey, we strive to consider and communicate this versatility by describing the wide variety of problems in the cancer imaging domain that can be approached with adversarial networks. For example, we highlight GAN and adversarial training solutions that range from unsupervised domain adaptation to patient privacy preserving distributed data synthesis, to adversarial segmentation mask discrimination, to multi-modal radiation dose estimation, amongst others.

Before reviewing and describing GAN and adversarial training solutions, we surveyed the literature to understand the current challenges in the field of cancer imaging with a focus on radiology, but without excluding non-radiology modalities common to cancer imaging. After screening and analysing the cancer imaging challenges, we grouped them into the challenge categories Data Scarcity and Usability, Data Access and Privacy, Data Annotation and Segmentation, Detection and Diagnosis, and Treatment and Monitoring. After categorisation, we surveyed the literature for adversarial networks applied to the field of cancer imaging and found 164 relevant publications, each of which we assigned to its respective cancer imaging challenge category. Finally, we provide a comprehensive analysis for each challenge and its assigned GAN-related publications to determine to what extent it has and can be solved using GANs and adversarial training. We further establish the *SynTRUST* framework for assessing the trustworthiness of medical image synthesis studies. Based on *SynTRUST*, we analyse 16 carefully selected cancer imaging challenge solutions. Notwithstanding the overall high level of rigour and validity of these studies, we are able to recommend a set of unaddressed trustworthiness improvements in order to guide future studies. To this end, we also highlight research potential for challenges where we were able to propose data synthesis or adversarial training solutions that have not yet been fully explored by the literature.

With our work, we strive to uncover and motivate promising lines of research in data synthesis and adversarial networks that we envision to ultimately benefit the field of cancer imaging in clinical practice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952103.

References

AAPM, 2016. 2016 NIH-AAPM-Mayo clinic low dose CT grand challenge. URL: https://www.aapm.org/grandchallenge/lowdosect/.

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al., 2020. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. arXiv preprint arXiv:2011.06225.

Abhishek, K., Hamarneh, G., 2019. Mask2lesion: Mask-constrained adversarial skin lesion image synthesis. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 71–80.

Abramian, D., Eklund, A., 2019. Refacing: reconstructing anonymized facial features using GANs. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 1104–1108.

Adamson, A.S., Smith, A., 2018. Machine learning and health care disparities in dermatology. JAMA Dermatol. 154 (11), 1247–1248.

Addepalli, S., Nayak, G.K., Chakraborty, A., Radhakrishnan, V.B., 2020. DeGAN: Data-Enriching GAN for retrieving representative samples from a trained classifier. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. pp. 3130–3137.

Ahmed, K.T., Sun, J., Yong, J., Zhang, W., 2021. Multi-omics data integration by generative adversarial network. BioRxiv http://dx.doi.org/10.1101/2021.03.13.435251.

Almond, D., Chay, K.Y., Lee, D.S., 2004. The Costs of Low Birth Weight. Working Paper 10552, National Bureau of Economic Research, http://dx.doi.org/10.3386/w10552.

Alshehhi, R., Alshehhi, A., 2021. Quantification of uncertainty in brain tumor segmentation using generative network and Bayesian active learning. In: VISIGRAPP (4: VISAPP). pp. 701–709.

Alyafi, B., Diaz, O., Martí, R., 2020. DCGANs for realistic breast mass augmentation in x-ray mammography. In: Medical Imaging 2020: Computer-Aided Diagnosis, Vol. 11314. International Society for Optics and Photonics, 1131420.

Angermann, Q., Bernal, J., Sánchez-Montes, C., Hammami, M., Fernández-Esparrach, G., Dray, X., Romain, O., Sánchez, F.J., Histace, A., 2017. Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In: Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures. Springer, pp. 29–41.

Argenziano, G., Soyer, H., De Giorgi, V., Piccolo, D., Carli, P., Delfino, M., et al., 2002. Dermoscopy: A Tutorial, Vol. 16. EDRA, Medical Publishing & New Media.

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 214–223.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684.

Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al., 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. 38 (2), 915–931.

Arnold, M., Rutherford, M.J., Bardot, A., Ferlay, J., Andersson, T.M., Myklebust, T.Å., Tervonen, H., Thursfield, V., Ransom, D., Shack, L., et al., 2019. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. Lancet Oncol. 20 (11), 1493–1505.

Arora, S., Risteski, A., Zhang, Y., 2018. Do GANs learn the distribution? some theory and empirics. In: International Conference on Learning Representations.

Babier, A., Boutilier, J.J., Sharpe, M.B., McNiven, A.L., Chan, T.C., 2018. Inverse optimization of objective function weights for treatment planning using clinical dose-volume histograms. Phys. Med. Biol. 63 (10), 105004.

Bae, H., Jung, D., Yoon, S., 2020. AnomiGAN: Generative adversarial networks for anonymizing private medical data. In: Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, Vol. 25. World Scientific, pp. 563–574.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data 4 (1), 1–13.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al., 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629.

Balagurunathan, Y., Kumar, V., Gu, Y., Kim, J., Wang, H., Liu, Y., Goldgof, D.B., Hall, L.O., Korn, R., Zhao, B., et al., 2014. Test–retest reproducibility analysis of lung CT image features. J. Digit. Imaging 27 (6), 805–823.

Ballerini, L., Fisher, R.B., Aldridge, B., Rees, J., 2013. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In: Color Medical Image Analysis. Springer, pp. 63–86.

Barbaro, B., Leccisotti, L., Vecchio, F.M., Di Matteo, M., Serra, T., Salsano, M., Poscia, A., Coco, C., Persiani, R., Alfieri, S., et al., 2017. The potential predictive value of MRI and PET-CT in mucinous and nonmucinous rectal cancer to identify patients at high risk of metastatic disease. Br. J. Radiol. 90 (1069), 20150836.

Baur, C., Albarqouni, S., Navab, N., 2018. MelanoGANs: high resolution skin lesion synthesis with GANs. arXiv preprint arXiv:1804.04338.

Beaulieu-Jones, B.K., Wu, Z.S., Williams, C., Lee, R., Bhavnani, S.P., Byrd, J.B., Greene, C.S., 2019. Privacy-preserving generative deep neural networks support clinical data sharing. Circ.: Cardiovasc. Qual. Outcomes 12 (7), e005122.

Beck, A.H., Sangoi, A.R., Leung, S., Marinelli, R.J., Nielsen, T.O., van de Vijver, M.J., West, R.B., van de Rijn, M., Koller, D., 2011. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci. Transl. Med. 3 (108), 108ra113.

Becker, A.S., Jendele, L., Skopek, O., Berger, N., Ghafoor, S., Marcon, M., Konukoglu, E., 2019. Injecting and removing suspicious in breast imaging with CycleGAN: A pilot study of automated adversarial attacks using neural networks on small images. Eur. J. Radiol. 120, 108649.

Ben-Cohen, A., Klang, E., Raskin, S.P., Amitai, M.M., Greenspan, H., 2017. Virtual PET images from CT data using deep convolutional networks: initial results. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 49–57.

Benhammou, Y., Achchab, B., Herrera, F., Tabik, S., 2020. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. Neurocomputing 375, 9–24.

Benson, S., Beets-Tan, R., 2020. GAN-based anomaly detection in multi-modal MRI images. BioRxiv http://dx.doi.org/10.1101/2020.07.10.197087.

Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F., 2015. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. 43, 99–111.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11), 2514–2525.

Beutel, A., Chen, J., Zhao, Z., Chi, E.H., 2017. Data decisions and theoretical implications when adversarially learning fair representations. arXiv preprint arXiv:1707.00075.

Bi, W.L., Hosny, A., Schabath, M.B., Giger, M.L., Birkbak, N.J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I.F., et al., 2019. Artificial intelligence in cancer imaging: clinical challenges and applications. CA: Cancer J. Clin. 69 (2), 127–157.

Bi, L., Kim, J., Kumar, A., Feng, D., Fulham, M., 2017. Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In: Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Springer, pp. 43–51.

Bica, I., Jordon, J., van der Schaar, M., 2020. Estimating the effects of continuous-valued interventions using generative adversarial networks. arXiv preprint arXiv:2002.12326.

Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.-W., Han, X., Heng, P.-A., Hesser, J., et al., 2019. The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056.

Billot, B., Robinson, E., Dalca, A.V., Iglesias, J.E., 2020. Partial volume segmentation of brain MRI scans of any resolution and contrast. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 177–187.

Bischoff-Grethe, A., Ozyurt, I.B., Busa, E., Quinn, B.T., Fennema-Notestine, C., Clark, C.P., Morris, S., Bondi, M.W., Jernigan, T.L., Dale, A.M., et al., 2007. A technique for the deidentification of structural brain MR images. Hum. Brain Mapp. 28 (9), 892–903.

Bissoto, A., Perez, F., Valle, E., Avila, S., 2018. Skin lesion synthesis with generative adversarial networks. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Springer, pp. 294–302.

Blake, C., 1998. UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html.

Borgli, H., Thambawita, V., Smedsrud, P.H., Hicks, S., Jha, D., Eskeland, S.L., Randel, K.R., Pogorelov, K., Lux, M., Nguyen, D.T.D., et al., 2020. Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Sci. Data 7 (1), 1–14.

Borji, A., 2019. Pros and cons of GAN evaluation measures. Comput. Vis. Image Underst. 179, 41–65.

Borji, A., 2021. Pros and cons of GAN evaluation measures: New developments. arXiv preprint arXiv:2103.09396.

Brady, A.P., 2017. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging 8 (1), 171–182.

van den Brandt, P.A., Goldbohm, R.A., Veer, P.V., Volovics, A., Hermus, R.J., Sturmans, F., 1990. A large-scale prospective cohort study on diet and cancer in The Netherlands. J. Clin. Epidemiol. 43 (3), 285–295.

Brennan, P., Silman, A., 1992. Statistical methods for assessing observer variability in clinical measures. BMJ: Br. Med. J. 304 (6840), 1491.

Brock, A., Donahue, J., Simonyan, K., 2018. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096.

Bromley, J., Guyon, I., LeCun, Y., 1994. Signature verification using a "siamese" time delay neural network.

Bu, T., Yang, Z., Jiang, S., Zhang, G., Zhang, H., Wei, L., 2020. 3D conditional generative adversarial network-based synthetic medical image augmentation for lung nodule detection. Int. J. Imaging Syst. Technol..

Buda, M., Saha, A., Walsh, R., Ghate, S., Li, N., Swiecicki, A., Lo, J.Y., Mazurowski, M.A., 2021. A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images. JAMA Netw. Open 4 (8), e2119100.

Caballo, M., Pangallo, D.R., Mann, R.M., Sechopoulos, I., 2020. Deep learning-based segmentation of breast masses in dedicated breast CT imaging: radiomic feature stability between radiologists and artificial intelligence. Comput. Biol. Med. 118, 103629.

Casanova, A., Careil, M., Verbeek, J., Drozdzal, M., Romero Soriano, A., 2021. Instance-conditioned gan. Adv. Neural Inf. Process. Syst. 34, 27517–27529.

Castro, D.C., Walker, I., Glocker, B., 2020. Causality matters in medical imaging. Nature Commun. 11 (1), 1–10.

Cem Birbiri, U., Hamidinekoo, A., Grall, A., Malcolm, P., Zwiggelaar, R., 2020. Investigating the performance of generative adversarial networks for prostate tissue detection and segmentation. J. Imaging 6 (9), 83.

Chaitanya, K., Karani, N., Baumgartner, C.F., Erdil, E., Becker, A., Donati, O., Konukoglu, E., 2021. Semi-supervised task-driven data augmentation for medical image segmentation. Med. Image Anal. 68, 101934.

Chan, E.R., Monteiro, M., Kellnhofer, P., Wu, J., Wetzstein, G., 2020. Pi-GAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis. arXiv preprint arXiv:2012.00926.

Chang, Q., Qu, H., Zhang, Y., Sabuncu, M., Chen, C., Zhang, T., Metaxas, D.N., 2020a. Synthetic learning: Learn from distributed asynchronized discriminator GAN without sharing medical image data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13856–13866.

Chang, Q., Yan, Z., Baskaran, L., Qu, H., Zhang, Y., Zhang, T., Zhang, S., Metaxas, D.N., 2020b. Multi-modal AsynDGAN: Learn from distributed medical image data without sharing private information. arXiv preprint arXiv:2012.08604.

Chaudhari, P., Agrawal, H., Kotecha, K., 2019. Data augmentation using MG-GAN for improved cancer classification on gene expression data. Soft Comput. 1–11.

Chen, S., Jia, R., Qi, G.-J., 2020a. Improved techniques for model inversion attack. arXiv preprint arXiv:2010.04092.

Chen, J., Konrad, J., Ishwar, P., 2018a. Vgan-based image representation learning for privacy-preserving facial expression recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1570–1579.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020b. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Chen, X., Li, Y., Yao, L., Adeli, E., Zhang, Y., 2021. Generative adversarial U-net for domain-free medical image augmentation. arXiv preprint arXiv:2101.04793.

Chen, X., Pawlowski, N., Rajchl, M., Glocker, B., Konukoglu, E., 2018b. Deep generative models in the real-world: An open challenge from medical imaging. arXiv preprint: arXiv:1806.05452.

Chen, L., Song, H., Wang, C., Cui, Y., Yang, J., Hu, X., Zhang, L., 2019. Liver tumor segmentation in CT volumes using an adversarial densely connected network. BMC Bioinformatics 20 (16), 1–13.

Chen, L., Yang, Q., Bao, J., Liu, D., Huang, X., Wang, J., 2017. Direct comparison of PET/CT and MRI to predict the pathological response to neoadjuvant chemotherapy in breast cancer: a meta-analysis. Sci. Rep. 7 (1), 1–10.

Chi, Y., Bi, L., Kim, J., Feng, D., Kumar, A., 2018. Controlled synthesis of dermoscopic images via a new color labeled generative style transfer network to enhance melanoma segmentation. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, pp. 2591–2594.

Choi, Y., Uh, Y., Yoo, J., Ha, J.-W., 2020. Stargan v2: Diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8188–8197.

Chong, M.J., Forsyth, D., 2020. Effectively unbiased FID and inception score and where to find them. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6070–6079.

Choyke, P., Turkbey, B., Pinto, P., Merino, M., Wood, B., 2016. Data from PROSTATE-MRI. Cancer Imaging Arch. 9, Available Online: http://doi.org/10.7937/K9/TCIA.2016.6046GUDv.

Chuquicusma, M.J., Hussein, S., Burt, J., Bagci, U., 2018. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 240–244.

CireAan, D., Meier, U., Masci, J., Schmidhuber, J., 2012. Multi-column deep neural network for traffic sign classification. Neural Netw. 32, 333–338.

Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2013. Mitosis detection in breast cancer histology images with deep neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 411–418.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al., 2015. Comprehensive molecular portraits of invasive lobular breast cancer. Cell 163 (2), 506–519.

Cirillo, M.D., Abramian, D., Eklund, A., 2020. Vox2Vox: 3D-GAN for brain tumour segmentation. arXiv preprint arXiv:2003.13653.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging 26 (6), 1045–1057.

Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Pike, G.B., Evans, A.C., 1997. Brainweb: Online interface to a 3D MRI simulated brain database. In: NeuroImage. Citeseer.

Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al., 2018. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 168–172.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al., 2019. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1902.03368.

Cohen, J.P., Luck, M., Honari, S., 2018a. Distribution matching losses can hallucinate features in medical image translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 529–536.

Cohen, J.P., Luck, M., Honari, S., 2018b. How to cure cancer (in images) with unpaired image translation. In: International Conference on Medical Imaging with Deep Learning (MIDL 2018)–Abstract Track.

Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., et al., 2019. BCN20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288.

Creswell, A., Pouplin, A., Bharath, A.A., 2018. Denoising adversarial autoencoders: classifying skin lesions using limited labelled training data. IET Comput. Vis. 12 (8), 1105–1111.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., D'Eustachio, P., 2014. The reactome pathway knowledgebase. Nucl. Acids Res. 42 (Database issue), D472–477.

Cuocolo, R., Caruso, M., Perillo, T., Ugga, L., Petretta, M., 2020. Machine learning in oncology: A clinical appraisal. Cancer Lett. 481, 55–62.

Dalmaz, O., Yurt, M., Çukur, T., 2021. Resvit: Residual vision transformers for multi-modal medical image synthesis. arXiv preprint arXiv:2106.16031.

Dashtban, M., Li, W., 2020. Predicting risk of hospital readmission for comorbidity patients through a novel deep learning framework. In: Proceedings of the 53rd Hawaii International Conference on System Sciences.

Denker, A., Schmidt, M., Leuschner, J., Maass, P., Behrmann, J., 2020. Conditional normalizing flows for low-dose computed tomography image reconstruction. arXiv preprint arXiv:2006.06270.

Desai, S.D., Giraddi, S., Verma, N., Gupta, P., Ramya, S., 2020. Breast cancer detection using GAN for limited labeled dataset. In: 2020 12th International Conference on Computational Intelligence and Communication Networks. CICN, IEEE, pp. 34–39.

DeVries, T., Romero, A., Pineda, L., Taylor, G.W., Drozdzal, M., 2019. On the evaluation of conditional GANs. arXiv preprint arXiv:1907.08175.

Dhariwal, P., Nichol, A., 2021. Diffusion models beat GANs on image synthesis. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems, Vol. 34. Curran Associates, Inc., pp. 8780–8794, URL: https://proceedings.neurips.cc/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf.

Diaz, O., Kushibar, K., Osuala, R., Linardos, A., Garrucho, L., Igual, L., Radeva, P., Prior, F., Gkontra, P., Lekadir, K., 2021. Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools. Phys. Medica 83, 25–37. http://dx.doi.org/10.1016/j.ejmp.2021.02.007.

Dimitriou, N., Arandjelović, O., Harrison, D.J., Caie, P.D., 2018. A principled machine learning framework improves accuracy of stage II colorectal cancer prognosis. NPJ Digit. Med. 1 (1), 1–9.

Dinh, L., Krueger, D., Bengio, Y., 2014. Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516.

Dinh, L., Sohl-Dickstein, J., Bengio, S., 2016. Density estimation using real nvp. arXiv preprint arXiv:1605.08803.

Doman, K., Konishi, T., Mekada, Y., 2020. Lesion image synthesis using DCGANs for metastatic liver cancer detection. Deep Learn. Med. Image Anal. 95–106.

Dong, X., Lei, Y., Wang, T., Thomas, M., Tang, L., Curran, W.J., Liu, T., Yang, X., 2019. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med. Phys. 46 (5), 2157–2168.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Drew, T., Võ, M.L.-H., Wolfe, J.M., 2013. The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. Psychol. Sci. 24 (9), 1848–1853.

Durán, J.M., Jongsma, K.R., 2021. Who is afraid of black box algorithms? on the epistemological and ethical basis of trust in medical AI. J. Med. Ethics 47 (5), 329–335.

Dwork, C., 2006. Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (Eds.), Automata, Languages and Programming. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12.

Edgar, R., Domrachev, M., Lash, A.E., 2002. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30 (1), 207–210.

Edupuganti, V., Mardani, M., Cheng, J., Vasanawala, S., Pauly, J., 2019. Uncertainty analysis of VAE-GANs for compressive medical imaging. arXiv preprint arXiv:1901.11228.

Elazab, A., Wang, C., Gardezi, S.J.S., Bai, H., Hu, Q., Wang, T., Chang, C., Lei, B., 2020. GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR images. Neural Netw. 132, 321–332. http://dx.doi.org/10.1016/j.neunet.2020.09.004.

Elazar, Y., Goldberg, Y., 2018. Adversarial removal of demographic attributes from text data. arXiv preprint arXiv:1808.06640.

Elmore, J.G., Wells, C.K., Lee, C.H., Howard, D.H., Feinstein, A.R., 1994. Variability in radiologists' interpretations of mammograms. N. Engl. J. Med. 331 (22), 1493–1499.

Esser, P., Rombach, R., Ommer, B., 2021. Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12873–12883.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J., 2019. A guide to deep learning in healthcare. Nat. Med. 25 (1), 24–29.

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., Hermjakob, H., 2017. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 18 (1), 142.

Farnia, F., Ozdaglar, A., 2020. GANs may have no nash equilibria. arXiv preprint arXiv:2002.09124.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136 (5), E359–E386.

Fernandes, K., Cardoso, J.S., Fernandes, J., 2017. Transfer learning with partial observability applied to cervical cancer screening. In: Iberian Conference on Pattern Recognition and Image Analysis. Springer, pp. 243–250.

Fischer, B.M., Olsen, M.W., Ley, C.D., Klausen, T.L., Mortensen, J., Højgaard, L., Kristjansen, P.E., 2006. How few cancer cells can be detected by positron emission tomography? a frequent question addressed by an in vitro study. Eur. J. Nucl. Med. Mol. Imaging 33 (6), 697–702.

Fitzpatrick, J.M., 1998. RIRE - retrospective image registration evaluation. URL: https://www.insight-journal.org/rire/.

Foroozandeh, M., Eklund, A., 2020. Synthesizing brain tumor images and annotations by combining progressive growing GAN and SPADE. arXiv preprint arXiv:2009.05946.

Fossen-Romsaas, S., Storm-Johannessen, A., Lundervold, A.S., 2020. Synthesizing skin lesion images using CycleGANs–a case study.

Frangioni, J.V., 2008. New technologies for human cancer imaging. J. Clin. Oncol. 26 (24), 4012.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321, 321–331.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybernet. 36 (4), 193–202. http://dx.doi.org/10.1007/BF00344251.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: International Conference on Machine Learning. PMLR, pp. 1050–1059.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning. PMLR, pp. 1180–1189.

Gao, C., Clark, S., Furst, J., Raicu, D., 2019. Augmenting LIDC dataset using 3D generative adversarial networks to improve lung nodule detection. In: Medical Imaging 2019: Computer-Aided Diagnosis, Vol. 10950. International Society for Optics and Photonics, 109501K.

GDPR, 2016. European Parliament and Council of European Union (2016) regulation (EU) 2016/679. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679. [Online; accessed 26-November-2020].

Ge, Q., Huang, X., Fang, S., Guo, S., Liu, Y., Lin, W., Xiong, M., 2020. Conditional generative adversarial networks for individualized treatment effect estimation and treatment selection. Front. Genet. 11, 1578. http://dx.doi.org/10.3389/fgene.2020.585804.

Ghorbani, A., Natarajan, V., Coz, D., Liu, Y., 2020. DermGAN: synthetic generation of clinical skin images with pathology. In: Machine Learning for Health Workshop. PMLR, pp. 155–170.

Ghosal, S.S., Sarkar, I., El Hallaoui, I., 2020. Lung nodule classification using convolutional autoencoder and clustering augmented learning method (CALM). In: HSDM@WSDM. pp. 19–26.

Giacomello, E., Loiacono, D., Mainardi, L., 2020. Brain MRI tumor segmentation with adversarial networks. In: 2020 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Gilles, F.H., Tavaré, C.J., Laurence, E.B., Burger, P.C., Yates, A.J., Pollack, I.F., Finlay, J.L., 2008. Pathologist interobserver variability of histologic features in childhood brain tumors: results from the CCG-945 study. Pediatr. Dev. Pathol. 11 (2), 108–117.

Gohagan, J.K., Marcus, P.M., Fagerstrom, R.M., Pinsky, P.F., Kramer, B.S., Prorok, P.C., Ascher, S., Bailey, W., Brewer, B., Church, T., et al., 2005. Final results of the lung screening study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. Lung Cancer 47 (1), 9–15.

Gohagan, J., Marcus, P., Fagerstrom, R., Pinsky, P., Kramer, B., Prorok, P., Group, L.S.S.R., et al., 2004. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute. Chest 126 (1), 114–121.

Goldsborough, P., Pawlowski, N., Caicedo, J., Singh, S., Carpenter, A., 2017. CytoGAN: Generative modeling of cell images. http://dx.doi.org/10.1101/227645.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning, Vol. 1. MIT Press, http://www.deeplearningbook.org.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Grall, A., Hamidinekoo, A., Malcolm, P., Zwiggelaar, R., 2019. Using a conditional generative adversarial network (cGAN) for prostate segmentation. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 15–25.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al., 2020. Bootstrap your own latent-a new approach to self-supervised learning. Adv. Neural Inf. Process. Syst. 33, 21271–21284.

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., Staudt, L.M., 2016. Toward a shared vision for cancer genomic data. N. Engl. J. Med. 375 (12), 1109–1112.

Grover, A., Chute, C., Shu, R., Cao, Z., Ermon, S., 2020. Alignflow: Cycle consistent learning from multiple domains via normalizing flows. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. pp. 4028–4035.

Grover, A., Dhar, M., Ermon, S., 2018. Flow-gan: Combining maximum likelihood and adversarial learning in generative models. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32.

Gu, Y., Zeng, Z., Chen, H., Wei, J., Zhang, Y., Chen, B., Li, Y., Qin, Y., Xie, Q., Jiang, Z., et al., 2020. MedSRGAN: medical images super-resolution using generative adversarial networks. Multimedia Tools Appl. 79, 21815–21840.

Guan, S., Loew, M., 2019. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. J. Med. Imaging 6 (3), 031411.

Guerraoui, R., Guirguis, A., Kermarrec, A.-M., Merrer, E.L., 2020. FeGAN: Scaling distributed GANs. In: Proceedings of the 21st International Middleware Conference. pp. 193–206.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of wasserstein GANs. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., URL: https://proceedings.neurips.cc/paper/2017/file/892c3b1c6dccd52936e27cbd0ff683d6-Paper.pdf.

Hadjiiski, L., Cho, H.-c., Chan, H.-P., Sahiner, B., Helvie, M.A., Paramagul, C., Nees, A.V., 2012. Inter-and intra-observer variability of radiologists evaluating CBIR systems. In: International Workshop on Digital Mammography. Springer, pp. 482–489.

Halling-Brown, M.D., Warren, L.M., Ward, D., Lewis, E., Mackenzie, A., Wallis, M.G., Wilkinson, L.S., Given-Wilson, R.M., McAvinchey, R., Young, K.C., 2020. OPTIMAM mammography image database: A large-scale resource of mammography images and clinical data. Radiol.: Artif. Intell. e200103.

Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., Nakayama, H., 2018. GAN-based synthetic brain MR image generation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 734–738.

Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., Umemoto, K., Li, Y., Nakayama, H., 2019a. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. arXiv:1906.04962.

Han, C., Murao, K., Satoh, S., 2019b. Learning more with less: GAN-based medical image augmentation. arXiv:1904.00838.

Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y., 2021. Deep symmetric adaptation network for cross-modality medical image segmentation. arXiv preprint arXiv:2101.06853.

Han, C., Rundo, L., Araki, R., Furukawa, Y., Mauri, G., Nakayama, H., Hayashi, H., 2020. Infinite brain MR images: PGGAN-based data augmentation for tumor detection. In: Neural Approaches to Dynamics of Signal Exchanges. Springer, pp. 291–303.

Han, C., Rundo, L., Araki, R., Nagano, Y., Furukawa, Y., Mauri, G., Nakayama, H., Hayashi, H., 2019c. Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection. IEEE Access 7, 156966–156977.

Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. Cell 100 (1), 57–70.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144 (5), 646–674.

Hardy, C., Le Merrer, E., Sericola, B., 2019. Md-gan: Multi-discriminator generative adversarial networks for distributed datasets. In: 2019 IEEE International Parallel and Distributed Processing Symposium. IPDPS, IEEE, pp. 866–877.

Hasani, N., Morris, M.A., Rhamim, A., Summers, R.M., Jones, E., Siegel, E., Saboury, B., 2022. Trustworthy artificial intelligence in medical imaging. PET Clin. 17 (1), 1–12.

Haubold, J., Hosch, R., Umutlu, L., Wetter, A., Haubold, P., Radbruch, A., Forsting, M., Nensa, F., Koitka, S., 2021. Contrast agent dose reduction in computed tomography with deep learning using a conditional generative adversarial network. Eur. Radiol. 1–9.

He, X., Yang, S., Li, G., Li, H., Chang, H., Yu, Y., 2019. Non-local context encoder: Robust biomedical image segmentation against adversarial attacks. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. pp. 8417–8424.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P., 2001. The digital database for screening mammography, IWDM-2000. In: Fifth International Workshop on Digital Mammography. Medical Physics Publishing, ISBN: 1-930524-00-5, pp. 212–218.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500.

HIPAA, 1996. The health insurance portability and accountability act of 1996. Public Law 104, 191, https://www.govinfo.gov/content/pkg/PLAW-104publ191/html/PLAW-104publ191.htm.

Hirano, H., Minagi, A., Takemoto, K., 2021. Universal adversarial attacks on deep neural networks for medical image classification. BMC Med. Imaging 21 (1), 1–13.

Hitaj, B., Ateniese, G., Perez-Cruz, F., 2017. Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 603–618.

Hjelm, R.D., Jacob, A.P., Che, T., Trischler, A., Cho, K., Bengio, Y., 2017. Boundary-seeking generative adversarial networks. arXiv preprint arXiv:1702.08431.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hognon, C., Tixier, F., Gallinato, O., Colin, T., Visvikis, D., Jaouen, V., 2019. Standardization of multicentric image datasets with generative adversarial networks. In: IEEE Nuclear Science Symposium and Medical Imaging Conference 2019.

Hopper, K.D., Kasales, C., Van Slyke, M.A., Schwartz, T.A., TenHave, T.R., Jozefiak, J.A., 1996. Analysis of interobserver and intraobserver variability in CT tumor measurements. AJR. Am. J. Roentgenol. 167 (4), 851–854.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J., 2018. Artificial intelligence in radiology. Nat. Rev. Cancer 18 (8), 500–510.

Hoyer, L., Dai, D., Van Gool, L., 2022a. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9924–9935.

Hoyer, L., Dai, D., Van Gool, L., 2022b. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. arXiv preprint arXiv:2204.13132.

Hu, X., Chung, A.G., Fieguth, P., Khalvati, F., Haider, M.A., Wong, A., 2018b. Prostategan: Mitigating data bias via prostate diffusion imaging synthesis with generative adversarial networks. arXiv preprint arXiv:1811.05817.

Hu, X., Guo, R., Chen, J., Li, H., Waldmannstetter, D., Zhao, Y., Li, B., Shi, K., Menze, B., 2020. Coarse-to-fine adversarial networks and zone-based uncertainty analysis for NK/T-cell lymphoma segmentation in CT/PET images. IEEE J. Biomed. Health Inf. 24 (9), 2599–2608.

Hu, J., Shen, L., Sun, G., 2018a. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141.

Huang, X., Liu, M.-Y., Belongie, S., Kautz, J., 2018. Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 172–189.

Huang, S., Yang, J., Fong, S., Zhao, Q., 2020. Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. Cancer Lett. 471, 61–71.

Hukkelås, H., Mester, R., Lindseth, F., 2019. Deepprivacy: A generative adversarial network for face anonymization. In: International Symposium on Visual Computing. Springer, pp. 565–578.

Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.-Y., Yang, M.H., 2019. Adversarial learning for semi-supervised semantic segmentation. In: 29th British Machine Vision Conference, BMVC 2018.

Huo, Y., Xu, Z., Bao, S., Bermudez, C., Moon, H., Parvathaneni, P., Moyo, T.K., Savona, M.R., Assad, A., Abramson, R.G., et al., 2018. Splenomegaly segmentation on multi-modal MRI using deep convolutional networks. IEEE Trans. Med. Imaging 38 (5), 1185–1196.

Huq, A., Pervin, M.T., 2020. Analysis of adversarial attacks on skin cancer recognition. In: 2020 International Conference on Data Science and Its Applications (ICoDSA). IEEE, pp. 1–4.

Huynh, E., Hosny, A., Guthier, C., Bitterman, D.S., Petit, S.F., Haas-Kogan, D.A., Kann, B., Aerts, H.J., Mak, R.H., 2020. Artificial intelligence in radiation oncology. Nat. Rev. Clin. Oncol. 17 (12), 771–781.

Hwang, U., Choi, S., Lee, H.-B., Yoon, S., 2017. Adversarial training for disease prediction from electronic health records with missing data. arXiv preprint arXiv:1711.04126.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.
Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.
Itri, J.N., Tappouni, R.R., McEachern, R.O., Pesch, A.J., Patel, S.H., 2018. Fundamentals of diagnostic error in imaging. Radiographics 38 (6), 1845–1865.
IXI Dataset, 2007. IXI Dataset by brain-development.org. URL: http://brain-development.org/ixi-dataset/.
Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S., et al., 2013. Automatic tuberculosis screening using chest radiographs. IEEE Trans. Med. Imaging 33 (2), 233–245.
JAMIT Japanese Society of Medical Imaging Technology, JAMIT CAD Contest. URL: http://www.jamit.jp/meetinginfo/cad.html.
Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D., 2011. Global cancer statistics. CA: Cancer J. Clin. 61 (2), 69–90.
Jendele, L., Skopek, O., Becker, A.S., Konukoglu, E., 2019. Adversarial augmentation for enhancing classification of mammography images. arXiv preprint arXiv:1902.07762.
Jeong, J., Shin, J., 2021. Training GANs with stronger augmentations via contrastive discriminator. arXiv preprint arXiv:2103.09742.
Jiang, Y., Chang, S., Wang, Z., 2021. TransGAN: Two transformers can make one strong GAN. arXiv:2102.07074.
Jiang, J., Hu, Y.-C., Tyagi, N., Zhang, P., Rimner, A., Mageras, G.S., Deasy, J.O., Veeraraghavan, H., 2018. Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 777–785.
Jin, D., Xu, Z., Tang, Y., Harrison, A.P., Mollura, D.J., 2018. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 732–740.
Johnson, A.E., Pollard, T.J., Shen, L., Li-Wei, H.L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G., 2016. MIMIC-III, a freely accessible critical care database. Sci. Data 3 (1), 1–9.
Jolicoeur-Martineau, A., 2018. The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734.
Jordon, J., Yoon, J., Van Der Schaar, M., 2018. PATE-GAN: Generating synthetic data with differential privacy guarantees. In: International Conference on Learning Representations.
Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., Zhavoronkov, A., 2017a. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget 8 (7), 10883–10890.
Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., Zhavoronkov, A., 2017b. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Mol. Pharm. 14 (9), 3098–3104.
Kaiser, B., Albarqouni, S., 2019. MRI to CT translation with GANs. arXiv preprint arXiv:1901.05259.
Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F., 2020. Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. 2 (6), 305–311.
Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al., 2017. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 597–609.
Kanayama, T., Kurose, Y., Tanaka, K., Aida, K., Satoh, S., Kitsuregawa, M., Harada, T., 2019. Gastric cancer detection from endoscopic images using synthesis by GAN. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 530–538.
Kang, M., Park, J., 2020. Contragan: Contrastive learning for conditional image generation. Adv. Neural Inf. Process. Syst. 33, 21357–21369.
Kang, D., Park, J.E., Kim, Y.-H., Kim, J.H., Oh, J.Y., Kim, J., Kim, Y., Kim, S.T., Kim, H.S., 2018. Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. Neuro-Oncology 20 (9), 1251–1261.
Kansal, S., Goel, S., Bhattacharya, J., Srivastava, V., 2020. Generative adversarial network–convolution neural network based breast cancer classification using optical coherence tomographic images. Laser Phys. 30 (11), 115601.
Kapil, A., Meier, A., Zuraw, A., Steele, K.E., Rebelatto, M.C., Schmidt, G., Brieu, N., 2018. Deep semi supervised generative learning for automated tumor proportion scoring on NSCLC tissue needle biopsies. Sci. Rep. 8 (1), 1–10.
Karras, T., Aila, T., Laine, S., Lehtinen, J., 2017. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T., 2021. Alias-free generative adversarial networks. Adv. Neural Inf. Process. Syst. 34, 852–863.
Karras, T., Laine, S., Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4401–4410.
Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2020. Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8110–8119.
Kather, J.N., Halama, N., Marx, A., 2018. 100,000 Histological images of human colorectal cancer and healthy tissue. http://dx.doi.org/10.5281/zenodo.1214456.
Kazemifar, S., Barragán Montero, A.M., Souris, K., Rivas, S.T., Timmerman, R., Park, Y.K., Jiang, S., Geets, X., Sterpin, E., Owrangi, A., 2020. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. J. Appl. Clin. Med. Phys. 21 (5), 76–86.
Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., Mukhopadhyay, A., 2020. GANs for medical image analysis. Artif. Intell. Med. 101938.
Kazuhiro, K., Werner, R.A., Toriumi, F., Javadi, M.S., Pomper, M.G., Solnes, L.B., Verde, F., Higuchi, T., Rowe, S.P., 2018. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. Tomography 4 (4), 159.
Ke, J., Shen, Y., Liang, X., Shen, D., 2021. Contrastive learning based stain normalization across multiple tumor in histopathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 571–580.
Kearney, V., Chan, J.W., Wang, T., Perry, A., Descovich, M., Morin, O., Yom, S.S., Solberg, T.D., 2020a. DoseGAN: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation. Sci. Rep. 10 (1), 1–8.
Kearney, V., Ziemer, B.P., Perry, A., Wang, T., Chan, J.W., Ma, L., Morin, O., Yom, S.S., Solberg, T.D., 2020b. Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks. Radiol.: Artif. Intell. 2 (2), e190027.
Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D., 2019. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 17 (1), 1–9.
Kim, M., Byun, H., 2020. Learning texture invariant representation for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12975–12984.
Kim, K.H., Do, W.-J., Park, S.-H., 2018a. Improving resolution of MR images with an adversarial network incorporating images with different contrast. Med. Phys. 45 (7), 3120–3131.
Kim, B.N., Dolz, J., Jodoin, P.-M., Desrosiers, C., 2019a. Privacy-net: An adversarial approach for identity-obfuscated segmentation of medical images. arXiv preprint arXiv:1909.04087.
Kim, H.S., Jung, J., Jeong, C., Kwak, J., Choi, E., Lee, S., Yoon, S., Cho, B., 2019b. Prediction of hepatic parenchymal change in gd-EOB-DTPA MR images after stereotactic body radiation therapy by cycle GAN deep neural network. Int. J. Radiat. Oncol.*Biol.*Phys. 105, E135. http://dx.doi.org/10.1016/j.ijrobp.2019.06.2171.
Kim, S., Kim, B., Park, H., 2020. Synthesis of brain tumor MR images for learning data augmentation. arXiv preprint arXiv:2003.07526.
Kim, R., Ock, C.-Y., Keam, B., Kim, T.M., Kim, J.H., Paeng, J.C., Kwon, S.K., Hah, J.H., Kwon, T.-K., Kim, D.-W., et al., 2016. Predictive and prognostic value of PET/CT imaging post-chemoradiotherapy and clinical decision-making consequences in locally advanced head & neck squamous cell carcinoma: a retrospective study. BMC Cancer 16 (1), 1–9.
Kim, M., Oh, I., Ahn, J., 2018b. An improved method for prediction of cancer prognosis by network learning. Genes 9, 478. http://dx.doi.org/10.3390/genes9100478.
Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
Klaver, C.E., Bulkmans, N., Drillenburg, P., Grabsch, H.I., van Grieken, N.C., Karrenbeld, A., Koens, L., van Lijnschoten, I., Meijer, J., Nagtegaal, I.D., et al., 2020. Interobserver, intraobserver, and interlaboratory variability in reporting pt4a colon cancer. Virchows Archiv 476 (2), 219–230.
Kodali, N., Abernethy, J., Hays, J., Kira, Z., 2017. On convergence and stability of GANs. arXiv preprint arXiv:1705.07215.
Kohl, S., Bonekamp, D., Schlemmer, H.-P., Yaqubi, K., Hohenfellner, M., Hadaschik, B., Radtke, J.-P., Maier-Hein, K., 2017. Adversarial networks for the detection of aggressive prostate cancer. arXiv preprint arXiv:1702.08014.
Koike, Y., Anetai, Y., Takegawa, H., Ohira, S., Nakamura, S., Tanigawa, N., 2020. Deep learning-based metal artifact reduction using cycle-consistent adversarial network for intensity-modulated head and neck radiation therapy treatment planning. Phys. Medica 78, 8–14.
Korkinof, D., Harvey, H., Heindl, A., Karpati, E., Williams, G., Rijken, T., Kecskemethy, P., Glocker, B., 2020. Perceived realism of high resolution generative adversarial network derived synthetic mammograms. Radiol.: Artif. Intell. e190181.
Kornblau, S.M., Tibes, R., Qiu, Y.H., Chen, W., Kantarjian, H.M., Andreeff, M., Coombes, K.R., Mills, G.B., 2009. Functional proteomic profiling of AML predicts response and survival. Blood 113 (1), 154–164.
Korpihalkola, J., Sipola, T., Puuska, S., Kokkonen, T., 2020. One-pixel attack deceives automatic detection of breast cancer. arXiv preprint arXiv:2012.00517.
Krause, J., Grabsch, H.I., Kloor, M., Jendrusch, M., Echle, A., Buelow, R.D., Boor, P., Luedde, T., Brinker, T.J., Trautwein, C., et al., 2021. Deep learning detects genetic alterations in cancer histology generated by adversarial networks. J. Pathol. 254 (1), 70–79.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105.
Kuang, Y., Lan, T., Peng, X., Selasi, G.E., Liu, Q., Zhang, J., 2020. Kuang. IEEE Access 8, 77725–77734.
Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al., 2019. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. IEEE Trans. Med. Imaging 38 (11), 2556–2568.
Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A., 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging 36 (7), 1550–1560.
Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Veta, M., 2019. Learning domain-invariant representations of histological images. Front. Med. 6, 162.
Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474.
Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R.G., Granton, P., Zegers, C.M., Gillies, R., Boellard, R., Dekker, A., et al., 2012. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer 48 (4), 441–446.
LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., et al., 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv http://dx.doi.org/10.1101/2019.12.13.19014902.
Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., Zhou, X., 2020. Generative adversarial networks and its applications in biomedical informatics. Front. Public Health 8, 164.
Larrazabal, A.J., Nieto, N., Peterson, V., Milone, D.H., Ferrante, E., 2020. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc. Natl. Acad. Sci. 117 (23), 12592–12594.
Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O., 2016. Autoencoding beyond pixels using a learned similarity metric. In: International Conference on Machine Learning. PMLR, pp. 1558–1566.
LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1 (4), 541–551.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690.
Lee, R.S., Gimenez, F., Hoogi, A., Rubin, D., 2016. Curated breast imaging subset of DDSM. Cancer Imaging Arch. 8, 2016.
Lee, H., Jo, J., Lim, H., 2020. Study on optimal generative network for synthesizing brain tumor-segmented MR images. Math. Probl. Eng. 2020.
Lee, J., Nishikawa, R.M., 2020. Simulating breast mammogram using conditional generative adversarial network: application towards finding mammographically-occult cancer. In: Medical Imaging 2020: Computer-Aided Diagnosis, Vol. 11314. International Society for Optics and Photonics, 1131418.
Lekadir, K., Osuala, R., Gallin, C., Lazrak, N., Kushibar, K., Tsakou, G., Aussó, S., Alberich, L.C., Marias, K., Tsiknakis, M., et al., 2021. FUTURE-AI: Guiding principles and consensus recommendations for trustworthy artificial intelligence in medical imaging. arXiv preprint arXiv:2109.09658.
Levine, A.B., Peng, J., Farnell, D., Nursey, M., Wang, Y., Naso, J.R., Ren, H., Farahani, H., Chen, C., Chiu, D., et al., 2020. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. J. Pathol. 252 (2), 178–188.
Levy, M.A., Rubin, D.L., 2008. Tool support to enable evaluation of the clinical response to treatment. In: AMIA Annual Symposium Proceedings, Vol. 2008. American Medical Informatics Association, p. 399.
Li, X., Cui, Z., Wu, Y., Gu, L., Harada, T., 2021a. Estimating and improving fairness with adversarial learning. arXiv preprint arXiv:2103.04243.
Li, H., Giger, M.L., Huynh, B.Q., Antropova, N.O., 2017. Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. J. Med. Imaging 4 (4), 041304.
Li, Y., Han, G., Wu, X., Li, Z.H., Zhao, K., Zhang, Z., Liu, Z., Liang, C., 2021b. Normalization of multicenter CT radiomics by a generative adversarial network method. Phys. Med. Biol. 66 (5), 055030.
Li, T., Lin, L., 2019. Anonymousnet: Natural face de-identification with measurable privacy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M.J., et al., 2019b. Privacy-preserving federated brain tumour segmentation. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 133–141.
Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019a. DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 795–803.
Li, H., Pan, S.J., Wang, S., Kot, A.C., 2018. Domain generalization with adversarial feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5400–5409.
Li, M., Tang, H., Chan, M.D., Zhou, X., Qian, X., 2020a. DC-AL GAN: pseudoprogression and true tumor progression of glioblastoma multiform image classification based on DCGAN and AlexNet. Med. Phys. 47 (3), 1139–1150.
Li, C., Wand, M., 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European Conference on Computer Vision. Springer, pp. 702–716.
Li, P., Wang, S., Li, T., Lu, J., HuangFu, Y., Wang, D., 2020b. A large-scale CT and PET/CT dataset for lung cancer diagnosis [dataset]. Cancer Imaging Arch..
Li, Y., Yuan, L., Vasconcelos, N., 2019c. Bidirectional learning for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6936–6945.
Liberman, L., Menell, J.H., 2002. Breast imaging reporting and data system (BI-RADS). Radiol. Clin. 40 (3), 409–430.
Liew, S.-L., Anglin, J.M., Banks, N.W., Sondag, M., Ito, K.L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., et al., 2018. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci. Data 5 (1), 1–11.
Lim, Z.W., Lee, M.L., Hsu, W., Wong, T.Y., 2019. Building trust in deep learning system towards automated disease detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. pp. 9516–9521.
Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., et al., 2018. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience 7 (6), giy065.
Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.
Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., 2014. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med. Image Anal. 18 (2), 359–373.
Liu, R., Ge, Y., Choi, C.L., Wang, X., Li, H., 2021. Divco: Diverse conditional image synthesis via contrastive generative adversarial network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16377–16386.
Liu, X., Guo, S., Zhang, H., He, K., Mu, S., Guo, Y., Li, X., 2019a. Accurate colorectal tumor segmentation for CT scans based on the label assignment generative adversarial network. Med. Phys. 46 (8), 3532–3542.
Liu, S., Setio, A.A.A., Ghesu, F.C., Gibson, E., Grbic, S., Georgescu, B., Comaniciu, D., 2020b. No surprises: Training robust lung nodule detection for low-dose ct scans by augmenting with adversarial attacks. IEEE Trans. Med. Imaging 40 (1), 335–345.
Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., Wang, Z., 2019b. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology. Engineering 5 (1), 156–163.
Liu, B., Zhu, Y., Song, K., Elgammal, A., 2020a. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In: International Conference on Learning Representations.
Ljosa, V., Sokolnicki, K.L., Carpenter, A.E., 2012a. Annotated high-throughput microscopy image sets for validation. Nature Methods 9 (7), 637.
Ljosa, V., Sokolnicki, K.L., Carpenter, A.E., 2012b. Annotated high-throughput microscopy image sets for validation. Nature Methods 9 (7), 637.
Loeb, S., Bjurlin, M.A., Nicholson, J., Tammela, T.L., Penson, D.F., Carter, H.B., Carroll, P., Etzioni, R., 2014. Overdiagnosis and overtreatment of prostate cancer. Eur. Urol. 65 (6), 1046–1055.
Lopez, M., Posada, N., Moura, D.C., Pollán, R.R., Valiente, J.M.F., Ortega, C.S., Solar, M., Diaz-Herrero, G., Ramos, I., Loureiro, J., et al., 2012. BCDR: a breast cancer digital repository. In: 15th International Conference on Experimental Mechanics, Vol. 1215.
Luc, P., Couprie, C., Chintala, S., Verbeek, J., 2016. Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408.
Lyu, M., Su, D., Li, N., 2016. Understanding the sparse vector technique for differential privacy. arXiv preprint arXiv:1603.01699.
MacKay, D.J., 1992. A practical Bayesian framework for backpropagation networks. Neural Comput. 4 (3), 448–472.
Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A., 2017. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083.
Mahajan, V., Venugopal, V., 2020. Audit of artificial intelligence algorithms and its impact in relieving shortage of specialist doctors. In: Artificial Intelligence: Applications in Healthcare Delivery. CRC Press, p. 207.
Mahmood, R., Babier, A., McNiven, A., Diamant, A., Chan, T.C., 2018. Automated treatment planning in radiation therapy using generative adversarial networks. In: Machine Learning for Healthcare Conference. PMLR, pp. 484–499.
Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269.
Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders. arXiv preprint arXiv:1511.05644.

Mardani, M., Gong, E., Cheng, J.Y., Vasanawala, S., Zaharchuk, G., Alley, M., Thakur, N., Han, S., Dally, W., Pauly, J.M., et al., 2017. Deep generative adversarial networks for compressed sensing automates MRI. arXiv preprint arXiv:1706.00051.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al., 2011. The parkinson progression marker initiative (PPMI). Prog. Neurobiol. 95 (4), 629–635.

Martin, B., Schäfer, E., Jakubowicz, E., Mayr, P., Ihringer, R., Anthuber, M., Schenkirsch, G., Schaller, T., Märkl, B., 2018. Interobserver variability in the H&E-based assessment of tumor budding in pT3/4 colon cancer: does it affect the prognostic relevance? Virchows Arch. 473 (2), 189–197.

Maspero, M., Savenije, M.H., Dinkla, A.M., Seevinck, P.R., Intven, M.P., Jurgenliemk-Schulz, I.M., Kerkmeijer, L.G., van den Berg, C.A., 2018. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Phys. Med. Biol. 63 (18), 185001.

Mathew, S., Nadeem, S., Kumari, S., Kaufman, A., 2020. Augmenting colonoscopy using extended and directional cyclegan for lossy image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4696–4705.

Maximov, M., Elezi, I., Leal-Taixé, L., 2020. Ciagan: Conditional identity anonymization generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5447–5456.

McCreadie, G., Oliver, T., 2009. Eight CT lessons that we learned the hard way: an analysis of current patterns of radiological error and discrepancy with particular emphasis on CT. Clin. Radiol. 64 (5), 491–499.

McDonald, R.J., Schwartz, K.M., Eckel, L.J., Diehn, F.E., Hunt, C.H., Bartholmai, B.J., Erickson, B.J., Kallmes, D.F., 2015. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad. Radiol. 22 (9), 1191–1198.

McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics. PMLR, pp. 1273–1282.

Melas-Kyriazi, L., Rupprecht, C., Laina, I., Vedaldi, A., 2021. Finding an unsupervised image segmenter in each of your deep generative models. arXiv preprint arXiv:2105.08127.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34 (10), 1993–2024.

Mescheder, L., Geiger, A., Nowozin, S., 2018. Which training methods for GANs do actually converge? In: International Conference on Machine Learning. PMLR, pp. 3481–3490.

Messiou, C., Hillengass, J., Delorme, S., Lecouvet, F.E., Moulopoulos, L.A., Collins, D.J., Blackledge, M.D., Abildgaard, N., Østergaard, B., Schlemmer, H.-P., et al., 2019. Guidelines for acquisition, interpretation, and reporting of whole-body MRI in myeloma: myeloma response assessment and diagnosis system (MY-RADS). Radiology 291 (1), 5–13.

Milchenko, M., Marcus, D., 2013. Obscuring surface anatomy in volumetric imaging data. Neuroinformatics 11 (1), 65–75.

Mirsky, Y., Mahler, T., Shelef, I., Elovici, Y., 2019. CT-GAN: Malicious tampering of 3D medical imagery using deep learning. In: 28th {USENIX} Security Symposium ({USENIX} Security 19). pp. 461–478.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Modanwal, G., Vellal, A., Mazurowski, M.A., 2019. Normalization of breast MRIs using cycle-consistent generative adversarial networks. arXiv preprint arXiv:1912.08061.

Mok, T.C., Chung, A.C., 2018. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: International MICCAI Brainlesion Workshop. Springer, pp. 70–80.

Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S., 2012. Inbreast: toward a full-field digital mammographic database. Acad. Radiol. 19 (2), 236–248.

Mullick, S.S., Datta, S., Das, S., 2019. Generative adversarial minority oversampling. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1695–1704.

Munawar, F., Azmat, S., Iqbal, T., Grönlund, C., Ali, H., 2020. Segmentation of lungs in chest X-ray image using generative adversarial networks. IEEE Access 8, 153535–153545.

Murakami, Y., Magome, T., Matsumoto, K., Sato, T., Yoshioka, Y., Oguchi, M., 2020. Fully automated dose prediction using generative adversarial networks in prostate cancer patients. PLoS One 15 (5), e0232697.

Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., Kubo, T., Togashi, K., Fujita, H., 2020. Improving breast mass classification by shared data with domain transformation using a generative adversarial network. Comput. Biol. Med. 119, 103698.

Nash, J.F., et al., 1950. Equilibrium points in n-person games. Proc. Natl. Acad. Sci. 36 (1), 48–49.

National Cancer Institute, 2018. Radiology data from the clinical proteomic tumor analysis consortium glioblastoma multiforme [cptac-gbm] collection [data set]. Cancer Imaging Arch..

Neal, R.M., 2012. Bayesian Learning for Neural Networks, Vol. 118. Springer Science & Business Media.

Nearchou, I.P., Soutar, D.A., Ueno, H., Harrison, D.J., Arandelovic, O., Caie, P.D., 2021. A comparison of methods for studying the tumor microenvironment's spatial heterogeneity in digital pathology specimens. J. Pathol. Inform..

Negi, A., Raj, A.N.J., Nersisson, R., Zhuang, Z., Murugappan, M., 2020. RDA-UNET-WGAN: An accurate breast ultrasound lesion segmentation using wasserstein generative adversarial networks. Arab. J. Sci. Eng. 45 (8), 6399–6410.

Nehmeh, S., Erdi, Y., Ling, C., Rosenzweig, K., Squire, O., Braban, L., Ford, E., Sidhu, K., Mageras, G., Larson, S., et al., 2002. Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer. Med. Phys. 29 (3), 366–371.

Network, C.G.A.R., et al., 2011. Integrated genomic analyses of ovarian carcinoma. Nature 474 (7353), 609–615.

Newman-Toker, D.E., Wang, Z., Zhu, Y., Nassery, N., Tehrani, A.S.S., Schaffer, A.C., Yu-Moe, C.W., Clemens, G.D., Fanai, M., Siegal, D., 2021. Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the "Big Three". Diagnosis 8 (1), 67–84.

Nie, D., Shen, D., 2020. Adversarial confidence learning for medical image segmentation and synthesis. Int. J. Comput. Vis. 1–20.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2017. Medical image synthesis with context-aware generative adversarial networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 417–425.

Nishio, M., Muramatsu, C., Noguchi, S., Nakai, H., Fujimoto, K., Sakamoto, R., Fujita, H., 2020. Attribute-guided image generation of three-dimensional computed tomography images of lung nodules using a generative adversarial network. Comput. Biol. Med. 126, 104032.

NLST Research Team, 2011. The national lung screening trial: overview and study design. Radiology 258 (1), 243–253.

Norton, L., Simon, R., Brereton, H.D., Bogden, A.E., 1976. Predicting the course of Gompertzian growth. Nature 264 (5586), 542–545.

Odena, A., Olah, C., Shlens, J., 2017. Conditional image synthesis with auxiliary classifier GANs. In: International Conference on Machine Learning. PMLR, pp. 2642–2651.

Oleszkiewicz, W., Kairouz, P., Piczak, K., Rajagopal, R., Trzciński, T., 2018. Siamese generative adversarial privatizer for biometric data. In: Asian Conference on Computer Vision. Springer, pp. 482–497.

Oliveira, D.A.B., 2020a. Controllable skin lesion synthesis using texture patches, Bézier curves and conditional GANs. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1798–1802.

Oliveira, D.A.B., 2020b. Implanting synthetic lesions for improving liver lesion segmentation in CT exams. arXiv preprint arXiv:2008.04690.

Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., Imaizumi, K., Fujita, H., 2019. Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. BioMed Res. Int. 2019.

Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., Imaizumi, K., Fujita, H., 2020. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. Int. J. Comput. Assist. Radiol. Surg. 15 (1), 173–178.

Pandey, S., Singh, P.R., Tian, J., 2020. An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation. Biomed. Signal Process. Control 57, 101782.

Pang, T., Wong, J.H.D., Ng, W.L., Chan, C.S., 2021. Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification. Comput. Methods Programs Biomed. 203, 106018.

Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., Talwar, K., 2016. Semi-supervised knowledge transfer for deep learning from private training data. arXiv preprint arXiv:1610.05755.

Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., Erlingsson, Ú., 2018. Scalable private learning with pate. arXiv preprint arXiv:1802.08908.

Park, H., Bayat, A., Sabokrou, M., Kirschke, J.S., Menze, B.H., 2020. Robustification of segmentation models against adversarial perturbations in medical imaging. In: International Workshop on PRedictive Intelligence in MEdicine. Springer, pp. 46–57.

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2337–2346.

Park, H., Yoo, Y., Kwak, N., 2018. Mc-gan: Multi-conditional generative adversarial network for image synthesis. arXiv preprint arXiv:1805.01123.

Parmar, C., Grossmann, P., Bussink, J., Lambin, P., Aerts, H.J., 2015. Machine learning methods for quantitative radiomic biomarkers. Sci. Rep. 5 (1), 1–11.

Patel, P., Kumari, N., Singh, M., Krishnamurthy, B., 2021. Lt-gan: Self-supervised gan with latent transformation detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3189–3198.

Paul, R., Schabath, M., Gillies, R., Hall, L., Goldgof, D., 2020. Mitigating adversarial attacks on medical image understanding systems. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1517–1521.

Pawlowski, N., Coelho de Castro, D., Glocker, B., 2020. Deep structural causal models for tractable counterfactual inference. Adv. Neural Inf. Process. Syst. 33, 857–869.

Peng, Y., Chen, S., Qin, A., Chen, M., Gao, X., Liu, Y., Miao, J., Gu, H., Zhao, C., Deng, X., et al., 2020. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiother. Oncol. 150, 217–224.

Phoulady, H.A., Mouton, P.R., 2018. A new cervical cytology dataset for nucleus detection and image classification (Cervix93) and methods for cervical nucleus detection. arXiv preprint arXiv:1811.09651.

Pinaya, W.H.L., Tudosiu, P.-D., Gray, R., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J., 2021. Unsupervised brain anomaly detection and segmentation with transformers. arXiv preprint arXiv:2102.11650.

Pittaluga, F., Koppal, S., Chakrabarti, A., 2019. Learning privacy preserving encodings through adversarial training. In: 2019 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 791–799.

Poorneshwaran, J., Kumar, S.S., Ram, K., Joseph, J., Sivaprakasam, M., 2019. Polyp segmentation using generative adversarial network. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, pp. 7201–7204.

Prior, F., Smith, K., Sharma, A., Kirby, J., Tarbox, L., Clark, K., Bennett, W., Nolan, T., Freymann, J., 2017. The public cancer radiology imaging collections of The Cancer Imaging Archive. Sci. Data 4 (1), 1–7.

Prokopenko, D., Stadelmann, J.V., Schulz, H., Renisch, S., Dylov, D.V., 2019. Unpaired synthetic image generation in radiology using GANs. In: Workshop on Artificial Intelligence in Radiation Therapy. Springer, pp. 94–101.

Pusey, E., Lufkin, R.B., Brown, R., Solomon, M.A., Stark, D.D., Tarr, R., Hanafee, W., 1986. Magnetic resonance imaging artifacts: mechanism and clinical significance. Radiographics 6 (5), 891–911.

Qasim, A.B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J.C., Sekuboyina, A., Kofler, F., Lipkova, J., Li, H., Menze, B., 2020. Red-GAN: Attacking class imbalance via conditioned generation. Yet another medical imaging perspective. In: Medical Imaging with Deep Learning. PMLR, pp. 655–668.

Qin, Z., Liu, Z., Zhu, P., Xue, Y., 2020. A GAN-based image synthesis method for skin lesion classification. Comput. Methods Programs Biomed. 195, 105568.

Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D., 2009. Dataset Shift in Machine Learning. The MIT Press.

Quiros, A.C., Murray-Smith, R., Yuan, K., 2019. PathologyGAN: Learning deep representations of cancer tissue. arXiv preprint arXiv:1907.02644.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Rahman, M.M., Fookes, C., Baktashmotlagh, M., Sridharan, S., 2019. Multi-component image translation for deep domain generalization. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 579–588.

Rashid, H., Tanveer, M.A., Khan, H.A., 2019. Skin lesion classification using GAN based data augmentation. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 916–919.

Rasouli, M., Sun, T., Rajagopal, R., 2020. Fedgan: Federated generative adversarial networks for distributed data. arXiv preprint arXiv:2006.07228.

Rau, A., Edwards, P.E., Ahmad, O.F., Riordan, P., Janatka, M., Lovat, L.B., Stoyanov, D., 2019. Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. Int. J. Comput. Assist. Radiol. Surg. 14 (7), 1167–1176.

Raval, N., Machanavajjhala, A., Cox, L.P., 2017. Protecting visual secrets using adversarial nets. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 1329–1332.

Razavi, A., Van den Oord, A., Vinyals, O., 2019. Generating diverse high-fidelity images with vq-vae-2. Adv. Neural Inf. Process. Syst. 32.

Rezaei, M., Harmuth, K., Gierke, W., Kellermeier, T., Fischer, M., Yang, H., Meinel, C., 2017. A conditional adversarial network for semantic segmentation of brain tumor. In: International MICCAI Brainlesion Workshop. Springer, pp. 241–252.

Rezende, D., Mohamed, S., 2015. Variational inference with normalizing flows. In: International Conference on Machine Learning. PMLR, pp. 1530–1538.

Rimmer, A., 2017. Radiologist shortage leaves patient care at risk, warns royal college. BMJ: Br. Med. J. (Online) 359.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Rubin, M., Stein, O., Turko, N.A., Nygate, Y., Roitshtain, D., Karako, L., Barnea, I., Giryes, R., Shaked, N.T., 2019. TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set. Med. Image Anal. 57, 176–185.

Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M., 2021a. Palette: Image-to-image diffusion models. arXiv preprint arXiv:2111.05826.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M., 2021b. Image super-resolution via iterative refinement. arXiv preprint arXiv:2104.07636.

Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., Barfett, J., 2018. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 990–994.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., 2016. Improved techniques for training GANs. arXiv preprint arXiv:1606.03498.

Samangouei, P., Kabkab, M., Chellappa, R., 2018. Defense-gan: Protecting classifiers against adversarial attacks using generative models. arXiv preprint arXiv:1805.06605.

Sampath, V., Maurtua, I., Martín, J.J.A., Gutierrez, A., 2021. A survey on generative adversarial networks for imbalance problems in computer vision tasks. J. Big Data 8 (1), 1–59.

Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M., 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci. Rep. 9 (1), 1–9.

Santurkar, S., Schmidt, L., Madry, A., 2018. A classification-based study of covariate shift in GAN distributions. In: International Conference on Machine Learning. PMLR, pp. 4480–4489.

Sarker, M., Kamal, M., Rashwan, H.A., Abdel-Nasser, M., Singh, V.K., Banu, S.F., Akram, F., Chowdhury, F.U., Choudhury, K.A., Chambon, S., et al., 2019. MobileGAN: Skin lesion segmentation using a lightweight generative adversarial network. arXiv preprint arXiv:1907.00856.

Sattigeri, P., Hoffman, S.C., Chenthamarakshan, V., Varshney, K.R., 2018. Fairness GAN. arXiv preprint arXiv:1805.09910.

Sauer, A., Schwarz, K., Geiger, A., 2022. Stylegan-xl: Scaling stylegan to large diverse datasets. p. 1, arXiv preprint arXiv:2202.00273.

Schimke, N., Kuehler, M., Hale, J., 2011. Preserving privacy in structural neuroimages. In: IFIP Annual Conference on Data and Applications Security and Privacy. Springer, pp. 301–308.

Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U., 2019. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. 54, 30–44.

Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G., 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 146–157.

Schmainda, K., Prah, M., 2018. Data from brain-tumor-progression. Cancer Imaging Arch..

Schonfeld, E., Schiele, B., Khoreva, A., 2020. A u-net based discriminator for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8207–8216.

Schwarz, C.G., Kremers, W.K., Therneau, T.M., Sharp, R.R., Gunter, J.L., Vemuri, P., Arani, A., Spychalla, A.J., Kantarci, K., Knopman, D.S., et al., 2019. Identification of anonymous MRI research participants with face-recognition software. N. Engl. J. Med. 381 (17), 1684–1686.

Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22 (3), 1060–1075.

Selvan, R., Faye, F., Middleton, J., Pai, A., 2020. Uncertainty quantification in medical image segmentation with normalizing flows. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 80–90.

Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., et al., 2017. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med. Image Anal. 42, 1–13.

Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., et al., 2014. The Cambridge centre for ageing and neuroscience (cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol. 14 (1), 1–25.

Shaham, T.R., Dekel, T., Michaeli, T., 2019. Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4570–4580.

Shahidi, F., 2021. Breast cancer histopathology image super-resolution using wide-attention GAN with improved wasserstein gradient penalty and perceptual loss. IEEE Access 9, 32795–32809.

Sharma, N., Aggarwal, L.M., 2010. Automated medical image segmentation techniques. J. Med. Phys./Assoc. Med. Phys. India 35 (1), 3.

Sharpe, M.B., Moore, K.L., Orton, C.G., 2014. Within the next ten years treatment planning will become fully automated without the need for human intervention. Med. Phys. 41 (12), 120601. http://dx.doi.org/10.1118/1.4894496.

Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R., et al., 2020. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci. Rep. 10 (1), 1–12.

Shen, T., Hao, K., Gou, C., Wang, F.-Y., 2021. Mass image synthesis in mammogram with contextual information based on GANs. Comput. Methods Programs Biomed. 202, 106019.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248.

Shi, Z., Hu, Q., Yue, Y., Wang, Z., AL-Othmani, O.M.S., Li, H., 2020. Automatic nodule segmentation method for CT images using aggregation-u-net generative adversarial networks. Sens. Imaging 21 (1), 1–16.

Shimodaira, H., 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Statist. Plann. Inference 90 (2), 227–244.

Shin, Y., Qadir, H.A., Balasingham, I., 2018b. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. IEEE Access 6, 56007–56017.

Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018a. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 1–11.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.-i., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. Am. J. Roentgenol. 174 (1), 71–74.

Shokri, R., Shmatikov, V., 2015. Privacy-preserving deep learning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. pp. 1310–1321.

Shyamala, K., Girish, H., Murgod, S., 2014. Risk of tumor cell seeding through biopsy and aspiration cytology. J. Int. Soc. Prev. Community Dent. 4 (1), 5.

Siddiquee, M.M.R., Zhou, Z., Tajbakhsh, N., Feng, R., Gotway, M.B., Bengio, Y., Liang, J., 2019. Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 191–200.

Silva, J., Histace, A., Romain, O., Dray, X., Granado, B., 2014. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int. J. Comput. Assist. Radiol. Surg. 9 (2), 283–293.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al., 2019. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063.

Singh, S., Bray, M.A., Jones, T.R., Carpenter, A.E., 2014. Pipeline for illumination correction of images for high-throughput microscopy. J. Microsc. 256 (3), 231–236.

Singh, N.K., Raza, K., 2020. Medical image generation using generative adversarial networks. arXiv preprint arXiv:2005.10687.

Singh, V.K., Romani, S., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M.M.K., Abdulwahab, S., Torrents-Barrena, J., Saleh, A., Arquez, M., et al., 2018. Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 833–840.

Sogancioglu, E., Murphy, K., van Ginneken, B., 2021. NODE21. http://dx.doi.org/10.5281/zenodo.5548363.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. PMLR, pp. 2256–2265.

Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M., Marescaux, J., 3D Image Reconstruction for Comparison of Algorithm Database. URL: https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/,

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. Adv. Neural Inf. Process. Syst. 32.

Sorin, V., Barash, Y., Konen, E., Klang, E., 2020. Creating artificial images for radiology applications using generative adversarial networks (GANs)–a systematic review. Acad. Radiol..

Stadler, T., Oprisanu, B., Troncoso, C., 2021. Synthetic data – anonymisation groundhog day. arXiv:2011.07018.

Stein, C.M., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 1135–1151.

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 240–248.

Sun, B., Liu, F., Zhou, Y., Jin, S., Li, Q., Jin, X., 2020a. Classification of lung nodules based on GAN and 3D CNN. In: Proceedings of the 4th International Conference on Computer Science and Application Engineering. pp. 1–5.

Sun, L., Wang, J., Huang, Y., Ding, X., Greenspan, H., Paisley, J., 2020b. An adversarial learning approach to medical image synthesis for lesion detection. IEEE J. Biomed. Health Inf. 24 (8), 2303–2314.

Sushko, V., Schönfeld, E., Zhang, D., Gall, J., Schiele, B., Khoreva, A., 2020. You only need adversarial supervision for semantic image synthesis. arXiv preprint arXiv:2012.04781.

Swiderski, B., Gielata, L., Olszewski, P., Osowski, S., Kołodziej, M., 2021. Deep neural system for supporting tumor recognition of mammograms using modified GAN. Expert Syst. Appl. 164, 113968.

Swiecicki, A., Konz, N., Buda, M., Mazurowski, M.A., 2021. A generative adversarial network-based abnormality detection using only normal images for model training with application to digital breast tomosynthesis. Sci. Rep. 11 (1), 1–13.

Szafranowska, Z., Osuala, R., Breier, B., Kushibar, K., Lekadir, K., Diaz, O., 2022. Sharing generative models instead of private data: a simulation study on mammography patch classification 12286. pp. 169–177. http://dx.doi.org/10.1117/12.2625781.

Tang, Y.-B., Tang, Y.-X., Xiao, J., Summers, R.M., 2019. Xlsor: A robust and accurate lung segmentor on chest X-Rays using criss-cross attention and customized radiorealistic abnormalities generation. In: International Conference on Medical Imaging with Deep Learning. PMLR, pp. 457–467.

Tanner, C., Ozdemir, F., Profanter, R., Vishnevsky, V., Konukoglu, E., Goksel, O., 2018. Generative adversarial networks for mr-ct deformable image registration. arXiv preprint arXiv:1807.07349.

Tanno, R., Worrall, D.E., Kaden, E., Ghosh, A., Grussu, F., Bizzi, A., Sotiropoulos, S.N., Criminisi, A., Alexander, D.C., 2021. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. NeuroImage 225, 117366.

Teh, S., Ranguis, S., Fagan, P., 2017. Inter-observer variability between radiologists reporting on cerebellopontine angle tumours on magnetic resonance imaging. J. Laryngol. Otol. 131 (S1), S47–S49.

Teramoto, A., Tsukamoto, T., Yamada, A., Kiriyama, Y., Imaizumi, K., Saito, K., Fujita, H., 2020. Deep learning approach to classification of lung cytological images: Two-step training using actual and synthesized images by progressive growing of generative adversarial networks. PLoS One 15 (3), e0229951.

Thambawita, V., Salehi, P., Sheshkal, S.A., Hicks, S.A., Hammer, H.L., Parasa, S., Lange, T.d., Halvorsen, P., Riegler, M.A., 2022. SinGAN-Seg: Synthetic training data generation for medical image segmentation. PLoS One 17 (5), e0267976.

Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp. Oncol. (Pozn) 19 (1A), 68–77.

Torfi, A., Fox, E.A., Reddy, C.K., 2020. Differentially private synthetic medical data generation using convolutional GANs. arXiv preprint arXiv:2012.11774.

Jimenez-del Toro, O., Müller, H., Krenn, M., Gruenberg, K., Taha, A.A., Winterstein, M., Eggel, I., Foncubierta-Rodríguez, A., Goksel, O., Jakab, A., et al., 2016. Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. IEEE Trans. Med. Imaging 35 (11), 2459–2475.

Treeby, B.E., Cox, B.T., 2010. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. J. Biomed. Opt. 15 (2), 021314.

Troyanskaya, O., Trajanoski, Z., Carpenter, A., Thrun, S., Razavian, N., Oliver, N., 2020. Artificial intelligence and cancer. Nature Cancer 1 (2), 149–152.

Tschandl, P., Rosendahl, C., Kittler, H., 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5 (1), 1–9.

Tschuchnig, M.E., Oostingh, G.J., Gadermayr, M., 2020. Generative adversarial networks in digital pathology: a survey on trends and future potential. Patterns 1 (6), 100089.

Vallières, M., Kayrivest, E., Perrin, L., et al., 2017. Data from head-neck-pet-ct. Cancer Imaging Arch..

Van Den Oord, A., Vinyals, O., et al., 2017. Neural discrete representation learning. Adv. Neural Inf. Process. Syst. 30.

Van der Goten, L.A., Hepp, T., Akata, Z., Smith, K., 2021. Adversarial privacy preservation in MRI scans of the brain. URL: https://openreview.net/forum?id=2NHl-ETnHxk.

Van Tulder, G., de Bruijne, M., 2015. Why does synthesized data improve multi-sequence classification? In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 531–538.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. arXiv preprint arXiv:1706.03762.

Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A., 2017. A benchmark for endoluminal scene segmentation of colonoscopy images. J. Healthc. Eng. 2017.

Verma, N., Cowperthwaite, M.C., Burnett, M.G., Markey, M.K., 2013. Differentiating tumor recurrence from treatment necrosis: a review of neuro-oncologic imaging strategies. Neuro-Oncol. 15 (5), 515–534.

Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A., 2008. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. pp. 1096–1103.

Volpi, R., Namkoong, H., Sener, O., Duchi, J., Murino, V., Savarese, S., 2018. Generalizing to unseen domains via adversarial data augmentation. arXiv preprint arXiv:1805.12018.

Vu, Q.D., Kim, K., Kwak, J.T., 2020a. Unsupervised tumor characterization via conditional generative adversarial networks. IEEE J. Biomed. Health Inf..

Vu, T., Li, M., Humayun, H., Zhou, Y., Yao, J., 2020b. A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer. Exp. Biol. Med. 245 (7), 597–605.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wang, T., Lei, Y., Curran, W.J., Liu, T., Yang, X., 2021. Contrast-enhanced MRI synthesis from non-contrast MRI using attention CycleGAN. In: Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging, Vol. 11600. International Society for Optics and Photonics, 116001L.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2018a. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8798–8807.

Wang, Z., She, Q., Ward, T.E., 2019b. Generative adversarial networks in computer vision: A survey and taxonomy. arXiv preprint arXiv:1906.01529.

Wang, H., Shen, T., Zhang, W., Duan, L.-Y., Mei, T., 2020a. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 642–659.

Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., Shoemaker, B.A., Gindulyte, A., Bryant, S.H., 2014. PubChem BioAssay: 2014 update. Nucl. Acids Res. 42 (Database issue), D1075–1082.

Wang, Q., Zhang, X., Chen, W., Wang, K., Zhang, X., 2020b. Class-aware multi-window adversarial lung nodule synthesis conditioned on semantic features. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 589–598.

Wang, T., Zhao, J., Yatskar, M., Chang, K.-W., Ordonez, V., 2019a. Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5310–5319.

Wang, Z., Zheng, H., He, P., Chen, W., Zhou, M., 2022. Diffusion-GAN: Training GANs with diffusion. http://dx.doi.org/10.48550/ARXIV.2206.02262, URL: https://arxiv.org/abs/2206.02262.

Wang, Y., Zhou, L., Wang, M., Shao, C., Shi, L., Yang, S., Zhang, Z., Feng, M., Shan, F., Liu, L., 2020c. Combination of generative adversarial network and convolutional neural network for automatic subcentimeter pulmonary adenocarcinoma classification. Quant. Imaging Med. Surg. 10 (6), 1249.

Wang, Y., Zhou, L., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., 2018b. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans. Med. Imaging 38 (6), 1328–1339.

Wei, L., Lin, Y., Hsu, W., 2020. Using a generative adversarial network for CT normalization and its impact on radiomic features. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 844–848.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, Jr., C.R., Jagust, W., Morris, J.C., et al., 2017. The Alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. Alzheimer's Dement. 13 (5), 561–571.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., 2013. The cancer genome atlas pan-cancer analysis project. Nature Genet. 45 (10), 1113–1120.

Wetstein, S.C., González-Gonzalo, C., Bortsova, G., Liefers, B., Dubost, F., Katramados, I., Hogeweg, L., van Ginneken, B., Pluim, J.P., de Bruijne, M., et al., 2020. Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. arXiv preprint arXiv:2006.06356.

Wilson, M.K., Chawla, T., Wang, L., Friedlander, M., Oza, A.M., Lheureux, S., 2018. Inter and intra-observer variability with the assessment of RECIST in ovarian cancer.

Wolterink, J.M., Dinkla, A.M., Savenije, M.H., Seevinck, P.R., van den Berg, C.A., Išgum, I., 2017. Deep MR to CT synthesis using unpaired data. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 14–23.

Wolterink, J.M., Kamnitsas, K., Ledig, C., Išgum, I., 2018. Generative adversarial networks and adversarial methods in biomedical image analysis. arXiv preprint arXiv:1810.10352.

Woo, M., Lowe, S.C., Devane, A.M., Gimbel, R.W., 2020. Intervention to reduce interobserver variability in computed tomographic measurement of cancer lesions among experienced radiologists. Curr. Probl. Diagn. Radiol..

World Health Organization, 2018. Cancer. URL: https://www.who.int/news-room/fact-sheets/detail/cancer.

Wu, Z., Wang, Z., Wang, Z., Jin, H., 2018c. Towards privacy-preserving visual recognition via adversarial training: A pilot study. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 606–624.

Wu, E., Wu, K., Cox, D., Lotter, W., 2018a. Conditional infilling GANs for data augmentation in mammogram classification. In: Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, pp. 98–106.

Wu, E., Wu, K., Lotter, W., 2020. Synthesizing lesions using contextual GANs improves breast cancer classification on mammograms. arXiv preprint arXiv:2006.00086.

Wu, Y., Yang, F., Ling, H., 2018b. Privacy-protective-GAN for face de-identification. arXiv preprint arXiv:1806.08906.

Wyman, B.T., Harvey, D.J., Crawford, K., Bernstein, M.A., Carmichael, O., Cole, P.E., Crane, P.K., DeCarli, C., Fox, N.C., Gunter, J.L., et al., 2013. Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimer's Dement. 9 (3), 332–337.

Xia, T., Chartsias, A., Wang, C., Tsaftaris, S.A., Alzheimer's Disease Neuroimaging Initiative, et al., 2021. Learning to synthesise the ageing brain without longitudinal data. Med. Image Anal. 73, 102169.

Xiao, X., Zhao, J., Qiang, Y., Chong, J., Yang, X., Kazihise, N.G.-F., Chen, B., Li, S., 2019. Radiomics-guided GAN for segmentation of liver tumor without contrast agents. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 237–245.

Xie, X., Chen, J., Li, Y., Shen, L., Ma, K., Zheng, Y., 2020. MI2GAN: Generative adversarial network for medical image domain adaptation using mutual information constraint. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 516–525.

Xie, L., Lin, K., Wang, S., Wang, F., Zhou, J., 2018. Differentially private generative adversarial network. arXiv preprint arXiv:1802.06739.

Xin, B., Yang, W., Geng, Y., Chen, S., Wang, S., Huang, L., 2020. Private fl-gan: Differential privacy synthetic data generation based on federated learning. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 2927–2931.

Xu, Z., Wang, X., Shin, H.-C., Yang, D., Roth, H., Milletari, F., Zhang, L., Xu, D., 2020. Correlation via synthesis: End-to-end image generation and radiogenomic learning based on generative adversarial network. In: Medical Imaging with Deep Learning. PMLR, pp. 857–866.

Xu, D., Yuan, S., Zhang, L., Wu, X., 2018. Fairgan: Fairness-aware generative adversarial networks. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 570–575.

Xue, Y., Xu, T., Zhang, H., Long, L.R., Huang, X., 2018. SEGAN: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics 16 (3), 383–392.

Yan, K., Wang, X., Lu, L., Summers, R.M., 2018. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging 5 (3), 036501.

Yang, T.-Y., Brinton, C., Mittal, P., Chiang, M., Lan, A., 2018c. Learning informative and private representations via generative adversarial networks. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 1534–1543.

Yang, J., Liu, S., Grbic, S., Setio, A.A.A., Xu, Z., Gibson, E., Chabin, G., Georgescu, B., Laine, A.F., Comaniciu, D., 2019. Class-aware adversarial lung nodule synthesis in CT images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 1348–1352.

Yang, J., Veeraraghavan, H., Armato, III, S.G., Farahani, K., Kirby, J.S., Kalpathy-Kramer, J., van Elmpt, W., Dekker, A., Han, X., Feng, X., et al., 2018b. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. Med. Phys. 45 (10), 4568–4581.

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., Firmin, D., 2018a. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans. Med. Imaging 37 (6), 1310–1321. http://dx.doi.org/10.1109/TMI.2017.2785879.

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: A review. Med. Image Anal. 58, 101552.

Yoon, J., Drumright, L.N., Van Der Schaar, M., 2020. Anonymization through data synthesis using generative adversarial networks (ads-gan). IEEE J. Biomed. Health Inf. 24 (8), 2378–2388.

Yoon, J., Jordon, J., Schaar, M., 2018. GANITE: Estimation of individualized treatment effects using generative adversarial nets. In: ICLR.

You, C., Li, G., Zhang, Y., Zhang, X., Shan, H., Li, M., Ju, S., Zhao, Z., Zhang, Z., Cong, W., et al., 2019. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). IEEE Trans. Med. Imaging 39 (1), 188–203.

Yu, X., Cai, X., Ying, Z., Li, T., Li, G., 2018b. Singlegan: Image-to-image translation by a single-generator network using multiple generative adversarial learning. In: Asian Conference on Computer Vision. Springer, pp. 341–356.

Yu, H., Hong, S., Yang, X., Ni, J., Dan, Y., Qin, B., 2013. Recognition of multiple imbalanced cancer types based on DNA microarray data using ensemble classifiers. BioMed Res. Int. 2013.

Yu, H., Zhang, X., 2020. Synthesis of prostate MR images for classification using capsule network-based GAN model. Sensors 20 (20), 5736.

Yu, B., Zhou, L., Wang, L., Fripp, J., Bourgeat, P., 2018a. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 626–630.

Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P., 2019. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging 38 (7), 1750–1762.

Yurt, M., Dar, S.U.H., Erdem, A., Erdem, E., Çukur, T., 2019. mustGAN: Multi-stream generative adversarial networks for MR image synthesis. arXiv preprint arXiv:1909.11504.

Zaman, A., Park, S.H., Bang, H., Park, C.-w., Park, I., Joung, S., 2020. Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images. Int. J. Comput. Assist. Radiol. Surg. 15, 931–941.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019a. Self-attention generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 7354–7363.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018c. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595.

Zhang, B.H., Lemoine, B., Mitchell, M., 2018a. Mitigating unwanted biases with adversarial learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 335–340.

Zhang, Y., Qu, H., Chang, Q., Liu, H., Metaxas, D., Chen, C., 2021. Training federated GANs with theoretical guarantees: A universal aggregation approach. arXiv preprint arXiv:2102.04655.

Zhang, Q., Wang, H., Lu, H., Won, D., Yoon, S.W., 2018b. Medical image synthesis with generative adversarial networks for tissue recognition. In: 2018 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, pp. 199–207.

Zhang, H., Zhang, Z., Odena, A., Lee, H., 2019b. Consistency regularization for generative adversarial networks. arXiv preprint arXiv:1910.12027.

Zhang, Z., Zhao, T., Gay, H., Sun, B., Zhang, W., 2020. Semi-supervised semantic segmentation of organs at risk on 3D pelvic CT images. arXiv preprint arXiv:2009.09571.

Zhao, J., Li, D., Kassam, Z., Howey, J., Chong, J., Chen, B., Li, S., 2020a. Tripartite-GAN: synthesizing liver contrast-enhanced MRI to improve tumor detection. Med. Image Anal. 63, 101667.

Zhao, B., Tan, Y., Tsai, W.-Y., Qi, J., Xie, C., Lu, L., Schwartz, L.H., 2016. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci. Rep. 6 (1), 1–7.

Zhao, K., Zhou, L., Gao, S., Wang, X., Wang, Y., Zhao, X., Wang, H., Liu, K., Zhu, Y., Ye, H., 2020b. Study of low-dose PET image recovery using supervised learning with CycleGAN. PLoS One 15 (9), e0238455.

Zhao, D., Zhu, D., Lu, J., Luo, Y., Zhang, G., 2018. Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale VGG16. Symmetry 10 (10), 519.

Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L., 2020. Hi-net: hybrid-fusion network for multi-modal MR image synthesis. IEEE Trans. Med. Imaging 39 (9), 2772–2781.

Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., van Ginneken, B., Madabhushi, A., Prince, J.L., Rueckert, D., Summers, R.M., 2021. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proc. IEEE 1–19.

Zhou, M., Leung, A., Echegaray, S., Gentles, A., Shrager, J.B., Jensen, K.C., Berry, G.J., Plevritis, S.K., Rubin, D.L., Napel, S., et al., 2018. Non–small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. Radiology 286 (1), 307–315.

Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2021. Domain generalization: A survey. arXiv preprint arXiv:2103.02503.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232.