Healthcare Analytics 5 (2024) 100340

Contents lists available at ScienceDirect

Healthcare Analytics

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

# A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions

Tauhidul Islam [a], Md. Sadman Hafiz [b], Jamin Rahman Jim [d], Md. Mohsin Kabir [c], M.F. Mridha [d,*]

[a] Department of Computer Science & Engineering, International Islamic University Chittagong, Chittagong 4318, Bangladesh
[b] Institute of Information and Communication Technology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
[c] Department of Computer Science & Engineering, Bangladesh University of Business & Technology, Dhaka 1216, Bangladesh
[d] Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh

ARTICLE INFO

*Keywords:*
Deep learning
Data augmentation
Image transformation
Medical imaging augmentation
Data synthesis
Systematic review

ABSTRACT

Data augmentation involves artificially expanding a dataset by applying various transformations to the existing data. Recent developments in deep learning have advanced data augmentation, enabling more complex transformations. Especially vital in the medical domain, deep learning-based data augmentation improves model robustness by generating realistic variations in medical images, enhancing diagnostic and predictive task performance. Therefore, to assist researchers and experts in their pursuits, there is a need for an extensive and informative study that covers the latest advancements in the growing domain of deep learning-based data augmentation in medical imaging. There is a gap in the literature regarding recent advancements in deep learning-based data augmentation. This study explores the diverse applications of data augmentation in medical imaging and analyzes recent research in these areas to address this gap. The study also explores popular datasets and evaluation metrics to improve understanding. Subsequently, the study provides a short discussion of conventional data augmentation techniques along with a detailed discussion on applying deep learning algorithms in data augmentation. The study further analyzes the results and experimental details from recent state-of-the-art research to understand the advancements and progress of deep learning-based data augmentation in medical imaging. Finally, the study discusses various challenges and proposes future research directions to address these concerns. This systematic review offers a thorough overview of deep learning-based data augmentation in medical imaging, covering application domains, models, results analysis, challenges, and research directions. It provides a valuable resource for multidisciplinary studies and researchers making decisions based on recent analytics.

## 1. Introduction

Data augmentation is a fundamental technique utilized in machine learning, which plays a crucial role in enhancing models' performance and generalization capabilities. It is essential for machine learning training dataset diversity without data collection. By rotating, translating, flipping, and more, current data points are transformed into new data instances with the same semantic labels but different appearances. In cases of sparse or expensive data, this artificially expands the training set and models a more generalizable learning process. Data augmentation aims to reduce overfitting, a typical machine learning problem when models trained on limited data perform well on training data but badly on unseen data. Models learn to reject unimportant variations and focus on decision-making patterns by introducing realistic transformations. [1]. The implementation of this technique aids in mitigating overfitting tendencies and enhances the efficacy of machine learning models. Perez et al. described various data augmentation techniques commonly employed in image processing, including random cropping, flipping, rotating, color jittering, and introducing noise. The generation of new training instances through systematic data augmentation involves modifying existing data while maintaining the original labels' accuracy and consistency. This enables models to recognize and generalize patterns, features, and structures across various situations. Embedded in foundational research, systematic data augmentation remains a crucial strategy for training robust and effective deep learning models, facilitating the learning of complex and comprehensive data representations which are essential for high variability

* Corresponding author.
*E-mail addresses:* tauhid.cse.iiuc@gmail.com (T. Islam), hafiz.sust333@gmail.com (Md. Sadman Hafiz), jaminjim.aiub@gmail.com (J.R. Jim), mdmkabi@gmail.com (Md. Mohsin Kabir), firoz.mridha@aiub.edu (M.F. Mridha).

https://doi.org/10.1016/j.health.2024.100340
Received 25 December 2023; Received in revised form 25 April 2024; Accepted 30 April 2024
Available online 8 May 2024

tasks [2]. While various established data augmentation methods exist, utilizing deep learning (DL)-based methods is essential in improving model performance. This is achieved by harnessing the capabilities of neural networks to generate task-specific and semantically significant variations in training data [3]. The technology facilitates adaptive augmentation, minimizes manual intervention, and is a robust regularization tool, promoting model generalization in many real-world circumstances. Deep neural networks have recently been employed to autonomously acquire intricate augmentation policies or transformations. These policies or transformations are designed to produce data that is more authentic and varied in comparison to conventional augmentations based on basic rules [4]. Neural augmentation networks receive the initial data and generate the parameters necessary to perform affine transformations and illumination modifications. The parameters are acquired through gradient descent throughout the training process mentioned by Ratner et al. [5]. Deep learning-based data augmentation generates meaningful and task-specific training data changes using neural networks. This makes models more robust and adaptable to different and difficult real-world settings. As deep learning advances, data augmentation is crucial to model performance in modern machine learning. However, using deep learning for data augmentation also presents novel issues and challenges, creating opportunities for further research [1]. Zhang et al. [6] explored neuroimaging multimodal data fusion issues, applications, and evaluation methods. The authors examined over 450 references to assess neurology applications, neuroimaging strengths and weaknesses, fusion rules, and assessment methods. The benefits of multimodal fusion include enhanced spatial and temporal resolution, contrast, and image quality for therapeutic and research purposes. Interdisciplinary education and research are essential for enhancing neuroimaging fusion methods. In their survey on biomedical data fusion strategies, Wang et al. [7] aimed to combine diverse data kinds to enhance information consistency and illness understanding. The authors suggested approaches to overcome preprocessing challenges such noise, missing data, and excessive complexity, including data alignment, registration, and dimensionality reduction. The study emphasized the need of pretreatment for successful data fusion, improving our understanding of complex biomedical data. Despite improvements, hurdles remain in standardizing and improving procedures, highlighting the need for continued innovation in this field. While both studies are similar to ours, they are older, done in 2021 and 2020. However, our analysis focuses on recent changes between 2022 and 2023.

While there are existing surveys related to Deep learning-based data augmentation, none specifically covers all these areas with a dedicated emphasis on medical imaging, As illustrated in Table 1, we conducted a comparative analysis to determine whether these studies include specific sections and to elucidate why our investigation provides a more comprehensive examination.

Therefore, for a better understanding of Deep learning-based data augmentation for medical imaging, a state-of-the-art comprehensive study is necessary with an in-depth discussion of the Deep Learning algorithms, preprocessing methods, datasets, evaluation metrics, diverse applications, and their challenges with future research direction to overcome those challenges. Hence, to gain a deeper insight into the contemporary progress of Deep learning-based data augmentation in medical imaging and to aid researchers in informed decision-making through recent analytics, we undertook this systematic review exclusively focusing on the latest state-of-the-art research articles dedicated to Deep learning-based data augmentation for medical imaging. This study thoroughly explores diverse applications of Deep learning-based data augmentation in medical imaging, commonly used datasets, evaluation metrics, conventional data augmentation methods, Deep Learning algorithms, a detailed results analysis of state-of-the-art research articles, and the challenges encountered.

Furthermore, it suggests potential future research directions to address these challenges. The primary objective of this article is to offer

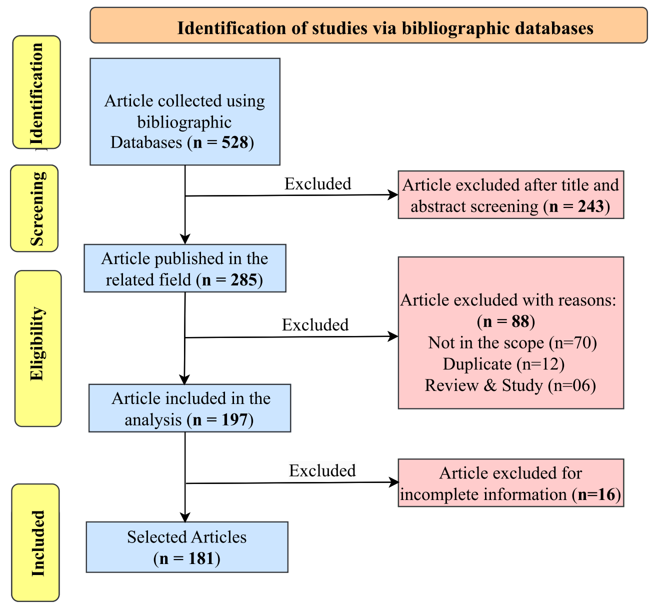


**Fig. 1.** PRISMA diagram illustrating the article selection process for this systematic review.

recommendations to investors and academics engaged in advancing data augmentation approaches rooted in deep learning. Additionally, it aims to enhance their comprehension of the most recent advancements in this discipline. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [13] flow diagram for the article selection process is outlined in Fig. 1. The survey is limited to high-quality academic articles in databases such as ScienceDirect, SpringerLink, ACM Digital Library and IEEE Xplore. The keywords that we used for article selection in different databases are: Data Augmentation on Medical Imaging, Deep learning-based Data Augmentation, Generative Models in Medical Imaging, Data Augmentation on Neurology Imaging, Data Augmentation on Xray imaging, Data Augmentation on MRI Analysis, Data Augmentation on CT analysis, Data Augmentation on Cancer Detection, Data Augmentation on Tumor detection and Data Augmentation on Skin Disease Detection. We've employed specific keywords to locate papers and established inclusion and exclusion criteria to select high-quality ones. Then, we followed some inclusion and exclusion criteria to select articles for our survey. (see Tables 2 and 3).

Briefly, the main contributions of this paper are as follows:

- Present a systematic review of the deep learning data augmentation, datasets, and evaluation metrics in medical imaging.
- A discussion of traditional data augmentation methods and a comprehensive analysis of the commonly used Deep learning-based data augmentation methods in medical imaging
- A comprehensive examination and critical evaluation of recent state-of-the-art research articles on Deep learning-based data augmentation in medical imaging.
- An in-depth analysis of the various challenges associated with Deep learning-based data augmentation in medical imaging and potential avenues for future research to overcome and alleviate these hurdles.

The subsequent sections of the paper are organized in the following manner: Section 2 provides an overview of commonly used datasets and evaluation metrics and Section 3.1 discusses some of the commonly used traditional data augmentation methods in this field of study. Subsequently. in Section 3.2, an analysis is conducted on commonly employed deep learning methods in data augmentation. The experimental results of some of the recent state-of-the-art research articles

**Table 1**
Comparison of recent review papers that have been done focusing on Deep learning-based data augmentation methods.

| Ref. | Datasets | Conventional augmentation techniques | DL-based augmentation techniques | Results analysis | Challenges & Future work | Contribution |
|---|---|---|---|---|---|---|
| Goceri et al. [8] | ✗ | ✓ | ✓ | ✓ | ✗ | Emphasized on the challenge of limited medical image datasets for Deep learning-based diagnosis, examines how data augmentation techniques affect organ images from different modalities, and concludes that optimal performance requires tailoring augmentation methods to image types. |
| Celard et al. [9] | ✓ | ✓ | ✗ | ✗ | ✓ | Provided a comprehensive survey of deep learning concepts, particularly generative models, in medical image analysis, provides concise overviews and encourages broader use of deep learning in medicine for segmentation, data augmentation, classification, and field challenges. |
| Xu et al. [10] | ✗ | ✗ | ✓ | ✗ | ✓ | A comprehensive survey of picture augmentation for deep learning categorizes algorithms into model-free, model-based, and optimal policy-based techniques, improving comprehension and supporting computer vision task creation was presented |
| Mumuni et al. [11] | ✗ | ✓ | ✓ | ✓ | ✗ | Advanced data augmentation methods in computer vision, including deeply learned, feature-level, and meta-learning-based methods, as well as data synthesis methods using 3D graphics modeling, neural rendering, and generative adversarial networks, are reviewed and compared in diverse datasets and tasks. |
| Khalifa et al. [12] | ✗ | ✓ | ✓ | ✗ | ✗ | Addressed the challenge of limited datasets in various fields, especially medical image processing and COVID-19, by surveying classical and Deep learning-based image data augmentation methods across research and application domains to increase dataset size and model accuracy. |
| This Paper | ✓ | ✓ | ✓ | ✓ | ✓ | Provides a comprehensive state-of-the-art review on recent advancements and challenges in data augmentation techniques within deep learning, with a specific focus on their crucial application in the medical domain, offering valuable observation for researchers and experts in navigating the evolving landscape of Deep learning-based data augmentation. |

**Table 2**
Criteria for selecting articles on deep learning-based data augmentation for medical imaging.

| Criteria type | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Study Type | Original and review articles focused on deep learning-based data augmentation | Studies not related to deep learning-based data augmentation |
| Language | Research articles written in English | Non-English articles |
| Publication Year | Articles published from 2021 onwards | Articles published before 2021 |
| Source | Peer-reviewed academic journals and conferences | Non-peer-reviewed sources like blogs or commercial websites |
| Intervention | Use of deep learning techniques for data augmentation | Purely traditional or non-deep learning techniques for data augmentation |
| Region | Not restricted to a particular region | – |
| Settings | Studies that apply data augmentation in image, text, and audio domains | Studies outside the context of data augmentation for machine learning |

**Table 3**
Search keywords used to find paper related to deep learning-based data augmentation in medical images.

| Search string | Number of papers |
|---|---|
| Deep learning data augmentation in medical imaging | 57 |
| Advanced data augmentation techniques for medical image analysis | 51 |
| Using deep learning for enhanced data augmentation in medical image processing | 57 |
| Deep neural networks and data augmentation in medical image diagnosis | 58 |
| Review of data augmentation impacts on deep learning in medical imaging | 53 |
| AI-driven data augmentation for medical image datasets | 50 |
| Data augmentation strategies in deep learning for medical imaging applications | 52 |
| The role of data augmentation in deep learning models for medical image classification | 53 |
| Improving medical image segmentation with deep learning-based data augmentation | 48 |
| Deep learning data augmentation for MRI/CT/X-ray image enhancement | 49 |

**Table 4**
Acronyms and their full forms used in this section.

| Acronym | Full form |
|---|---|
| AUC | Area Under the Curve |
| AUROC | Area Under the Receiver Operating Characteristic Curve |
| cGAN | Conditional Generative Adversarial Network |
| CNN | Convolutional Neural Network |
| DSC | Dice Similarity Coefficient |
| GAN | Generative Adversarial Network |
| MAE | Mean Absolute Error |
| MRI | Magnetic Resonance Imaging |
| RMSE | Root Mean Square Error |
| ROC | Receiver Operating Characteristic |
| ROI | Region of Interest |
| VAE | Variational Autoencoder |

and their applications are discussed in Section 4. Section 5 delineates the areas for additional investigation and the potential avenues for enhancement. Finally, the study concludes in Section 6.

In Table 4 we provided full forms of the acronyms we used throughout this article for better readability.

## 2. Datasets and evaluation metrics

In this section, we briefly discussed publicly available and most commonly used datasets, preprocessing methods, and evaluation metrics of some state-of-the-art papers published in this field of study within 2022–2023.

### 2.1. Datasets

Medical imaging datasets are limited in size and costly to obtain. Data augmentation techniques help artificially expand small datasets. This enables more effective training of deep learning models for medical imaging applications with scarce real-world data. We indexed some of the most used and publicly available datasets in this research area in Table 5:

### 2.2. Evaluation metrics

Evaluation metrics are crucial when it comes to assessing the performance and effectiveness of deep learning models. The importance of performance metrics lies in their ability to provide quantitative measures that help stakeholders, researchers, and practitioners understand the effectiveness of a model in achieving a specific goal. Table 6 contains data we picked up regarding some of the most used evaluation metrics in this field.

## 3. Data augmentation methods

We classified data augmentation techniques into two categories: traditional and Deep learning-based. To extend datasets, traditional and basic approaches change incoming data through transformations like flipping, cropping, and noise injection. Beyond simple data warping, advanced data augmentation approaches utilize generative models to create new data instances. These instances are not just manipulations of existing data, but whole new samples that are generated based on learnt data distributions. Generative Adversarial Networks (GANs) and autoencoders are currently leading the way in this method. GANs employ two neural networks, namely the generator and the discriminator, in a competitive framework. The generator's objective is to produce data that is indistinguishable from real data, while the discriminator's task is to differentiate between real and created data. This configuration allows for the generation of exceptionally lifelike data samples. Autoencoders, in contrast, acquire a compact representation of the data in a latent space with fewer dimensions and subsequently reconstruct the input data from this condensed form. This compressed representation can be adjusted to generate novel data points. These methods enable a deliberate and subtle creation of data, resulting in notable advantages in terms of improving the variety and excellence of training datasets [66]. In Sections 3.1 and 3.2 we discussed frequently used methods of both types.

### 3.1. Traditional methods

Some of the most common traditional augmentation methods include cropping, padding, horizontal flipping, rotations, permutations, scaling, translations, the addition of noise, etc. For image data, techniques like cropping, flipping, rotations, and scaling spatial transformations aim to mimic variations that could conceivably occur when taking photos. Things like permutations, noise injection, and pixel value transformations operationally diversify data without changing semantics. Overall, traditional techniques focus on expansions that plausibly increase diversity while maintaining the essential statistical properties and content of the data. We gathered some of these methods of this research area in Table 7.

### 3.2. Deep learning-based methods

Deep learning-based data augmentation techniques leverage neural networks such as GANs or variational autoencoders (VAEs) to acquire knowledge about the underlying data distribution. By doing so, these models can generate novel synthetic training instances with similar characteristics to real-world examples. These algorithms afford greater control and variability in comparison to rudimentary augmentation techniques. Generative models like GANs and autoencoders generate novel synthetic training instances that exhibit similar characteristics to real-world examples. These models afford greater control and variability in the data augmentation process compared to more rudimentary techniques. By understanding and manipulating the underlying data distributions, these advanced models can introduce changes to fundamental attributes such as style and texture, thus enhancing the diversity and quality of training data available for deep learning models. This capability is particularly valuable in domains where data collection is challenging or privacy concerns restrict the use of real data [178]. Deep learning-based data augmentation techniques have the potential to mitigate overfitting and enhance the robustness and generalization performance of models. This is particularly beneficial in scenarios when the availability of original training data is constrained. The frequently used methods used for Deep learning-based data augmentation are briefly classified in Fig. 2.

When data is scarce, data augmentation is commonly used in computer vision and medical imaging duties. Data augmentation amplifies the resilience and efficacy of models by incorporating valuable diversity. In general, these algorithms augment datasets without necessitating the acquisition of supplementary data. Throughout this section, we discussed different types of these methods and their variants, discussed their use cases, and explained their steps.

#### 3.2.1. Generative adversarial networks

GANs can be used to artificially generate new data samples that mimic real data. This allows the training data set to be enlarged and improves the model performance. A GAN consists of a generator network that produces synthetic samples and a discriminator network that distinguishes between accurate and false data. As long as the generator cannot outwit the discriminator, the two networks are trained in competition. Models can be better generalized if the training data is supplemented with GAN-generated examples.

**Table 5**
Frequently used datasets and their usage in recent research articles in deep learning-based data augmentation for medical imaging.

| Name | Type | Domain | Feature |
|---|---|---|---|
| MSD dataset [14] | 3d images | Hepatology | The dataset (MSD) contains 2633 3D medical images, including 443 CT scans for the Hepatic Vessel task, offering diverse anatomical perceptions. |
| LC2000 [15] | Image | Pulmonology | There are 25,000 images in the color image dataset (LC25000), distributed across 5 categories. In each class, there are five thousand pictures that represent the histologic entities colon cancer, benign colonic tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. |
| COVID-19 Xray [16] | images | Pulmonology | The data set consists of 5232 radiographs of children, of which 1349 had normal pneumonia, and 3883 had pneumonia (of which 2538 were bacterial and 1345 were viral), for a total of 5856 patients. |
| BT [17] | Image | Neurology | A brain T1-weighted CE-MRI dataset was collected from the General Hospital of Tian-jin Medical University and Nanfang Hospital in Guangzhou, China, between 2005 and 2010. From 233 patients, 1111 slices were collected, including 708 meningiomas, 1426 gliomas, and 930 pituitary tumors. |
| ChestX-ray NIHCC [18] | Images | pulmonology | The ChestX-ray-8 and ChestX-ray-14 datasets include radiographs labeled with disease names (e.g., 30,805 individuals' histories of pneumonia, infiltration, and atelectasis. |
| MIMIC-CXR [19] | Electronic reports-Image-report | Pulmonology | The dataset contains 227,835 chest x-ray imaging exams performed on 65,379 patients who attended the Beth Israel Deaconess Medical Center emergency room from 2011 to 2016. There are 377,110 images in the dataset overall. |
| COVID-19 NY-SBU [20] | Image | Pulmonology | The New York-Stony Brook University dataset contains information on diagnosis, treatments, lab tests, and symptoms for every patient and radiographs of individuals with COVID-19 lung disease. |
| CheXpert [21] | Image | Pulmonology | The dataset contains 224,316 chest radiographs of 65,240 patients. |
| ADNI [22] | Image | Neurology | The Alzheimer's Disease Neuroimaging Initiative dataset includes Alzheimer's disease patients, subjects with mild cognitive impairment, and elderly controls totaling 850,000 cases. |
| HNSCC-TCIA [23] | Image | Pulmonology | This is a multimodal (CT, MR, PT, RT, RTDOSE, RTPLAN, RTSTRUCT) Dataset containing a total of 537942 images from 627 patients. |
| TCGA-TCIA [24] | Image | Neurology | As part of The Cancer Genome Atlas Low-Grade Glioma project, it supports multimodal diagnosis, treatment planning, and follow-up imaging studies on low-grade gliomas. It contains 241,183 de-identified CT and MR images from 219 people. |
| UCI [25] | Image | Oncology | The collection consists of 76 pictures taken from 76 brief colonoscopy clips, showing a variety of lesions and polyps, including 21 hyperplastic lesions, 15 serrated adenomas, and 40 adenomatous polyps. |
| PICCOLO [26] | Image | Oncology | The collection includes 3433 total photos from recordings of 46 human patients undergoing clinical colonoscopy procedures. White light (WL) and narrow-band imaging (NBI) techniques collected pictures of the 76 distinct lesions. |
| CAMUS [27] | Image | Cardiology | The dataset comprises clinical examinations from 500 patients at the University Hospital of St Etienne, France. It was intentionally left unprocessed to reflect real-world clinical variability, including challenging cases. |
| MIT-BIH [28] | time-series | Cardiology | 48 two-channel, half-hour ambulatory ECG recordings from 47 individuals are included in the collection. Digitalized at 360 samples per second per channel, the recordings have an 11-bit resolution. 109,000 tagged beats total, comprising timed, fusion, supraventricular, and ventricular beats. |
| Heart Failure Prediction Dataset [29] | Multi- Modal | Cardiology | The dataset consists of 11 clinical characteristics, including age, blood pressure, cholesterol level, and ECG results. It also includes factors such as type of chest pain and exercise-induced angina. There are a total of 918 samples with a balanced distribution between healthy people (410) and people with heart disease (508). |

*3.2.1.1. Deep convolutional generative adversarial network.* Radford et al. [179] were the first to use a convolutional decoder neural network as the generator in a generative adversarial network. Their architecture uses deep convolutional neural networks for the generator and discriminator models, named 'Deep Convolutional Generative Adversarial Network'. DCGAN introduces a class of CNNs called deep generative convolutional networks for unsupervised learning. The architecture is based on recent advances such as eliminating deterministic spatial pooling, eliminating fully connected layers, and using batch normalization and LeakyReLU activations. The generator uses fractional convolutions to learn upsampling, which enables end-to-end training. The architecture of the discriminator follows a pure convolution based design. These architectural constraints have improved stability during training across different datasets. Experiments show that DCGAN can learn good representations for discriminative tasks - a DCGAN discriminator trained on Imagenet can achieve competitive accuracy on CIFAR-10 classification with as few as 1000 labels. Visualizations of filters and features show that the discriminator learns to recognize objects and scene parts. The generator can learn interesting vector arithmetic properties that allow easy manipulation of visual concepts. Overall, the architectural constraints allow for robust training of deeper CNN GANs, and the visualizations and experiments show interesting, learned representations that suggest applicability as a general-purpose feature learner.

The DCGAN generator starts by sampling a 100-dimensional uniform noise vector Z as input. This Z vector is projected through a dense layer onto a convolutional representation with a small spatial extent and many feature maps (typically 1024), followed by stack normalization and ReLU. A series of four fractional transposition convolutions then upsample this representation to produce a $64 \times 64$ pixel image by upsampling the feature maps. Each transposition convolution halves the number of feature maps and finally produces a 3-channel RGB image. Batch normalization and ReLU activations are used between

**Table 6**
Some commonly used evaluation metrics and their implementation in recent research articles in deep learning-based data augmentation for medical imaging.

| Metrics | Description | Formula | Studies |
|---|---|---|---|
| Accuracy | Although accuracy is a useful measure for assessing the correctness of a model, it can be misleading for unbalanced data sets. In scenarios characterized by an uneven class distribution, a viable approach to achieving high accuracy might involve direct prediction of the majority class. | $\frac{TP+TN}{TP+FP+TN+FN}$ | [30–37] |
| Precision | In situations where the consequences of false positive predictions are significant, the need for precision cannot be overemphasized. In medical testing, a strong emphasis on accuracy is crucial as it guarantees that the positive predictions of the model are very accurate. | $\frac{TP}{TP+FP}$ | [37–42] |
| Sensitivity | Taking sensitivity into account are particularly important when the potential consequences of false negatives are costly. In scenarios such as disease detection, a high recall figure guarantees that the model actually identifies a significant proportion of true positive cases. | $\frac{TP}{TP+FN}$ | [37,39,40, 43–47] |
| F1-score | The F1 score is a metric that effectively combines precision and sensitivity, proving its usefulness in scenarios with an unbalanced class distribution. This is particularly important in scenarios where the impact of false positive and false negative errors is significant. | $\frac{2\ *\ \text{Precision}\ *\ \text{sensitivity}}{\text{Precision}\ +\ \text{sensitivity}}$ | [32,37,40,48, 49] |
| Specificity | The importance of specificity becomes clear in situations where the precise definition of unfavorable events is paramount. In a cybersecurity application, precisely identifying non-malicious behavior is very important. | $\frac{TN}{TN+FP}$ | [45,50,51, 51–55] |
| AUC score | The AUC (Area Under the Curve) is a statistic used to evaluate the performance of a model in binary classification tasks. The area under the curve (AUC) ranges from 0.5 to 1, with a higher value indicating superior performance in differentiating between positive and negative examples. | $\int_0^1 \mathrm{TPR}(\mathrm{fpr}^{-1}(t))\,dt$ | [37,56–65] |

**Table 7**
Some frequently used traditional data augmentation methods and their usage in recent research articles in medical imaging.

| Method | Description | Studies |
|---|---|---|
| Rotation | Images are rotated by various angles (e.g., 90, 180, 270 degrees) to increase viewpoint variability. Useful for object recognition tasks. The initial introduction of the concept may be traced back to the work of [67] as documented in their publication. | [56,68–75] |
| Flip (Horizontal/Vertical) | Mirroring images horizontally or vertically. This is effective for tasks where orientation does not matter. It was first introduced in [76] | [30,56,77–85] |
| Scaling and Resizing | Images are scaled up or down to simulate varying distances or resolutions. This feature was first used in [86] | [43,74,87–93] |
| Translation | Shifting images horizontally and vertically, useful for simulating changes in object position. | [39,56,71,94–100] |
| Brightness and Contrast Adjustment | Altering image brightness and contrast levels to simulate different lighting conditions. | [37,40,78,101–107] |
| Noise Addition | Injecting random noise into images, simulating noisy sensor data or variations in image quality. | [34,35,78,78, 108–113] |
| Cropping | Randomly cropping portions of an image help models focus on relevant details. | [74,77,114–121] |
| Shearing | Distorting the shape of objects in the image to simulate various angles and perspectives. | [87,122–130] |
| Color Space Transformation | Changing the color space of images, for instance, from RGB to grayscale. | [90,123,131–138] |
| Elastic Transformations | Applying deformations to the image can simulate tissue distortions in medical images. | [125,139–148] |
| Cutout | Removing square or rectangular patches from images, encouraging the model to learn from other parts of the image. | [104,135,149–151] |
| Mixup | Combining two or more images and their labels to create new samples, promoting smoother decision boundaries. | [91,152–160] |
| Resampling | altering image dimensions to standardize input for deep learning, often applied in medical and satellite imaging | [34,45,49,161–166] |
| Normalization | Scaling features to a common range (e.g., [0, 1]) for consistent model training. | [167–174] |
| ROI localization | Identifying regions of interest in images for targeted analysis, common in object detection and medical imaging. | [40,45,49,74, 175–177] |

layers, except for the output, which uses tanh. The discriminator takes a $64 \times 64$ image as input. This is passed through a series of layer-by-layer convolutions to reduce the image representation from $64 \times 64$ to $4 \times 4$ spatial dimensions. With each layer-by-layer convolution, the number of feature maps doubles, typically starting with 64 filters and ending with 512 in the final layer. Leaky RLU activations and batch normalization are used between convolutional layers. After the last convolutional layer, the $4 \times 4$ 512-fold feature representation is flattened and fed into a sigmoid output to produce a probability indicating whether the input is real or fake. The pure convolutional architecture, constraints such as stack normalization, and activation decisions enable stable training of the high-resolution deep CNN GAN on complex image datasets. The generator learns upsampling while the discriminator learns downsampling, aided by the (fractionally) stepped convolutions in their CNN stacks.

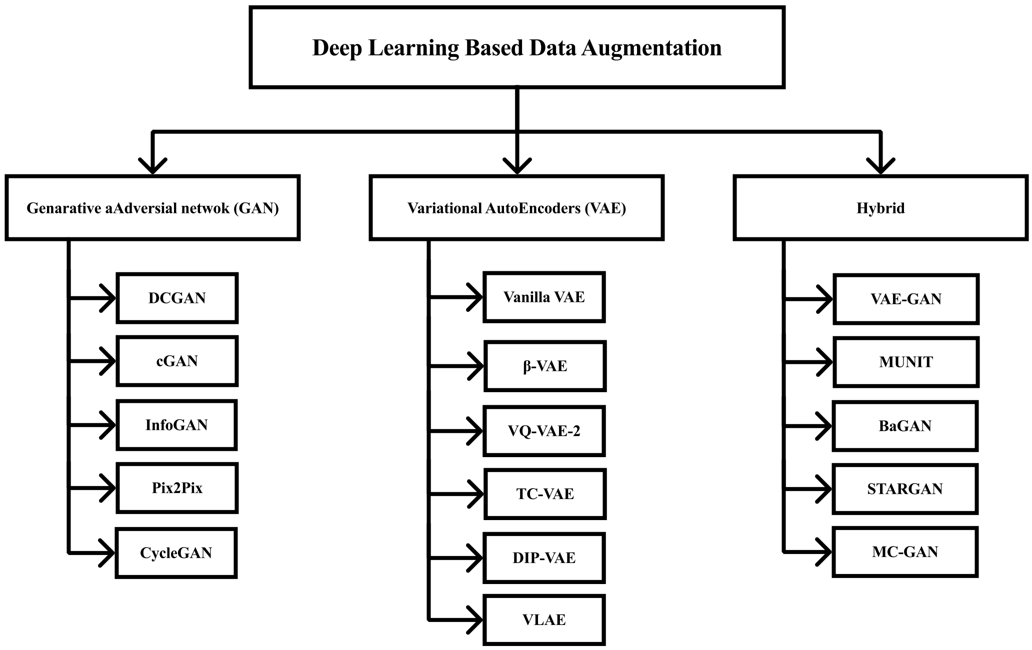


**Fig. 2.** Example taxonomy of deep-learning-based data augmentation algorithms.

*3.2.1.2. Conditional generative adversarial network.* Conditional Generative Adversarial Network (cGAN) are an extension of GANs proposed by Mirza et al. [180] in 2014 to enable conditional generative modeling. In a cGAN, random noise p(z) and conditioning information $y$ are fed to the generator and combined in a common hidden representation. This allows p(z) and $y$ to interact flexibly to generate the output. The conditioning information $y$ can be any auxiliary information such as class labels or data from other modalities. The discriminator is also conditioned on $y$ and the generated/real samples x. The objective function is modified to maximize logD(x|y) for the discriminator and minimize log(1 - D(G(z|y))) for the generator. This framework allows the generation of samples that depend on the additional information y. Mirza and Osindero demonstrate cGANs on MNIST digits conditioned on class labels, where the generator produces digit images based on the specified label. They also demonstrate an application to multimodal modeling by training a cGAN on images and text labels from Flickr to generate descriptive tags for images. Conditioning allows control over the data generation process. cGANs are flexible models that can learn complex multimodal distributions and generate patterns that depend on additional information. Preliminary results illustrate the potential of cGANs for conditional and multimodal generative modeling. This is the loss function of cGAN as seen in Eq. (1)

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

To train a cGAN, a generator, and a discriminator network are first defined. The generator receives both the random noise z and the conditioning data $y$ as input. These are passed through the generator network, which combines them to produce a synthetic output. The discriminator receives natural and synthetic samples along with the conditioning data y. It outputs a probability that the sample is real or fake. The cGAN is adversarially trained by performing alternating gradient descent steps for the generator and the discriminator. The generator attempts to minimize log(1 - D(G(z|y))) to fool the discriminator, while the discriminator attempts to maximize logD(x|y) to classify genuine and spurious samples correctly. After training, the cGAN can be used for conditioned generative modeling by feeding the conditioned input data $y$ and random noise z into the trained generator. The generator then outputs a synthetic sample conditioned on y. Some examples of conditioned generation enabled by cGANs include generating images based on class labels, generating captions based on image features, or generating missing image sections based on surrounding pixels. Through adversarial training, cGANs can learn complex conditional distributions for various generative modeling tasks.

*3.2.1.3. InfoGAN.* InfoGAN is an extension of generative adversarial networks proposed by Chen et al. that can learn interpretable and disentangled representations in a completely unsupervised manner [181]. In InfoGAN, the input noise vector is decomposed into incompressible noise z and latent code c, which targets salient structured semantic features of the data distribution. The InfoGAN target contains the original GAN minimax game target plus an additional term that maximizes the mutual information between the latent codes c and the generator distribution G(z,c). More specifically, it maximizes a lower bound on the variation of the mutual information using an auxiliary distribution Q(c|x). InfoGAN induces the generator to maximize the information about c encoded in G(z,c) by maximizing this mutual information. This information-theoretic regularization causes semantic features in c to appear without any monitoring. InfoGAN can decode variation factors, such as style, lighting, pose, emotion, etc., in various image datasets. In 3D facial images, pose, lighting and other aspects have been decomposed into interpretable latent factors completely unsupervised. GANs can discover meaningful latent representations comparable to supervised methods by simply adding a target to maximize mutual information. InfoGAN provides an approach for unsupervised learning of disentangled representations on complex image datasets.

*3.2.1.4. Pix2pix.* Phillip Isola et al. [182] presents a universal method for performing image-to-image translation tasks using cGAN. This is a pioneering approach where cGANs learn a task-specific loss function, allowing the same framework to be used for different problems without needing to manually evolve a loss function. The work makes a notable contribution to the field of computer vision. The pix2pix model uses a U-grid generator to transport low-level data directly through the grid and a PatchGAN discriminator that only applies penalties to patch-scale structures. This produces sharp, high-resolution results. The objective combines an adversarial loss, which promotes realism, with an L1 loss, which leads to ground truth output. Experiments are conducted to demonstrate the effectiveness of Pix2pix in graphical tasks such as facade generation and aerial image translation, image processing tasks such as semantic segmentation, and artistic applications such as synthesizing sketches into images. The authors analyze various design decisions, such as generator architecture, patch size, and loss functions. Human trials and classifier accuracy are considered to verify the realism of the output. The release of the code has led to creative applications developed by the online community. Overall, pix2pix provides a simple and coherent framework for translating images into pictures with cGANs.

$$G^* = arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \tag{2}$$

Where,

$G^*$ is the optimal generator

$D$ is the discriminator

$L_{cGAN}$ is the conditional GAN adversarial loss

$L_{L1}$ is the L1 reconstruction loss

$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z)))]$

$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]$

$\lambda$ controls the relative weight of the losses

The pix2pix framework follows the standard approach for training cGANs. First, the generator (G) and discriminator (D) networks are initialized. G uses a U-Net architecture to transport low-level information directly through the network. D uses a PatchGAN classifier to distinguish image patches as real or fake. Next, the networks are trained in opposite ways - G to minimize the objective function and D to maximize it. This involves alternating gradient descent steps on D and G. When training G, batch normalization and dropout are used for regularization. The combined objective function includes an adversarial cGAN loss, which promotes realistic outputs, and an L1 loss, which returns outputs to ground truth. Real input–output pairs are passed through D to classify them as real or fake at each iteration. G takes the source image and random noise as input and tries to produce outputs to fool D. After training, G can generate output images for new inputs by a simple forward pass. Multiple losses and architectural decisions allow pix2pix to produce sharp, realistic results with a unified framework for various image-to-image translation problems.

*3.2.1.5. CycleGAN.* Zhu et al. [183] presented CycleGAN, an unpaired training set, as part of an unsupervised learning technique for image-to-image translation between two visual domains, X and Y. It incorporates knowledge about a discriminator F that maps $Y \to X$ and a generator G that maps $X \to Y$. The generated distributions G(X) and F(Y) are made to match the target distributions Y and X with the help of the opponent discriminators DX and DY. G and F are able to translate the domains into each other thanks to this adversarial loss. On the other hand, random mappings between inputs and outputs can occur if only adversarial training is used. CycleGAN includes a cycle consistency loss, which is given by $F(G(x)) \approx x$ and $G(F(y)) \approx y$ to further regulate the mapping functions. This loss utilizes the intuitive belief that the translation should be "cycle consistent" i.e., if we go from X to Y and back again, we should arrive at the original. For the overall goal, Adversarial Loss and Cycle Consistency Loss are combined. Without paired training data, CycleGAN showed convincing qualitative results in studies on the translation of paintings into pictures, the transformation of objects, the translation of seasons, and the translation of styles. It performed better than weight-sharing techniques, pixel-level loss, and feature-level loss, among other unsupervised image translation techniques. Avoiding mode collapses and unrealistic domain mappings was largely dependent on cycle consistency loss. Two generator networks, $G : X \to Y$ and $F : Y \to X$, are present in the model. G tries to generate images G(x) that an adversarial discriminator DY cannot distinguish from the domain Y. In the same way, F attempts to generate images F(y) that DX cannot distinguish from area X.

*3.2.2. Variational autoencoders (VAEs)*

One type of deep generative model useful for data augmentation is variational autoencoders (VAEs). VAEs can generate new sample points in a latent space by learning to encode inputs. A random sample from the latent space leads to different results. Thus, VAEs trained on small data sets can artificially increase sample diversity. The mode collapse of VAE outputs is lower than that of GANs. Adding VAE-generated samples to the actual training data reduces overfitting and increases the robustness of the model. However, VAE outputs may not appear as realistic as GANs. VAEs offer a versatile approach to combining new, credible data.

*3.2.2.1. Vector quantized VAE.* Van den Oord et al. [184] (2019) introduced the autoregressive image generation model, vector-quantized variational autoencoder, as an enhancement to the VQ-VAE framework. It contributes in multiple important ways. It independently models global and local visual information using a hierarchical encoder–decoder structure with multi-scale latent code mappings. Compared to single-scale models, this better represents long-range dependencies. Secondly, it models the accurate distribution more accurately by fitting flexible PixelCNN priors over the discrete latent. Third, it allows for faster sampling and high-resolution image modeling since it operates in the compressed latent space instead of pixel space. Test results show that VQ-VAE-2 avoids mode collapse and produces high-fidelity $256 \times 256$ ImageNet samples with quality comparable to BigGAN. The variety results from using likelihood based training to model every mode. Fast encoding/decoding is also made possible by the straightforward feedforward architecture. Overall, VQ-VAE-2 is an improvement over the original VQ-VAE and earlier autoregressive models since it successfully balances diversity, sample quality, and speed. Two significant developments that were the basis for many other generative models that followed were the hierarchical multi-scale approach and compressed domain modeling. The promise of using strong priors and discrete representations to create realistic and practical picture synthesis is fulfilled by VQ-VAE-2.

VQ-VAE-2 first encodes a picture into a hierarchy of discrete latent coding maps using a vector quantized autoencoder. The encoder network transforms the image into multi-scale feature maps, and each level is vector-quantized into discrete codes. The compressed forms of these codes store both local and global image data. PixelCNN priors are then fitted on the discrete latent to model the distribution. Creating an image involves sampling codes in an autoregressive manner at each level from the priors, conditioning the local details code on the global structure code. Ultimately, the decoder network receives the hierarchical codes and uses the compressed representations to reassemble the image. Working in the discrete latent space enables quick sampling from strong priors and high-resolution modeling.

*3.2.2.2. Total correlation VAE.* Chen et al. [185] proposed a new VAE variant called $\beta$-TCVAE, which penalizes the total correlation (TC) component found in a novel ELBO decomposition. Reducing TC promotes statistical independence among patients, resulting in disentangled representations. Stochastic minibatch training without additional hyperparameters is how $\beta$-TCVAE carries this out. Mutual Information Gap (MIG), a brand-new metric, is presented to assess disentanglement. Test results indicate that $\beta$-TCVAE outperforms VAE and InfoGAN regarding MIG scores and more comprehensible representations. A strong negative association between TC and MIG is confirmed by analysis, supporting their methodology. This study provides a valuable understanding of disentanglement in VAEs and a practical approach using $\beta$-TCVAE and the MIG metric.

Its purpose is to automatically learn disentangled representations. To encourage each dimension of the learned representation to capture independent and significant determinants of variation, it works by reducing the overall correlation between latent variables. The crucial phases include encoding the input data into a probabilistic latent space, breaking down the overall correlation into component parts, and refining the model to lower the overall correlation. In doing so, TC-VAE hopes to better understand and provide control over the latent factors regulating the input data by learning a compact and interpretable representation in which each dimension corresponds to a unique and statistically independent feature.

*3.2.2.3. Disentangled inferred prior VAE.* A variational autoencoder framework called Disentangled Inferred Prior (DIP) Variational Autoencoder is proposed by Kumar et al. [186] for the unsupervised learning of disentangled representations from unlabeled data. Adding a regularizer that compares the moments of the disentangled prior p(z) to the inferred latent distribution q$\phi$(z) promotes disentanglement. This eliminates the conflict between disentanglement and reconstruction quality in earlier techniques like $\beta$-VAE. The SAP metric correlates more strongly with qualitative outcomes than earlier metrics, is another quantification of disentanglement that the research suggests. Experiments on multiple datasets demonstrate that DIP-VAE provides stronger disentanglement with comparable or superior reconstruction quality than VAE and $\beta$-VAE. The main contributions are a more robust disentanglement assessment metric and a principled VAE framework for disentangled representation learning without sacrificing reconstruction.

The main idea behind DIP-VAE is to match a factorized disentangled prior p(z) to the aggregate posterior distribution q$\phi$(z), which promotes disentanglement of the inferred latent representations. To do this, add a regularizer (e.g., decorrelating the dimensions by matching covariances) that matches moments of q$\phi$(z) and p(z). Most importantly, this avoids the disentanglement and reconstruction quality trade-off that approaches like $\beta$-VAE, which upweight the KL divergence term, encounter. DIP-VAE optimizes the additional regularizer and the standard evidence lower bound (ELBO) aim of the VAE. Using the unlabeled data, the system first trains a normal VAE. Next, it adds the moment matching regularizer and aggregates the approximate posteriors over the data distribution to compute q$\phi$(z). Next, we optimize the updated VAE objective to learn disentangled representations while maintaining the quality of the reconstruction. The suggested SAP measure, which has a stronger correlation with the qualitative traversal results, is used to quantify disentanglement.

### 3.2.3. Hybrid algorithms

Hybrid deep learning architectures have proven to be highly efficient in the context of data augmentation and synthesis of medical images. The integration of Generative Adversarial Networks and variational autoencoders has been implemented in hybrid models to solve the limited availability of annotated medical data. This approach enables synthesizing realistic, improved data from datasets with scarce information. Hybrid deep learning models, such as MUNIT, StarGAN, BAGAN, MC-GAN, etc, have shown significant potential in data augmentation and synthesis in medical imaging. The MUNIT system was created to translate unpaired medical pictures [187]. This method separates content and style to create a variety of improved medical visuals. MUNIT combines its content code with a random style code from the destination domain's style space to translate a picture to a new domain. MUNIT requires less assumption than other approaches like CycleGAN. The StarGAN model has demonstrated its ability to translate images into numerous target domains, enabling various augmentation possibilities for medical imaging in different modalities or situations [188]. The main focus of BaGAN (Bayesian GAN) has been the augmentation of MRI data, particularly in developing synthetic brain MRIs containing many abnormalities [189]. This approach aimed to address the limited availability of training data for brain disease classification models. The use of MC-GAN (Multi-Conditioned GAN) led to the synthesis of 3D CT scans containing liver lesions of different sizes and locations [190]. This enhancement of the training data was intended to improve the performance of the lesion detection algorithms. The hybrids discussed in this study address the challenges of limited data availability and diversity in medical imaging by effectively combining realistic enhanced scans and images from different modalities. These models' flexible structures and training methods enable the effective enhancement of medical images, a capability not present in previous GAN models that relied on a single network. The data synthesis capabilities of MUNIT, StarGAN, BAGAN, MC-GAN and other hybrid models have been shown to improve medical imaging algorithms by utilizing a wider range of training data.

## 4. Results analysis

Results must be carefully analyzed in every experiment or research. Good result analysis yields useful information. Trends, correlations, and statistical significance are interpreted. A good analysis links findings to hypotheses and research goals. It finds mistakes and areas for improvement. Careful result analysis turns unclear data into meaningful findings. Thorough analysis and sharing of results build credibility and help others to build on them. Analyzing results ensures experiments are valuable. An automated diagnosis model [191] using a Convolutional Neural Network with ResNet50 architecture was proposed by Sivakumar et al. to detect malignant melanoma cancer. The dataset comprised 3300 images from the International Skin Image Collaboration (ISIC), of which 80% were used for training and 20% for testing. Pre-processing involved resizing the images to 224 × 224 × 3 and applying data augmentation techniques. The evaluation metrics for the test data yielded an accuracy of 94%, an F1 score of 93.9% and a loss of 41.9%, while precision and recall were 81.1% and 80.6%, respectively. The proposed ResNet50 model outperformed the baseline methods, achieving higher accuracy, higher F1 score and lower loss. Statistical analysis confirmed a significant performance improvement over other skin cancer classification models. Priya et al. [192] presented a fetal brain abnormality (DFBA) detection method using a 22-slice CNN model and data augmentation on a dataset of 875 MRI images of the fetal brain. The dataset included 23 types of anomalies, and pre-processing included normalization, reshaping, and data augmentation for balance. The model achieved a test accuracy of 83%, outperforming other methods with higher precision, recall, F1 score, and accuracy. Diagnosis of ophthalmological diseases using deep learning on the STARE dataset, which had been extended to 5965 images, was presented by Almustafa et al. [193] using five models — EfficientNet, InceptionV2, VGG-16, 3-Layers CNN, and ResNet-50. They emphasized pre-processing techniques such as augmentation. EfficientNet outperformed the other models with an accuracy of 98.43% and outperformed them in precision, recall, and F1 score, representing a remarkable advance in the detection of retinal defects.

In Table 8, we further analyzed some other recent state-of-the-art research articles published in this field between 2022–2023.

## 5. Challenges and future research directions

The process of data augmentation poses several challenges in terms of maintaining the properties of the data, achieving diversity, and managing computational costs. When implementing simple augmentation techniques, there is a risk of introducing biases that can compromise the integrity of the data. Generative models also struggle with balancing the specificity of the training set with the diversity of the generated results. All techniques expand data sets and increase training time and resources. A careful balance between bias, diversity gain and efficiency is crucial when expanding to positively impact model performance. It is a balancing act to maintain data properties, achieve sample diversity, and control computational effort.

**Table 8**
Analysis of the experimental results of some state-of-the-art research papers that published in 2022 and 2023.

| Ref. | Application | Dataset | Preprocessing and data augmentation methods | Model | Results | Limitations |
|---|---|---|---|---|---|---|
| Alirr et al. [45] | Hepatology | MSD dataset | Image Resampling, Feature Extraction, Image Augmentation | FCN | DSC: 79%, Sensitivity: 82.2%, Specificity: 95.1% | The study focuses solely on CT scans, using a single dataset (Medical Segmentation Decathlon). It only segments hepatic arteries, not the entire vasculature, and is susceptible to errors at vessel boundaries and branch tips, which can impact surgical planning. Errors in liver segmentation can propagate through the process. |
| Hille et al. [34] | Hepatology | 48 MRI patient cases with 157 hepatic lesions | Adaptive histogram equalization, Z-score normalization, N4 bias correction, Image Augmentation | Hybrid CNN-transformer network (SWTR-Unet) | DSC: 0.98 ± 0.02, HD: 1.02 ± 0.18 mm | The study conveys its narrow focus on a single MRI sequence, its 48-case small dataset, its lack of comparisons with semi-automatic or interactive segmentation techniques, and its uncertainty about performance with different MRI modalities and lesion types. |
| kovacset al. [49] | Urology | 774 prostate MRI exams with biopsy-confirmed lesions | Image Augmentation, Segmentation, Normalization | 3D nnU-Net segmentation model | ROC: 59.92 ± 13.27% , F1 score: 61.64 ± 3.61%, FROC: 59.55 ± 5.97% | The study shows the evaluation of a single institution dataset, uncertainty for unlabeled lesions, potential variation across model architectures, the need for additional organ segmentations, skewed deformation simulations, a lack of computational performance analysis, and the absence of comparisons with other deformation techniques. The evaluation of qualitative realism in the study and clinical workflow integration were highlighted as possible drawbacks. |
| Liet al. [194] | Urology | 300 prostate cancer patients, 19,200 CT slices, 10 patients, 640 CT slices | HU normalization, Resampling, Image Augmentation | U-Net + Variational Autoencoder (VAE) | DSC: 0.90, HD: 5.14 mm, COM: 1.03 mm, NCC: 0.6 | The generalizability of the study is uncertain due to the evaluation of a single internal data set, the limited data between observers, and the organ-specific focus. The effects of repeated outlier detection, 3D networks, and comparisons with newer architectures must be further investigated. |
| Wilsonet al. [195] | Urology | 391 subjects, 1028 total biopsy cores | Resampling, Image Augmentations | VICReg | AUROC: 91% | The generalizability of the study is limited to two clinical centers with 391 patients. Biopsy cores with less cancer involvement were not considered, and ground truth labeling may lack precision. The computational burden and label loss were not assessed. |
| Anayaet al. [71] | Neurology | 110 patients from TCGA-LGG Axial FLAIR MRI images | Resampling, Normalization Augmentations | resNet50 | F1 score: 92% Accuracy: 92% | Limited to 1.5T FLAIR MRI images in one data set, generalization to other datasets and image types is essential. Evaluation of PCA data enhancement with different noise levels and exploration of different network architectures is required. |
| Kim et al. [196] | Neurology | BCI Competition IV 2a and 2b datasets | Normalization, Image Augmentations | ShallowConvNet, DeepConvNet, EEGNet | Accuracy: 0.863 ± 0.003% | This study was limited to two public datasets and motor imaging tasks, where conventional methods faced challenges in optimizing the hyperparameters. The optimal pruning ratios of the temporal pruning strategy and possible overfitting for small datasets were not investigated. Real-world validation is required. |

(continued on next page)

**Table 8** (*continued*).

| Ref. | Application | Dataset | Preprocessing and data augmentation methods | Model | Results | Limitations |
|---|---|---|---|---|---|---|
| Shamrat et al. [197] | Neurology | ADNI | Data Format Conversion, Image Augmentation, | Inception v3 | Accuracy: 98.68%, Precision: 98.68%, Recall: 98.68%, Specificity: 99.74% F1-score: 98.68% | The study's limited data set, which focused exclusively on T2-weighted MRI scans, may not be transferable to other modalities. Only classification accuracy was assessed. Additional metrics were not considered. Modeling biases and the need for further validation with real clinical data sets remain. |
| Mishra et al. [198] | Neurology | Brain MRI 2-class dataset Brain MRI 4-class dataset | Color Space Conversion, Resampling, | SSCLNet | Accuracy: 72 ± 3% | The study focuses exclusively on contrastive self-supervised learning and does not include comparisons with other pre-training methods. It uses a small dataset with limited clinical validation and mainly examines accuracy metrics. |
| Tripathy et al. [199] | Neurology | Brain tumor detection 2020 dataset from Kaggle | Color Space Conversion, Resampling, Image Augmentations | EfficientNet-B2, EfficientNet-B3, EfficientNet-B4 | Accuracy: EfficientNet-B2: 100%, EfficientNet-B3: 99%, EfficientNet-B4: 99.5% | The promising results of the study with a small data set need to be validated using more extensive and more diverse clinical data. The lack of explainability of the model and the limited evaluation metrics require further analyses and testing. |
| Shen et al. [200] | Pulmonology | Chest X-ray images from collaborative hospitals | Image Augmentation | GAN, CNN | MAE: 0.0296, PSNR: 28.9815, SSIM: 0.7555 | The study does not explore using additional nodule characteristics for augmentation. It also faces challenges in quantifying image quality and altering nodule properties. The method's potential applications beyond augmentation are unexplored, highlighting the need to enhance controllability and generalizability. |
| Van et al. [201] | Pulmonology | MIMIC-CXR, CheXpert | Image Resampling, | PubMedCLIP | AUC: 0.82 ± 0.3 | The use of domain-specific data to fine-tune the general CLIP model has not been studied. The method's low transferability to other datasets underscores the necessity of expanding its application to other medical imaging modalities to improve generalizability. |
| Obayya et al. [202] | Pulmonology | LC25000 | noise removal | GhostNet TSA | Accuracy: 99.33%, Precision: 98.28%, Recall: 98.26%, F1-score: 98.27%, AUC: 98.97% | The paper is limited to lung and colon cancer detection. Extending the methodology to encompass additional forms of cancer may corroborate its efficacy. There was just one dataset used for evaluation. Therefore, it would be helpful to test on more datasets. The model's computational complexity was not examined. |
| Gardner et al. [203] | Pulmonology | The Cancer Imaging Archive (TCIA) HNSCC | Image Augmentation | cGAN | Mean centroid error: 1.1 ± 0.8 mm, Mean DSC: 0.90 ± 0.03, Mean MSD: 1.6 ± 0.5 mm | Small dataset from two clinical centers (391 patients) limits generalizability. Biopsy cores with over 40% cancer involvement were used; performance on low involvement cores is unclear. More advanced neural network models could be tested for better architectures. |
| Grama et al. [204] | Pulmonology | 6 lung cancer patients with small tumors | Image augmentations, Image synthesis, Image Segmentation | Siamese neural network | MAD: 0.66 ± 0.6 mm, RMSE: 0.89 mm, Max error: 4.02 mm | The study employs a small dataset of six patients and one phantom. Tumor tracking accuracy is assessed indirectly, limiting it to 2D tracking, and real-time clinical integration and 3D localization validation are required. |

(*continued on next page*)

**Table 8** (*continued*).

| Ref. | Application | Dataset | Preprocessing and data augmentation methods | Model | Results | Limitations |
|---|---|---|---|---|---|---|
| Younaset al. [205] | Oncology | UCI, PICCOLO | Image resampling, Image augmentations | GoogLeNet, ResNet50, Xception, DenseNet-201, SqueezeNet | Accuracy: 96.3%, Precision: 95.5%, Recall: 97.2%, F1-score: 96.3% | The study's reliance on a small UCI dataset of 76 photos limits the deep learning models' performance and generalizability due to inadequate training data. Limited model exploration and lack of comparisons with advanced networks pose challenges. Annotated data scarcity remains a key obstacle. |
| Liu et al. [206] | Oncology | BreaKHis, IDC, UCSB | Image augmentation, Normalization | AlexNet, ImageNet | Accuracy: 85.7% | The suggested model only performs binary cancer classification without tumor localization, and the study employed just one dataset for development and testing. |
| Moreno et al. [109] | Oncology | TCGA | feature selection | ModCGAN | Accuracy: 85% | Limited to 18 gene expression datasets, no in-depth assessment of synthetic data, minimal model comparisons (CGAN and ModCGAN only) and limited to classification tasks without exploring other bioinformatic analyses. |
| Ham et al. [207] | Oncology | Gastroscopy images from 158 patients | Image Resampling, Intensity Normalization, Grad-CAM | Xception | Accuracy: 79.9%, Sensitivity: 76.8%, Specificity: 78.2%, F1-Score: 79.2%, AUC 88.5% | The study used a small data set from a single medical center. Only white light endoscopy images were included, not narrow band images. The image enhancement method is simple, without advanced techniques such as GANs. The evaluation is limited to classification metrics. |
| Caselles et al. [208] | Oncology | 204 CESM images | Data augmentations | U-net with Resnet block | Dice: +10.2% | Small dataset from a single medical center with only 204 exams, limited to only breast lesions in contrast-enhanced mammography, Basic U-Net architecture used for segmentation, not state-of-the-art model, limited evaluation to only Dice metric for segmentation performance. |
| Ling et al. [209] | Cardiology | 45 patient apical 3-chamber view color Doppler echocardiography sequences (1338 aliased frames, 2379 non-aliased frames) | Extracted Doppler velocity and power data, normalization, plane conversion | nnUNet | Cosine similarity: 0.99, Accuracy: 0.96, Recall: 0.91, Precision: 0.89 | The low color Doppler frame rate (10–15 fps) prevented them from utilizing temporal information. Although they focused on single aliasing artifacts, certain valvular illnesses can cause multiple aliasing. There were proprietary/masked settings in the clinical ultrasonography system used. The performance in very turbulent flows like valvular jets is unknown. |
| Chao et al. [210] | Cardiology | Apical 4-chamber (A4C) echocardiographic view of 381 patients | Video frame sampling, color space conversion, basic data augmentation | ResNet50 | AUC: 0.97, Precision: >90%, Recall: >90%, F1-score: >90% | The limited sample size for this uncommon ailment poses hazards of overfitting and absence of external validation. The frame based technique may inadequately capture significant temporal correlations in comparison to video. The frame based approach may not fully capture temporal relationships compared to video clips. |

### 5.1. Clarity of the impact of data augmentation

The lack of clarity in this context is primarily due to the intricate nature of data augmentation techniques. These techniques involve various methods and processes to improve the quality and utility of data, but pinpointing the exact techniques that have the most significant impact on enhancing outcomes can be challenging. The complexity arises from the interplay of multiple factors and methods, making it challenging to isolate the most influential ones in the context of improving results or achieving specific objectives. Ablation studies and contribution analyses can be used to systematically dissect the impact of each enrichment technique. By understanding the specific contributions, researchers can optimize the selection of augmentation methods based on the characteristics of different medical imaging tasks.

### 5.2. Evaluation metrics consensus

The lack of common evaluation criteria leads to uncertainty in assessing the effectiveness of different techniques. Without standardized metrics, it becomes difficult to objectively compare and measure the impact of different augmentation methods on the quality and outcomes of the data. To address this challenge, researchers can explore and establish quantitative metrics and qualitative measures that comprehensively assess the quality and variety of augmented medical images. This may include developing consensus on metrics tailored to the specific nuances of medical imaging tasks.

### 5.3. Generative adversarial networks mode collapsing

GANs encounter problems such as "mode collapsing", where the generator produces limited patterns, limiting their usefulness in medical imaging. This phenomenon limits the diversity and richness of the data generated, potentially affecting the GAN's ability to accurately represent the complexity and variability of medical images. Investigating constraints and regularizers for GANs can mitigate mode collapsing. By imposing specific constraints during training, researchers can increase the diversity of patterns, making GANs more reliable in producing diverse and realistic medical images.

### 5.4. Modality-specific augmentation

Modality-specific enhancements are essential in medical imaging because of the different anatomical and pathological characteristics of the various imaging modalities. For example, an MRI of the brain differs significantly from other modalities, such as X-ray or CT scans, regarding tissue contrast, resolution, and the type of abnormalities detected. This ensures that machine learning models are robust and well adapted to the nuances of the medical data they encounter, ultimately leading to more accurate diagnoses and treatment recommendations. Researchers can explore and develop augmentation techniques tailored to the unique characteristics of each modality. This requires understanding the specific challenges of different modalities and developing augmentation methods that effectively address these challenges.

### 5.5. Limited labeled data for specific diseases

The challenge arises from the limited availability of labeled data to train models tailored to specific diseases. It can be challenging to obtain a sufficient amount of accurately labeled data for rare or specific diseases. This scarcity hinders the development of effective deep-learning models for accurate disease detection and diagnosis and highlights the importance of data collection and curation in healthcare. The widespread adoption and implementation of federated learning algorithms in various domains is still limited. Efficient annotation interfaces and the integration of pre-trained models offer potential solutions to the problem of limited labeled data. Streamlined annotation processes and already trained models can optimize available annotations. Developing semi-supervised and weakly supervised learning techniques also allows models to use the labeled data better and improve their performance even in data scarcity scenarios. These strategies contribute to more efficient and effective machine learning in healthcare applications.

### 5.6. Complexity of hyperparameter tuning

Hyperparameter tuning is an empirical and complicated process without a standardized methodology, making it difficult to determine the ideal model configurations. A trial-and-error approach is often required, and the effectiveness of specific hyperparameters may vary across different tasks and datasets. This complexity makes it challenging to find universally optimal settings. Therefore, careful experimentation and expertise are required to fine-tune machine learning models for specific applications. Deriving generalized principles for hyperparameter optimization through comparative studies can streamline the tuning process. Research can focus on developing methods that apply to different medical imaging tasks and provide greater consistency in model performance.

### 5.7. Opacity and computational overhead in hybrid deep learning structures

Hybrid Deep Learning models, particularly those that integrate convolutional neural networks (CNNs) exhibit improved performance in specific scenarios. However, their complicated structures present challenges for understanding and fine-tuning parameters. The opaqueness of these models makes them difficult to interpret, which prevents researchers and practitioners from optimizing and adapting them for different applications. In addition, the computational requirements of hybrid models, which increase with increasing image size, require high-performance hardware, limiting accessibility and efficiency. Future research in hybrid Deep Learning should focus on improving interpretability through attention mechanisms and explainable AI techniques. Optimizing computational efficiency is critical, and advances in algorithms, model compression, and hardware innovations should be explored. Adapting hybrid Deep Learning to specific applications is needed, with domain-specific architectures to be explored for balancing performance and interpretability. Joint research initiatives between academia and industry can accelerate progress and foster a collective understanding and broad adoption of hybrid Deep Learning models.

### 5.8. Temporal dynamics gap

The temporal Dynamics Gap is the gap in generative models primarily designed to represent single images. Current models, such as those using LSTM blocks and 3D convolution, show promise for generating video and sequential images, but their application in the medical domain for evolving datasets has not been adequately explored. The challenge is to adapt these models to capture nuanced temporal changes and states across a range of medical images, which hinders their broader applicability. Future research may extend generative models to capture dynamic states over time, similar to video and sequence generation. Building on the foundations of LSTM blocks and 3D convolution, further exploration holds significant potential for advancing medical applications and enabling a more comprehensive understanding of evolving states and temporal changes in medical data.

### 5.9. Model for understanding accountability

Complex deep learning models are often considered 'black boxes', meaning their internal workings are difficult to interpret or explain. This lack of transparency and interpretability can be a significant limitation, especially in critical applications such as healthcare, where understanding the model's decision-making process and providing accountability is essential. Addressing this challenge is critical to ensuring that AI systems in medical contexts are trustworthy, safe, and ethical. Developing techniques for model introspection involves developing methods for deciphering the decision-making processes in deep learning models. Visualization schemes can provide incremental insights, improve model transparency, and build trust in the medical community.

### 5.10. Addressing class imbalances

The imbalance of classes in disease datasets can challenge model training, especially when the minority class is underrepresented. This imbalance can cause the models to be biased towards the majority class and may have lower sensitivity in detecting the disease in question. Dealing with imbalanced classes often requires special techniques such as oversampling, undersampling, or cost-sensitive learning to ensure accurate and balanced model performance. Developing advanced data augmentation techniques and sampling strategies specifically designed to address class imbalances will ensure that models are effectively trained across disease categories, improving overall diagnostic performance.

### 5.11. Multimodal data integration

Integrating multimodal data is challenging due to the various modalities' different data formats, scales, and information content. Combining data from sources such as medical images, clinical records, and genetic data requires addressing data representation and quality differences. Addressing these challenges is critical to developing comprehensive and informative healthcare models that use multiple data sources effectively. Accounting for these inter-modality differences is difficult. Designing neural network architectures that can effectively merge multimodal medical data involves developing models capable of extracting meaningful features from multiple data sources. This can improve the holistic understanding of medical conditions by integrating complementary information from different imaging modalities.

### 5.12. Explainability of models

The lack of explainability and interpretability of deep learning models is an obstacle to their use in medical applications. Understanding the decision-making process of these complex models is critical to ensuring patient safety and gaining the trust of medical professionals. Developing methods for model interpretation is critical to closing this gap and unlocking the full potential of Deep Learning in healthcare. Incorporating techniques such as attention mechanisms, concept vectors, and gradient-based methods can improve the explainability of models. These methods provide insight into the decision-making processes of the model, making it more interpretable by medical professionals and providing confidence in the predictions.

### 5.13. Model bias and fairness

Model bias leading to unfair predictions is a major problem, especially in healthcare applications where unbiased and equitable decisions are paramount. Biased models can lead to inequalities in patient care and outcomes, which is unacceptable in the medical field. Addressing and mitigating model bias through rigorous testing and fairness-aware algorithms are critical to ensuring equitable and fair decisions in healthcare. Investigating and mitigating biases in model predictions requires the development of mitigation techniques. These techniques should be tailored to the specific challenges of medical data to ensure fair and unbiased results for different patient populations.

### 5.14. Computational efficiency for 3D images

The analysis of 3D medical images often involves high computational costs, which can hinder the scalability and wider application of such models. Processing volumetric data requires significant computing resources that are not readily available in all healthcare settings. To overcome this challenge, algorithms must be optimized, and efficient hardware must be deployed to make 3D medical image analysis more accessible and cost-effective. Researching efficient model compression and pruning techniques is critical to reducing computational overhead. This includes developing methods that preserve the essential information in 3D medical images while optimizing the computational efficiency of the models, making them more amenable to practical applications.

### 5.15. Conditional image generation

Conditional image generation for medical images is difficult since clinical relevance must be maintained. Throughout augmentation, medical images must retain their anatomical structures, pathogenic traits, and modality characteristics. Keeping these crucial characteristics while adding variability for model training in generated images is challenging. The semantics of medical picture data must be understood and respected to avoid biases that could jeopardize the clinical accuracy of the supplemented dataset. To facilitate practical deep-learning model training for medical applications, approaches must balance variety and accuracy. The challenge can be solved using advanced methods that capture and improve clinical relevance. Exploring advanced conditional augmentation methods that account for anatomical components, pathology, and imaging modalities is critical. This includes creating more interpretable models and using domain-specific information to assure clinical accuracy in generated images. To overcome healthcare deep learning problems, exploring adaptive techniques that dynamically modify conditional generation to medical data complexity is essential.

## 6. Conclusions

Integrating deep learning-based data augmentation is a pivotal strategy in enhancing the efficacy of medical image analysis and diagnostic models. Its ability to generate diverse and realistic variations in medical datasets bolsters deep learning algorithms' robustness and is paramount in advancing precision medicine and healthcare outcomes. Despite the scarcity of research on recent advancements in deep learning-based data augmentation, this paper conducts a systematic review to understand the subject comprehensively within diverse medical application domains. The study also explores prevalent datasets, preprocessing techniques, and evaluation metrics, emphasizing recent developments. It assesses popular deep learning algorithms and conducts a detailed analysis of recent state-of-the-art research. Furthermore, the paper addresses challenges associated with deep learning-based data augmentation and outlines future research directions, positioning itself as a valuable resource for interdisciplinary studies and fostering innovation in this rapidly evolving field.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. M.F. Mridha reports was provided by American International University Bangladesh. Dr. M.F. Mridha reports a relationship with American International University Bangladesh that includes: employment.

## Data availability

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

## Acknowledgment

The authors would like to thank the Advanced Machine Intelligence Research Lab - AMIR Lab for Supervision and Resources.

## References

[1] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[2] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, 2017, arXiv preprint arXiv:1712.04621Baad.
[3] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, A. Haworth, A review of medical image data augmentation techniques for deep learning applications, J. Med. Imaging Radiat. Oncol. 65 (5) (2021) 545–563.
[4] J. Lemley, S. Bazrafkan, P. Corcoran, Smart augmentation learning an optimal data augmentation strategy, IEEE Access 5 (2017) 5858–5869.
[5] A.J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, C. Ré, Learning to compose domain-specific transformations for data augmentation, Adv. Neural Inf. Process. Syst. 30 (2017).
[6] Y.-D. Zhang, Z. Dong, S.-H. Wang, X. Yu, X. Yao, Q. Zhou, H. Hu, M. Li, C. Jiménez-Mesa, J. Ramirez, et al., Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation, Inf. Fusion 64 (2020) 149–187.
[7] S. Wang, M.E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J.M. Gorriz, et al., Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects, Inf. Fusion 76 (2021) 376–421.

[8] E. Goceri, Medical image data augmentation: techniques, comparisons and interpretations, Artif. Intell. Rev. (2023) 1–45.
[9] P. Celard, E. Iglesias, J. Sorribes-Fdez, R. Romero, A.S. Vieira, L. Borrajo, A survey on deep learning applied to medical images: from simple artificial neural networks to generative models, Neural Comput. Appl. 35 (3) (2023) 2291–2323.
[10] M. Xu, S. Yoon, A. Fuentes, D.S. Park, A comprehensive survey of image augmentation techniques for deep learning, Pattern Recognit. (2023) 109347.
[11] A. Mumuni, F. Mumuni, Data augmentation: A comprehensive survey of modern approaches, Array (2022) 100258.
[12] N.E. Khalifa, M. Loey, S. Mirjalili, A comprehensive survey of recent trends in deep learning for digital images augmentation, Artif. Intell. Rev. (2022) 1–27.
[13] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, et al., The prisma 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.
[14] A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, et al., A large annotated medical image dataset for the development and evaluation of segmentation algorithms, 2019, arXiv preprint arXiv:1902.09063.
[15] Andrew A. Borkowski, Marilyn M. Bui, L. Brannon Thomas, Catherine P. Wilson, Lauren A. DeLand, Stephen M. Mastorides, Lung and colon cancer histopathological image dataset (lc25000), 2019, arXiv preprint arXiv:1912.12142.
[16] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, cell 172 (5) (2018) 1122–1131.
[17] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng, Enhanced performance of brain tumor classification via tumor region augmentation and partition, PLoS One 10 (10) (2015) e0140381.
[18] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.
[19] A.E. Johnson, T.J. Pollard, S.J. Berkowitz, N.R. Greenbaum, M.P. Lungren, C. y. Deng, R.G. Mark, S. Horng, MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports, Sci. Data 6 (1) (2019) 317.
[20] J. Saltz, M. Saltz, P. Prasanna, R. Moffitt, J. Hajagos, E. Bremer, J. Balsamo, T. Kurc, Stony brook university covid-19 positive cases [data set]. the cancer imaging archive, 2021.
[21] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 590–597.
[22] ADNI — Alzheimerś Disease neuroimaging initiative — adni.loni.usc.edu, 2020, https://adni.loni.usc.edu/.
[23] A. Grossberg, H. Elhalawani, A. Mohamed, S. Mulder, B. Williams, A.L. White, J. Zafereo, A.J. Wong, J.E. Berends, S. AboHashem, J.M. Aymard, A. Kanwar, S. Perni, C.D. Rock, S. Chamchod, M. Kantor, T. Browne, K. Hutcheson, G.B. Gunn, S.J. Frank, D.I. Rosenthal, A.S. Garden, C.D. Fuller, M.D. Anderson cancer center head and neck quantitative imaging working group, 2020, http://dx.doi.org/10.7937/k9/tcia.2020.a8sh-7363, HNSCC - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki — doi.org.
[24] N. Pedano, A.E. Flanders, L. Scarpace, T. Mikkelsen, J.M. Eschbacher, B. Hermes, V. Sisneros, J. Barnholtz-Sloan, Q. Ostrom, The cancer genome atlas low grade glioma collection (TCGA-LGG) - the cancer imaging archive (TCIA) public access - cancer imaging archive Wiki — doi.org, 2016, http://dx.doi.org/10.7937/K9/TCIA.2016.L4LTD3TK.
[25] Colonoscopic dataset — depeca.uah.es, 2016, http://www.depeca.uah.es/colonoscopy/.
[26] Basque biobank - PICCOLO RGB/NBI (WIDEFIELD) image collection — biobancovasco.bioef.eus, 2020, https://www.biobancovasco.bioef.eus/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html. (Accessed 18 December 2023).
[27] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A.R. Berg, P.-M. Jodoin, T. Grenier, et al., Deep learning for segmentation using an open large-scale dataset in 2d echocardiography, IEEE Trans. Med. Imaging 38 (9) (2019) 2198–2210.
[28] G.B. Moody, R.G. Mark, The impact of the mit-bih arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[29] Heart failure prediction dataset — kaggle.com, 2021, https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction, (Accessed 29 October 2023).
[30] M. Chhabra, R. Kumar, An advanced vgg16 architecture-based deep learning model to detect pneumonia from medical images, in: Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021, Springer, 2022, pp. 457–471.
[31] J. Meng, Z. Tan, Y. Yu, P. Wang, S. Liu, Tl-med: A two-stage transfer learning recognition model for medical images of covid-19, Biocybern. Biomed. Eng. 42 (3) (2022) 842–855.
[32] D.M. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, 2020, arXiv preprint arXiv:2010.16061.
[33] P. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (10) (2012) 78–87.
[34] G. Hille, S. Agrawal, P. Tummala, C. Wybranski, M. Pech, A. Surov, S. Saalfeld, Joint liver and hepatic lesion segmentation in MRI using a hybrid CNN with transformer layers, Comput. Methods Programs Biomed. (2023) 107647.
[35] W. El-Shafai, A.A. Mahmoud, A.M. Ali, E.-S.M. El-Rabaie, T.E. Taha, A.S. El-Fishawy, O. Zahran, F.E.A. El-Samie, Efficient classification of different medical image multimodalities based on simple CNN architecture and augmentation algorithms, J. Opt. (2023) 1–13.
[36] Y. Hu, Z. Zhong, R. Wang, H. Liu, Z. Tan, W.-S. Zheng, Data augmentation in logit space for medical image classification with limited training data, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part V, vol. 24, Springer, 2021, pp. 469–479.
[37] M.M.A. Monshi, J. Poon, V. Chung, F.M. Monshi, Covidxraynet: Optimizing data augmentation and CNN hyperparameters for improved covid-19 detection from cxr, Comput. Biol. Med. 133 (2021) 104375.
[38] M. Ajmal, M.A. Khan, T. Akram, A. Alqahtani, M. Alhaisoni, A. Armghan, S.A. Althubiti, F. Alenezi, Bf2sknet: Best deep learning features fusion-assisted framework for multiclass skin lesion classification, Neural Comput. Appl. (2022) 1–17.
[39] M. Platscher, J. Zopes, C. Federau, Image translation for medical image generation: Ischemic stroke lesion segmentation, Biomed. Signal Process. Control 72 (2022) 103283.
[40] Q. Guan, Y. Chen, Z. Wei, A.A. Heidari, H. Hu, X.-H. Yang, J. Zheng, Q. Zhou, H. Chen, F. Chen, Medical image augmentation for lesion detection using a texture-constrained multichannel progressive gan, Comput. Biol. Med. 145 (2022) 105444.
[41] M. Sajjad, F. Ramzan, M.U.G. Khan, A. Rehman, M. Kolivand, S.M. Fati, S.A. Bahaj, Deep convolutional generative adversarial network for Alzheimer's disease classification using positron emission tomography (pet) and synthetic data augmentation, Microsc. Res. Tech. 84 (12) (2021) 3023–3034.
[42] I. Chouat, A. Echtioui, R. Khemakhem, W. Zouch, M. Ghorbel, A.B. Hamida, Covid-19 detection in CT and CXR images using deep learning models, Biogerontology 23 (1) (2022) 65–84.
[43] Y. Zhang, Q. Wang, B. Hu, Minimalgan: diverse medical image synthesis for data augmentation using minimal training data, Appl. Intell. 53 (4) (2023) 3899–3916.
[44] M. Rana, M. Bhushan, Machine learning and deep learning approach for medical image analysis: diagnosis to detection, Multimedia Tools Appl. 82 (17) (2023) 26731–26769.
[45] O.I. Alirr, A.A.A. Rahni, Hepatic vessels segmentation using deep learning and preprocessing enhancement, J. Appl. Clin. Med. Phys. 24 (5) (2023) e13966.
[46] L. Gaur, U. Bhatia, N. Jhanjhi, G. Muhammad, M. Masud, Medical image-based detection of covid-19 using deep convolution neural networks, Multimedia Syst. 29 (3) (2023) 1729–1738.
[47] M. Mishra, U.C. Pati, A classification framework for autism spectrum disorder detection using smri: Optimizer based ensemble of deep convolution neural network with on-the-fly data augmentation, Biomed. Signal Process. Control 84 (2023) 104686.
[48] F.B. Mofrad, G. Valizadeh, Densenet-based transfer learning for lv shape classification: Introducing a novel information fusion and data augmentation using statistical shape/color modeling, Expert Syst. Appl. 213 (2023) 119261.
[49] B. Kovacs, N. Netzer, M. Baumgartner, C. Eith, D. Bounias, C. Meinzer, P.F. Jäger, K.S. Zhang, R. Floca, A. Schrader, et al., Anatomy-informed data augmentation for enhanced prostate cancer detection, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 531–540.
[50] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, A. Liotta, Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, Comput. Methods Programs Biomed. 223 (2022) 106951.
[51] S. Jain, G. Seth, A. Paruthi, U. Soni, G. Kumar, Synthetic data augmentation for surface defect detection and classification using deep learning, J. Intell. Manuf. (2022) 1–14.
[52] Y. Zhang, Z. Wang, J. Zhang, C. Wang, Y. Wang, H. Chen, L. Shan, J. Huo, J. Gu, X. Ma, Deep learning model for classifying endometrial lesions, J. Transl. Med. 19 (2021) 1–13.
[53] Y. Deng, L. Lu, L. Aponte, A.M. Angelidi, V. Novak, G.E. Karniadakis, C.S. Mantzoros, Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients, NPJ Digit. Med. 4 (1) (2021) 109.
[54] R. Li, L. Wang, P. Suganthan, O. Sourina, Sample-based data augmentation based on electroencephalogram intrinsic characteristics, IEEE J. Biomed. Health Inf. 26 (10) (2022) 4996–5003.
[55] C. Ding, R. Xiao, D.H. Do, D.S. Lee, R.J. Lee, S. Kalantarian, X. Hu, Log-spectral matching gan: Ppg-based atrial fibrillation detection can be enhanced by GAN-based data augmentation with integration of spectral loss, IEEE J. Biomed. Health Inf. 27 (3) (2023) 1331–1341.
[56] K. Üreten, H.H. Maraş, Automated classification of rheumatoid arthritis, osteoarthritis, and normal hand radiographs with deep learning methods, J. Digit. Imaging 35 (2) (2022) 193–199.

[57] D.R. Nayak, N. Padhy, P.K. Mallick, A. Singh, A deep autoencoder approach for detection of brain tumor images, Comput. Electr. Eng. 102 (2022) 108238.
[58] Y. Zhang, Z. Wang, Z. Zhang, J. Liu, Y. Feng, L. Wee, A. Dekker, Q. Chen, A. Traverso, Gan-based one dimensional medical data augmentation, Soft Comput. (2023) 1–11.
[59] S.K. Mahanta, D. Kaushik, H. Van Truong, S. Jain, K. Guha, Covid-19 diagnosis from cough acoustics using convnets and data augmentation, in: 2021 First International Conference on Advances in Computing and Future Communication Technologies, ICACFCT, IEEE, 2021, pp. 33–38.
[60] S. Cohen, N. Dagan, N. Cohen-Inger, D. Ofer, L. Rokach, Icu survival prediction incorporating test-time augmentation to improve the accuracy of ensemble-based models, IEEE Access 9 (2021) 91584–91592.
[61] R.G. Poola, L. Pl, et al., Covid-19 diagnosis: A comprehensive review of pre-trained deep learning models based on feature extraction algorithm, Results Eng. (2023) 101020.
[62] A.M. Al-Ssulami, R.S. Alsorori, A.M. Azmi, H. Aboalsamh, Improving coronary heart disease prediction through machine learning and an innovative data augmentation technique, Cogn. Comput. (2023) 1–16.
[63] A.u. Haq, J.P. Li, R. Kumar, Z. Ali, I. Khan, M.I. Uddin, B.L.Y. Agbley, Mcnn: a multi-level CNN model for the classification of brain tumors in iot-healthcare system, J. Ambient Intell. Humaniz. Comput. 14 (5) (2023) 4695–4706.
[64] A. Sharma, R. Kumar, P. Garg, Deep learning-based prediction model for diagnosing gastrointestinal diseases using endoscopy images, Int. J. Med. Inform. 177 (2023) 105142.
[65] H.-C. Park, I.-P. Hong, S. Poudel, C. Choi, Data augmentation based on generative adversarial networks for endoscopic image classification, IEEE Access (2023).
[66] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[67] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, 2012, arXiv preprint arXiv:1207.0580.
[68] G. Zhang, H. Dang, Y. Xu, Epistemic and aleatoric uncertainties reduction with rotation variation for medical image segmentation with convnets, SN Appl. Sci. 4 (2) (2022) 56.
[69] S. Shukla, L. Birla, A.K. Gupta, P. Gupta, Trustworthy medical image segmentation with improved performance for in-distribution samples, Neural Netw. 166 (2023) 127–136.
[70] A.u. Haq, J.P. Li, S. Khan, M.A. Alshara, R.M. Alotaibi, C. Mawuli, Dacbt: Deep learning approach for classification of brain tumors using MRI data in iot healthcare environment, Sci. Rep. 12 (1) (2022) 15331.
[71] A. Anaya-Isaza, L. Mera-Jiménez, Data augmentation and transfer learning for brain tumor detection in magnetic resonance imaging, IEEE Access 10 (2022) 23217–23233.
[72] X. Li, X. Hu, X. Qi, L. Yu, W. Zhao, P.-A. Heng, L. Xing, Rotation-oriented collaborative self-supervised learning for retinal disease diagnosis, IEEE Trans. Med. Imaging 40 (9) (2021) 2284–2294.
[73] S. Motamed, P. Rogalla, F. Khalvati, Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images, Inf. Med. Unlocked 27 (2021) 100779.
[74] T. Sadad, A.R. Khan, A. Hussain, U. Tariq, S.M. Fati, S.A. Bahaj, A. Munir, Internet of medical things embedding deep learning with data augmentation for mammogram density classification, Microsc. Res. Tech. 84 (9) (2021) 2186–2194.
[75] A. Rasheed, A.I. Umar, S.H. Shirazi, Z. Khan, S. Nawaz, M. Shahzad, Automatic eczema classification in clinical images based on hybrid deep neural network, Comput. Biol. Med. 147 (2022) 105807.
[76] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[77] R. Raj, J. Mathew, S.K. Kannath, J. Rajan, Crossover based technique for data augmentation, Comput. Methods Programs Biomed. 218 (2022) 106716.
[78] S. AbuSalim, S.A. Mostafa, N. Zakaria, S.J. Abdulkadir, N. Mokhtar, Data augmentation on intra-oral images using image manipulation techniques, in: 2022 International Conference on Digital Transformation and Intelligence, ICDI, IEEE, 2022, pp. 117–120.
[79] T. Nemoto, N. Futakami, E. Kunieda, M. Yagi, A. Takeda, T. Akiba, E. Mutu, N. Shigematsu, Effects of sample size and data augmentation on u-net-based automatic segmentation of various organs, Radiol. Phys. Technol. 14 (2021) 318–327.
[80] S.R. Syed, S.D. Ma, A diagnosis model for detection and classification of diabetic retinopathy using deep learning, Netw. Model. Anal. Health Inf. Bioinformatics 12 (1) (2023) 37.
[81] K. Dabre, S.L. Varma, P.B. Patil, Rapid-net: Reduced architecture for pneumonia in infants detection using deep convolutional framework using chest radiograph, Biomed. Signal Process. Control 87 (2024) 105375.
[82] P. Rastogi, K. Khanna, V. Singh, Classification of single-cell cervical pap smear images using efficientnet, Expert Syst. (2023) e13418.
[83] Z.E. Jouibari, H.N. Moakhkhar, Y. Baleghi, Emergency covid-19 detection from chest x-rays using deep neural networks and ensemble learning, Multimedia Tools Appl. (2023) 1–29.
[84] Y. Wang, Y. Ji, H. Xiao, A data augmentation method for fully automatic brain tumor segmentation, Comput. Biol. Med. 149 (2022) 106039.
[85] A. Biswas, P. Bhattacharya, S.P. Maity, R. Banik, Data augmentation for improved brain tumor segmentation, IETE J. Res. 69 (5) (2023) 2772–2782.
[86] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[87] W. He, M. Liu, Y. Tang, Q. Liu, Y. Wang, Differentiable automatic data augmentation by proximal update for medical image segmentation, IEEE/CAA J. Autom. Sin. 9 (7) (2022) 1315–1318.
[88] J. An, R.E. Gregg, S. Borhani, Effective data augmentation, filters, and automation techniques for automatic 12-lead ECG classification using deep residual neural networks, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 1283–1287.
[89] B. Shetty, R. Fernandes, A.P. Rodrigues, R. Chengoden, S. Bhattacharya, K. Lakshmanna, Skin lesion classification of dermoscopic images using machine learning and convolutional neural network, Sci. Rep. 12 (1) (2022) 18134.
[90] M. Tariq, V. Palade, Y. Ma, A. Altahhan, Diabetic retinopathy detection using transfer and reinforcement learning with effective image preprocessing and data augmentation techniques, in: Fusion of Machine Learning Paradigms: Theory and Applications, Springer, 2023, pp. 33–61.
[91] C. Rajeev, K. Natarajan, Data augmentation in classifying chest radiograph images (CXR) using DCGAN-CNN, in: GANs for Data Augmentation in Healthcare, Springer, 2023, pp. 91–110.
[92] S.G. Paul, A. Saha, M. Assaduzzaman, A real-time deep learning approach for classifying cervical spine fractures, Healthc. Anal. 4 (2023) 100265.
[93] A. Ijaz, S. Akbar, B. AlGhofaily, S.A. Hassan, T. Saba, Deep learning for pneumonia diagnosis using CXR images, in: 2023 Sixth International Conference of Women in Data Science At Prince Sultan University, WiDS PSU, IEEE, 2023, pp. 53–58.
[94] W. Wang, X. Yu, B. Fang, D.-Y. Zhao, Y. Chen, W. Wei, J. Chen, Cross-modality lge-cmr segmentation using image-to-image translation based data augmentation, IEEE/ACM Trans. Comput. Biol. Bioinform. (2022).
[95] R. Toda, A. Teramoto, M. Kondo, K. Imaizumi, K. Saito, H. Fujita, Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation, Sci. Rep. 12 (1) (2022) 12867.
[96] D.O. Oyewola, E.G. Dada, S. Misra, R. Damaševičius, A novel data augmentation convolutional neural network for detecting malaria parasite in blood smear images, Appl. Artif. Intell. 36 (1) (2022) 2033473.
[97] J. Hua, X. Cui, X. Li, K. Tang, P. Zhu, Multimodal fake news detection through data augmentation-based contrastive learning, Appl. Soft Comput. 136 (2023) 110125.
[98] Y. Lin, Z. Wang, K.-T. Cheng, H. Chen, Insmix: Towards realistic generative data augmentation for nuclei instance segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 140–149.
[99] W. Zouch, D. Sagga, A. Echtioui, R. Khemakhem, M. Ghorbel, C. Mhiri, A.B. Hamida, Detection of covid-19 from CT and chest x-ray images using deep learning models, Ann. Biomed. Eng. 50 (7) (2022) 825–835.
[100] H. Huhtanen, M. Nyman, T. Mohsen, A. Virkki, A. Karlsson, J. Hirvonen, Automated detection of pulmonary embolism from CT-angiograms using deep learning, BMC Med. Imaging 22 (1) (2022) 43.
[101] M. Aamir, Z. Rahman, Z.A. Dayo, W.A. Abro, M.I. Uddin, I. Khan, A.S. Imran, Z. Ali, M. Ishfaq, Y. Guan, et al., A deep learning approach for brain tumor classification using MRI images, Comput. Electr. Eng. 101 (2022) 108105.
[102] Y. Masmoudi, M. Ramzan, S.A. Khan, M. Habib, Optimal feature extraction and ulcer classification from wce image data using deep learning, Soft Comput. 26 (16) (2022) 7979–7992.
[103] L. Nanni, D. Cuza, A. Lumini, A. Loreggia, S. Brahman, Polyp segmentation with deep ensembles and data augmentation, in: Artificial Intelligence and Machine Learning for Healthcare: 1: Image and Data Analytics, Springer, 2022, pp. 133–153.
[104] S. Niu, Y. Peng, B. Li, Y. Qiu, T. Niu, W. Li, A novel deep learning motivated data augmentation system based on defect segmentation requirements, J. Intell. Manuf. (2023) 1–15.
[105] M.A.K. Raiaan, K. Fatema, I.U. Khan, S. Azam, M.R. ur Rashid, M.S.H. Mukta, M. Jonkman, F. De Boer, A lightweight robust deep learning model gained high accuracy in classifying a wide range of diabetic retinopathy images, IEEE Access (2023).
[106] J. Yao, Z. Guo, W. Yu, Enhanced deep residual network for bone classification and abnormality detection, Med. Phys. 49 (11) (2022) 6914–6929.
[107] Z. Zhang, Y. Li, B.-S. Shin, Robust color medical image segmentation on unseen domain by randomized illumination enhancement, Comput. Biol. Med. 145 (2022) 105427.
[108] Q. Zhou, H. Yin, A u-net based progressive gan for microscopic image augmentation, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2022, pp. 458–468.
[109] F.J. Moreno-Barea, J.M. Jerez, L. Franco, Gan-based data augmentation for prediction improvement using gene expression data in cancer, in: International Conference on Computational Science, Springer, 2022, pp. 28–42.
[110] Z. He, M. Lin, Z. Xu, Z. Yao, H. Chen, A. Alhudhaif, F. Alenezi, Deconv-transformer (dect): A histopathological image classification model for breast cancer based on color deconvolution and transformer architecture, Inform. Sci. 608 (2022) 1093–1112.

[111] M. Li, C. Li, C. Peng, L. Liu, B. Lovell, End to end generative meta curriculum learning for medical data augmentation, in: 2023 IEEE International Conference on Image Processing, ICIP, IEEE, 2023, pp. 2155–2159.

[112] V. Corbetta, R. Beets-Tan, W. Silva, Interpretability-guided data augmentation for robust segmentation in multi-centre colonoscopy data, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2023, pp. 330–340.

[113] J. Levy, D. Álvarez, F. Del Campo, J.A. Behar, Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry, Nature Commun. 14 (1) (2023) 4881.

[114] R. Gulakala, B. Markert, M. Stoffel, Generative adversarial network based data augmentation for CNN based detection of covid-19, Sci. Rep. 12 (1) (2022) 19186.

[115] X. Chen, Y. Li, L. Yao, E. Adeli, Y. Zhang, X. Wang, Generative adversarial u-net for domain-free few-shot medical diagnosis, Pattern Recognit. Lett. 157 (2022) 112–118.

[116] J. Zhang, Y. Zhang, X. Xu, Objectaug: object-level data augmentation for semantic image segmentation, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, 2021, pp. 1–8.

[117] A. Arora, N. Shoeibi, V. Sati, A. González-Briones, P. Chamoso, E. Corchado, Data augmentation using Gaussian mixture model on csv files, in: Distributed Computing and Artificial Intelligence, 17th International Conference, Springer, 2021, pp. 258–265.

[118] J. Zhou, Z. Wu, Z. Jiang, K. Huang, K. Guo, S. Zhao, Background selection schema on deep learning-based classification of dermatological disease, Comput. Biol. Med. 149 (2022) 105966.

[119] R. Liu, Z.-A. Huang, Y. Hu, Z. Zhu, K.-C. Wong, K.C. Tan, Attention-like multimodality fusion with data augmentation for diagnosis of mental disorders using MRI, IEEE Trans. Neural Netw. Learn. Syst. (2022).

[120] P. Whig, P. Sharma, R.R. Nadikattu, A.B. Bhatia, Y.J. Alkali, Gan for augmenting cardiac MRI segmentation, in: GANs for Data Augmentation in Healthcare, Springer, 2023, pp. 207–222.

[121] W. Li, J. Chen, J. Cao, C. Ma, J. Wang, X. Cui, P. Chen, Eid-gan: Generative adversarial nets for extremely imbalanced data augmentation, IEEE Trans. Ind. Inform. 19 (3) (2022) 3208–3218.

[122] M.M. Auzine, M.H.-M. Khan, S. Baichoo, N.G. Sahib, X. Gao, P. Bissoonauth-Daiboo, Endoscopic image analysis using deep convolutional gan and traditional data augmentation, in: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME, IEEE, 2022, pp. 1–6.

[123] S. Hamida, O. El Gannour, A. Maafiri, Y. Lamalem, I. Haddou-Oumouloud, B. Cherradi, Data balancing through data augmentation to improve transfer learning performance for skin disease prediction, in: 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET, IEEE, 2023, pp. 1–7.

[124] D.L. Krishna, N.V.D. Padmanabhuni, G. JayaLakshmi, Data augmentation based brain tumor detection using CNN and deep learning, in: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions, CISES, IEEE, 2023, pp. 317–321.

[125] Z. Yang, Y. Li, G. Zhou, Ts-gan: Time-series gan for sensor-based health data augmentation, ACM Trans. Comput. Healthc. 4 (2) (2023) 1–21.

[126] Y. Xu, I. Klyuzhin, S. Harsini, A. Ortiz, S. Zhang, F. Bénard, R. Dodhia, C.F. Uribe, A. Rahmim, J.L. Ferres, Automatic segmentation of prostate cancer metastases in psma pet/ct images using deep neural networks with weighted batch-wise dice loss, Comput. Biol. Med. 158 (2023) 106882.

[127] S. Kollem, K.R. Reddy, C.R. Prasad, A. Chakraborty, J. Ajayan, S. Sreejith, S. Bhattacharya, L.L. Joseph, R. Janapati, Alexnet-ndtl: Classification of MRI brain tumor images using modified alexnet with deep transfer learning and lipschitz-based data augmentation, Int. J. Imaging Syst. Technol. (2023).

[128] H. Tan, X. Lang, B. He, Y. Lu, Y. Zhang, Gan-based medical image augmentation for improving CNN performance in myositis ultrasound image classification, in: 2023 6th International Conference on Electronics Technology, ICET, IEEE, 2023, pp. 1329–1333.

[129] A. Anaya-Isaza, M. Zequera-Diaz, Detection of diabetes mellitus with deep learning and data augmentation techniques on foot thermography, IEEE Access 10 (2022) 59564–59591.

[130] S.-S. Park, H.-J. Kwon, J.-W. Baek, K. Chung, Dimensional expansion and time-series data augmentation policy for skeleton-based pose estimation, IEEE Access 10 (2022) 112261–112272.

[131] M.E. Salman, G.Ç. Çakar, J. Azimjonov, M. Kösem, İ.H. Cedimoğlu, Automated prostate cancer grading and diagnosis system using deep learning-based yolo object detection algorithm, Expert Syst. Appl. 201 (2022) 117148.

[132] M. Tasyurek, E. Gul, A new deep learning approach based on grayscale conversion and dwt for object detection on adversarial attacked images, J. Supercomput. (2023) 1–34.

[133] F. Pérez-García, R. Sparks, S. Ourselin, Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning, Comput. Methods Programs Biomed. 208 (2021) 106236.

[134] N. Goel, S. Kaur, D. Gunjan, S. Mahapatra, Investigating the significance of color space for abnormality detection in wireless capsule endoscopy images, Biomed. Signal Process. Control 75 (2022) 103624.

[135] X. Tang, C. Zhou, L. Chen, Y. Wen, Enhancing medical image classification via augmentation-based pre-training, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2021, pp. 1538–1541.

[136] S. Tajjour, S. Garg, S.S. Chandel, D. Sharma, A novel hybrid artificial neural network technique for the early skin cancer diagnosis using color space conversions of original images, Int. J. Imaging Syst. Technol. 33 (1) (2023) 276–286.

[137] R. Thakur, P. Maheshwari, S.K. Datta, S.K. Dubey, Smartphone-based, automated detection of urine albumin using deep learning approach, Measurement 194 (2022) 110948.

[138] R. Ramadan, S. Aly, M. Abdel-Atty, Color-invariant skin lesion semantic segmentation based on modified u-net deep convolutional neural network, Health Inf. Sci. Syst. 10 (1) (2022) 17.

[139] A.M. Güneş, W. van Rooij, S. Gulshad, B. Slotman, M. Dahele, W. Verbakel, Impact of imperfection in medical imaging data on deep learning-based segmentation performance: An experimental study using synthesized data, Med. Phys. (2023).

[140] D. Bar-David, L. Bar-David, Y. Shapira, R. Leibu, D. Dori, A. Gebara, R. Schneor, A. Fischer, S. Soudry, Elastic deformation of optical coherence tomography images of diabetic macular edema for deep-learning models training: how far to go? IEEE J. Transl. Eng. Health Med. (2023).

[141] Y. Gao, F. Song, P. Zhang, J. Liu, J. Cui, Y. Ma, G. Zhang, J. Luo, Improving the subtype classification of non-small cell lung cancer by elastic deformation based machine learning, J. Digit. Imaging 34 (3) (2021) 605–617.

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

[143] Y. Gao, Z. Tang, M. Zhou, D. Metaxas, Enabling data diversity: efficient automatic augmentation via regularized adversarial training, in: International Conference on Information Processing in Medical Imaging, Springer, 2021, pp. 85–97.

[144] S. Nallamolu, H. Nandanwar, A. Singh, C. Subalalitha, A cnn-based approach for multi-classification of brain tumors, in: 2022 2nd Asian Conference on Innovation in Technology, ASIANCON, IEEE, 2022, pp. 1–6.

[145] L. Nanni, M. Paci, S. Brahnam, A. Lumini, Feature transforms for image data augmentation, Neural Comput. Appl. 34 (24) (2022) 22345–22356.

[146] S. Kalaivani, N. Asha, A. Gayathri, Geometric transformations-based medical image augmentation, in: GANs for Data Augmentation in Healthcare, Springer, 2023, pp. 133–141.

[147] H. Chen, Y. Jiang, H. Ko, M. Loew, A teacher–student framework with fourier transform augmentation for covid-19 infection segmentation in CT images, Biomed. Signal Process. Control 79 (2023) 104250.

[148] E. Castro, J.C. Pereira, J.S. Cardoso, Symmetry-based regularization in deep breast cancer screening, Med. Image Anal. 83 (2023) 102690.

[149] P. Liu, W. Qian, J. Cao, D. Xu, Semi-supervised medical image classification via increasing prediction diversity, Appl. Intell. 53 (9) (2023) 10162–10175.

[150] K. Hammoudi, A. Cabani, B. Slika, H. Benhabiles, F. Dornaika, M. Melkemi, Superpixelgridmasks data augmentation: Application to precision health and other real-world data, J. Healthc. Inf. Res. 6 (4) (2022) 442–460.

[151] H. Cho, Y. Han, W.H. Kim, Anti-adversarial consistency regularization for data augmentation: Applications to robust medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 555–566.

[152] Ö. Özdemir, E.B. Sönmez, Attention mechanism and mixup data augmentation for classification of covid-19 computed tomography images, J. King Saud Univ.-Comput. Inf. Sci. 34 (8) (2022) 6199–6207.

[153] T. Bdair, B. Wiestler, N. Navab, S. Albarqouni, Roam: Random layer mixup for semi-supervised learning in medical images, IET Image Process. 16 (10) (2022) 2593–2608.

[154] M. Gadermayr, L. Koller, M. Tschuchnig, L.M. Stangassinger, C. Kreutzer, S. Couillard-Despres, G.J. Oostingh, A. Hittmair, Mixup-mil: Novel data augmentation for multiple instance learning and a study on thyroid cancer diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 477–486.

[155] Y. Li, Q. Yang, F.L. Wang, L.-K. Lee, Y. Qu, T. Hao, Asymmetric cross-modal attention network with multimodal augmented mixup for medical visual question answering, Artif. Intell. Med. 144 (2023) 102667.

[156] C. Ambroise, V. Frouin, B. Dufumier, E. Duchesnay, A. Grigis, Mixup brain-cortical augmentations in self-supervised learning, in: International Workshop on Machine Learning in Clinical Neuroimaging, Springer, 2023, pp. 102–111.

[157] D.H. Hong, S.Y. Kim, M.K. Park, S.J. Oh, I.C. Doo, Object detection improvements on skin burn image data via data augmentation and semi-supervised learning, in: International Conference on Computer Science and Its Applications and the International Conference on Ubiquitous Information Technologies and Applications, Springer, 2022, pp. 247–252.

[158] Z. Yao, W. Mao, Y. Yuan, Z. Shi, G. Zhu, W. Zhang, Z. Wang, G. Zhang, Fuzzy-VGG: A fast deep learning method for predicting the staging of Alzheimer's disease based on brain MRI, Inform. Sci. 642 (2023) 119129.

[159] H.-T. Joo, I.-C. Baek, K.-J. Kim, A swapping target $q$-value technique for data augmentation in offline reinforcement learning, IEEE Access 10 (2022) 57369–57382.

[160] J. Zhang, L. Zhang, D. Li, A unified search framework for data augmentation and neural architecture on small-scale image datasets, IEEE Trans. Cogn. Dev. Syst. (2023).

[161] H. Li, X. Dong, W. Shen, F. Ge, H. Li, Resampling-based cost loss attention network for explainable imbalanced diabetic retinopathy grading, Comput. Biol. Med. 149 (2022) 105970.

[162] X. Guo, Z. Chen, J. Liu, Y. Yuan, Non-equivalent images and pixels: Confidence-aware resampling with meta-learning mixup for polyp segmentation, Med. Image Anal. 78 (2022) 102394.

[163] Z. Yu, J. Chen, Y. Liu, Y. Chen, T. Wang, R. Nowak, Z. Lv, Ddcnn: A deep learning model for af detection from a single-lead short ECG signal, IEEE J. Biomed. Health Inf. 26 (10) (2022) 4987–4995.

[164] Y. Deng, H. Li, Deep learning for few-shot white blood cell image classification and feature learning, Comput. Methods Biomech. Biomed. Eng. Imaging Vis. (2023) 1–11.

[165] C. Tiago, A. Gilbert, A.S. Beela, S.A. Aase, S.R. Snare, J. Šprem, K. McLeod, A data augmentation pipeline to generate synthetic labeled datasets of 3d echocardiography images using a gan, IEEE Access 10 (2022) 98803–98815.

[166] C. Ouyang, C. Chen, S. Li, Z. Li, C. Qin, W. Bai, D. Rueckert, Causality-inspired single-source domain generalization for medical image segmentation, IEEE Trans. Med. Imaging 42 (4) (2022) 1095–1106.

[167] A. Anaya-Isaza, M. Zequera-Diaz, Fourier transform-based data augmentation in deep learning for diabetic foot thermograph classification, Biocybern. Biomed. Eng. 42 (2) (2022) 437–452.

[168] N. Marini, S. Otalora, M. Wodzinski, S. Tomassini, A.F. Dragoni, S. Marchand-Maillet, J.P.D. Morales, L. Duran-Lopez, S. Vatrano, H. Müller, et al., Data-driven color augmentation for h & e stained images in computational pathology, J. Pathol. Inform. 14 (2023) 100183.

[169] Z. Li, K. Kamnitsas, Q. Dou, C. Qin, B. Glocker, Joint optimization of class-specific training-and test-time data augmentation in segmentation, IEEE Trans. Med. Imaging (2023).

[170] C. Sfakianakis, G. Simantiris, G. Tziritas, Gudu: Geometrically-constrained ultrasound data augmentation in u-net for echocardiography semantic segmentation, Biomed. Signal Process. Control 82 (2023) 104557.

[171] Z. Ren, X. Kong, Y. Zhang, S. Wang, Ukssl: Underlying knowledge based semi-supervised learning for medical image classification, IEEE Open J. Eng. Med. Biol. (2023).

[172] Y.-C. Chou, C.-C. Chen, Improving deep learning-based polyp detection using feature extraction and data augmentation, Multimedia Tools Appl. 82 (11) (2023) 16817–16837.

[173] W. Yu, Z. Huang, J. Zhang, H. Shan, San-net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization, Comput. Biol. Med. 156 (2023) 106717.

[174] M.T. Akram, S. Asghar, A.R. Shahid, Effective data augmentation for brain tumor segmentation, Int. J. Imaging Syst. Technol. (2023).

[175] F. Prinzi, M. Insalaco, S. Gaglio, S. Vitabile, Breast cancer localization and classification in mammograms using yolov5, in: Applications of Artificial Intelligence and Neural Systems To Data Science, Springer, 2023, pp. 73–82.

[176] B. Guan, J. Yao, S. Wang, G. Zhang, Y. Zhang, X. Wang, M. Wang, Automatic detection and localization of thighbone fractures in x-ray based on improved deep learning method, Comput. Vis. Image Underst. 216 (2022) 103345.

[177] C. Chen, K. Hammernik, C. Ouyang, C. Qin, W. Bai, D. Rueckert, Cooperative training and latent space data augmentation for robust medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part III, vol. 24, Springer, 2021, pp. 149–159.

[178] D.P. Kingma, M. Welling, Auto-encoding variational bayes, 2013, arXiv preprint arXiv:1312.6114.

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

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

[181] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, Adv. Neural Inf. Process. Syst. 29 (2016).

[182] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

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

[184] A. Razavi, A. Van den Oord, O. Vinyals, Generating diverse high-fidelity images with VQ-VAE-2, Adv. Neural Inf. Process. Syst. 32 (2019).

[185] R.T. Chen, X. Li, R.B. Grosse, D.K. Duvenaud, Isolating sources of disentanglement in variational autoencoders, Adv. Neural Inf. Process. Syst. 31 (2018).

[186] A. Kumar, P. Sattigeri, A. Balakrishnan, Variational inference of disentangled latent concepts from unlabeled observations, 2017, arXiv preprint arXiv:1711.00848.

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

[188] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8789–8797.

[189] Y. Saatci, A.G. Wilson, Bayesian gan, Adv. Neural Inf. Process. Syst. 30 (2017).

[190] H. Park, Y. Yoo, N. Kwak, MC-GAN: Multi-conditional generative adversarial network for image synthesis, 2018, arXiv preprint arXiv:1805.01123.

[191] M.S. Sivakumar, L.M. Leo, T. Gurumekala, V. Sindhu, A.S. Priyadharshini, Deep learning in skin lesion analysis for malignant melanoma cancer identification, Multimedia Tools Appl. (2023) 1–21.

[192] M. Priya, M. Nandhini, Detection of fetal brain abnormalities using data augmentation and convolutional neural network in internet of things, Measurement: Sensors (2023) 100808.

[193] K.M. Almustafa, A.K. Sharma, S. Bhardwaj, Starc: Deep learning algorithms' modeling for structured analysis of retina classification, Biomed. Signal Process. Control 80 (2023) 104357.

[194] X. Li, H. Bagher-Ebadian, S. Gardner, J. Kim, M. Elshaikh, B. Movsas, D. Zhu, I.J. Chetty, An uncertainty-aware deep learning architecture with outlier mitigation for prostate gland segmentation in radiotherapy treatment planning, Med. Phys. 50 (1) (2023) 311–322.

[195] P.F. Wilson, M. Gilany, A. Jamzad, F. Fooladgar, M.N.N. To, B. Wodlinger, P. Abolmaesumi, P. Mousavi, Self-supervised learning with limited labeled data for prostate cancer detection in high frequency ultrasound, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2023).

[196] S.-J. Kim, D.-H. Lee, Y.-W. Choi, Cropcat: Data augmentation for smoothing the feature distribution of eeg signals, in: 2023 11th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2023, pp. 1–4.

[197] F.J.M. Shamrat, S. Akter, S. Azam, A. Karim, P. Ghosh, Z. Tasnim, K.M. Hasib, F. De Boer, K. Ahmed, Alzheimernet: An effective deep learning based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images, IEEE Access 11 (2023) 16376–16395.

[198] A. Mishra, R. Jha, V. Bhattacharjee, Ssclnet: A self-supervised contrastive loss-based pre-trained network for brain MRI classification, IEEE Access 11 (2023) 6673–6681.

[199] S. Tripathy, R. Singh, M. Ray, Automation of brain tumor identification using efficientnet on magnetic resonance images, Procedia Comput. Sci. 218 (2023) 1551–1560.

[200] Z. Shen, X. Ouyang, B. Xiao, J.-Z. Cheng, D. Shen, Q. Wang, Image synthesis with disentangled attributes for chest x-ray nodule augmentation and detection, Med. Image Anal. 84 (2023) 102708.

[201] T. van Sonsbeek, M. Worring, X-tra: Improving chest x-ray tasks with cross-modal retrieval augmentation, in: International Conference on Information Processing in Medical Imaging, Springer, 2023, pp. 471–482.

[202] M. Obayya, M.A. Arasi, N. Alruwais, R. Alsini, A. Mohamed, I. Yaseen, Biomedical image analysis for colon and lung cancer detection using tuna swarm algorithm with deep learning model, IEEE Access (2023).

[203] M. Gardner, Y.B. Bouchta, A. Mylonas, M. Mueller, C. Cheng, P. Chlap, R. Finnegan, J. Sykes, P.J. Keall, D.T. Nguyen, Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kv images acquired during radiation therapy, Med. Phys. (2023).

[204] D. Grama, M. Dahele, W. van Rooij, B. Slotman, D.K. Gupta, W.F. Verbakel, Deep learning-based markerless lung tumor tracking in stereotactic radiotherapy using siamese networks, Med. Phys. (2023).

[205] F. Younas, M. Usman, W.Q. Yan, A deep ensemble learning method for colorectal polyp classification with optimized network parameters, Appl. Intell. 53 (2) (2023) 2410–2433.

[206] M. Liu, L. Hu, Y. Tang, C. Wang, Y. He, C. Zeng, K. Lin, Z. He, W. Huo, A deep learning method for breast cancer classification in the pathology images, IEEE J. Biomed. Health Inf. 26 (10) (2022) 5025–5032.

[207] H.-S. Ham, H.-S. Lee, J.-W. Chae, H.C. Cho, H.-C. Cho, Improvement of gastroscopy classification performance through image augmentation using a gradient-weighted class activation map, IEEE Access 10 (2022) 99361–99369.

[208] L. Caselles, C. Jailin, S. Muller, Data augmentation for breast cancer mass segmentation, in: Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis, Springer, 2022, pp. 228–237.

[209] H.J. Ling, O. Bernard, N. Ducros, D. Garcia, Phase unwrapping of color doppler echocardiography using deep learning, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2023).

[210] C.-J. Chao, J. Jeong, R. Arsanjani, K. Kim, Y.-L. Tsai, W.-C. Yu, J.M. Farina, A.K. Mahmoud, C. Ayoub, M. Grogan, et al., Echocardiography-based deep learning model to differentiate constrictive pericarditis and restrictive cardiomyopathy, JACC: Cardiovasc. Imaging (2023).