Beyond Traditional Augmentation: Lesion Classification Refined with LesionGAN

Advaith Menon
advaithmenon6612@gmail.com
Dunbar High School

***Abstract*—Computed Tomography (CT) scans are the preferred imaging modality for lesion detection. On average, the daily human error rate of lesion detection by radiologists from CT scans is 4%, and with there being approximately 1 billion scans annually, this means 40 million patients are misdiagnosed annually. To reduce this high error rate of missed lesions, many have turned to deep learning algorithms to assist in the automation of lesion detection. However, obtaining large-scale annotated datasets in the medical domain is challenging. While traditional data augmentation is a commonly adopted remedy, it often falls short of introducing sufficient variability into the dataset. This research explores the use of a specialized Wasserstein generative adversarial network with gradient penalty (WGAN-GP) to synthesize artificial CT scans of lesions. The aim is to generate more diverse data points, potentially enhancing the performance of deep learning neural networks in lesion detection compared to traditional augmentation methods. The research methodology encompasses data acquisition, image preprocessing, and the formulation of traditionally augmented, GAN-augmented, and combined datasets. Subsequently, each dataset is evaluated using a deep neural network classifier. Evaluation of the test results indicates that the combined dataset outperformed the others, achieving an accuracy of 78.95% alongside superior average recall and precision metrics. In contrast, the least effective dataset, which was solely preprocessed without any augmentation, has a test accuracy of 71.93%. Additionally, the traditionally-augmented dataset and the GAN-augmented dataset performed equally well, yielding accuracies of 77.19% and 75.44%. The findings underscore the potential benefits of employing synthetic augmentation in medical imaging tasks, paving the way for more precise lesion detection methodologies.**

***Index Terms*—GAN, Lesion, Classification, Traditional Augmentation, Preprocessed, ResNet, UNet, DenseNet**

I. Introduction

A lesion is an area of abnormal tissue usually caused by disease or trauma. Lesions can be malignant or benign [2]. The localization of lesions allows radiologists to identify those that are enlarged, malignant, and at risk for tumor spread.

Computed Tomography (CT) scans have emerged as the preferred imaging modality for lesion detection in clinical practice. CT scans utilize a combination of X-rays and computer algorithms to generate detailed cross-sectional images of the body. This imaging technique provides radiologists with a comprehensive view of internal structures, enabling them to identify and analyze lesions with a high level of precision.

One of the key advantages of CT scans for lesion detection is their ability to capture images with excellent spatial resolution. The high-resolution images produced by CT scanners allow radiologists to visualize lesions with great clarity, enabling them to discern even subtle abnormalities in the tissues. This level of detail is particularly crucial when it comes to the detection and characterization of lesions, as it aids in differentiating between benign and malignant lesions and assessing the extent of lesion progression.

However, it is important to acknowledge that radiological interpretation is not infallible. Radiologists, who are responsible for interpreting the CT images and providing diagnostic reports, are susceptible to errors due to the complexity and subjective nature of the process. On average, the daily error rate of lesion detection by radiologists from CT scans is approximately 4% [3]. This value is largely a result of human errors. A missed lesion can ultimately grow larger and progress into a malignant tumor, if not treated [4].

II. Related Work

The use of deep learning techniques in the medical domain has gained significant attention due to their remarkable performance in the form of accuracy gains [5]. In an effort to extend these advancements to lesion detection and mitigate the high error rate associated with missed lesions, numerous researchers have explored the utilization of machine learning and deep learning algorithms to automate lesion detection. Recently, deep learning algorithms have performed significantly better than conventional approaches when applied to CT scans of lesions [7], [41]. Many convolutional neural network (CNN) models have also been used in combination with different techniques to better detect lesions. Zlocha et al. [11] used a deep neural network RetinaNet and the segmentation technique of GrabCut to detect lesions. Lavanya et al. [12] also combined the segmentation techniques and detection from CNN for an improved result. Many other modifications to CNN have been introduced, such as multi-headed CNNs [13], 3D context-enhanced feature maps [14], and feature-based pyramids [15]. These solutions implement efficient algorithms for lesion detection.

Nonetheless, achieving these performance gains often necessitates access to a substantial volume of training data accompanied by high-quality annotated labels for supervised learning [8]. Generally, algorithms exhibit superior performance when trained on larger datasets that encompass images with greater diversity than mere geometric or color transformations of previous images. However, a significant challenge arises in the medical field, as datasets of medically screened

images are typically unavailable to the broader public due to two primary reasons:

A) The conventional method for large-scale medical dataset annotations, which is through crowd-sourcing, is not readily applicable to the medical image domain because medical image annotation requires extensive clinical expertise
B) High variability in results, even between experts, frequently occurs and thus may compromise reliable labeling of a large number of medical lesion images.

This issue can be overcome when training samples can be synthesized with more intrinsic and extrinsic variations using a technique such as generative modeling. GANs, or generative adversarial networks are most effective at generating image samples, compared to other generative models such as Variational AutoEncoders (VAEs) [25]. GANs for generating medical images are an active research area with rapid advancements in the GAN architecture and losses for its efficiency. An alternative architecture for GANs to train effectively on augmented training data has been introduced in [6] which can be very useful in medical fields where data points are scarce. Several applications such as generating chest X-ray scans, brain tumor MRI images, etc. have been thoroughly developed through research such as Bozorgtabar et al. [9]. High-resolution skin lesion images have been generated by Bissoto et al. [10], where instead of noise, the authors make use of a semantic label map (an image where each pixel value represents the object class) and an instance map (an image where the pixels combine information from its object class and its instance). Frid-Adar et al. [26] introduce GAN augmented Liver Lesion images to improve classification on a very small training dataset. Instead of generating whole images, their approach involves generating high-quality Region Of Interest (ROIs).

The GAN aspect of this research is essential as it enables the generation of additional images, supplementing the existing dataset. This augmentation process helps address the limitation of insufficient image availability for training CNN models. By leveraging the power of GANs, the model aims to improve the accuracy and robustness of deep learning-assisted radiology systems, ultimately assisting radiologists in their decision-making processes and reducing their workload. The paper makes the following key contributions:

1) Design of a CNN-based solution for the lesion classification task
2) Evaluate the effectiveness of augmentation of the CNN training set using the GAN-generated synthetic data
3) Compares the effectiveness of traditional augmentation versus GAN augmentation

III. Methodology

In this section, we outline the methodology employed in our research. We begin with an overview of the data collection process, followed by data refinement techniques. Next, we describe the design of a data augmentation filter and traditional augmentation methods. Subsequently, we introduce the Data Augmentation Generative Adversarial Network (DAGAN) model and elaborate on the design of the generator and discriminator. Finally, we discuss the design of the test classifiers using transfer learning. A flow chart describing the pipeline of our methodology is shown in Figure 1.

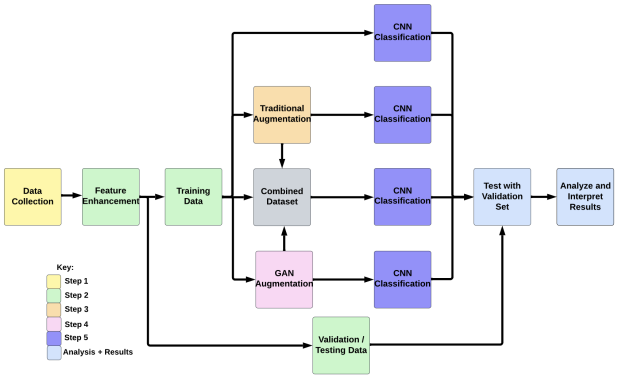


Fig. 1: Flow chart describing the methodology

A. Data Collection

The data used for the experiments was sourced from the DeepLesion Dataset [1]. We used a total of 32000 images to conduct our experiments. The dataset includes eight categories of various lesion types, including bone, abdomen, mediastinum, liver, lung, kidney, soft tissue, and pelvis, which are used in this research.

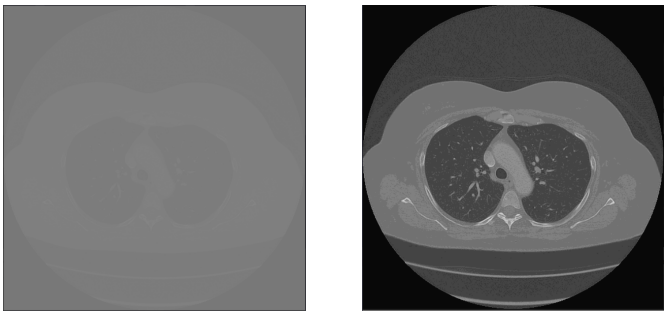
The dataset was split into train, validating, and test subsets. This resulted in 23205, 5775, and 3020 images in the train, validation, and test subsets. To avoid data leakage, the splitting was done before data augmentation on the training dataset.

B. Data/Feature Refinement

To reduce the large file size of the dataset, the NIHCC compiled the images in a monochrome 64-bit image, resulting in barely defined edges and regions. Within the dataset itself, several steps to re-improve data quality were recommended, from subtracting pixel intensity to upscaling image bits. However, in this research, neither of these methods proved to be effective, as the resulting images were not refined enough for feature extraction using deep learning. Instead, we created an algorithm to contrast the separate shades of gray within the image, and after tuning, this proved to be far more effective in creating higher-quality, sharp-edged images directly from the DeepLesion dataset as compared to the other methods mentioned earlier. The input and output images have been compared in Figure 2a and 2b.

C. Data Augmentation Filter Design

Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points [22], [27], [28]. Data augmentation methods, specifically geometric



(a) Before Feature Refinement (b) After Feature Refinement

Fig. 2: Comparison of Feature Refinement

transformations to images are easy to apply and have been shown to improve deep learning model's accuracy [23]. Hence, a data augmentation filter was designed for the same. As the processed data passes through the data augmentation filter, the filter performs a randomized series of simple augmentation techniques such as rotation and trimming. This randomized set of operations creates another image, that is typically similar in structure to the original processed images but can be differentiated by the classification algorithm. The entire dataset of originally processed images is then run through this filter 3 times, creating a dataset 3 times larger than the original dataset.

D. DAGAN

GANs are a popular approach to generative modeling using deep learning methods, such as convolutional neural networks. Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset [29].

GANs work by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated) [30]. The two models are trained together in a zero-sum game, which means that they compete against each other until the discriminator model is fooled approximately half the time, meaning the generator model is generating plausible examples [31].

A GAN utilizes a computed loss to improve both models. Currently, the traditional method of computing GAN losses is through the minimax function, which can be represented by the equation 1

$$\mathbb{E}_x[log(D(x))] + \mathbb{E}_z[log(1 - D(G(z)))] \tag{1}$$

The goal of the generator is to minimize the minimax loss, while the discriminator attempts to maximize it. While effective, the minimax loss function is packaged with several issues, such as mode collapse and vanishing gradients. The

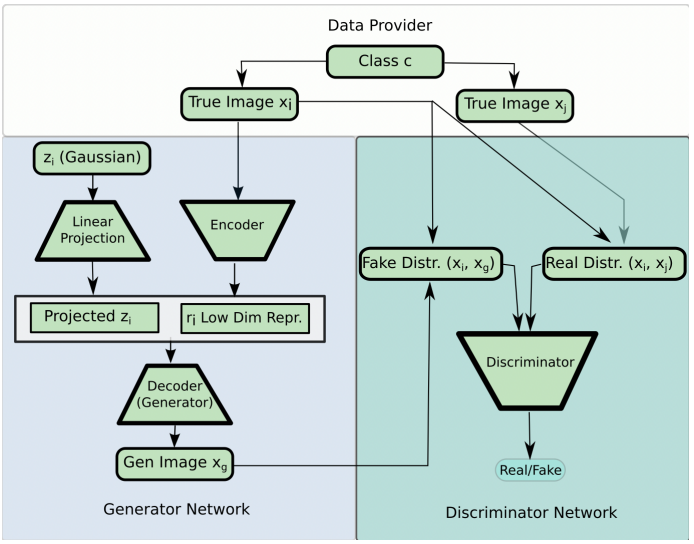


Fig. 3: Model Diagram

proposed model instead uses the penalized Wasserstein loss, which can be represented with the equations 2 and 3 respectively corresponding to discriminator loss and generator loss.

$$\mathbb{E}_x[D(x)] - \mathbb{E}_z[D(G(z))] \tag{2}$$

$$\mathbb{E}_z[D(G(z))] \tag{3}$$

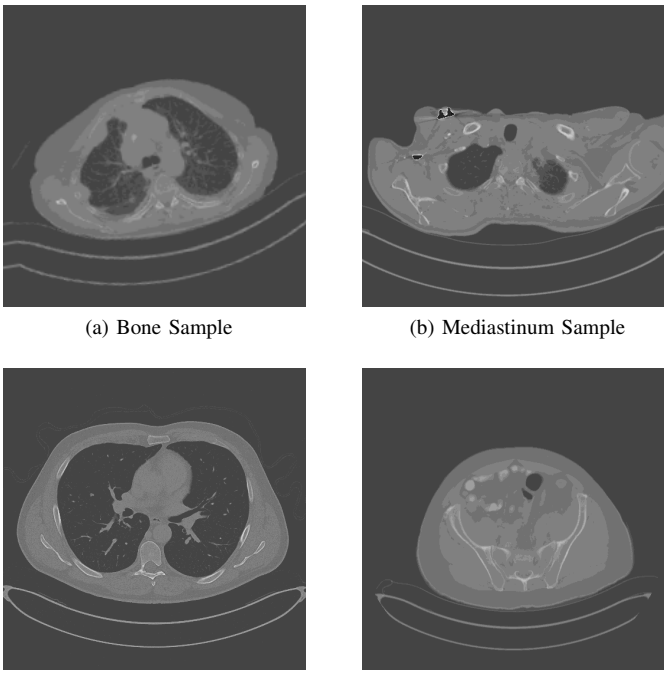
The formula derives from the cross-entropy between the real and generated distributions with the penalized norm of the gradient of the discriminator concerning its input [33]. Both the discriminator and generator attempt to maximize their respective functions during the training process [32].

WGAN-GP offers an alternative to clipping weights, penalizing the norm of the gradient of the critic concerning its input. This makes it significantly more stable than traditional GAN networks and even other WGANs. Data augmentation generative adversarial network (DAGAN) [40], which is the model used in this research, uses this WGAN-GP critic in which the generator takes a real image x_i from a class c. It then encodes it into a lower dimension before adding Gaussian noise. Finally, the generator decodes this lower dimension vector to generate a fake image x_g. The critic or discriminator takes three inputs x_i, x_g, and another real image from class c x_j as shown in Figure 3.

Alternatively, the critic takes either an input data point x_i and a second data point from the same class x_j such that their target scores or critic scores should be equal, or an input data point x_i and the output of the current generator x_g which takes x_i as an input. The critic tries to discriminate the scores of the generated inputs (scores from x_i and x_g) from the real inputs (scores from x_i and x_j). The generator is trained to minimize this discriminative capability as measured by the Wasserstein distance [38].

E. Generator and Discriminator Design

The GAN in use here consists of two different models for the generator and discriminator. The generator is a combi-



(a) Bone Sample (b) Mediastinum Sample

(c) Lung Sample (d) Pelvis Sample

Fig. 4: Samples of GAN-generated CT scans

nation of UNet [34] and ResNet [36] while DenseNet [37] is used as the discriminator. In the generator, a striped 1x1 convolution layer is used to pass information between each downsampling or upsampling block. This helps with proper gradient flow as it bypasses the non-linearity between adjacent blocks. Skip connections between equivalent-sized filters at each end of the network (UNet) have been used which help with the reconstruction of features in the generated image. To add the filter values and the number of blocks, a Gaussian noise layer of the size of 100 was added to the projected lower dimension of the input image. The discriminator uses naive DenseNet architecture consisting of 4 dense blocks and 4 transition layers with each block using a filter size of 128 and having a growth rate k=64. This helped each block access features from all previous blocks helping with better gradient flow and learning refined features for discrimination. This architecture was also chosen to keep the discriminator powerful enough to give meaningful feedback to the generator to have stable GAN training. The model was trained in 100 epochs, using an Adam optimizer with a learning rate of $1{*}e^{-4}$ and $\beta_1 = 0.9$ and $\beta_2 = 0.99$. A batch size of 4 was used due to computation constraints.

F. Image Classification

The image classification experiment was conducted using the standard transfer learning technique where a pre-trained model with ImageNet weights is chosen as the base model and dense layers are added to customize it to a given application. For our experiments, we chose MobileNetV2 and added a dense layer with 100 neurons followed by a classification layer with softmax activation. MobileNetV2, optimized for mobile devices, offers a lightweight and computationally efficient architecture, making it a prime choice for image classification tasks, especially under constraints. Each of the test classifiers was created using transfer learning for MobileNetV2 [35].

The image classification result is the ultimate measure of augmentation effectiveness. it is tested on the same set of real CT scan images to ensure a fair comparison. The classification algorithm is trained on their respective datasets for a total of 50 epochs, using an Adam optimizer with a learning rate of $1 * e^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and a batch size of 32.

IV. EXPERIMENTS

Datasets	Classifier Metrics			
	Accuracy	***Precision***	***Recall***	***Loss***
Preprocessed	71.93%	0.7344	0.6782	0.9059
Traditionally Augmented	77.19%	0.7423	0.6985	0.7416
GAN-Augmented	75.44%	0.7508	0.7067	0.8333
Combined	78.95%	0.7717	0.7732	0.6404

TABLE I: Training results from different data augmentation techniques

Experiments using the MobileNetV2 model were conducted across all 4 datasets. In each case, hyper-parameter tuning was done, where the learning rate was varied between 0.0001-0.00019, and the epochs were varied between 20-100. Models with the best validation accuracy were chosen to conduct experiments on the test dataset. Care was taken that there was no data leakage between the train and validation dataset when data augmentation was conducted.

Overall, the metrics of the 4 models were close as shown in Table I. The MobileNetV2 model trained with the preprocessed dataset performed the worst, with a significantly lower accuracy of 71.93% and a higher loss of 0.9059. The combined dataset performed the best, with a higher accuracy of 78.95% and a lower loss of 0.6404.

As the models approached epoch 50, the accuracies of all four models began to stagnate, suggesting there was no further room for improvement. All four classifiers performed the worst on the kidney and soft tissue dataset, due to the significantly lower accuracies, as demonstrated in table II. The category $Bone$ had a 100% accuracy in all four cases of data pre-processing and augmentation. In the case of pre-processed images, the $Lung$ category had the lowest accuracy of 25%. For the traditional augmentation, the $Kidney$ category performed the worst with an accuracy of 33%. On the other hand, GA and combined models had the worst performance in the $SoftTissue$ category with an accuracy of 50%.

Table II shows the accuracy values for each category in the dataset. Note that TA stands for Traditional Augmentation and GA stands for GAN-Augmentation. Figure 5 show the confusion matrix obtained from the test dataset after training the models with different pre-processing and data augmentation techniques. The figures have numbered labels (1, 2, etc.) to indicate the index of each corresponding lesion

Index	Classifier Metrics				
	Datasets	***Preprocessed***	***TA***	***GA***	***Combined***
1	Bone	100%	100%	100%	100%
2	Abdomen	88.2%	82.4%	88.2%	94.1%
3	Mediastinum	90%	90%	dr%	90%
4	Liver	50%	66.7%	66.7%	85.7%
5	Lung	25%	83.3%	66.7%	85.7%
6	Kidney	33.3%	33.3%	66.7%	66.7%
7	Soft Tissue	50%	50%	50%	50%
8	Pelvis	80%	80%	80%	80%

TABLE II: Accuracy for each lesion category with different augmentation techniques

category. The specific indices associated with each category are provided in Table II, allowing for easy reference and understanding of the lesion categories being depicted in the figures showing the confusion matrix.

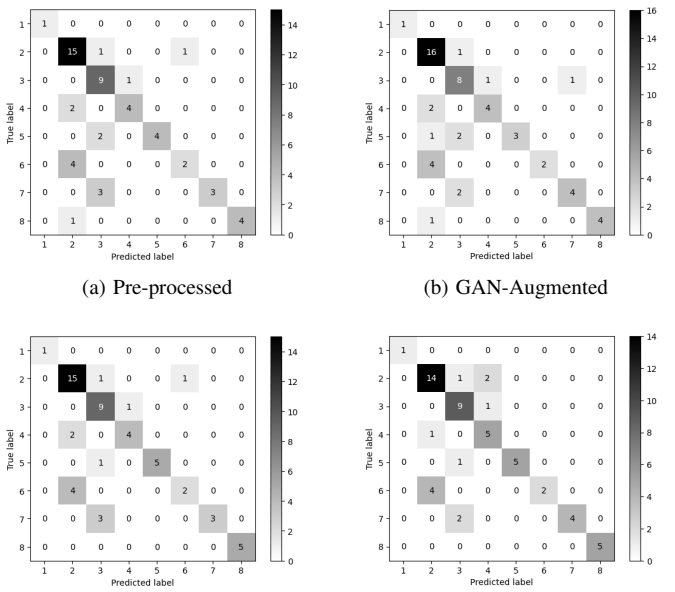
The confusion matrix in the case of the pre-processed dataset shown in Figure 5 shows that the model was most confused between kidney and abdomen classes. The model predicted the lesion to be an abdomen lesion, even though it was a kidney lesion 4 times, which can be considered significant. This result is consistent between all 4 models, with the individual accuracies of both classes being significantly worse. This shows that CNN-based neural networks have difficulty differentiating between these two categories. A similar issue can be seen with soft-tissue lesions

Overall, the results obtained in this study show the importance of data preprocessing techniques in achieving optimal performance in classification tasks. The combined dataset model outperformed the others, highlighting the benefits of utilizing multiple augmentation techniques, including the generation of synthetic images using a GAN. Future research could focus on exploring additional augmentation techniques or investigating alternative models to further improve classification accuracy and address the limitations observed in this study.

A. MS-SSIM Analysis

Structural Similarity Index Measure (SSIM), is used for measuring the structural resemblance between two images. SSIM is calculated from the average brightness of each image (μ_x and μ_y), variance of the brightness (σ_x^2 and σ_y^2), covariance (σ_{xy}), and stabilizing variables (c_1 and c_2).

Rather than employing a conventional SSIM Analysis, we utilized the MS-SSIM (Multi-Scale SSIM) [39] method, offering enhanced structural similarity evaluations between two images. When comparing the GAN to the preprocessed dataset using this method, the GAN-generated image dataset achieved a score of 0.8942. This score underscores that the GAN-produced lesion slices closely resemble the original images while ensuring the necessary variability ideal for augmentation. It suggests that upon appropriate color processing and accounting for other influencing factors, the GAN images hold significant potential to boost accuracy.



(a) Pre-processed (b) GAN-Augmented

(c) Traditional Augmentation (d) Combined

Fig. 5: Confusion matrix on test data: (a) Pre-processed Dataset, (b) GAN- Augmented dataset, (c) Traditional Augmentation Dataset, (d) Combined Dataset

V. Conclusion

In this study, we delved into data refinement, augmentation strategies, and the application of Generative Adversarial Networks (GANs) in the realm of medical images for lesion detection collected using the computed tomography scans.

Our findings illustrated that the compressed monochrome images could be improved upon with simple pre-processing steps to enhance image contrast. Data augmentation methods, both traditional and GAN-based, showcased a marked improvement in model performance over the original pre-processed dataset. While the traditionally augmented dataset marginally outperformed the GAN-augmented dataset, this difference is slight and may be attributed to the inherent randomness in training a deep neural network.

The effectiveness of the MobileNetV2-based classifier, trained on the augmented datasets, reiterates the significance of data augmentation in medical imaging. The combined dataset, which amalgamated both traditional and GAN-based augmentation, exhibited the most promising results, emphasizing the merit of diversity and volume in training data.

This research underscores the potential of synthetic augmentation. With the dual challenges of data scarcity and the critical need for accuracy in healthcare applications, techniques like GAN can pave the way for more robust radiology support systems. As we look to the future, refining these approaches with colored or textured images can lead to further advancements and concrete improvements in deep learning models tailored for medical imaging.

REFERENCES

[1] Yan, Ke, et al. "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning." Journal of Medical Imaging 5.3 (2018): 036501-036501.

[2] "NCI Dictionary of Cancer Terms." National Cancer Institute, www.cancer.gov/publications/dictionaries/cancer-terms/def/lesion. Accessed 21 May 2023.

[3] Brady, Adrian P. "Error and Discrepancy in Radiology: Inevitable or Avoidable?" Insights into Imaging, Feb. 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5265198/.

[4] National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/books/NBK9963/. Accessed 22 May 2023.

[5] Lee, June-Goo, et al. "Deep Learning in Medical Imaging: General Overview." Korean Journal of Radiology, 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5447633/.

[6] Cao, Jie, et al. "ScoreMix: A Scalable Augmentation Strategy for Training Gans with Limited Data." arXiv.Org, 4 Nov. 2022, arxiv.org/abs/2210.15137.

[7] Chan, Heang-Ping, et al. "Deep Learning in Medical Image Analysis." Advances in Experimental Medicine and Biology, 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7442218/.

[8] Deo, Rahul C. "Machine Learning in Medicine." Circulation, 17 Nov. 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC5831252/.

[9] Behzad Bozorgtabar a b 1, et al. "Informative Sample Generation Using Class Aware Generative Adversarial Networks for Classification of Chest Xrays." Computer Vision and Image Understanding, 10 May 2019, www.sciencedirect.com/science/article/abs/pii/S107731421930061X.

[10] Bissoto, Alceu, et al. "Skin lesion synthesis with generative adversarial networks." OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5. Springer International Publishing, 2018.

[11] Zlocha, Martin, Qi Dou, and Ben Glocker. "Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels." Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer International Publishing, 2019.

[12] Lavanya M, Kannan PM. Lung Lesion Detection in CT Scan Images Using the Fuzzy Local Information Cluster Means (FLICM) Automatic Segmentation Algorithm and Back Propagation Network Classification. Asian Pac J Cancer Prev. 2017 Dec 29;18(12):3395-3399. doi: 10.22034/APJCP.2017.18.12.3395. PMID: 29286609; PMCID: PMC5980900.

[13] Yan, Ke, et al. "Mulan: Multitask Universal Lesion Analysis Network for Joint Lesion Detection, Tagging, and Segmentation." arXiv.Org, 12 Aug. 2019, arxiv.org/abs/1908.04373.

[14] Yan, Ke, Mohammadhadi Bagheri, et al. "3D Context Enhanced Region-Based Convolutional Neural Network for End-to-End Lesion Detection." arXiv.Org, 29 July 2018, arxiv.org/abs/1806.09648.

[15] Lavanya, M, and P Muthu Kannan. "Lung Lesion Detection in CT Scan Images Using the Fuzzy Local Information Cluster Means (FLICM) Automatic Segmentation Algorithm and Back Propagation Network Classification." Asian Pacific Journal of Cancer Prevention: APJCP, 29 Dec. 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5980900/.

[16] "Lesion...What Does the Doctor Mean?" MedicineNet, 7 July 2004, www.medicinenet.com/script/main/art.asp?articlekey=9695.

[17] "Computed Tomography (CT)." National Institute of Biomedical Imaging and Bioengineering, www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct. Accessed 22 May 2023.

[18] RSNA Publications Online — Home, pubs.rsna.org/doi/full/10.1148/radiol.2018180547. Accessed 22 May 2023.

[19] Das, Angel. "Convolution Neural Network for Image Processing-Using Keras." Medium, 11 Jan. 2021, towardsdatascience.com/convolution-neural-network-for-image-processing-using-keras-dc3429056306.

[20] "A Comprehensive Guide to Convolutional Neural Networks - the Eli5 Way." Saturn Cloud Blog, 17 Apr. 2023

[21] Dinesh. "CNN vs MLP for Image Classification." Medium, 28 Nov. 2019, medium.com/analytics-vidhya/cnn-convolutional-neural-network-8d0a292b4498.

[22] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." Journal of Big Data 6.1 (2019): 1-48.)

[23] The Effectiveness of Data Augmentation in Image ... - Stanford University, vision.stanford.edu/teaching/cs231n/reports/2017/pdfs/300.pdf. Accessed 29 May 2023.

[24] Zhong, Zhun, et al. "Random erasing data augmentation." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020.)

[25] Pan, Zhaoqing, et al. "Recent progress on generative adversarial networks (GANs): A survey." IEEE Access 7 (2019): 36322-36333.

[26] Maayan Frid-Adar a, et al. "Gan-Based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification." Neurocomputing, 21 Sept. 2018, www.sciencedirect.com/science/article/abs/pii/S0925231218310749 preview-section-introduction.

[27] Awan, Abid Ali. "A Complete Guide to Data Augmentation." DataCamp, 23 Nov. 2022, www.datacamp.com/tutorial/complete-guide-data-augmentation.

[28] "What Is Data Augmentation? Techniques amp; Examples in 2023." AIMultiple, research.aimultiple.com/data-augmentation/. Accessed 21 May 2023.

[29] Brownlee, Jason. "A Gentle Introduction to Generative Adversarial Networks (Gans)." MachineLearningMastery.Com, 19 July 2019, machinelearningmastery.com/what-are-generative-adversarial-networks-gans/.

[30] Singh, Anushka. "Gan: Generative Adversarial Network." Medium, 15 July 2021, medium.com/analytics-vidhya/gan-generative-adversarial-network fbef2a96e183: :text=GANs%20are%20a%20clever%20way,)%20or%20fake%20(generated).

[31] Patil, Pratik. "What Are Gan's? An Introduction to Generative Adversarial Networks." Medium, 31 May 2020, medium.com/analytics-vidhya/what-are-gans-an-introduction-to-generative-adversarial-networks-92efd00623d5.

[32] "Loss Functionsnbsp; —nbsp; Machine Learningnbsp; —nbsp; Google for Developers." Google, developers.google.com/machine-learning/gan/loss. Accessed 22 May 2023.

[33] Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in neural information processing systems 30 (2017).

[34] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[35] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[36] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[37] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[38] Arjovsky, Martin, et al. "Wasserstein Gan." arXiv.Org, 6 Dec. 2017, arxiv.org/abs/1701.07875.

[39] Wang, Zhou, Eero P. Simoncelli, and Alan C. Bovik. "Multiscale structural similarity for image quality assessment." The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003. Vol. 2. Ieee, 2003.

[40] Antoniou, Antreas, Amos Storkey, and Harrison Edwards. "Data augmentation generative adversarial networks." arXiv preprint arXiv:1711.04340 (2017).

[41] Chlap, Phillip, et al. "A review of medical image data augmentation techniques for deep learning applications." Journal of Medical Imaging and Radiation Oncology 65.5 (2021): 545-563.