International Journal of Multidisciplinary Innovation and Research Methodology (IJMIRM)
ISSN: 2960-2068, Volume 2, Issue 4, Oct-Dec, 2023, Available online at: https://ijmirm.com

Generative Adversarial Networks for Anomaly Detection in Medical Images

Bhuman Vyas[1], Rajashree Manjulalayam Rajendran[2]

[1]Credit Acceptance Corporation, USA
[2]HomeASAP LLC, USA

ABSTRACT

In computer vision, anomaly detection (AD) is a challenging task. AD presents additional difficulties, especially in the realm of medical imaging, for several reasons, one of which being the dearth of ground truth (annotated) data. AD models built on generative adversarial networks (GANs) have advanced significantly in the last several years. Their usefulness in biological imaging is still not well understood, though. In this study, we provide an overview of the use of GANs for AD and a detailed analysis of the difficulties faced in implementing the most advanced GAN-based AD techniques for biomedical imaging. Additionally, we have explicitly examined the benefits and constraints of AD approaches on medical image datasets, conducting tests on 2 medical imaging datasets from various modalities, organs, and tissues using 3 AD methods. We examined the outcomes from the perspectives of both data and models, given the strikingly disparate results obtained in these studies. The outcomes demonstrated that no technique could consistently identify anomalies in medical imaging. A few of the phenomena that have a significant influence on the AD models' performance include the quantity of training samples, the subtlety of the anomaly, and the anomaly's distribution throughout the images. We also anticipate significant research paths and offer recommendations for the application of AD models in medical imaging.

Keywords: Artificial intelligence, Anomaly Detection, Computer Vision, Generative Adversarial Networks (GANs), Medical Imaging, Unsupervised Learning, Deep Learning, Auto encoders (AEs), Variational Auto encoders (VAEs), GAN-Based Anomaly Detection, CT Images, Mammogrpahy Images.

INTRODUCTION

Finding data samples (also known as out-of-distribution samples) that do not fit the overall data distribution is the main goal of anomaly detection (AD) [1]. Anomalies can arise from a multitude of reasons, such as noise in the data collection process or novel or undiscovered elements in the context that was collected [2]. Finding structural and functional abnormalities in exposed organs and tissues is the general goal of medical imaging, or AD. However, challenges arise because gathering annotations, or labels, is often costly, time-consuming, and impossible without a reliable ground truth [3]. Because of this, unsupervised and semi-supervised techniques have drawn a lot of interest in medical imaging in recent years. A more recent development in unsupervised learning techniques is the ability of generative adversarial networks (GANs) [4] to estimate the distribution of extremely complicated medical imaging data. Therefore, in an ideal world, these techniques would not be significantly hampered by imbalanced datasets, a known problem in the field of medical imaging [5]. Because of these qualities, GANs have become a well-reputed option for creating AD techniques in applications involving medical imaging.

Currently, one of the most widely used unsupervised AD techniques is GAN-based AD [6]. The primary function of GANs is to learn a dataset's distribution, which enables them to create new samples that are difficult for the discriminator to distinguish from the original dataset. But in GAN architectures, the discriminator plays a more important role in anomaly detection than the generator, along with structural alterations. GAN-based AD has been created in numerous studies [7] using natural datasets and medical pictures. Determining the reliability of contemporary unsupervised AD methods—particularly GAN-based AD—on medical images with varying abnormalities, pathologies, modalities, and resolutions is a complex and tough task, despite their partial success and promising outcomes. However, it is challenging to evaluate how well these techniques function in terms of identifying different anomalies in medical imaging and selecting the most effective technique. This is due to the fact that the majority of these techniques have been tried and evaluated on real datasets or on medical imaging with particular pathologies, modalities, and resolutions. The goal of this work is to demonstrate and contrast the effectiveness of current unsupervised AD, particularly GAN-based AD for medical image datasets.

ANOMALY DETECTION

For several years, data mining and statistics experts have been drawn to the dynamic study field of finding data outliers, also known as anomalies, novelties, and out-of-distribution data [1], [8]. The main causes of this are the difficulties in the AD process as well as the important knowledge and insights that can be gained from dataset abnormalities [9].Data points that do not fit into the distribution of the majority of the dataset that is known to be normal are considered anomalies in technical terms [9]. Furthermore, it can be challenging to classify anomalous data samples as normal because they are typically uncommon and unfamiliar. In other words, normal samples are thought to all approximately follow the same distribution, whereas anomalous samples are thought to originate from distinct distributions. For various datasets, like medical images, this assumption might not always be true. Finding aberrant data in medical imaging is a difficult endeavour because there is a wide range of normal cases. To be more precise, an AD model experiences reduced sensitivity due to a high number of false positive samples, as well as lower specificity due to a high number of false positive samples. Notwithstanding the various obstacles, possessing a model capable of identifying irregularities or novelties in medical images can be crucial in offering a decision support system for the identification of unidentified or uncommon illnesses [10].

Conventional unsupervised techniques include One-Class Support Vector Machines (OC-SVM) [12] and Support Vector Data Description (SVDD) [11]. These techniques look for a hyperplane with normal training samples in order to identify rich normal features [13]. On complex, high-dimensional datasets, however, their performance is said to suffer [14]. While it may not always be the best method, choosing features based on expert judgement can help reduce the dimensionality of data [15]. Alternatively, different feature extraction and selection techniques, such as principal component analysis (PCA) [16], multidimensional scaling (MDS) [17], and deep auto-encoder networks, can be applied. Without requiring any prior knowledge from human experts, end-to-end training of more discriminative feature representations from the usual input photos can be accomplished through the application of deep learning (DL)-based AD approaches. Due to these essential characteristics, DL-based AD techniques are more resilient and generalizable [15], especially when dealing with high dimensionality datasets like medical images, where manually developed feature engineering pipelines are insufficient and excessively time-consuming.

The most widely used deep AD technique depends on understanding the distribution of normal images and deriving latent representations from them in order to accurately rebuild normal images [20]. Only normal images can be reconstructed using techniques such autoencoders (AEs) [21], [22], variational autoencoders (VAEs) [23], [24], [25], and GANs [26], [27], [28]. As a result, poor reconstruction is anticipated for anomalous photos. In order to better identify anomalous samples, information from latent space [30], [31], and discriminator [32] can also be used. As a result, the reconstructed error can be thought of as the anomaly score [29].Reconstruction-based AD techniques can be troubled by issues such mode collapse, instability, non-convergence, and computational expense for picture reconstruction, while being popular and intuitive as well [33].

OVERVIEW OF GAN-BASED

In order to produce new samples that convincingly appear to come from the same (source/input) dataset, generative modeling is an unsupervised machine learning approach that automatically finds and learns the underlying patterns in the source dataset [32]. GANs are a subset of generative modeling techniques that are built on convolutional or deep neural networks. A generator network "" and a discriminator network "" are the two sub-modules that a GAN model [4] uses to reformulate the training process of a generative model as a supervised learning job. While the discriminator module works to categorize the generated samples as Real (originating from the source data domain) or Fake (fabricated samples), the generator module creates new samples. These two modules are trained together in the manner of an adversarial zero-sum game in order to develop a GAN model [32]. In other words, in a competitive setting, the discriminator network, the opponent of the generator network, faces off against it. The discriminator network, the competitor of the generator network, attempts to distinguish between the generated samples and the training dataset samples while the generator network continuously generates bogus examples [4]. The ultimate objective of a GAN is to create new (fake) samples from the observed data that the discriminator network cannot tell apart from the genuine samples that come straight from the input dataset. A generator network gains knowledge during training in the form of a latent vector, which is a projection of the actual data distribution into a latent space. Following a successful training session, the learned latent vector is utilized to produce new images that closely resemble the input dataset's distribution. Furthermore, as the generator network is being trained, it continuously gets signals from the discriminator, which determines whether the generated samples are sufficiently similar to the samples from the source dataset, and adjusts its parameters accordingly. In order for the generator module to employ these projected points in the latent space to create new samples that are sufficiently similar to the

distribution of the input samples, the learned latent vector is therefore referred to as a compression of the observed input data during training [32].

Consequently, especially with complicated datasets, GANs are able to capture the broad idea of most of the input dataset. It is assumed that a well-trained generator can produce more normal results than abnormal ones, assuming that most of the input data correspond to the Normal samples. This characteristic of the GAN's discriminator and generator modules can be utilized to identify anomalous occurrences within a dataset. Typically, an anomaly score is measured to accomplish this [28]. Many GAN-based AD techniques are now widely employed in a variety of fields, including infrastructure, industry, medicine, and other fields .

Generally speaking, GANs are made to produce artificial data that appears realistic, whether it is normal or abnormal. Therefore, the task at hand is to figure out how to apply GANs to AD directly. As previously stated, a trained GAN's generator module's latent space holds a high-level conceptual representation of the training data that the network is trained on. Finding the ideal latent space for the sample test image during the differencing process is the primary difficulty when using GAN on an AD job, though. This is mostly due to the lack of inverse mapping techniques from the image space to the latent space in standard GAN systems. The latent space and the image space have a significant association. Stated differently, one of the high-level characteristics of the data conveyed to the latent space is the distinction between normal and aberrant pictures [28]. In order to address the inverse mapping from the image space to the latent space, there is a rise in effort.

The goal of GANomaly [31] is to jointly learn latent and image representations. The suggested model is composed of an encoder to learn the latent space representations, an adversarial auto-encoder (Encoder-Decoder) as a generator to learn genuine picture representations, and a discriminator sub-module to distinguish between real and fake images. To address learning instability, GANomaly leverages feature matching and a DCGAN architecture. In order to produce realistic-looking images and optimize the encoding process, the model is trained using three separate losses: an adversarial loss, a contextual loss, and an encoder loss. In this case, the encoder loss is used to define the anomaly score. Another GAN-based AD model that has been suggested is called f-AnoGAN [32]. It employs a WGAN to capture more smooth representations by using an encoder for inverse mapping that has been trained using the generator and the discriminator following the training phase. While using more networks, such as encoders or decoders for GANs, may result in a more effective reconstruction of images, there are drawbacks to this as well. For example, training more networks means learning more parameters, which leads to the well-known problems of overfitting or data memorization [32]. The efficacy of GAN as an unsupervised anomaly detection method has been demonstrated. It has conquered several obstacles, including imbalanced datasets, insufficiently labeled datasets, and a deficiency of anomalous data. In the realm of medical imaging, it is particularly challenging to overcome these restrictions. It is still necessary to look at the success rate of GAN-based AD techniques on medical images. In order to further explore their potential and overcome their shortcomings in medical imaging anomaly identification, we carried out a number of experiments in the following section utilizing GAN-based AD techniques for anomaly detection on a variety of publically accessible medical imaging datasets. More precisely, we used two public imaging datasets—a headhaemorrhage CT images dataset and a mammography images dataset,—to test three different GAN-based AD techniques.

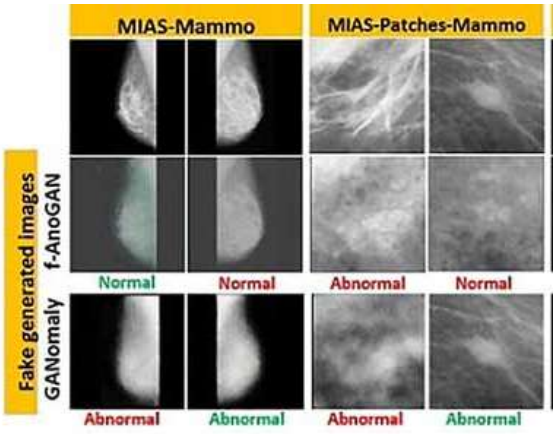

Figure 1: Normal and abnormal examples of bimedical image datasets (the first row) and their reconstructed images by f-AnoGAN (the second row) and GANomaly (the third row) along with their predicted labels

METHODS

In this part, we discuss the experiments we conducted using two publicly accessible medical images datasets to compare three DL-based AD approaches. We examined these three models' performances from two key angles: a model-centric and a data-centric one. To ensure a fair comparison of the approaches, these two datasets were chosen, each with unique characteristics. The amount of samples, picture sizes, interested organ or tissue, disease type, modality type, and features of the abnormality to be detected vary throughout the target datasets. On the other hand, three cutting-edge AD models—each with unique conceptual and structural features—are chosen for comparison [32]. Our intention in presenting these comparisons was to shed light on the potential obstacles that AD techniques may encounter in the field of medical imaging, from several angles.

A. Collections of Data

We collected two medical imaging datasets from various modalities and concentrated on various organs and tissues for our work. The brief description of these datasets and their specifications is as follows:

1) CT Images of Head Haemorrhage (Head-CT)

In this dataset, there are 100 2D CT single slices of a normal head and 100 2D CT image slices with evident brain haemorrhage in each. Every 2D slice is associated with a distinct patient. All 100 slices have the haemorrhage irregularity visible to the naked eye. The dimensions and pixel sizes of these photos vary [33].

2) MIAS-Patches-MAMMO & Mammographic Image Analysis Society (MIAS-MAMMO)

The three kinds of mammography images in this collection [33] are normal, benign, and malignant. Therefore, we regarded both benign and malignant classes as abnormal, leaving the remaining classes as normal, in order to build the dataset for AD. This resulted in 207 normal images and 115 abnormal images with dimensions of 1024 x 1024. In terms of visual observability, this dataset differs from the previously described one because the irregularities in the photos are relatively modest. Based on the provided coordinates of the region of interest, an image patch of 120 by 120 pixels was extracted for each sample from the MIAS-Mammo dataset to create the MIAS-PATCHES-MAMMO dataset. That equates to 117 patches of aberrant samples and 207 patches of normal images.Three unsupervised AD methods were trained on the medical images and they were trained on normal samples and on portion containing both normal and anomalous samples. In all these settings, the training set contained normal images and the test set contained both normal and anomalous images.

RESULTS AND DISCUSSION

The obtained results demonstrate that the models' performance on the Head-CT dataset, is quite low. Over fitting occurs when the discriminator module merely memorizes the training data during training because the training dataset is not large enough compared to the AD model's capability. As a result, the model will break. As a result, the produced photos become of worse quality. This might be a significant setback, particularly in the field of medical imaging where gathering data is a costly endeavor.

Data augmentation is one potential way to address this issue and raise the resilience and performance of GANs. Research has shown that augmenting data for both generated and real images can improve the performance of GANs. However, if it had only been applied to real photos, the result might not have been the same. This suggests that data augmentation can help to improve the performance of the AD models even if it is performed only on the input data and not on the constructed data.When using GANs, the quality of the images produced by both normal and abnormal samples can be used as a heuristic to assess the AD model's performance and, in turn, provide an explanation for the model's final conclusion. The figure below shows some normal and abnormal samples from the Head-CT datasets, as well as the corresponding reconstructions made with the two GAN-based AD models that were looked at. CT provides a comprehensive, higher contrast image of the cortical bones. As was previously mentioned, it is evident that when training sizes are less, GANs may not be able to learn all the intricacies and will instead just be able to provide the general structure of the brain. For example, the cortical bone is well rebuilt in the CT pictures (Head-CT dataset), mostly because of the strong contrast of the region [34]. Notably, GAN was able to learn and reconstruct images with better information, especially normal samples, by using data augmentation just on genuine samples.

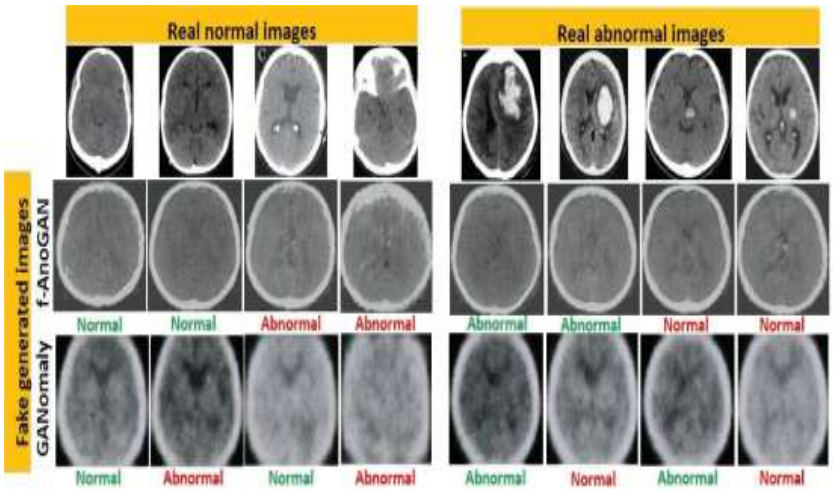

Figure 2: Examples of the Head-CT dataset (the first row) and their reconstructed images by f-AnoGAN (the second row) and GANomaly (the third row) along with their predicted labels.

MIAS-MAMMO and MIAS-Patches-MAMMO Datasets

The MIAS-Mammo dataset yielded somewhat poor performance for all three AD approaches, even with uniformly high image resolutions. Furthermore, in order to better handle the dispersion of anomalies in the breast tissue, we isolated patches from the regions of interest in the pictures due to the high resolution of the images in this dataset. Although we created the patches dataset with the intention of improving the AD models' accuracy, the results we got were still quite poor. One possibility could be that the model's capacity to fully simulate tissue composition is limited when feeding the AD networks with extracted patches instead of the entire slice during training, and global information may not be taken into account. On the MIAS-Patches-Mammo dataset, the Multi-KD technique has performed marginally better than the GAN-based algorithms. Comparable to the findings published in [33] on this method's excellent texture AD performance on the MVTecAD dataset. The MIAS-Mammo and MIAS-Patches-Mammo image samples and their reconstructions are demonstrated with a few image examples. The diagnosis of aberrant tissue in breast cancer using mammography pictures is error-prone, in contrast to brain abnormalities which are generally detectable. Breast cancer is difficult to detect and characterize due to variation in lesion appearance. Size, form, density, borders, subtlety, and location of lesions are among the critical variables that impact cancer diagnosis. Overall, these methods have not been very successful in detecting the abnormalities in breast cancer lesions because of their intricacy and closeness to normal tissue.In these situations, anomalies are conceptually and semantically related to a normal distribution and normality. Actually, the main premise of most AD approaches does not align with the different definition of abnormality used here. Recent research has tackled the problem of "near novelty" detection.

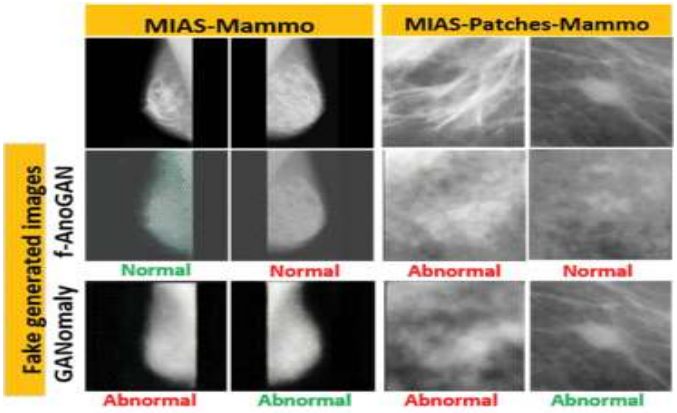

Figure 3: Normal and abnormal examples of the MIAS-Mammo, MIAS-Patches-Mammo ,dataset (the first row) and their reconstructed images by f-AnoGAN (the second row) and GANomaly (the third row) along with their predicted labels.

CONCLUSION

Using medical images, this study illustrated how unreliable modern unsupervised DL-based AD techniques are. In order to accomplish this, we used three unsupervised DL-based AD techniques with various loss functions and structural variations on two sets of medical picture datasets that varied in terms of diseases, abnormalities, modalities, and sample counts. As a result, we created an essentially exhaustive comparison between various techniques and demonstrated how their efficacy could differ depending on the type of medical imaging, which provided some useful information. The current issues were thoroughly examined from the perspectives of the model and the data. Overall, none of the techniques worked well enough to be applied in clinical settings. We ascribe this to difficulties with the generalizability of AI techniques as well as the variety of abnormalities in the field of biomedical imaging. Our research demonstrated that, in addition to the processes behind the chosen AD techniques, the effects of aberrant characteristics should be carefully taken into account. It is advised to take into account a number of factors when designing and developing AD algorithms for medical images, including the degree of subtlety of the anomaly, its spread, tissue-related anomalies like blood cell cancers or breast cancer in mammography images, and imaging modalities where the contrast difference between the abnormal and normal regions is comparable. Therefore, there is an urgent need for greater research and the implementation of more reliable, trustworthy, and generalizable AD models.

REFERENCES

[1]. Aggarwal, Charu C., and Charu C. Aggarwal. *An introduction to outlier analysis*. Springer International Publishing, 2017.
[2]. Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A unifying review of deep and shallow anomaly detection." *Proceedings of the IEEE* 109, no. 5 (2021): 756-795.
[3]. Ching, Travers, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero et al. "Opportunities and obstacles for deep learning in biology and medicine." *Journal of The Royal Society Interface* 15, no. 141 (2018): 20170387.
[4]. Creswell, Antonia, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. "Generative adversarial networks: An overview." *IEEE signal processing magazine* 35, no. 1 (2018): 53-65.
[5]. Roy, Sudipta, Tanushree Meena, and Se-Jung Lim. "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine." *Diagnostics* 12, no. 10 (2022): 2549.
[6]. Pang, Guansong, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. "Deep learning for anomaly detection: A review." *ACM computing surveys (CSUR)* 54, no. 2 (2021): 1-38.
[7]. Di Mattia, Federico, Paolo Galeone, Michele De Simoni, and Emanuele Ghelfi. "A survey on gans for anomaly detection." *arXiv preprint arXiv:1906.11632* (2019).
[8]. Grubbs, Frank E. "Procedures for detecting outlying observations in samples." *Technometrics* 11, no. 1 (1969): 1-21.
[9]. Chalapathy, Raghavendra, and Sanjay Chawla. "Deep learning for anomaly detection: A survey." *arXiv preprint arXiv:1901.03407* (2019).
[10]. Bi, Wenya Linda, Ahmed Hosny, Matthew B. Schabath, Maryellen L. Giger, Nicolai J. Birkbak, Alireza Mehrtash, Tavis Allison et al. "Artificial intelligence in cancer imaging: clinical challenges and applications." *CA: a cancer journal for clinicians* 69, no. 2 (2019): 127-157.
[11]. Turkoz, Mehmet, Sangahn Kim, Youngdoo Son, Myong K. Jeong, and Elsayed A. Elsayed. "Generalized support vector data description for anomaly detection." *Pattern Recognition* 100 (2020): 107119.
[12]. Scholkopf, Bernhard, and Alexander J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2018.
[13]. Salehi, Mohammadreza, Hossein Mirzaei, Dan Hendrycks, Yixuan Li, Mohammad Hossein Rohban, and Mohammad Sabokrou. "A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges." *arXiv preprint arXiv:2110.14051* (2021).
[14]. Chalapathy, Raghavendra, Aditya Krishna Menon, and Sanjay Chawla. "Anomaly detection using one-class neural networks." *arXiv preprint arXiv:1802.06360* (2018).
[15]. Hosny, Ahmed, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo JWL Aerts. "Artificial intelligence in radiology." *Nature Reviews Cancer* 18, no. 8 (2018): 500-510.
[16]. Jolliffe, Ian T., and Jorge Cadima. "Principal component analysis: a review and recent developments." *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 374, no. 2065 (2016): 20150202.
[17]. Ghojogh, Benyamin, Ali Ghodsi, Fakhri Karray, and Mark Crowley. "Multidimensional scaling, sammon mapping, and isomap: Tutorial and survey." *arXiv preprint arXiv:2009.08136* (2020).

[18]. Sullivan, Daniel C., Nancy A. Obuchowski, Larry G. Kessler, David L. Raunig, Constantine Gatsonis, Erich P. Huang, Marina Kondratovich et al. "Metrology standards for quantitative imaging biomarkers." *Radiology* 277, no. 3 (2015): 813-825.

[19]. O'connor, James PB, Eric O. Aboagye, Judith E. Adams, Hugo JWL Aerts, Sally F. Barrington, Ambros J. Beer, Ronald Boellaard et al. "Imaging biomarker roadmap for cancer studies." *Nature reviews Clinical oncology* 14, no. 3 (2017): 169-186.

[20]. Yang, Jie, Ruijie Xu, Zhiquan Qi, and Yong Shi. "Visual anomaly detection for images: A survey." *arXiv preprint arXiv:2109.13157* (2021).

[21]. Davletshina, Diana, Valentyn Melnychuk, Viet Tran, Hitansh Singla, Max Berrendorf, Evgeniy Faerman, Michael Fromm, and Matthias Schubert. "Unsupervised anomaly detection for X-ray images." *arXiv preprint arXiv:2001.10883* (2020).

[22]. Nguyen, Duc Tam, Zhongyu Lou, Michael Klar, and Thomas Brox. "Anomaly detection with multiple-hypotheses predictions." In *International Conference on Machine Learning*, pp. 4800-4809. PMLR, 2019.

[23]. Venkataramanan, Shashanka, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. "Attention guided anomaly localization in images." In *European Conference on Computer Vision*, pp. 485-503. Cham: Springer International Publishing, 2020.

[24]. Pham, Dang, and Tuan Le. "Auto-encoding variational bayes for inferring topics and visualization." *arXiv preprint arXiv:2010.09233* (2020).

[25]. Liu, Wenqian, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J. Radke, and Octavia Camps. "Towards visually explaining variational autoencoders." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8642-8651. 2020.

[26]. Sabokrou, Mohammad, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. "Adversarially learned one-class classifier for novelty detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3379-3388. 2018.

[27]. Pidhorskyi, Stanislav, Ranya Almohsen, and Gianfranco Doretto. "Generative probabilistic novelty detection with adversarial autoencoders." *Advances in neural information processing systems* 31 (2018).

[28]. Akçay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection." In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2019.

[29]. Gong, Dong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1705-1714. 2019.

[30]. Abati, Davide, Angelo Porrello, Simone Calderara, and Rita Cucchiara. "Latent space autoregression for novelty detection." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 481-490. 2019.

[31]. Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "Ganomaly: Semi-supervised anomaly detection via adversarial training." In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, pp. 622-637. Springer International Publishing, 2019.

[32]. Paudice, Andrea, Luis Muñoz-González, Andras Gyorgy, and Emil C. Lupu. "Detection of adversarial training examples in poisoning attacks through anomaly detection." *arXiv preprint arXiv:1802.03041* (2018).

[33]. Suckling, J., J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, and S. Kok. "Mammographic image analysis society (MIAS) database v1. 21 [Dataset]." (2015).

[34]. Esmaeili, Marzieh, AmirhoseinToosi, Arash Roshanpoor, Vahid Changizi, Marjan Ghazisaeedi, Arman Rahmim, and Mohammad Sabokrou. "Generative Adversarial Networks for Anomaly Detection in Biomedical Imaging: A Study on Seven Medical Image Datasets." *IEEE Access* 11 (2023): 17906-17921.