Towards Multimodal Deep Learning for Medical Image Analysis: Developing a Cross-Modality Data Augmentation Technique by Interpolating Modality-Specific Characteristics in Medical Images

Bachelor Thesis

Bachelor of Science in Computer Science: Software Systems Science

Julius Stutz
for the Alzheimer's Disease Neuroimaging Initiative*

October 20, 2024

Supervisor:

1st: Prof. Dr. Christian Ledig
2nd: Sebastian Dörrich M. Sc.

Chair of Explainable Machine Learning
Faculty of Information Systems and Applied Computer Sciences
Otto-Friedrich-University Bamberg

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

Training deep learning (DL) models requires large amounts of data, posing a challenge in the medical domain due to the particular sensitivity of medical data. To mitigate this data scarcity, especially in medical imaging, data augmentation is commonly employed to synthetically enlarge training data. However, existing solutions either lack specialization in the medical domain or have high computational requirements.

Therefore, this work proposes the Cross-Modality Data Augmentation (*CMDA*), an adaptive real-time data augmentation dealing with limited medical data. By translating images between different medical imaging modalities, it specifically addresses the cross-modality shift. To achieve this, it synthesizes new training samples that represent the target modality's distribution, using existing images from a given modality.
CMDA consists of four types of augmentations, focusing on color, artifacts, spatial resolution, and noise respectively. It supports the clinically relevant modalities positron emission tomography (PET), magnetic resonance imaging (MRI), and computed tomography (CT) and furthermore ensures compatibility with common data augmentations.
Quantitative experiments evaluated *CMDA*'s potential in improving model robustness and generalization. This was done by comparing the classification performance of NNs trained with *CMDA* and other commonly-used data augmentations. Results showed minimal improvements ($<2\%$) across all performance metrics in some experiments and a substantial ($<-9\%$) decrease in others. Qualitative assessments indicate *CMDA*'s success in aligning augmented images with the target modality. They compared *CMDA*-augmented images to images of the original and target modality by various approaches, with two experiments showing an average alignment improvement of 23.5%.

Despite remaining challenges in enhancing model generalization, *CMDA* demonstrates its potential in addressing the data scarcity in medical imaging. As such, it can be integrated into existing data augmentation pipelines and serve as a foundation for further research in cross-modality translation.

Abstract

Das Training von Deep-Learning (DL) Modellen erfordert große Datenmengen, was im medizinischen Bereich auf Grund der besonderen Sensibilität medizinischer Daten eine Herausforderung darstellt. Um dieser Datenknappheit, insbesondere in der medizinischen Bildgebung, entgegen zu wirken, werden üblicherweise Data Augmentations zur synthetischen Vergrößerung der Trainingsdaten eingesetzt. Existierende Lösungen hierfür sind jedoch entweder nicht auf den medizinischen Bereich spezialisiert oder erfordern viel Rechenleistung.

Deshalb wird in dieser Arbeit die Cross-Modality Data Augmentation (*CMDA*) vorgestellt, eine adaptive Echtzeit-Data Augmentation, die sich dem Problem begrenzter medizinischer Daten annimmt. Durch die Übersetzung von Bildern zwischen verschiedenen medizinischen Bildgebungsmodalitäten geht sie speziell auf den cross-modality shift ein. Um dies zu erreichen, werden vorhandene Bilder einer bestimmten Modalität verwendet um neue Trainingsbilder zu synthetisieren, die die Verteilung der Zielmodalität repräsentieren.
CMDA besteht aus vier kleineren Augmentierungen, die sich jeweils mit Farbe, Bildartefakten, räumlicher Auflösung und Rauschen befassen. Es unterstützt die klinisch relevanten Modalitäten Positronen-Emissions-Tomographie (PET), Magnetresonanztomographie (MRI) und Computertomographie (CT) und gewährleistet darüber hinaus die Kompatibilität mit gängigen Data Augmentations.
In quantitativen Experimenten wurde das Potenzial von *CMDA* zur Verbesserung der Modellrobustheit und Generalisierung untersucht. Dazu wurde die Klassifizierungsleistung von NNs, die mit *CMDA* trainiert wurden, mit anderen häufig verwendeten Data Augmentations verglichen. Die Ergebnisse zeigten in einigen Experimenten minimale Verbesserungen ($<2\%$) bei allen leistungsmessenden Metriken, in anderen Experimenten jedoch einen erheblichen Leistungsabfall ($<-9\%$). Qualitative Experimente deuten auf den Erfolg von *CMDA* bei der Angleichung von augmentierten Bildern an die Zielmodalität hin. Sie verglichen *CMDA*-augmentierte Bilder mit Bildern der Original- und Zielmodalität durch verschiedene Ansätze, wobei zwei Experimente eine durchschnittliche Verbesserung der Angleichung von 23,5% zeigten.

Trotz der verbleibenden Herausforderungen bezüglich der Verbesserung der Modellgeneralisierung, demonstriert *CMDA* sein Potenzial bei der Bekämpfung der Datenknappheit in der medizinischen Bildgebung. Als solches kann es in bestehende Data Augmentation Pipelines integriert werden und als Grundlage für weitere Forschung im Bereich der modalitätsübergreifenden Bildübersetzung dienen.

Acknowledgements

Above all, I'd like to thank everyone in my social environment for the interesting discussions and the emotional support.

Also, thanks to Sebastian for being a chill supervisor who always had fresh ideas and constructive criticism.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The results shown here are in whole or part based upon data generated by the TCGA Research Network: `http://cancergenome.nih.gov/`.

Contents

List of Figures

List of Tables

List of Acronyms

ADNI	Alzheimer's Disease Neuroimaging Initiative
AUPR-IN	Area Under the Precision-Recall Curve for In-Distribution-Samples
AUPR-OOD	Area Under the Precision-Recall Curve for Out-Of-Distribution-Samples
AUROC	Area Under the Receiver Operating Curve
CMDA	Cross-Modality Data Augmentation
CT	Computed Tomography
DA	Domain Adaptation
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DG	Domain Generalization
FID	Fréchet inception distance
FPR95TPR	False Positive Rate at 95% True Positive Rate
GAN	Generative Adversarial Network
GLCM	Gray Level Co-Occurrence Matrix
HU	Hounsfield Unit
IN	In-Distribution
MAE	Mean Absolute Error
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NLP	Natural Language Processing
NST	Neural Style Transfer
NN	Neural Network
OOD	Out-Of-Distribution
PCA	Principal Component Analysis
PET	Positron Emission Tomography
ResNet	Residual Neural Network
RMSE	Root Mean Square Error
RSNA	Radiological Society of North America
TCGA-BLCA	The Cancer Genome Atlas Urothelial Bladder Carcinoma Collection
TL	Transfer Learning
VAE	Variational Autoencoder
ViT	Vision Transformer

1 Introduction

Due to its ability to address complex challenges through data-driven approaches, DL has had substantial influence across various domains (Gheisari et al., 2023). Yet in healthcare, the full potential of DL is often limited by the scarcity of high-quality medical data. As a subtype of machine learning (ML), DL uses multilayered neural networks (NNs) to learn how to solve a task directly from the data. Therefore, large-scale data processing is needed, where the data can be labeled (supervised DL) or unlabeled (unsupervised DL). Instead of training each layer sequentially one after the other, end-to-end learning models are employed, where all parts of the network are trained simultaneously. These models are then able to automatically extract features from the inputs, which allows them to detect patterns in high-dimensional data (LeCun et al., 2015). Apart from the previously mentioned applications, it enabled major advancements in natural language processing (NLP), computer vision, and speech recognition (Goodfellow et al., 2016). Because NNs learn by themselves, minimal engineering is needed from the programming side. This makes DL versatile across many more domains.

One of these is the formerly stated medical domain, specifically medical imaging. There it can be utilized to detect patterns in complex image data, and consequently to detect, classify, and segment diseases. DL is also applicable for almost all anatomical areas, most prominently the brain, eye, abdomen, chest, or in digital pathology (Litjens et al., 2017). This further supports the clinical workflow by assisting with personalized treatment and clinical decision support systems (Shen et al., 2017). These benefits lead to enhanced diagnostic accuracy and the active support of medical staff.

However, the aforementioned NNs need lots of data to learn enough to actually be implemented into real clinical scenarios. Thus, the scarcity of medical imaging data still remains a major challenge to overcome. This scarcity is due to a manifold of reasons, most prominently because of privacy concerns as patients often don't want their medical data to be published, even if de-identified (Kagadis et al., 2013; Ziller et al., 2021; Vizitiu et al., 2019; Bansal et al., 2022). The General Data Protection Regulation (GDPR) and other comparable confidentiality regulations also restrict the sharing of patient data (Zhang et al., 2023). Meanwhile, legal obligations and ethical considerations like specifically signing an informed consent, not publishing data due to research ethics, or only providing request-based access, further reduce available data (Larson et al., 2020). The technical side also poses limitations due to high costs, time, and efforts required to gather high-quality medical data, specifically images. This often exceeds available resources (Hendee et al., 2010; Bansal et al., 2022) or results in scarce and weak (imprecise) image annotations (Tajbakhsh et al., 2020). Furthermore, accessible data is often not consistent due to a domain and a cross-modality shift. In this case, domain shift describes data inconsistencies owing to non-standardized equipment and imaging protocols (Smith et al., 2021). Meanwhile, the cross-modality shift is concerned with differences among images originating from different imaging techniques (modalities) (Chen et al., 2019).

1.1 Motivation

In comparison to domain shift, the cross-modality shift between images is severe (Chen et al., 2019). Since different modalities utilize different physics, scans of the same patient may thus look quite dissimilar for most modalities. These wide gaps in appearance make it particularly challenging for NNs to apply multimodal learning (learning that combines information from different data sources) (Liang et al., 2024; Gat et al., 2021; Ngiam et al., 2011). Therefore, a lot of data is needed to cross said gaps and allow the model to better generalize across modalities (Schmidt et al., 2018). In this context, generalizing refers to a model making accurate predictions on unseen data. But as mentioned before, medical imaging data is scarce, often making adequate multimodal learning impossible.
To address this problem, a commonly employed ML technique is data augmentation. Hereby, limited training data is artificially enlarged by synthetic sample generation. Simple image manipulations like rotation, flipping, blurring, or color changes are one option for this. However, these are domain-independent and thus fail to incorporate medical characteristics, constraining their full potential for this specific domain (Ratner et al., 2017). Another approach are DL-based data augmentations. They enable visually stunning image translation between modalities or generation of completely new data. Yet, existing implementations all suffer from the high computational requirements that DL-based methods bring along, as this limits their applicability for real-time augmentation (Mikołajczyk and Grochowski, 2018).
A cross-modality data augmentation that ensures real-time usability, retains medical relevance, and generates diverse training data is thus needed to overcome existing problems, address data scarcity, and support multimodal learning.

1.2 Contribution

Existing data augmentations lack modality-specific transformations that can be employed during runtime and help to combat the cross-modality shift. Therefore, this work proposes *CMDA* as an adaptive, real-time data augmentation to address said problems. The objective is to synthesize new training samples that better represent the distribution of the target modality, while also helping to improve the generalization performance of deep learning algorithms.
To assess *CMDA*'s potential in improving the robustness and generalization capabilities of models, quantitative experiments are carried out. They compare performance metrics of NNs trained with *CMDA* and other data augmentations and also partially combine them to analyze compatibility.
Furthermore, qualitative experiments evaluate the visual image quality and characteristics of augmented images. This is done by comparing CMDA-augmented images to images of the original and target modality. To provide information on how well *CMDA* works across anatomical structures, these experiments are conducted for two different anatomies.

2 Related Work

Since the advent of DL, the importance of data augmentation has risen, as its many benefits, such as more diverse data, improved generalization performance, and reduced overfitting (Shorten and Khoshgoftaar, 2019), have made it more relevant than ever. The domain of medical imaging is no exception to this, with certain established standards and interesting approaches.
PyTorch's transforms, imgaug, and Albumentations (Buslaev et al., 2020) are among the most popular basic data augmentation libraries used in medical imaging. They allow users to augment their data at runtime but only with modality-unspecific augmentations like flipping or rotating. In comparison, they therefore are not tailored to medical imaging while *CMDA* addresses modality unique characteristics and preserves medical image integrity, all while working in runtime as well.
Another approach when working with multiple modalities is image harmonization. In this case, it aims to standardize images to a common representation, such that they all have similar appearances and consistent characteristics after the translation. To do so, this technique attempts to remove domain-specific features. Ren et al. (2021) and Liu et al. (2021) both used image harmonization to translate between images of the same modality but different scanners. This works well, as the scans are rather similar to each other and not much detail has to be removed. With cross-modality translation, however, the harmonization has to fill too much of a distance, as the differences between modalities are comparably huge. Thus, a translation would lead to too much loss of information, which could become a problem for subsequent tasks such as disease classification or segmentation. Nevertheless, Seoni et al. (2024) give a good overview of why image harmonization is still especially interesting in the case of addressing the distribution shift across medical imaging scanners.
With rising from Schock and Baumgartner (2023) and SimpleITK from Lowekamp et al. (2013) there also exist specialized libraries for medical image processing, with the first one providing highly performant image processing and augmentation tools, while the latter simplifies the use of the Insight Segmentation and Registration Toolkit (ITK). Both provide basic but, unlike *CMDA*, lack cross-modality and modality-specific augmentations. Cardoso et al. (2022) developed the framework MONAI which covers the full medical imaging workflow and even includes some rare modality-specific, but also no cross-modality augmentations. The Eisen framework by Mancolo (2020) is very similar to MONAI, only that development has stopped and it is not available anymore. Thus, both frameworks do not offer the functionalities that *CMDA* provides. If interested, Chlap et al. (2021), Hussain et al. (2017), and Goceri (2023) provide an extensive survey of medical data augmentation using basic geometric transformations.
In addition to classic data augmentations, advancements in deep learning have also enabled novel approaches. In 2016, Gatys et al. (2016) introduced the idea of neural style transfer (NST) through convolutional neural networks (CNNs), where NST describes the process of transferring the style of an image to another image without changing its content. This foundational research led to many interesting ideas,

among others enabling a translation between modalities. But the proposed solution relied on a slow iterative optimization process and only had fixed implemented styles. Shortly after, Huang and Belongie (2017) refined this procedure by introducing adaptive instance normalization (AdaIN) to make the NST fast while giving the user more control by being able to choose a custom style and more settings such as style interpolation. Chandran et al. (2021) later extended AdaIN by Adaptive Convolutions (AdaConv) which allows transferring structural and statistical styles at the same time, yielding even better results than AdaIN. However, neither AdaIN nor AdaConv is specifically tailored for medical imaging and may thus produce inadequate samples when used for cross-modality translation. Furthermore, both require more computational resources than basic geometric transformations and can therefore not be applied at runtime. Jing et al. (2019) and Singh et al. (2021) both review the current progress of NST and may be referred to for further insight.
Alternative approaches use Generative Adversarial Networks (GANs) or other deep learning techniques to not only transfer style but also to generate entirely new content which is also of high interest in medical imaging. (Bowles et al., 2018) illustrates this interest by proposing a GAN that creates synthetic MRI and CT patches, while Shin et al. (2020) synthesized PET images from given MRI images by utilizing the Bidirectional Encoder Representation from Transformers (BERT). While both studies align with *CMDA* to address the challenge of modality shift, they offer a different methodology that may produce unrealistic results and is not deployable at runtime. Although more related to policy sampling than cross-modality translation, the Automated Augmentation for Domain Generalization (AADG) developed by Lyu et al. (2022) presents an interesting domain generalization strategy as the method is based on data manipulation, where new domains for retinal images can be created via sampled augmentation policies. More closely related to *CMDA* is the MedGAN framework which is capable of PET, MRI, and CT cross-modality translation. Therefore, Armanious et al. (2020), whose overall contribution to this sector is noteworthy, also utilize a GAN that takes an image as an input and transforms it to the target modality. While MedGAN produces more realistic images than *CMDA* does, it has the same downsides as all deep-learning approaches, mainly not being applicable during model training. Similarly, Yang et al. (2020) present a cross-modality generation framework that employs conditional GANs (cGANs) to translate between different T1, T2 (see Section 3.1), and T2-Flair MRI modalities, whereas TarGAN, a target-aware GAN introduced by Chen et al. (2021) can be used to translate between CT and MRI images where the target area with a possible disease is further enhanced. The latter is more concerned with target-aware augmentation than with the cross-modality aspect, but both methods also implement fewer modalities than *CMDA* does. Other deep learning based data augmentation and modality translation approaches can be found in Kebaili et al. (2023) and Kaji and Kida (2019).
Despite the presented options providing great opportunities for medical data augmentation, basic augmentations continue to be the most widely used augmentation techniques in practice (Chlap et al., 2021). This underlines *CMDA*'s value as it addresses the specific problem of modality translation at runtime.

3 Theoretical Background

3.1 Implemented Modalities

In order to be more effective than random augmentations, *CMDA* takes advantage of modality-specific characteristics. This leads to a targeted selection of implemented modalities where augmentations can be fine-tuned with precision. PET, MRI, and CT were chosen for this study due to several relevant factors. They ensure broad applicability and relevance by not only being prevalent in clinical practice but also providing consistent and comparable representations of anatomical structures across dimension and shape. Furthermore, these modalities partially overlap in their scanned anatomical structures which only makes sense in the context of a modality transformation. It is therefore necessary to explore the mode of operation, specific applications, and characteristics of the chosen modalities to better understand *CMDA* and utilize its full potential. Therefore, the following Section will introduce all implemented modalities.

PET Positron Emission Tomography is part of nuclear medicine where it is often conducted in combination with a CT scan for attenuation correction purposes. The procedure starts with a tracer being injected into the patient. In this case, a tracer refers to a radiopharmaceutical combined with a carrier, mostly sugar. The current standard for this is [^{18}F]Fluorodeoxyglucose (FDG) (Bailey et al., 2005). The goal of the injected tracer is to take part in the patient's metabolism or blood flow so that it can be used to monitor functional processes inside the human body. As diseases often consume abnormal (less or more) amounts of energy in order to exist, the tracer centers (or specifically does not appear) around the affected body part. This enables disease localization through certain reconstruction algorithms, indicating what medical problem the patient has and where it is situated (Rennie, 1999).
The physical processes exploited to create the images provide more information about the name of the modality. As radionuclides have an unstable connection, they tend to decay (Saha, 2015). In addition to a neutrino and a newly formed nuclide, this decay causes the emission of a positron. As the surrounding air is full of electrons, it is inevitable that the positron immediately hits one of them, causing an annihilation. The resulting energy is then emitted in the form of two photons being shot in opposite directions. These are then detected by scanners and can now be used to calculate the point of decay (Saha, 2015). This sequence of events is also shown in Figure 1.

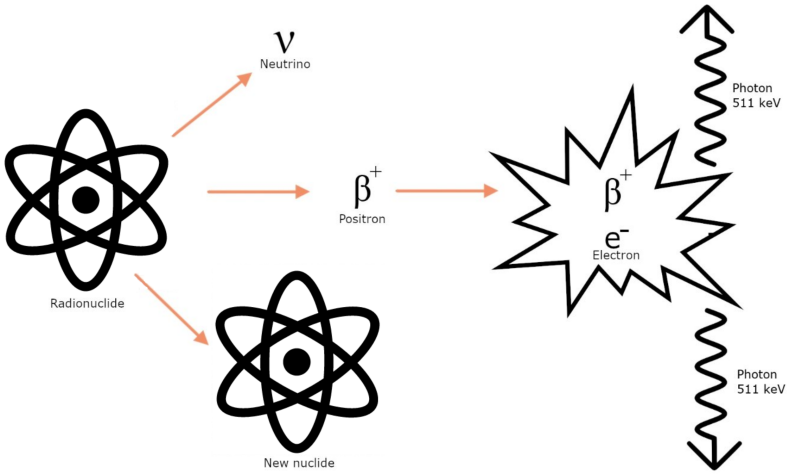
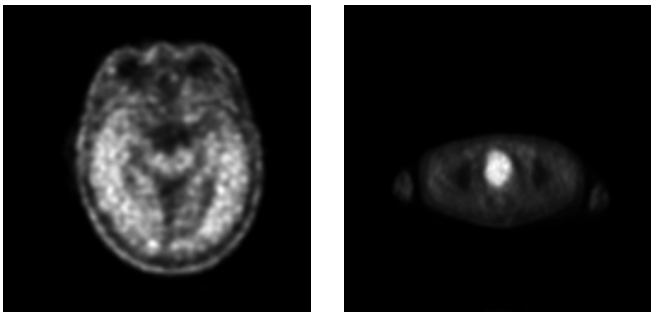


Figure 1: Physical processes following the decay of a tracer. A positron is emitted and subsequently collides with an electron, resulting in an annihilation. The photons thereby emitted photons can be used to calculate the point of decay and thus to localize the disease.

As randomly detected photons or a weakened signal through the interactions of the particles with human tissue can cause attenuation (Muehllehner and Karp, 2006), the created images have to carefully be checked for such.
PET is commonly used in oncology, cardiology, and neurology and, unlike the other implemented modalities, gives insights into the functional processes inside the body (Bailey et al., 2005). An entire scan typically takes 30-45 minutes. Sample PET images can be seen in Figure 2.

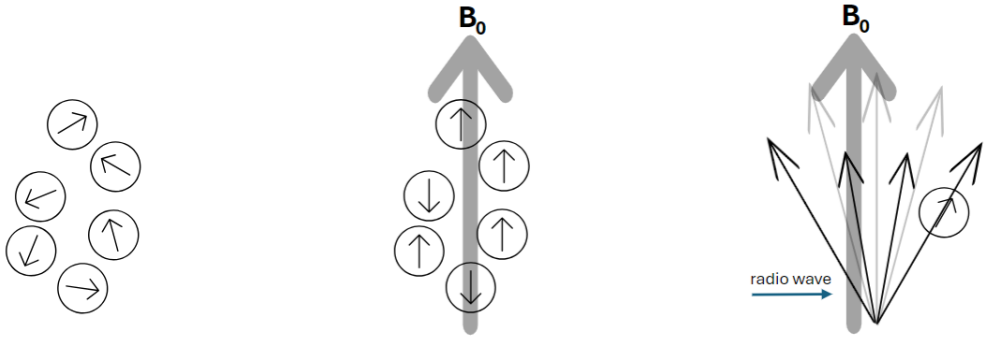


brain (ADNI, 2022) bladder (Kirk et al., 2016)

Figure 2: Sample PET images of two adult patients taken in the axial plane, viewed from the feet upwards. The left image depicts a healthy brain, while the right image depicts an abdominal scan with bladder cancer.

MRI Magnetic Resonance Imaging is a non-invasive imaging technique. Therefore, it uses strong magnetic fields and radio waves to produce detailed images of soft tissue, organs, and other internal body structures. This makes it suitable for

imaging the brain, spinal cord, joints, and muscles (Hashemi et al., 2012). Meanwhile, its absence of ionizing radiation expands its application to patients who must not be exposed to such.
On a physical level, the imaging procedure begins by exposing the patient to a strong and static magnetic field $\mathbf{B_0}$. As a result, hydrogen atoms in the patient's body align with $\mathbf{B_0}$. The patient is now pulsed by radio waves, which stimulate the atoms and make them push against $\mathbf{B_0}$'s influence. This disturbs the alignment, as can be seen in Figure 3. As soon as the pulse of radio waves stops, the atoms will realign with $\mathbf{B_0}$, thereby emitting the supplied energy as electromagnetic signals. This process is called relaxation and is the reason for different types of image contrast. As the intensity and amount of released energy differ depending on the tissue the hydrogen atoms are located in, these signals can then be used to create images of the scanned anatomy (Weishaupt et al., 2009). Additionally, pharmaceuticals, named contrast agents, taken by the patient can enhance the emitted signals.
Depending on the focus, MRI scans can either be T1- or T2-weighted. T1-weighted



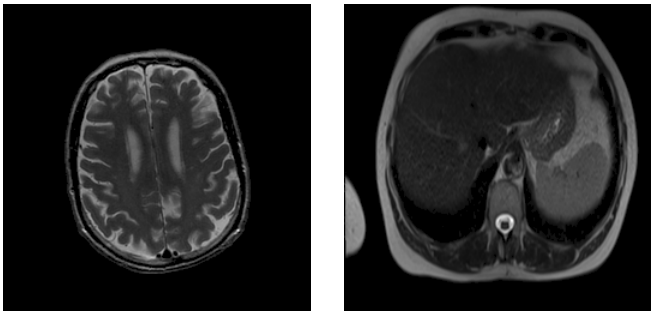
Hydrogen atoms inside the patient are initially randomly aligned.

Hydrogen atoms align with the applied magnetic field $\mathbf{B_0}$.

When stimulated by radio waves, the hydrogen atoms push against $\mathbf{B_0}$'s force.

Figure 3: Physical processes that take place during an MRI scan, in chronological order from left to right. It starts with randomly aligned hydrogen atoms which are subsequently influenced by a magnetic field and radio waves.

scans concentrate on signals emitted shortly after the relaxation of the tissue starts. They highlight fat and are thus useful for accurately imaging anatomical structures and detecting certain abnormalities. In turn, T2-weighted scans highlight fluid and can better be used to locate edema, tumors, and inflammation. In contrast to T1, they focus on signals emitted a longer time after relaxation starts (Katti et al., 2011).
As it can provide precise anatomical information, the diagnosis of tumors, spinal injuries, brain disorders, and joint abnormalities are the most common use-cases. With 30-90 minutes, an entire scan takes comparably long and can be unsettling for some patients (Hashemi et al., 2012). Sample MRI images can be seen in Figure 4.



brain (ADNI, 2022) bladder (Kirk et al., 2016)

Figure 4: Sample MRI images of two adult patients taken in the axial plane, viewed from the feet upwards. The left image depicts a healthy brain, while the right image depicts an abdominal scan with bladder cancer.

CT Computed Tomography combines X-ray images taken from different angles to create cross-sectional images (slices) of the body. The scanning procedure starts with the patient being placed on a bed. This bed will then be moved through the circular-shaped part of the CT scanner, called gantry. The structure of the gantry can be observed in Figure 5. It consists of an X-ray tube and corresponding X-ray detectors. When the patient is moved further into the gantry, the tube starts rotating around the patient, meanwhile constantly sending out X-ray beams at different angles. The detectors then measure the radiation absorbed in Hounsfield units (HUs). Here, each kind of tissue traversed has different HU values with high values indicating high, and low values low attenuation (Goldman, 2007; Mazonakis and Damilakis, 2016). These measurements can then be processed by computers to generate detailed images, hence the name. If interested, Koetzier et al. (2023) provide more information about commonly used reconstruction techniques. Each rotation of the tube results in one slice with 5mm - 1mm thickness (Thrower et al., 2021). The bed is then slightly pushed forward and the procedure repeats. In the end, all slices can be combined to create a three-dimensional representation of the scanned body parts.

Its use of ionizing radiation limits its applicability for some patients, but it produces detailed images of bones, blood vessels, organs, and other internal structures. This makes it excellent for detecting bone fractures, internal bleeding, infections, and tumors (Buzug, 2011).

The scan itself only takes 5-20 minutes. Sample CT images can be seen in Figure 6. Due to the speed of acquisition, it is often used in emergency situations where it can be utilized to assist in surgeries or biopsies.

Depending on the intended use-case, images can be reconstructed in multiple planes or 3D. In addition, contrast agents can be used to enhance the visibility of certain tissues or blood vessels (Buzug, 2011). CT is most commonly used in oncology, neurology, cardiology, and trauma care.

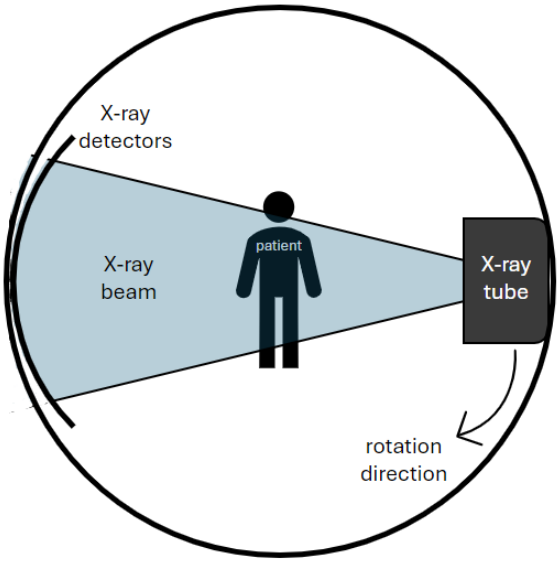
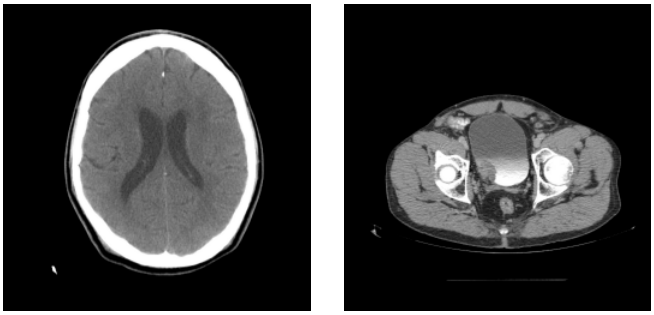


Figure 5: Structure of a gantry, the circular shaped part of a CT scanner. An X-ray tube rotates around the patient and sends out X-ray beams which are then used to measure the experienced attenuation with the X-ray detectors.



brain (A. Stein, 2019) bladder (Kirk et al., 2016)

Figure 6: Sample CT images of two adult patients taken in the axial plane, viewed from the feet upwards. The left image depicts a healthy brain, while the right image depicts an abdominal scan with bladder cancer.

3.2 Transfer Learning, Domain Adaptation, Domain Generalization

Each medical imaging modality provides different essential insights for the precise diagnosis and therapy of diseases. However, it's often not feasible for a patient to undergo scans of multiple modalities. This is due to various factors such as health constraints, costs, and time. The thereby created cross-modality shift leads to a data scarcity where high-quality data is often only available for a single rather than multiple modalities. As a result, the development of robust models across different modalities is limited. As the appearance of images changes heavily between diagnostic procedures the following concepts are the key to overcome this challenge. Understanding how these concepts can benefit from the larger and more diverse training data provided by the *CMDA* is important to create more robust and generalizable models.

Transfer Learning Weiss et al. (2016) define Transfer Learning (TL) as a technique that supports a model in the learning process by utilizing information learned from another, related domain. Therefore, a model trained on one task $\mathcal{T}_S$ and source domain $\mathcal{D}_S = (x_i^S, y_i^S)_{i=1}^{N_S}$ is used to assist in solving a different task $\mathcal{T}_T$ with a different target domain $\mathcal{D}_T = (x_j^T(, y_j^T))_{j=1}^{N_T}$. This often involves the use of models pre-trained on a large dataset.
It is commonly used when the target domain has limited labeled data as it utilizes information from $\mathcal{D}_S$. This approach helps to reduce training time and improve the model performance on the target task.
TL can be divided into homogeneous and heterogeneous learning, where the first covers types with $\mathcal{D}_T = \mathcal{D}_S$ while the latter one includes the cases where $\mathcal{D}_T \neq \mathcal{D}_S$ (Day and Khoshgoftaar, 2017).
Commonly used techniques listed by Donges (2024) are

- Feature extraction: Use learned features from $\mathcal{D}_S$ as input for $\mathcal{T}_T$
- Domain-specific pre-training: Keep $\mathcal{D}_S$ similar to $\mathcal{D}_T$
- Fine-tuning: Train a model on $\mathcal{D}_S$, then fine-tune it on $\mathcal{D}_T$

The applications involve image classification, natural language processing, object detection, or segmentation tasks (Torrey and Shavlik, 2010; Weiss et al., 2016; Csurka, 2017).
Example: Training a model on a large MRI brain dataset as $\mathcal{D}_S$. Here $\mathcal{T}_S$ is a binary classification problem deciding whether the image includes a tumor or not. Then train the model further on a small CT brain dataset $\mathcal{D}_T$ with $\mathcal{T}_T$ to classify different types of lesions.

Domain Adaptation The general concept of Domain Adaptation (DA) is adapting a model trained on a source domain $\mathcal{D}_S$ and task $\mathcal{T}_S$ to perform well on a different target domain $\mathcal{D}_T$, where the data distributions $P_S(X)$ and $P_T(X)$ differ. The tasks stay the same, such that $\mathcal{T}_T = \mathcal{T}_S$. Hence, DA is a variant of TL(Farahani et al., 2021).
DA is used for cases where no or limited labeled data is available for $\mathcal{D}_T$. The technique aims to minimize the distribution shift $distance(P_S(X), P_T(X))$ between source and target domains.
It can be categorized into three subtypes (Guan and Liu, 2022). Unsupervised DA presents the case where no labeled data is available from $\mathcal{D}_T$, while supervised DA is the opposite with $\mathcal{D}_T$ providing labeled data. If labeled and unlabeled data is mixed in $\mathcal{D}_T$, it is further referred to as semi-supervised DA.
Kundu (2022) and Farahani et al. (2021) enumerate the following, frequently employed techniques

- Domain-Invariant Feature Learning: Learn features invariant to domain changes
- Feature Based Adaptation: Align feature spaces between $\mathcal{D}_S$ and $\mathcal{D}_T$
- Instance Based Adaptation: Assign weights to samples from $\mathcal{D}_S$ to match $P_T(X)$
- Reconstruction Based Adaptation: Minimize $distance(P_S(X), P_T(X))$ though reconstruction of samples within a shared intermediate feature space
- Adversarial Training: Use adversarial networks to minimize $distance(P_S(X), P_T(X))$

Its uses span various fields, including cross-domain image segmentation and classification, language translation, and autonomous driving (Farahani et al., 2021; Guan and Liu, 2022; Csurka, 2017).
Example: Training a model on a large, high-quality MRI brain dataset as $\mathcal{D}_S$. Here, $\mathcal{T}_S$ is a binary classification problem deciding whether the image includes a tumor or not. Then train the model on low-quality custom MRI images with the same binary classification task such that $\mathcal{T}_T = \mathcal{T}_S$.

Domain Generalization While TL and DA require data from the target domain $\mathcal{D}_T$, Domain Generalization (DG) focuses on building models that generalize well to unseen target domains $\mathcal{D}_{T_1}, \mathcal{D}_{T_2}, ..., \mathcal{D}_{T_m}$. However, these models are trained without having access to data from the target domains. This typically involves training a model on several source domains $\mathcal{D}_{S_1}, \mathcal{D}_{S_2}, ..., \mathcal{D}_{S_n}$ which are similar to the target domains. This in turn helps the model to learn domain-invariant features, ensuring robustness and adaptability to new environments and conditions (Zhou et al., 2022).
DG uses the techniques outlined by Zhou et al. (2022)

- Data Augmentation: Generate diverse training samples to cover potential target domains (further explored in Section 3.3)

- Domain Alignment: Learn features that remain stable across different domains
- Meta-Learning: Train model on variety of tasks $\mathcal{T}_{S_1}, \mathcal{T}_{S_2}, ..., \mathcal{T}_{S_n}$ to improve its generalization ability
- Ensemble Learning: Combine predictions from multiple models trained on different source domains
- Regularization: Prevent overfitting by applying regularization

However, realizing a generalizing model is often challenging. It requires large and diverse datasets to cover $\mathcal{D}_{T_1}, \mathcal{D}_{T_2}, ..., \mathcal{D}_{T_m}$ while having to balance domain invariant features with task-specific performance.
When successfully implemented, applications cover autonomous systems, natural language processing, and general-purpose medical diagnosis models, among others (Zhou et al., 2022; Gulrajani and Lopez-Paz, 2020; Wang et al., 2022).
Example: Training a model on MRI ($\mathcal{D}_{S_1}$), PET ($\mathcal{D}_{S_2}$), and ultrasound (US) ($\mathcal{D}_{S_3}$) brain datasets. Here $\mathcal{T}_S$ is a binary classification problem deciding whether the image includes a tumor or not. Then test the model on a CT brain dataset $\mathcal{D}_{T_1}$ with the same binary classification task such that $\mathcal{T}_T = \mathcal{T}_S$.

3.3 Data Augmentation as a Concept

Data augmentation describes a DG technique to enhance the size and diversity of the training data. It does so by creating synthetic data samples, often being variations of the original training samples. Applicable domains along with their most commonly used manipulation techniques are

- Images: flipping, rotation, cropping, scaling, contrast adjustment, color space transformation, noise injection, mixup, erasing (Shorten and Khoshgoftaar, 2019)
- Text: back-translation, random insertion, synonym replacement, word dropout, random swap (Bayer et al., 2022)
- Audio: pitch shifting, time stretching, noise injection, mixup (Wei et al., 2020)

Figure 7 displays some of the mentioned image manipulations in action. As can be observed, all augmentations are domain-independent, which means that they can be applied to every possible image, no matter the content. In contrast, cross-modality augmentations utilize modality-specific features and are thus tailored for the medical imaging context. They assume medical images as an input, thereby trading domain independence for task-specific and meaningful augmentations. As an example, take Sharpening from Figure 7. While this augmentation is equally suited for every image, a cross-modality Sharpening augmentation could incorporate

information about spatial resolution relations across modalities. This may result in more realistic augmentations in that specific context.

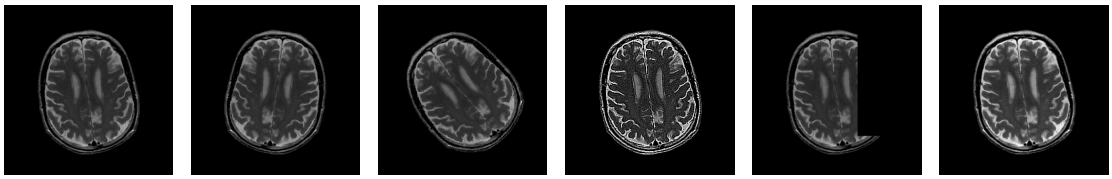


Figure 7: Brain MRI image (ADNI, 2022) of a healthy patient is transformed by various traditional data augmentation techniques. The caption always names the respective technique.

Such tailored augmentations, alongside traditional techniques, contribute to more, and more diverse training data. Additionally, they simulate real-world data variations and can thus mitigate overfitting and improve robustness. This is key for deep learning, as models have only limited real-world data at their disposal but can hereby improve their accuracy and generalization performance (Shorten and Khoshgoftaar, 2019).
As data is especially rare in the medical domain (see Section 1), many possible use-cases arise. With the most obvious application being the enlargement of the data size, it could also combat class imbalance by creating more samples of the underrepresented class. Both applications lead to enhanced diagnostic tool accuracy. More diverse data also increases the training data variability. This in turn improves the model generalization across heterogeneous patient populations and enables the simulation of rare diseases (Chlap et al., 2021; Hussain et al., 2017). Apart from the technical benefits, data augmentation also supports regulatory compliance by altering the original data which helps protect the patients' privacy (Shorten and Khoshgoftaar, 2019).
While more advanced data augmentations like synthetic data creation and style transfer via GANs (see Section 2) exist, geometric and pixel-level adjustments enable a more efficient approach as they are faster and require less data. They also take real data as a foundation which protects the data augmentations from hallucinating. This, as well as the outlined advantages of Data Augmentation, provides a basis for exploring the targeted approach of a cross-modality translation in Section 4.1.

4 Methods

4.1 Cross-Modality Data Augmentation

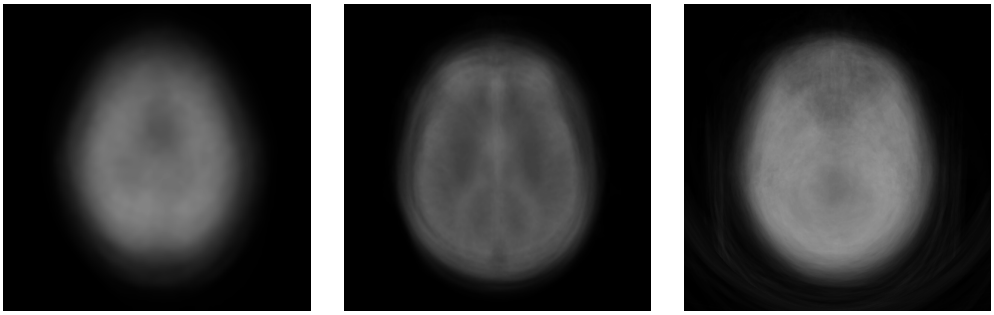
The Cross-Modality Data Augmentation *CMDA* takes gray-scale or color images as input and translates them from their initial to a given target modality. This translation helps to adapt the distribution of the training data to that of the target modality by keeping the original content but changing its style. It does so by utilizing and altering modality-specific characteristics. This is accomplished by sequentially applying special but resource-efficient augmentations to the input image, executed in the specified order presented in this Section. Since it is not always desired for each image to be augmented, *CMDA* also offers users the choice of setting a probability for the data augmentation to actually be applied to an image.

To increase randomness, and thus also the variety of the augmented data, users can choose a range of augmentations to be applied to each image. This includes the option to make certain augmentations more likely to be applied than others. As each augmentation might not prove useful in each context, this is frequently useful. Additionally, the intensity of each applied augmentation can again be defined, which further supports diversity. These adjustments provide full control and allow to customize *CMDA* for each specific use-case.

Basic geometric augmentations like flipping, rotating, scaling, shearing, or such are not implemented on purpose. Instead, the data augmentation can easily be added to existing augmentation pipelines which is very much recommended. For the implementation, more elaborate information, permitted parameter values, examples, and the use of custom reference images please refer to the corresponding GitHub repository. Additionally, Appendix A.1 provides randomly augmented example images created by *CMDA*.

4.1.1 Color Augmentation

In comparison to the other augmentations, color is of major importance as it can change the focus, contrast, intensity, and overall appearance of an image the most. Taking advantage of modality-specific characteristics, certain structures can be highlighted or adjusted by changing their brightness and color.
Thus, each image first undergoes the same initial modality-unspecific step which is a basic alignment to the target modality. Therefore, a reference image (see Figure 8) has been created for every implemented modality. This was done by iterating through a sufficiently large (≥ 200) dataset of each modality, calculating the mean color value for each pixel, and assembling these mean pixels to a new image. This procedure makes sure that essential features and areas of interest are correctly selected. The size of the dataset is further needed to make the creation of the reference image more robust to outliers.

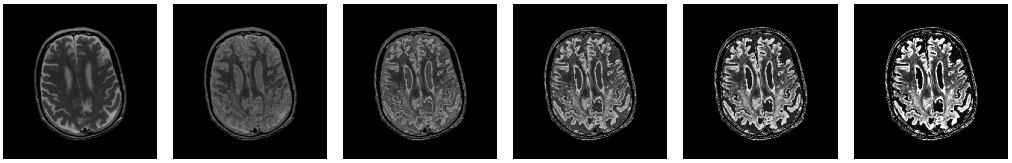


PET MRI CT

Figure 8: Reference images of the implemented modalities, created by iterating through a respective dataset and building an "average image". They are required for the color augmentation as they include modality-specific color information.

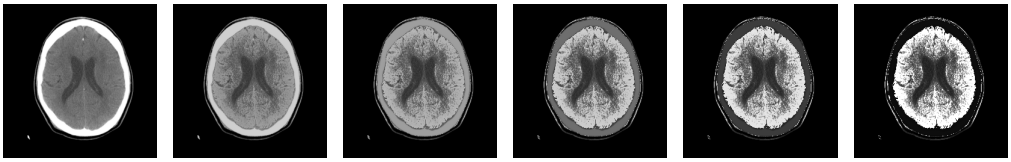
To now augment an image, the color histogram of that image is calculated and compared to the histogram of the reference image for the target modality. Histogram matching (scikit-image) is then used to adjust the image's color histogram to match that of the reference image. Thereby its colors are transformed to resemble those of the target modality. This technique aligns the cumulative distribution function of both images' color histograms, thus ensuring similar intensity and color. At the same time, it preserves the structural content of the input image.
CMDA also allows to create custom reference images for any desired modality, with sample results being shown in Section 4.1.5.
Additionally, certain refinements are performed based on the target modality:

PET PET is used to highlight functional processes inside the human body (see Section 3.1). Thus, bones are attenuated and attention is drawn to the active soft tissue by brightening it, resulting in black and white images. Sample results can be observed in Figure 9 and 10.



original 20% 40% 60% 80% 100%

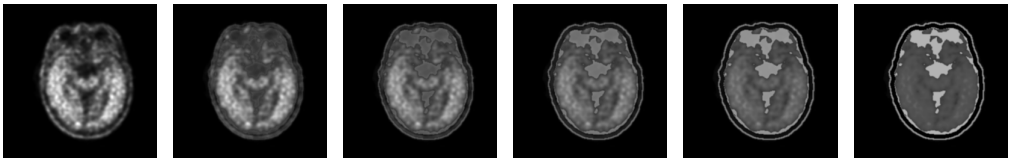
Figure 9: A brain MRI image is transformed to target modality PET, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from ADNI (2022).



original 20% 40% 60% 80% 100%

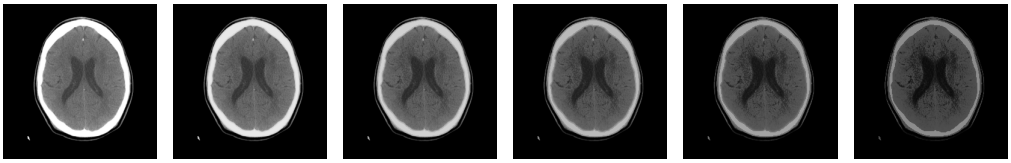
Figure 10: A brain CT image is transformed to target modality PET, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from A. Stein (2019).

MRI MRI gives a detailed view of soft tissue and its structure (see Section 3.1). Therefore, bones are also darkened, and texture is slightly added to the soft tissue. This leads to detailed, dark gray images. Sample results can be observed in Figure 11 and 12.



original 20% 40% 60% 80% 100%

Figure 11: A brain PET image is transformed to target modality MRI, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from ADNI (2022).

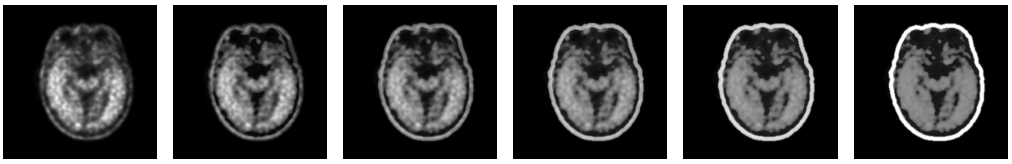


original 20% 40% 60% 80% 100%

Figure 12: A brain CT image is transformed to target modality MRI, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from A. Stein (2019).

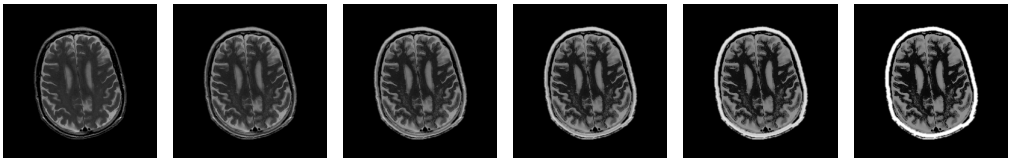
CT CT focuses on displaying the overall structure of the body with an emphasis on the bones (see Section 3.1). Consequently, these are illuminated while the

soft tissue structure is slightly blurred, producing rather detailed, light gray images. Sample results can be observed in Figure 13 and 14.



original 20% 40% 60% 80% 100%

Figure 13: A brain PET image is transformed to target modality CT, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from ADNI (2022).



original 20% 40% 60% 80% 100%

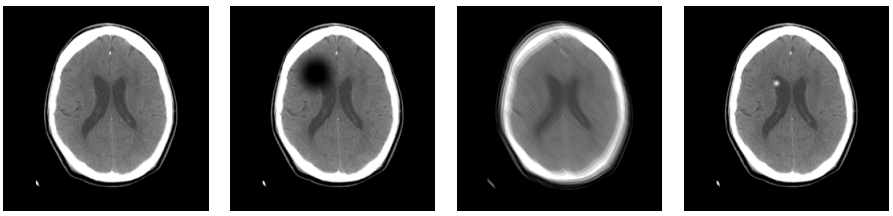
Figure 14: A brain MRI image is transformed to target modality CT, using *CMDA*'s color augmentation. The above images are augmented with different intensities that are displayed in the respective caption. This illustrates how the set intensity influences the severity of the augmentation. Original image taken from ADNI (2022).

4.1.2 Artifact Augmentation

Imaging artifacts are the second augmentation to be added to an image. They describe accidental, unwanted, and in reality often non-existent anomalies in the final medical images (Stanford), caused by technical or environmental factors. They therefore corrupt the actual data which is why this augmentation should be used with care. However, since normal datasets are usually not perfect either, it still helps to create a more accurate distribution of the target modality when used moderately. As each modality has different underlying physics, artifacts are modality-specific which might cause them to look different even if having the same cause. A brief explanation of said causes and effects is given in the following.

PET (Abouzied et al., 2005; Sureshbabu and Mawlawi, 2005; Cook et al., 2004)

Artifact	**Cause**	**Effect**
metal object	metal objects in patient	dark areas
motion	movement during scan	blurring
attenuation	random coincidence events	light dots

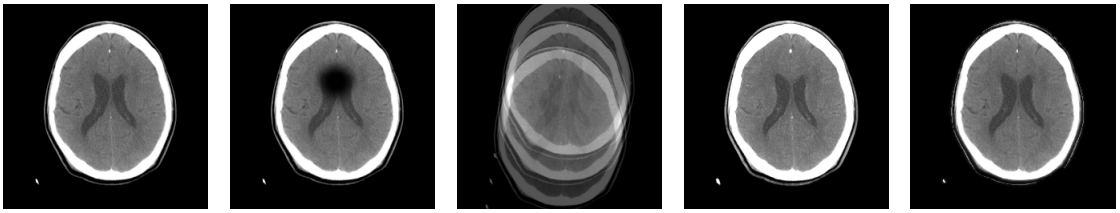


original metal object motion attenuation

Figure 15: A brain CT image is transformed to target modality PET, using *CMDA*'s artifact augmentation. The above images thus include certain target-modality-characteristic artifacts, with their names displayed in the respective caption. Original images taken from A. Stein (2019).

MRI (Krupa and Bekiesińska-Figatowska, 2015; Smith, 2010)

Artifact	**Cause**	**Effect**
metal object	metal objects in patient	dark areas
motion	movement during scan	ghosting
gibbs	inadequate sampling of frequencies for reconstruction	oscillations near sharp edges
chemical shift	different resonance frequencies between fat and water	double contours

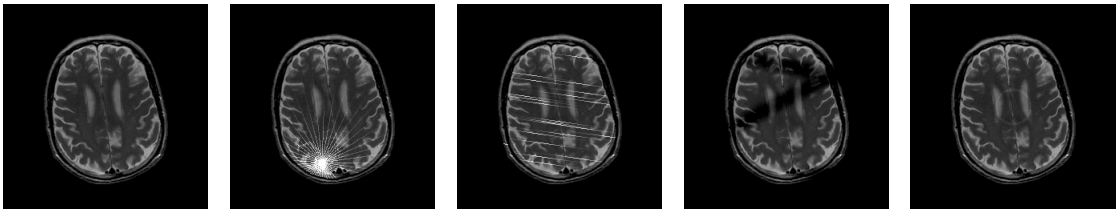


original metal object motion gibbs chemical shift

Figure 16: A brain CT image is transformed to target modality MRI, using *CMDA*'s artifact augmentation. The above images thus include certain target-modality-characteristic artifacts, with their names displayed in the respective caption. Original images taken from A. Stein (2019).

CT (Boas et al., 2012; Barrett and Keat, 2004; Cook et al., 2004)

Artifact	**Cause**	**Effect**
metal object	metal objects in patient	bright rays
motion	movement during scan	bright streaks
beam hardening	energy-absorbing objects	dark bands
ring	miscalibrated scanner	circular contour



original metal object motion beam hardening ring

Figure 17: A brain MRI image is transformed to target modality CT, using *CMDA*'s artifact augmentation. The above images thus include certain target-modality-characteristic artifacts, with their names displayed in the respective caption. Original images taken from ADNI (2022).

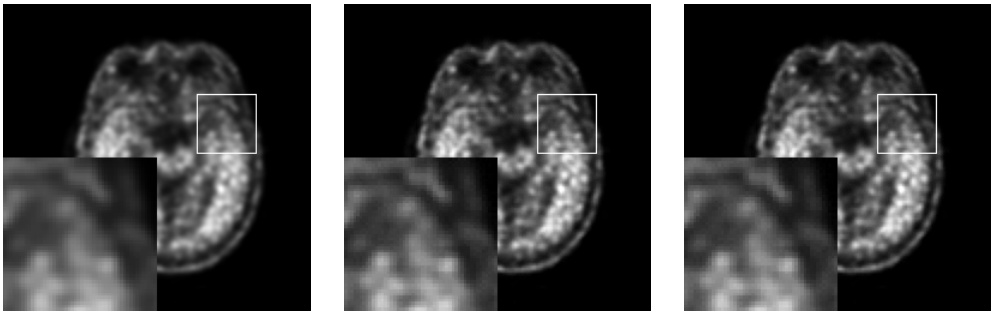
4.1.3 Spatial Resolution Augmentation

Because of the basic functionality of the scanners in use, each modality produces images of different quality. The third augmentation is thus concerned with spatial resolution. Adapting it is achieved by applying either a blur of sharpness filter, depending on the initial and target modality. Hereby, blurring takes while sharpening gives detail to the transformed image.
While this augmentation might not produce as visually noticeable results as the others (see Figure 18), it is still the second most important to be applied as it strongly contributes to changing the distribution of the input images. Great results can especially be achieved when paired with the color augmentation as it, depending on the refinements performed, may have slightly added or removed detail.

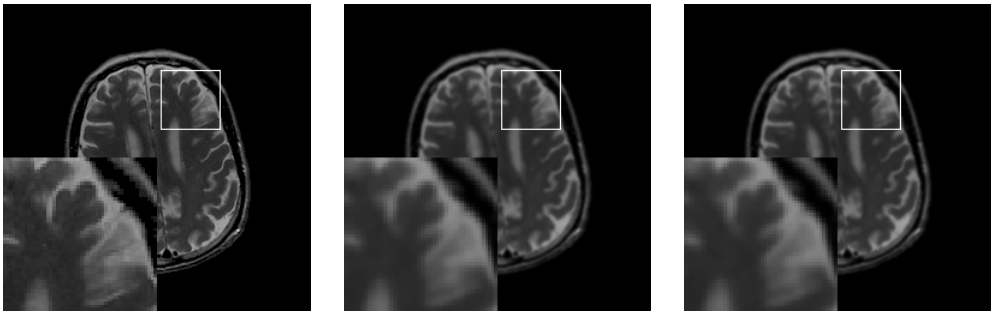
PET Since modalities including radionuclides rather concentrate on functional body processes than exact structures, their spatial resolution of 1000-3000μm is comparably poor (Kasban et al., 2015; Key and Leary, 2014; Yim et al., 2011).

MRI The high magnetic field strengths and advanced gradient coils, paired with a long screening time during which the patient has to stay still, leads to a superior image quality of 10-200μm (Kasban et al., 2015; Key and Leary, 2014; Yim et al., 2011).

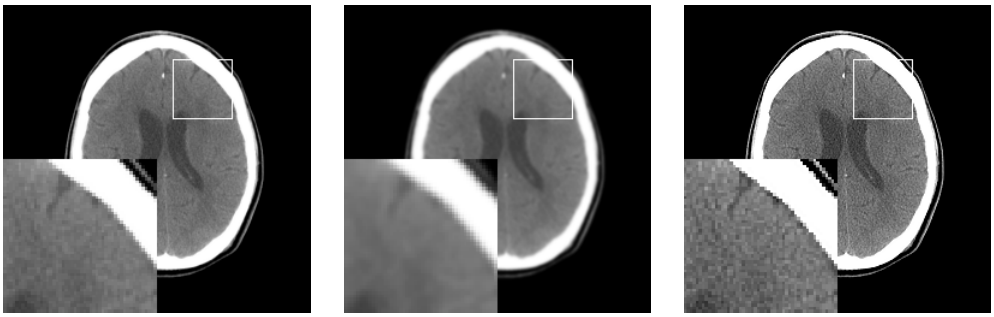
CT Even with x-ray beam divergence, the high-resolution x-ray detectors and rapid sequential imaging allow for a spatial resolution of 50-500μm (Kasban et al., 2015; Key and Leary, 2014; Yim et al., 2011).



original PET (ADNI, 2022) PET to MRI (sharpened) PET to CT (sharpened)



original MRI (ADNI, 2022) MRI to PET (blurred) MRI to CT (blurred)



original CT (A. Stein, 2019) CT to PET (blurred) CT to MRI (sharpened)

Figure 18: Brain images transformed to the respective target modalities, using *CMDA*'s spatial resolution augmentation. The above images compare their original and the spatial resolution transformed images, where blur and sharpen filters have been applied.

4.1.4 Noise Augmentation

According to Morin and Mahesh (2018), noise refers to the graininess of an image as it describes unintentionally added pixels all over said image. Similar to artifacts, it usually corrupts the data by hiding or concealing potentially useful information. However, noise augmentation can thereby support robustness as it introduces new variations to the training data which leads to better generalization. The model thus

learns to focus on the underlying patterns, ultimately improving its performance in real-world applications where data may be noisy and imperfect as well.
For the implemented modalities, the noise of each follows a certain statistical distribution. In turn, these can be utilized and applied to the images in the form of noise filters which can be seen in Figure 19.

PET The detection of coincidence events is subject to statistical fluctuations that often follow a normal distribution (Kim et al., 2013).

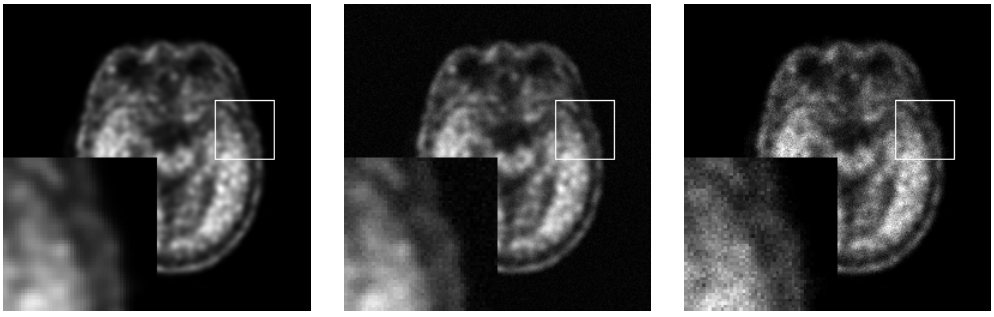
$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
where x is the random variable, $\mu = 0$ is the mean of the distribution, σ is the standard deviation and depends on the given augmentation ratio.

MRI As MRI signals are processed as complex numbers, a real and an imaginary noise is created. Thus, the noise level can be seen as the magnitude of their combination which can best be modeled by a Rician distribution (Aja-Fernández and Vegas-Sánchez-Ferrero, 2016).

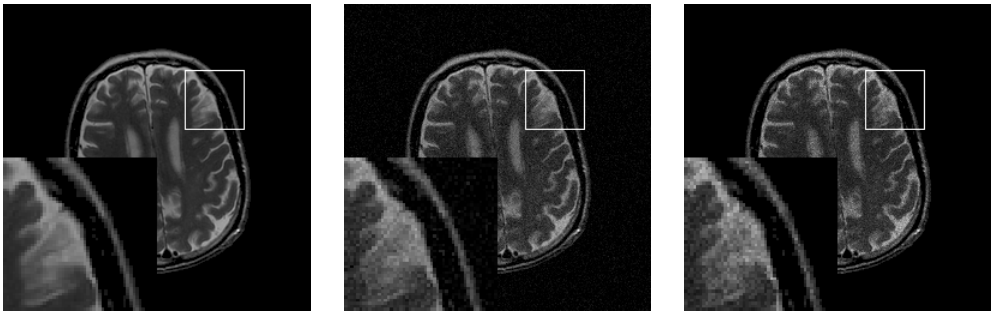
$f(x|\sigma,\nu) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2+\nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right)$
where x is the random variable, σ and ν are shape parameters dependent on the given augmentation ratio, I_0 is the zero-order modified Bessel function of the first kind.

CT The amount of detected X-ray photons varies because of the random nature of their radiation. This variation leads to characteristic noise which follows a Poisson distribution (Diwakar and Kumar, 2018; FAU; Wang et al., 2008).

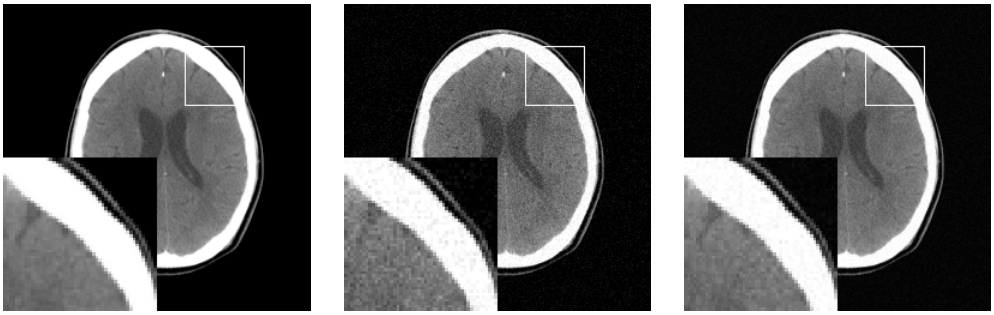
$f(k|\lambda) = \frac{\lambda^k \exp(-\lambda)}{k!}$
where k is the random variable, λ is the mean of the distribution and depends on the given augmentation ratio.



original PET (ADNI, 2022) PET to MRI (Rician) PET to CT (Poisson)



original MRI (ADNI, 2022) MRI to PET (Gaussian) MRI to CT (Poisson)



original CT (A. Stein, 2019) CT to PET (Gaussian) CT to MRI (Rician)

Figure 19: Brain images transformed to the respective target modalities, using *CMDA*'s noise augmentation. The above images compare their original and the noise-transformed images, where noise has been added using the respective distributions given in parentheses.

4.1.5 Custom Modalities

In addition to the implemented modalities, *CMDA* also provides the possibility to translate images to any desired modality. This is demonstrated in Figure 20 which shows an MRI brain scan (ADNI, 2022) being transformed to eight other modalities. Here, CT (A. Stein, 2019) and PET (ADNI, 2022) are implemented by default while the other six (Yang et al., 2021, 2023) are custom modalities.
This requires a dataset of sufficient size (≥ 200), used to create an adequate reference image. Due to the missing fine-tuning, the transformations are rather coarse, but certain scenarios may still benefit from it.
For more elaborate information on the use of custom reference images please refer to the corresponding GitHub repository.

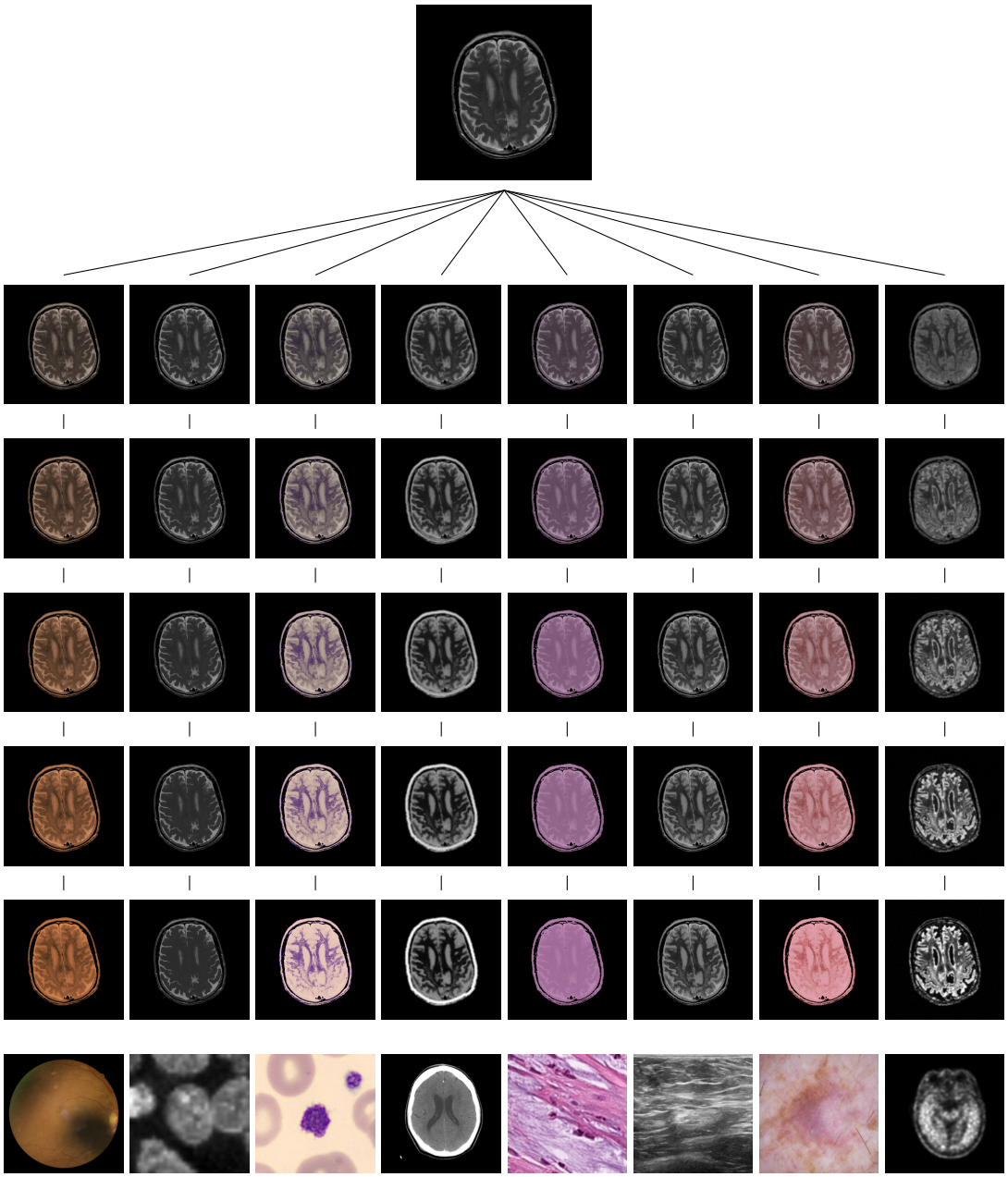


Figure 20: MRI image of the brain being gradually (20%, 40%, 60%, 80%, 100%) transformed by *CMDA* to eight different modalities. While CT and PET are implemented target modalities by default, the other six are custom target modalities that have been implemented by providing *CMDA* with a dataset of each custom modality. The last row shows sample images of the respective target modality for comparative purposes.

4.2 Evaluation Metrics

4.2.1 Quantitative Evaluation Metrics

Quantitative evaluation is concerned with hard numbers, assessing the performance and efficiency of *CMDA* by focusing on numerical, empirical, measurable, and objective results (Roessner, 2000; Garbarino, 2009; Dobrovolny and Fuentes, 2008). The evaluation of models in the quantitative experiments is thus carried out using suitable metrics. Therefore, the trained model is presented with data from the test set, measuring the amount of correct, also known as true positives (TP) and true negatives (TN). Additionally, incorrect predictions, termed false positives (FP) and false negatives (FN), are recorded. These values are then used to calculate the following metrics. Results for all may vary between 0 and 1 where higher values indicate better performance.

Balanced Accuracy $= \frac{Recall_{class1}+Recall_{class2}}{2}$
Useful accuracy metric for imbalanced datasets as it prevents bias towards a majority class. It does so by taking the average of recall (see further metrics) for each class to adjust for class imbalance.

Precision $= \frac{TP}{TP+FP}$
Measures the model's performance just looking at the degree of correct positive predictions, especially important when FP are dangerous.

Recall $= \frac{TP}{TP+FP}$
Measures the model's performance in finding all positive samples, especially important when FN are dangerous.

F1-Score $= 2\frac{Precision \cdot Recall}{Precision+Recall}$
Balances Precision and Recall and calculates their harmonic mean, combining both values in one metric.

ROC AUC
Measures the model's performance in distinguishing between classes over a range of decision thresholds. This is done by plotting the true positive rate (TPR) $= \frac{TP}{TP+FN}$ against the false positive rate (FPR) $= \frac{FP}{FP+TN}$ and calculating the area under the resulting curve.

4.2.2 Qualitative Evaluation Metrics

Qualitative evaluation pertains to soft insights, assessing quality and characteristics of the images *CMDA* generates. It is therefore concerned with visual, statistical,

perceptual, and subjective judgment (Roessner, 2000; Garbarino, 2009; Dobrovolny and Fuentes, 2008).

GLCM features In 1973, Haralick et al. (1973) first described the idea of textural analysis through gray level co-occurrence matrix (GLCM) features. These provide a powerful tool for the statistical analysis of spatial correlations between pixel values in images. In modern times, they are most commonly utilized for image classification, segmentation, or pattern recognition (Yogeshwari and Thailambal, 2023) by comparing learned features to those of the input images.
Their calculation first requires forming the GLCM. The GLCM is a matrix where each element $P(i, j)$ is representative of how often a pixel with intensity i is adjacent to a pixel with intensity j under a defined spatial relationship. For this evaluation the following GLCM features are of interest:

- Contrast:
  - Measures intensity variations and local contrast changes.
  - $\sum_{i,j=1}^{L}(i-j)^2 P(i,j)$
- Dissimilarity:
  - Measures degree of intensity variations and image roughness.
  - $\sum_{i,j=1}^{L}|i-j|P(i,j)$
- Homogeneity:
  - Measures local pixel similarity, uniformity, and texture smoothness.
  - $\sum_{i,j=1}^{L}\frac{P(i,j)}{1+|i-j|}$
- Energy:
  - Measures texture uniformity and complexity.
  - $\sum_{i,j=1}^{L}P(i,j)^2$
- Correlation:
  - Measures co-occurrence likelihood and linear dependencies between pixel pairs. Requires calculating means (μ_i, μ_j) and standard deviations (σ_i, σ_j) of pixel values in GLCM.
  - $\sum_{i,j=1}^{L}\frac{(ij)P(i,j)-(\mu_i\mu_j)}{(\sigma_i\sigma_j)}$

FID In search of evaluation metrics to assess the image quality of synthetically created images, Heusel et al. (2017) introduced the Fréchet inception distance (FID). The FID measures the Fréchet distance (FD) between two Gaussian distributions derived from the feature representations of real and generated images (Woodland et al., 2024). Its calculation requires feature extraction of the real and generated images. Therefore, the Inception-v3 model was used, as Woodland et al. (2024) showed that even in the medical domain feature extractors based on ImageNet constantly provide better results than those trained with medical datasets. The extracted feature vectors are then used to calculate the FID as follows

$$d^2 = ||\mu_r - \mu_g||^2 + Tr(C_r + C_g - 2\sqrt{C_r C_g})$$

Here, μ_r and μ_g refer to the mean, while C_r and C_g refer the the covariance matrix of the r(eal) and g(enerated) feature vectors. Tr refers to the trace operation defined as the sum of all elements on the main diagonal of a square matrix.

PCA In ML, it's common to work with high-dimensional data as in- and output data are often represented as vectors. Therefore, a principal component analysis (PCA) can be used as a dimensionality reduction technique, enabling the visualization of complex data by trading information against simplicity. To calculate the most important features, a PCA is conducted by following these steps (Ringnér, 2008; Maćkiewicz and Ratajczak, 1993):

- Standardize all given features to ensure equal contribution of every feature.
- Compute the covariance matrix of the standardized features to capture variance and correlation between features.
- Use the covariance matrix to compute the eigenvectors (principal components) and eigenvalues (variances).
- Rank the eigenvalues to see which principal component explains the highest variance.
- Multiply the original data by the highest-ranked principal components to obtain lower-dimensional data.

The resulting data can then be plotted or used for other purposes.

VAE Proposed by Kingma (2013), the variational autoencoder (VAE) is a probabilistic generative DL approach that can be used for data generation, feature learning, dimensionality reduction, and compression (Zhai et al., 2018). It does so by combining two different DL models. The first one takes data, transforms it into a latent space, and is called encoder. The second model is called decoder and tries to

reconstruct these latent representations back to their original content. The quality of the reconstructions is evaluated on the following metrics.

- Test Loss = $\frac{1}{n}\sum_{i=1}^{n}(BCE(y_i, \hat{y}_i) + KLD(\mu_i, log(\sigma_i^2)))$
  where n is the number of images in the test set, $BCE(y_i, \hat{y}_i)$ refers to the binary cross-entropy loss between original image y_i and reconstructed image $\hat{y}_i$, and $KLD(\mu_i, log(\sigma_i^2)))$ is the Kullback-Leibler divergence measuring the difference between a learned latent space distribution with mean μ_i and variance σ_i^2, and a normal distribution
- Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
  where n is the number of images in the test set, y_i is the original and $\hat{y}_i$ the reconstructed image
- Root Mean Square Error (RMSE) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
  where n is the number of images in the test set, y_i is the original and $\hat{y}_i$ the reconstructed image

OOD-Sample Detection Test data for model evaluation often closely resembles the training data to achieve optimal measures. These training samples are said to be in-distribution (ID). In real-world scenarios, however, the models' generalization capabilities are often tested by being confronted with previously unseen data, called out-of-distribution-samples (OOD-samples). To ensure reliability and robustness it is thus desired to be able to identify samples that do not belong to the training distribution (Yang et al., 2024). Therefore, the following measures are used.

- Area Under the Receiver Operating Curve (AUROC): Measures the model's ability to distinguish between OOD- and ID-samples. Higher values indicate better distinction.
- Area Under the Precision-Recall Curve for ID-samples (AUPR-IN): Measures the model's ability to detect ID-samples. Higher values indicate better precision and recall.
- Area Under the Precision-Recall Curve for OOD-samples (AUPR-OOD): Measures the model's ability to detect OOD-samples. Higher values indicate better precision and recall.
- False Positive Rate at 95% True Positive Rate (FPR95TPR): Measures the FPR when the TPR is at 95%. Lower values indicate precise detection.

5 Evaluation

5.1 Datasets

ADNI "The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD)" (ADNI, 2022). Among other information, it provides PET and MRI human brain scans, all of which are in gray-scale with intensity range [0, 1]. The scans gathered from the Image and Data Archive (IDA) (ADNI, 2022) are available in the Digital Imaging and Communications in Medicine (DICOM, .dcm) format and focus on the diagnosis of AD, the most common form of dementia.
This evaluation uses a subset of the original ADNI data with 847 PET and 914 MRI images, all acquired in the axial plane. The PET data was acquired from 2006 till 2021 and includes scans from the image series ADNI Brain PET: Raw FDG, ADNI Brain PET: Raw AV45, ADNI Brain PET: Raw, ADNI (AC) FDG, ADNI2 (AC) FDG, ADNIGO - FDG BRAIN STUDY, and ADNI3-AV45 (AC). Meanwhile, the MRI scans were collected between 2006 and 2018 with the images series Axial PD/T2 FSE and Axial T2-FLAIR being used. For preparation, each scan has been adapted to the .npy (numpy array) format, resized to dimensions 256x256, normalized to intensity range [0, 255], and denoised. Additionally, an own test set with distinct subjects from the train and validation set has been created. For more information see Table 1.
The ADNI dataset was selected due to its carefully annotated and well-established PET and MRI images. In contrast, alternative datasets were characterized by sparse or non-existent annotations, suboptimal image quality, limited scale, or lack of public accessibility. However, limitations arise as ADNI does not include CT scans. For evaluation purposes, this dataset therefore had to be merged with the RSNA dataset. Although this contains CT scans, it does not have the same classification task as ADNI. Hence, a custom classification task "healthy"/"not healthy" had to be established for the merged dataset. Furthermore, access to ADNI data has to be applied for, which, for legal reasons, makes it impossible to publish the created subset for reproduction purposes.

RSNA Published by the Radiological Society of North America (RSNA) (Flanders et al., 2020) in 2019 as part of the RSNA Intracranial Hemorrhage Detection challenge on Kaggle (A. Stein, 2019), the task in this dataset is to classify whether a patient has intracranial hemorrhage (bleeding inside cranial) and if true also what exact subtypes are present. Therefore, CT human brain scans are provided, all of them acquired in the axial plane, presented in gray-scale, expressed in HUs, and wrapped as DICOM files.
The subset used for the evaluation consists of 1075 CT images, all preprocessed by

being formatted to the .npy format, resized to dimensions 256x256, normalized to intensity range [0, 255], and denoised. Distribution details can be checked in Table 1.
Finding publicly accessible, large enough, well-annotated, and high-quality CT brain scans is difficult, thus choosing the RSNA dataset was a well-justified decision. This is because, despite the deviating classification task in comparison to the ADNI dataset, it fulfills all other requirements.

TCGA-BLCA Version 8 of The Cancer Genome Atlas Urothelial Bladder Carcinoma Collection (TCGA-BLCA) presents a data collection with a focus on urothelial bladder carcinoma (bladder cancer) published as part of TCGA. It provides grayscale PET, MRI, and CT scans of the human bladder with intensity range [0, 1]. All data is provided by The Cancer Imaging Archive (TCIA) (Kirk et al., 2016) in the DICOM format.
Here, a selection from the initial TCGA-BLCA dataset is utilized, consisting of 300 PET, 300 MRI, and 300 CT images, each captured in the axial or coronal plane and with the task of classifying cancer as stage II or stage III. The specifics of the data distribution are listed in Table 1. Data preprocessing was done by transforming every image to the .npy format, resizing them to dimensions 256x256, normalizing them to intensity range [0, 255], and denoising them. Similar to ADNI, an own test set with from training and validation data distinct subjects had to be created as well.
Picking the TCGA-BLCA dataset was not much of a choice as a dataset of a second anatomy was needed for the evaluation to suggest the applicability of *CMDA* for the whole body, instead of just the brain. However, with extreme effort, only this one dataset could be found when it came to searching for an anatomy for which scans of all three implemented modalities exist. Yet, TCGA-BLCA still brings many problems along. As the dataset did not provide negative samples, a different classification task had to be created. The chosen task was the only choice with sufficient data but may be hard to accomplish due to stage II and stage III cancer being difficult to distinguish. Still, the major problem was the distribution between modalities in the original dataset. As only very few patients had undergone PET scans, lots of PET images from each patient had to be included in the subset used, resulting in invariable data. Additionally, to avoid imbalances, this forced the number of images of the other modalities the be reduced as well.

5.2 Comparative Models and Data Augmentations

Some experiments require the use of different NNs, other data augmentations, or generally applied transformations. Therefore, this Section justifies their usage and presents their benefits.

Table 1: Dateset subsets used for the evaluation. "-" indicates that no or insufficient information was given.

Dataset	Anatomy	Modality	Diagnosis	# images	Age	Sex (F/M)
ADNI	brain	PET	AD	531 (62.69%)	56-96	102/100
			CN	316 (37.31%)		
		MRI	AD	498 (54.49%)	56-95	93/109
			CN	416 (45.51%)		
RSNA	brain	CT	Intracranial Hemorrhage	642 (59,72%)	-	-
			healthy	433 (40,28%)		
BLCA	bladder	PET	stage II	150 (50,00%)	-	-
			stage III	150 (50,00%)		
		MRI	stage II	150 (50,00%)	-	-
			stage III	150 (50,00%)		
		CT	stage II	150 (50,00%)	-	-
			stage III	150 (50,00%)		

5.2.1 Models

Both qualitative and quantitative evaluation include experiments that require training a NN. For each of these cases not just one, but two separate models are trained and implemented using the PyTorch library. This strengthens the reliability and robustness of the presented findings as their consistency across models can be evaluated. Furthermore, it might reveal architecture- and pretraining-specific behavior and interactions.

ResNet-18 Introduced by He et al. (2016), ResNet-18 is a version of the Residual NN (ResNet) architecture with 18 layers. It utilizes the deep residual learning framework which enables skip connections that allow for some layers to be skipped. This helps to mitigate the problem of vanishing or exploding gradients and thus enables more effective training of very deep NNs. It combines efficiency with the ability to capture low- and high-level features, leading to good performance when it comes to image classification tasks.
If used in an experiment and not otherwise stated, the model is not pretrained, set to two output classes, and the first convolutional layer only accepts gray-scale images.

ViT-B/16 Initially designed for NLP, Vision Transformers (ViTs) follow another paradigm as they split images into patches and treat them as sequences. The experiments employ the ViT-B/16 proposed by Dosovitskiy (2020) which uses a patch size of 16x16. Instead of convolutions, ViTs make use of self-attention mechanisms that allow them the capture global dependencies and complex patterns across images better than CNNs do. However, they require large datasets for their training while also being slower than CNNs.

If used in an experiment and not otherwise stated, the model is pretrained on ImageNET1k.

5.2.2 Comparative Data Augmentations and General Transformations

To better assess the results achieved by *CMDA*, a comparison to other data augmentations is appropriate for some experiments, establishing a performance benchmark. This could also help to justify the innovation of *CMDA* by highlighting its advantages over current standards, or by giving insights into augmentation synergies when data augmentations are combined. For information on the exact augmentation pipelines used for each experiment, please refer to the corresponding GitHub repository.

imgaug imgaug is a well-known and popular library for data augmentation in image processing. It supports complex geometric and color transformations, works efficiently, provides native flexibility in designing augmentation pipelines, and is transparent as its source code is made available on GitHub.

Albumentations Known for its speed and efficiency, Albumentations also provides extensive transformations that easily integrate with deep learning frameworks (Buslaev et al., 2020). It is designed to be easy to use and customizable while the source code is once again published on GitHub.

v2 The PyTorch library also supplies built-in data augmentations in the name of transforms v2. Due to their PyTorch integration, they are easy to implement and do not require additional libraries to be used. They provide standard transformations such as normalization, flipping, or rotating, and are suitable for provisional drafts and augmentation strategies.

RandAugment Unlike the previous data augmentations, RandAugment is not a library but rather an automated augmentation method presented by Cubuk et al. (2020). It works by randomly selecting augmentation from a given set and then sequentially applying them with random intensities. Its results have proved its effectiveness in improving robustness and generalization which, besides its simple implementation, makes it a popular choice.

General Transformations The following transformations are performed on every image if used as input for a model

- ResNet-18
    1. Convert image to Tensor
    2. Normalize Tensor using a calculated mean and standard deviation

- ViT-B/16
    1. Convert image to Tensor
    2. Convert Tensor from gray-scale to RGB
    3. Resize Tensor to 224x224 pixels
    4. Normalize Tensor using a given/calculated mean and standard deviation

5.3 Quantitative Evaluation

This evaluation allows to test whether *CMDA* can be used to enhance model generalization capabilities, improve model robustness under changing conditions, and provide efficient resource utilization. All experiments shown in this Section are conducted with the ADNI/RSNA dataset. Unfortunately, there are no comparative results with other anatomical structures as the acquired TCGA-BLCA dataset is too small. Implementations of the following experiments with exact parameters used can be found in the corresponding GitHub repository.

5.3.1 Experimental Setup

Generalization performance The following experiments aim to prove the effectiveness of *CMDA* when comparing the ability of models to generalize across medical imaging modalities. Therefore, ResNet-18 and ViT-B/16 models are trained with and without data augmentations, afterward being evaluated on the same test set. The hypothesis always sees the models trained with augmented data performing better than those without. However, because of unexpected results multiple experiments were conducted, trying to support the hypothesis. In the following, let $\mathcal{M} = \{\text{ResNet-18}, \text{ViT-B/16}\}$ be a set of a ResNet-18 and a ViT-B/16 model used in the experiments, where an operation on $\mathcal{M}$ denotes that the operation is applied to each element of $\mathcal{M}$. $\mathcal{M}^{\text{all}}$ is further defined as the set of $\mathcal{M}$ for each data augmentation (e.g. $\mathcal{M}^{\text{all}} = \{\mathcal{M}^{\text{None}}, \mathcal{M}^{\text{imgaug}}, \mathcal{M}^{\text{Albumentations}}, \mathcal{M}^{\text{v2}}, \mathcal{M}^{\text{RandAugment}}, \mathcal{M}^{CMDA}\}$. Additionally, a set with one or more modality names refers to a set of images from these modalities (e.g., {PET, CT} refers to a set of non-augmented PET and CT images). If not otherwise stated, these images are taken from the combined ADNI and RSNA dataset. If the set of images is augmented by some data augmentation it is denoted by adding it as superscript (e.g. $\{\text{PET}\}^{CMDA}$ refers to a set of PET images augmented by *CMDA*). Training is always performed with the training sets, testing with the prepared test sets.

Experiment 1
Training: $\mathcal{M}^{\text{None}}$ is trained on {PET, MRI, CT}, then $\mathcal{M}^{CMDA}$ is trained on $\{\text{PET, MRI, CT}\}^{CMDA}$ where each image is randomly augmented to look like one of the other two modalities (e.g., PET images are either transformed to MRI or CT). The

same procedure is conducted for the other comparative data augmentations, yielding $\mathcal{M}^{\text{all}}$. Furthermore, $\mathcal{M}^{\text{combined}}$ is created by combining *CMDA* with each data augmentation (e.g. $\mathcal{M}^{\text{imgaug} \circ \mathit{CMDA}} \in \mathcal{M}^{\text{combined}}$).

Testing: $\mathcal{M}^{\text{all}}$ and $\mathcal{M}^{\text{combined}}$ are evaluated on {PET, MRI, CT}.

Rationale: This experiment should demonstrate how *CMDA* can be used to create more diverse training data. This should lead the models to have a higher focus on disease-specific, rather than modality-specific characteristics.

Experiment 2

Training: $\mathcal{M}^{\text{None}}_{\text{CT}}$ is trained on {PET, MRI}, $\mathcal{M}^{\text{None}}_{\text{MRI}}$ on {PET, CT}, and $\mathcal{M}^{\text{None}}_{\text{PET}}$ on {MRI, CT}. Next, $\mathcal{M}^{\mathit{CMDA}}_{\text{CT}}$ is trained on $\{\text{PET, MRI}\}^{\mathit{CMDA}}$ where each image is augmented to {CT}, $\mathcal{M}^{\mathit{CMDA}}_{\text{MRI}}$ on $\{\text{PET, CT}\}^{\mathit{CMDA}}$ where each image is augmented to {MRI}, and $\mathcal{M}^{\mathit{CMDA}}_{\text{PET}}$ on $\{\text{MRI, CT}\}^{\mathit{CMDA}}$ where each image is augmented to {PET}. Again, the same happens for all other comparative data augmentations, yielding $\mathcal{M}^{\text{all}}$ where each data augmentation now provides three sets of models instead of one (e.g., $\{\mathcal{M}^{\text{None}}_{\text{CT}}, \mathcal{M}^{\text{None}}_{\text{MRI}}, \mathcal{M}^{\text{None}}_{\text{PET}}\} \subset \mathcal{M}^{\text{all}}$).

Testing: For all data augmentations and the non-augmented case, $\mathcal{M}_{\text{CT}}$, $\mathcal{M}_{\text{MRI}}$, and $\mathcal{M}_{\text{PET}}$ are evaluated on {PET, MRI, CT}.

Rationale: With this setup, it is expected that models trained with *CMDA*-augmented data perform better than those without. The augmented images should help the models to learn more about the characteristics of the missing modality. Thus, positive results would prove *CMDA*'s usefulness in improving the generalization ability of models.

Experiment 3

Training: $\mathcal{M}^{\text{None}}_{\text{CT}}$ is trained on {PET, MRI}, then fine-tuned on {CT}, $\mathcal{M}^{\text{None}}_{\text{MRI}}$ is trained on {PET, CT}, then fine-tuned on {MRI}, and $\mathcal{M}^{\text{None}}_{\text{PET}}$ is trained on {MRI, CT}, then fine-tuned on {PET}. For *CMDA*, the custom modalities shown in Figure 20 have been added. $\mathcal{M}^{\mathit{CMDA}}_{\text{CT}}$ is thus trained on $\{\text{PET, MRI}\}^{\mathit{CMDA}}$ where each image is randomly augmented to *mod* ={PET, MRI, CT, Retinal OCT, Kidney Cortex Microscopy, Blood Cell Microscopy, Colon Pathology, Breast Ultrasound, Dermatoscopy} \ {current image modality}, then fine-tuned on {CT}. Similarly, $\mathcal{M}^{\mathit{CMDA}}_{\text{MRI}}$ is trained on $\{\text{PET, CT}\}^{\mathit{CMDA}}$ where each image is randomly augmented to *mod*, then fine-tuned on {MRI}, and $\mathcal{M}^{\mathit{CMDA}}_{\text{PET}}$ is trained on $\{\text{MRI, CT}\}^{\mathit{CMDA}}$ where each image is randomly augmented to *mod*, then fine-tuned on {PET}. To obtain $\mathcal{M}^{\text{all}}$ where each data augmentation now provides three sets of models instead of one, the same is done for all other comparative data augmentations. For this experiment, the ViT-B/16 model is not pretrained as the pretraining is done in the first

step. Additionally, as the reference images of the custom modalities are in color, the first convolutional layer of the ResNet-18 model now accepts RGB images, and all gray-scale images are transformed to RGB.

Testing: For all data augmentations and the non-augmented case, $\mathcal{M}_{\text{CT}}$ is evaluated on {CT}, $\mathcal{M}_{\text{MRI}}$ is evaluated on {MRI}, and $\mathcal{M}_{\text{PET}}$ is evaluated on {PET}.

Rationale: Similar to Experiment 1, if successful, this protocol could illustrate that models trained with *CMDA* provide a better model base for other medical imaging related tasks as they are trained with more diverse data.

Training Details
All experiments have been conducted with learning rate = 0.001, batch size = 32, epochs = 25, 5 runs to perform a kind of cross-validation, the Adam optimizer, and a 70%/15%/15% train/validate/test split.

Execution time As previously mentioned, *CMDA* should function during runtime, making it integrable into existing pipelines. Hence, its execution time is tested and compared to other data augmentations where *CMDA* is hypothesized to perform in the same magnitude as such.
As *CMDA* has multiple possible modality combinations, the data augmentations are tested by augmenting a PET, MRI, and CT dataset, each containing 300 images. The augmentation pipelines are designed to be realistic and similar in complexity, exact values can be found in the corresponding GitHub repository. The time it takes each data augmentation to augment all 300 images is measured in milliseconds, with this experiment being repeated 1000 times to account for outliers or unpredicted results.

5.3.2 Results

Generalization performance
Experiment 1
The results presented in Table 2 show that the best values for the ResNet-18 model could be achieved with Albumentations, while *CMDA* partially performs slightly better, partially slightly worse than the case where no augmentation is applied. For the ViT-B/16 model, however, *CMDA* always achieves the best results across all comparative data augmentations. Yet, the differences to the values of the non-augmented case can not be proven to be significant.
When combining the comparative data augmentations with *CMDA* it can easily be seen that, apart from RandAugment, all others benefit from *CMDA*, as highlighted by the bold values in Table 3. As before, Albumentations performs best for the ResNet-18 model, while imgaug performs best for the ViT-B/16 model, both com-

bined with *CMDA*. The results in both tables support the hypothesis rather than invalidating it but are still less meaningful than expected.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.6193	0.6437	0.6516	0.6285	0.6282	0.6259
	ViT-B/16	0.7176	0.7203	0.7194	0.7095	0.7092	0.7229
Precision	ResNet-18	0.6360	0.6620	0.6676	0.6449	0.6505	0.6446
	ViT-B/16	0.7295	0.7322	0.7322	0.7237	0.7248	0.7355
Recall	ResNet-18	0.6407	0.6537	0.6639	0.6343	0.6310	0.6282
	ViT-B/16	0.7282	0.7306	0.7301	0.7241	0.7213	0.7338
F1-Score	ResNet-18	0.6356	0.6494	0.6613	0.6312	0.6250	0.6254
	ViT-B/16	0.7268	0.7292	0.7284	0.7216	0.7186	0.7321
ROC AUC	ResNet-18	0.6801	0.7078	0.7235	0.7025	0.6879	0.6880
	ViT-B/16	0.7787	0.7807	0.7798	0.7801	0.7813	0.7819

Table 2: Metrics measured for experiment 1 where each data augmentation is tested on its own. The best values for each row are highlighted by color.

Metric	Model	imgaug+	Albumentations+	v2+	RandAugment+
Balanced Accuracy	ResNet-18	**0.6482**	**0.6657**	**0.6494**	0.6118
	ViT-B/16	**0.7235**	0.7162	**0.7119**	0.7039
Precision	ResNet-18	**0.6795**	**0.6874**	**0.6677**	0.6415
	ViT-B/16	**0.7350**	0.7307	**0.7274**	0.7208
Recall	ResNet-18	**0.6593**	**0.6755**	**0.6556**	0.6204
	ViT-B/16	**0.7333**	0.7287	**0.7264**	0.7167
F1-Score	ResNet-18	0.6474	**0.6703**	**0.6522**	0.6050
	ViT-B/16	**0.7321**	0.7259	**0.7236**	0.7133
ROC AUC	ResNet-18	**0.7312**	**0.7393**	**0.7229**	**0.6910**
	ViT-B/16	**0.7830**	**0.7809**	**0.7808**	**0.7854**

Table 3: Metrics measured for experiment 1 where the + indicates that each data augmentation is tested in combination with *CMDA*. The best values for each row are highlighted by color while values that are better in comparison to the data augmentation's application without *CMDA* are put in bold.

Experiment 2

Tables 4, 5, and 6 illustrate that this experiment fails to confirm the hypothesis. Apart from some values in Table 4, *CMDA* performs notably worse than all other data augmentations and mostly also worse than the case where no data augmentation is applied. Instead, imgaug stands out through good results across all left-out modalities.

Experiment 3

Similar to experiment 2, Tables 7 and 8 show that results achieved by *CMDA* are partially substantially worse than comparative measures. Table 9 is the only case where all *CMDA* values are better than those where no data augmentation has been applied. Overall, Albumentations, v2, and RandAugment perform the best.

Execution time As can be seen in Table 10, *CMDA* takes just slightly longer than the comparative data augmentations, whereby Albumentations is a special case as it is specifically designed for being fast. While the standard deviations of *CMDA*

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet	0.5605	0.5958	0.5543	0.5661	0.5544	0.5579
	ViT	0.5662	0.5673	0.5680	0.5638	0.5764	0.5929
Precision	ResNet	0.5795	0.6297	0.5823	0.6165	0.5948	0.5810
	ViT	0.5875	0.5897	0.5892	0.5865	0.6049	0.6369
Recall	ResNet	0.5484	0.5976	0.5538	0.5830	0.5805	0.5864
	ViT	0.5995	0.6024	0.6010	0.6000	0.6107	0.6409
F1-Score	ResNet	0.5360	0.5788	0.5276	0.5425	0.5589	0.5736
	ViT	0.5851	0.5862	0.5869	0.5816	0.5933	0.6096
ROC AUC	ResNet	0.5875	0.6201	0.6000	0.5830	0.6085	0.5814
	ViT	0.6162	0.6214	0.6176	0.6187	0.6188	0.6135

Table 4: Metrics measured for experiment 2 where CT data is not included in the training. The best values for each row are highlighted by color.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.5752	0.5902	0.6041	0.6009	0.5358	0.5192
	ViT-B/16	0.6000	0.5976	0.5951	0.5972	0.5959	0.5513
Precision	ResNet-18	0.5950	0.6152	0.6250	0.6225	0.5456	0.5456
	ViT-B/16	0.6143	0.6129	0.6113	0.6137	0.6159	0.5706
Recall	ResNet-18	0.5538	0.5674	0.6112	0.5830	0.5669	0.5849
	ViT-B/16	0.6170	0.6175	0.6161	0.6190	0.6234	0.5805
F1-Score	ResNet-18	0.5501	0.5623	0.6053	0.5783	0.5300	0.5030
	ViT-B/16	0.6147	0.6135	0.6113	0.6136	0.6135	0.5691
ROC AUC	ResNet-18	0.6200	0.6164	0.6675	0.6459	0.6003	0.5656
	ViT-B/16	0.6287	0.6146	0.6171	0.6193	0.6467	0.5772

Table 5: Metrics measured for experiment 2 where MRI data is not included in the training. The best values for each row are highlighted by color.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.5926	0.6006	0.6090	0.6024	0.5507	0.5088
	ViT-B/16	0.6814	0.6863	0.6837	0.6830	0.6833	0.5975
Precision	ResNet-18	0.6106	0.6188	0.6255	0.6199	0.5676	0.5281
	ViT-B/16	0.7131	0.7210	0.7161	0.7040	0.7066	0.6169
Recall	ResNet-18	0.6024	0.6147	0.6127	0.6146	0.5645	0.5523
	ViT-B/16	0.7109	0.7178	0.7139	0.7022	0.7075	0.6165
F1-Score	ResNet-18	0.5979	0.6112	0.6107	0.6111	0.5547	0.5162
	ViT-B/16	0.6999	0.7058	0.7026	0.6966	0.7004	0.6098
ROC AUC	ResNet-18	0.6721	0.6802	0.6674	0.6593	0.6174	0.5373
	ViT-B/16	0.7689	0.7620	0.7637	0.7610	0.7733	0.6460

Table 6: Metrics measured for experiment 2 where PET data is not included in the training. The best values for each row are highlighted by color.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.7047	0.7160	0.7373	0.7413	0.6046	0.6510
	ViT-B/16	0.5861	0.5722	0.5876	0.6174	0.6020	0.5830
Precision	ResNet-18	0.7365	0.7509	0.7577	0.7650	0.7137	0.6891
	ViT-B/16	0.6389	0.6066	0.6534	0.6351	0.6262	0.6503
Recall	ResNet-18	0.7130	0.7354	0.7578	0.7466	0.6161	0.6857
	ViT-B/16	0.6124	0.6000	0.6137	0.6398	0.6261	0.6099
F1-Score	ResNet-18	0.7069	0.7267	0.7535	0.7430	0.5553	0.6628
	ViT-B/16	0.5704	0.5544	0.5728	0.6352	0.6150	0.5686
ROC AUC	ResNet-18	0.8111	0.8272	0.8248	0.8354	0.7714	0.7513
	ViT-B/16	0.6909	0.6780	0.6917	0.6881	0.6885	0.6880

Table 7: Metrics measured for experiment 3 where CT data is not included in the first and then fine-tuned on in the second training stage. The best values for each row are highlighted by color.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.6075	0.6108	0.6128	0.6550	0.6450	0.5881
	ViT-B/16	0.5222	0.5239	0.5222	0.5047	0.5261	0.5208
Precision	ResNet-18	0.6563	0.6507	0.6526	0.6965	0.6874	0.6293
	ViT-B/16	0.6528	0.6516	0.6529	0.5425	0.6097	0.6511
Recall	ResNet-18	0.6696	0.6625	0.6679	0.6750	0.6964	0.6661
	ViT-B/16	0.6286	0.6321	0.6287	0.6089	0.6179	0.6268
F1-Score	ResNet-18	0.6488	0.6429	0.6525	0.6678	0.6853	0.6292
	ViT-B/16	0.5524	0.5507	0.5524	0.5435	0.5621	0.5510
ROC AUC	ResNet-18	0.6924	0.7165	0.7092	0.7229	0.7344	0.6844
	ViT-B/16	0.5842	0.5975	0.5815	0.5864	0.5697	0.5751

Table 8: Metrics measured for experiment 3 where MRI data is not included in the first and then fine-tuned on in the second training stage. The best values for each row are highlighted by color.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
Balanced Accuracy	ResNet-18	0.6812	0.6855	0.6769	0.6930	0.6196	0.6903
	ViT-B/16	0.5244	0.5336	0.5261	0.5606	0.5688	0.5268
Precision	ResNet-18	0.6931	0.6958	0.6873	0.7051	0.6340	0.6994
	ViT-B/16	0.6148	0.6041	0.6057	0.5888	0.5782	0.6219
Recall	ResNet-18	0.6884	0.6870	0.6768	0.7000	0.6130	0.6971
	ViT-B/16	0.5493	0.5580	0.5493	0.5681	0.5797	0.5507
F1-Score	ResNet-18	0.6836	0.6835	0.6730	0.6950	0.5738	0.6935
	ViT-B/16	0.4555	0.4709	0.4636	0.5393	0.5638	0.4619
ROC AUC	ResNet-18	0.7435	0.7547	0.7297	0.7515	0.7063	0.7481
	ViT-B/16	0.5721	0.6152	0.5992	0.5989	0.6217	0.6058

Table 9: Metrics measured for experiment 3 where PET data is not included in the first and then fine-tuned on in the second training stage. The best values for each row are highlighted by color.

are notably higher than the others, the average execution times are still within an acceptable time range and thus integration viable, supporting the hypothesis.

Dataset	imgaug	Albumentations	v2	RandAugment	*CMDA* to MRI	*CMDA* to CT
PET	64.88 ± 3.67	11.37 ± 1.48	53.75 ± 9.08	104.04 ± 9.30	150.54 ± 52.08	103.64 ± 23.68

Dataset	imgaug	Albumentations	v2	RandAugment	*CMDA* to PET	*CMDA* to CT
MRI	64.27 ± 3.96	11.91 ± 1.64	53.89 ± 8.68	103.23 ± 8.63	121.63 ± 48.49	90.59 ± 19.34

Dataset	imgaug	Albumentations	v2	RandAugment	*CMDA* to PET	*CMDA* to MRI
CT	65.15 ± 3.65	12.13 ± 1.56	54.10 ± 8.71	103.40 ± 8.48	112.76 ± 44.67	89.11 ± 37.69

Table 10: Average execution times and their standard deviations in milliseconds (ms).

5.4 Qualitative Evaluation

The following investigations enable the examination of *CMDA* preserving content, aligning with the target modality distribution, as well as retaining medical image integrity. All experiments shown in this Section are conducted with the ADNI/RSNA dataset. To see results of the same experiments conducted with the TCGA-BLCA dataset, please refer to Appendix A.2.2. Implementations of the following experiments with exact parameters used can be found in the corresponding GitHub repository.

5.4.1 Experimental Setup

GLCM features The experimental setup begins by taking two datasets, one from the original modality (D_O) and one from the target modality (D_T). D_O is then randomly split into two distinct datasets (D_O^O, D_O^A) of the same size, such that $size(D_O) = size(D_O^O) + size(D_O^A)$. D_O^A is special in that it is augmented from the original modality to the target modality by *CMDA* with given intensities ranging from 0% to 100%, rising by 10% each run, resulting in eleven runs. A run starts by augmenting D_O^A and calculating the mentioned GLCM features for each element in each dataset and creating a mean GLCM feature vector with five values for every dataset ($V_{D_O^O}, V_{D_O^A}, V_{D_T}$). These vectors are now Z-score standardized to avoid high values dominating, as afterward the Euclidean distances $||V_{D_O^A} - V_{D_O^O}||$ and $||V_{D_O^A} - V_{D_T}||$ are calculated. The two distances are saved which ends a run. After all runs are completed, the results are plotted on a diagram, allowing for a direct visual comparison. The hypothesis is that with rising augmentation intensities the Euclidean distance $||V_{D_O^A} - V_{D_O^O}||$ should increase while $||V_{D_O^A} - V_{D_T}||$ should decrease. This experiment is conducted six times, once for each possible distinct combination of implemented modalities.

FID The experiment protocol is extremely similar to that of the GLCM features. It differs only in that instead of calculating the GLCM features, the Inception-v3 model is used to extract features $f_{D_O^O}, f_{D_O^A}, f_{D_T}$ for each dataset, whereupon the FIDs $d^2(f_{D_O^A}, f_{D_O^O})$ and $d^2(f_{D_O^A}, f_{D_T})$ are calculated. The hypothesis is adapted as with rising intensities $d^2(f_{D_O^A}, f_{D_O^O})$ should increase while $d^2(f_{D_O^A}, f_{D_T})$ should decrease.

PCA Once again, this experimental setup uses D_O^O, D_O^A, and D_T, where D_O^A is one non-augmented and once augmented with intensity 100%. However, this time a PCA to three dimensions is carried out for each image in each dataset. The corresponding lower-dimensional data is then plotted into one joint 3D plot with different symbols and colors depending on the current dataset, resulting in lots of data points. In the end, it is expected that with rising intensity the lower-dimensional data produced by D_O^A moves closer to the data points of D_T and further away from D_O^O. Again, the experiment is conducted six times, once for each possible distinct combination of implemented modalities.

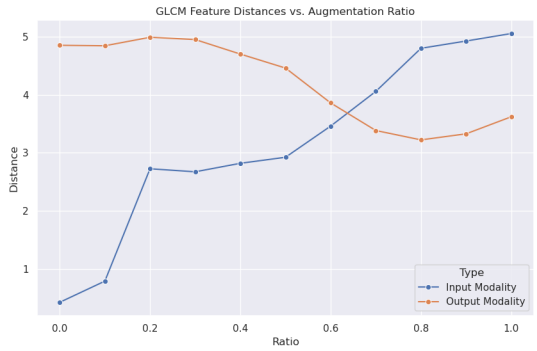
VAE The experiment starts with loading a dataset D_O^O with images from the original modality as well as a dataset D_T with images from the target modality. Additionally, D_O^O is augmented by *CMDA* with 100% intensity, creating a new augmented dataset D_O^A. Next, two independent VAEs (VAE_O, VAE_A) are created with input size 256x256, size of the hidden layer being 400, and size of the latent space being 20. VAE_O is then trained on D_O^O, while VAE_A is trained on D_O^A, both with a learning rate of 0.001 and for 50 epochs. Finally, both VAEs are evaluated on D_T. The hypothesis is that, in comparison to VAE_O, all three metrics decrease when VAE_A is evaluated on D_T. This would indicate that VAE_A is better at image reconstruction and provides more accurate latent space representations. The experiment is again repeated six times.

OOD-Sample Detection This experiment used the pytorch-ood library from Kirchheim et al. (2022) to create an energy-based OOD-detector det_{OOD} proposed by Liu et al. (2020).
The experimental protocol begins by loading a dataset $D_{train\&val}$ of the original modality and then splitting it into training (D_{train}) and validation (D_{val}) set, using a 70%/15%/15% split. Next, the test set D_{test} is created by merging another dataset D_{test}^{ID} of the original modality with a dataset D_{test}^{OUT} of the target modality. Now a run starts by training an untrained ResNet-18 model and an untrained ViT on $D_{train\&val}$ for 25 epochs, using a learning rate of 0.001 and a batch size of 32. det_{OOD} then tests both models on D_{test} for their OOD-detection capabilities.
This run is performed five times with random train/val splits and then takes the average for each of the measured metrics to implement some kind of cross-validation. The whole procedure is conducted six times, where D_{train} is not augmented, or augmented by imgaug, Albumentations, v2, RandAugment, or *CMDA*. To prove *CMDA*'s capabilities of aligning images with the distribution of the target modality,

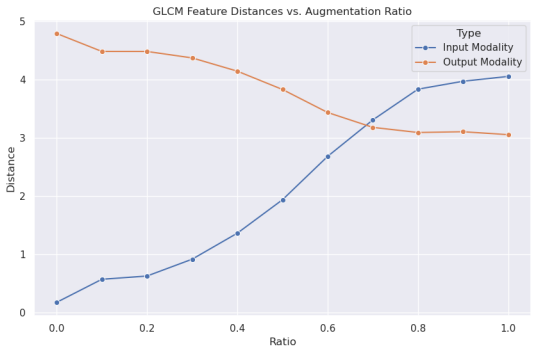
it is expected that AUROC, AUPR-IN, and AUPR-OOD decrease while FPR95TPR increases for both the ResNet-18 and the ViT model when D_{train} is augmented with *CMDA*. This is meant in comparison to all other data augmentations and the non-augmented case.

5.4.2 Results

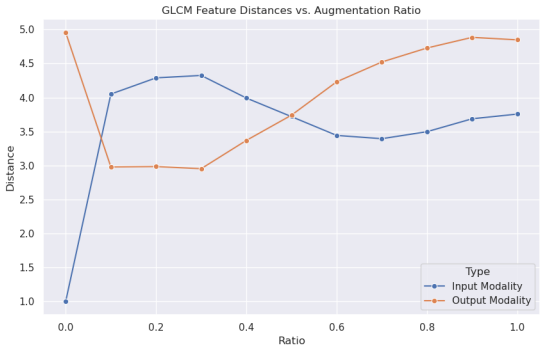
GLCM features Figure 21 presents the Euclidean distances $||V_{D_O^A} - V_{D_O^O}||$ and $||V_{D_O^A} - V_{D_T}||$ for a range [0, 1] (intensities [0%, 100%]) of augmentation ratios (intensities) that *CMDA* has been applied with to D_O^A. The orange graph displays the distances $||V_{D_O^A} - V_{D_T}||$ to the target (Output) modality while the blue graph illustrates the distances $||V_{D_O^A} - V_{D_O^O}||$ to the original (Input) modality.
It can be observed that for $Ratio = 0.0$, distance $||V_{D_O^A} - V_{D_T}||$ is constantly higher than for $Ratio = 1.0$. Furthermore, some cases also show that the smallest distance might as well be achieved at medium intensities. Complementary, for $Ratio = 0.0$, distance $||V_{D_O^A} - V_{D_O^O}||$ is constantly lower than for $Ratio = 1.0$. Thus, this experiment's hypothesis is supported by the presented findings.
The exact distances can be taken from Table 13 in Appendix A.2.1.



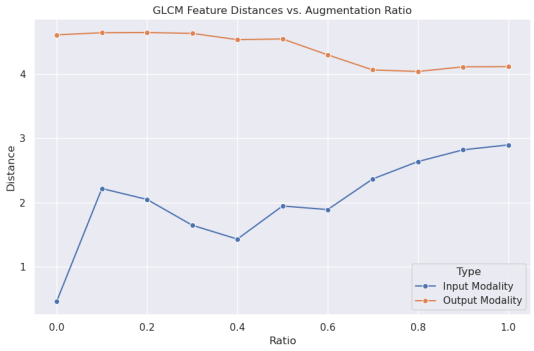
PET to MRI



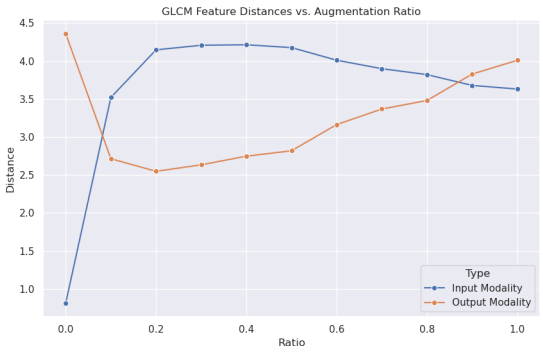
PET to CT



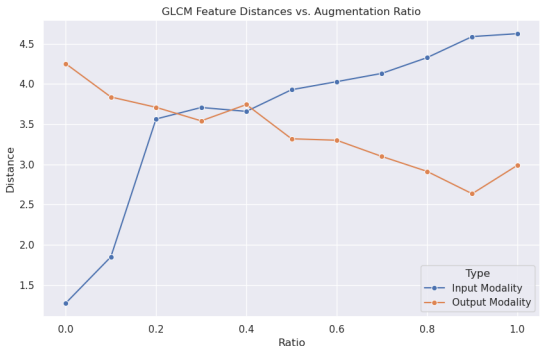
MRI to PET



MRI to CT



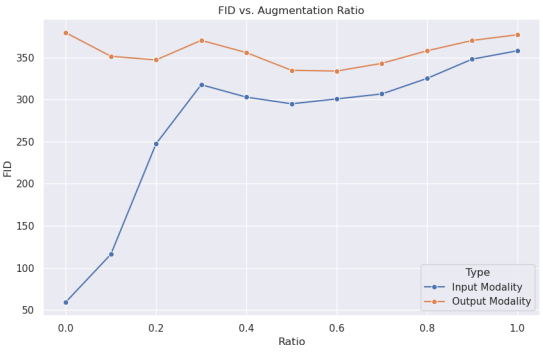
CT to PET



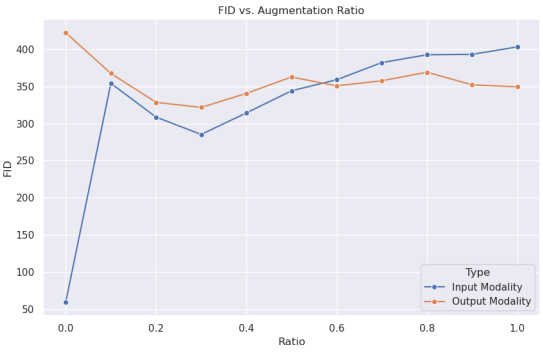
CT to MRI

Figure 21: Euclidean distances between GLCM features of augmented dataset to original and target modality dataset. The X-axis shows the augmentation intensity applied to a set of images from the original modality. The Y-axis displays the distance of this augmented set to a set of images from the original modality (blue graph) and the target modality (orange graph). The captions display the translation direction.

FID The FIDs $d^2(f_{D_O^A}, f_{D_O^O})$ and $d^2(f_{D_O^A}, f_{D_T})$, where D_O^A has been augmented by *CMDA* with ratio [0, 1], are illustrated in Figure 22. Again, the orange graph displays the distances $d^2(f_{D_O^A}, f_{D_T})$ to the target (Output) modality whereas the distance $d^2(f_{D_O^A}, f_{D_O^O})$ to the original (Input) modality is shown by the blue graph. Similar to the graphs of the GLCM features, it can be seen that for $Ratio = 0.0$, distance $d^2(f_{D_O^A}, f_{D_T})$ ($d^2(f_{D_O^A}, f_{D_O^O})$) is constantly higher (lower) than for $Ratio = 1.0$. However, it should be noted that the largest decreases in $d^2(f_{D_O^A}, f_{D_T})$ are attained with target modality PET while the other transformations produce rather small improvements. Nevertheless, the results shown are still supportive of the formulated hypothesis.
The exact FIDs can be taken from Table 14 in Appendix A.2.1.



PET to MRI



PET to CT

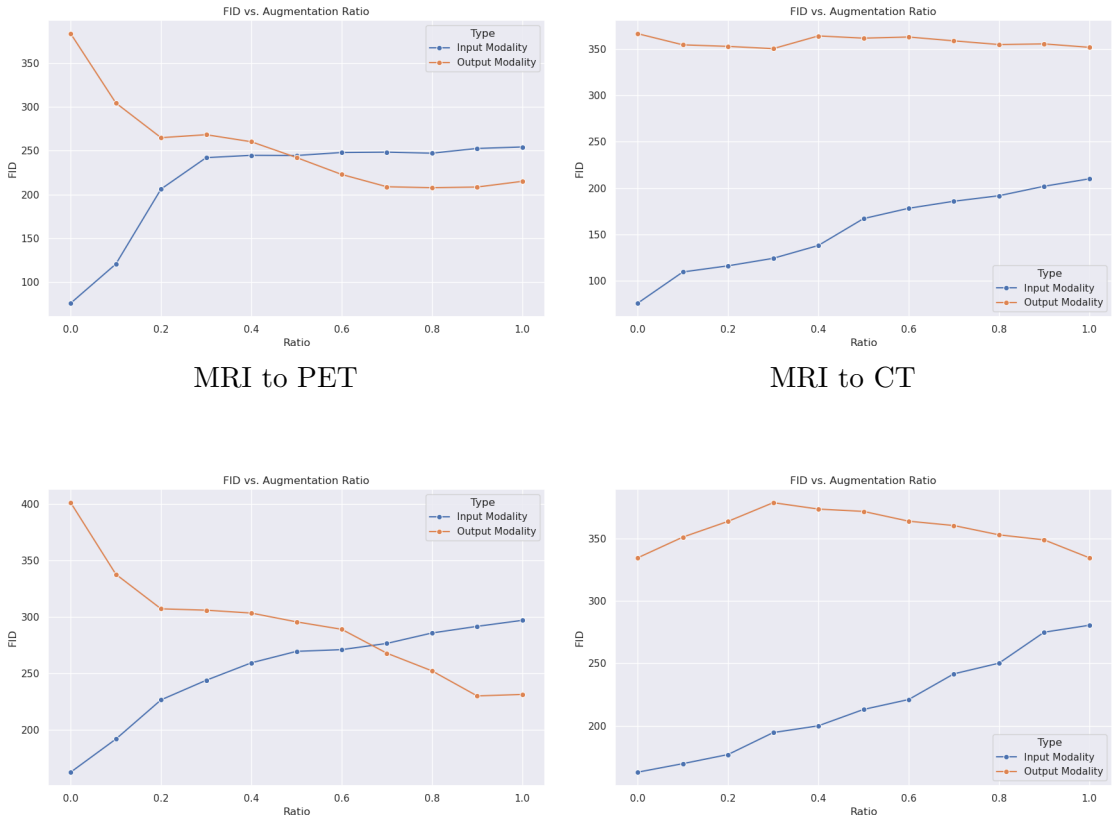
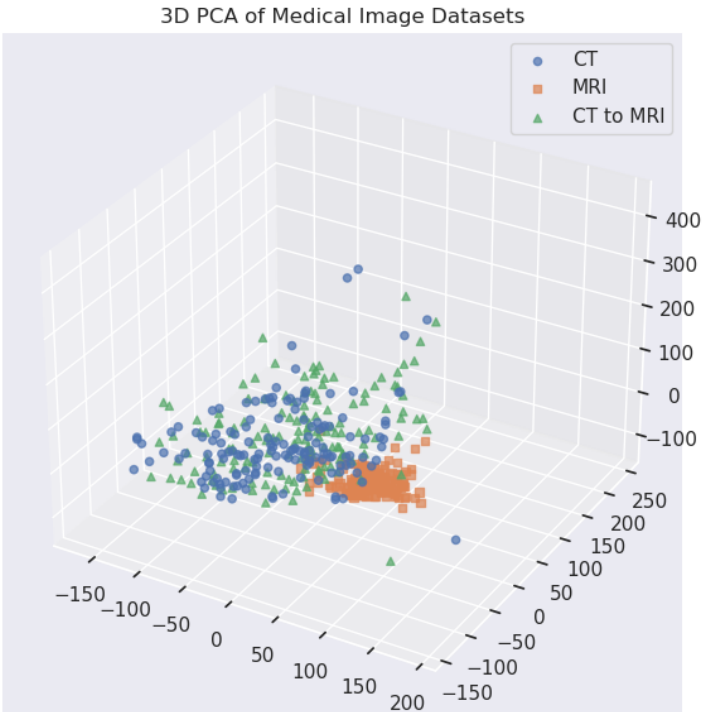


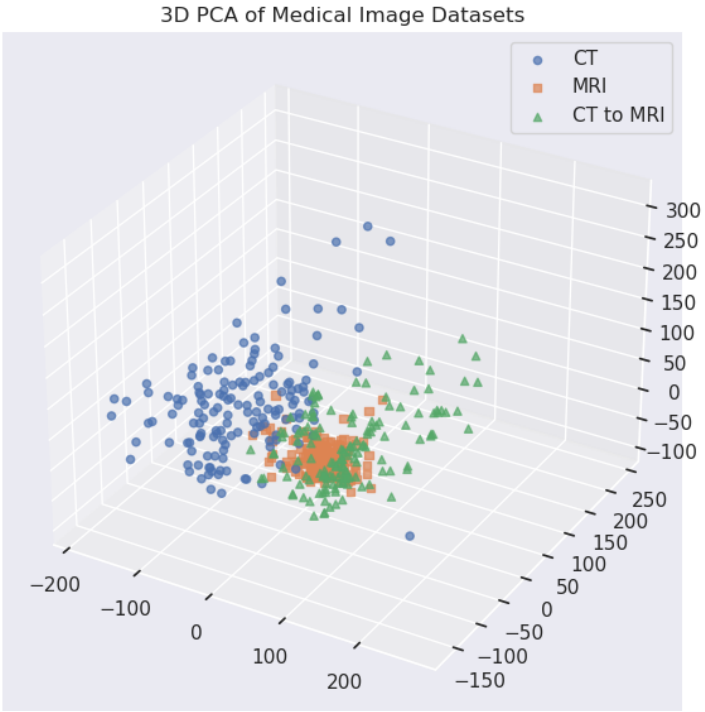
Figure 22: FIDs of augmented dataset to original and target modality dataset. The X-axis shows the augmentation intensity applied to a set of images from the original modality. The Y-axis displays the distance of this augmented set to a set of images from the original modality (blue graph) and the target modality (orange graph). The captions display the translation direction.

PCA The results illustrated in Figure 23 and the other Figures 32, 33, 34, 35, 36 located in Appendix A.2.1, provide visual support of the set hypothesis. They show that for almost all transformations the three most influential principal components of D_O^A seem to align with those of D_T when D_O^A is augmented by *CMDA* with the corresponding target modality and an intensity of 100%. Meanwhile, they align with D_O^O's principal components if not augmented (intensity 0%). The only exception to this is the transformation with original modality MRI and target modality CT, where the principal components seem to stay similar to those of the original modality, no matter how heavy D_O^A is augmented.

VAE As hypothesized, substantial results supporting the formulated claim are achieved by this experiment. The metrics listed in Table 11 demonstrate that Test Loss and RMSE are consistently lower for the VAE_A trained with D_O^A. The MAE follows that schema, only deviating for original modality PET and target modality MRI, where it is slightly lower for VAE_O.



CT to MRI 0% augmented



CT to MRI 100% augmented

Figure 23: Comparison of 3D-PCAs for CT dataset augmented to MRI. The upper image shows a PCA with a dataset of CT images, one of MRI images, and one of CT images that should be augmented. The lower image shows a PCA with the same datasets, just that the second CT dataset has been augmented by *CMDA* with target modality MRI and intensity 100%.

Modalities	$\boldsymbol{VAE}_O$			$\boldsymbol{VAE}_A$		
	Test Loss	MAE	RMSE	Test Loss	MAE	RMSE
PET to MRI	22492	0.0779	0.1409	17492	0.0791	0.1377
PET to CT	61032	0.1289	0.2616	26607	0.1217	0.2505
MRI to PET	13321	0.0750	0.1376	10812	0.0472	0.0924
MRI to CT	28423	0.1333	0.2685	27847	0.1279	0.2665
CT to PET	15072	0.1050	0.1611	10764	0.0401	0.0836
CT to MRI	22287	0.1507	0.2223	19441	0.0954	0.1664

Table 11: Test Losses, MAEs, and RMSEs for the VAE experiment.

OOD-Sample Detection Table 12 summarizes the collected results of this experiment by calculating the mean and standard deviation of each metric across all possible combinations of original and target modalities. It can be noted that all data augmentations can reduce the number of detected OOD-samples in the test set. Although the best values are distributed evenly across all data augmentations, *CMDA* is always close to the winning value if not winning itself. Additionally, performed t-tests showed that five of eight values are significantly better for *CMDA* in comparison to None, especially the experiments performed with the ViT-B/16 model ($\alpha = 0.05$, $n = 30$). The full data for each combination of original and target modalities is provided in Tables 15, 16, 17, 18, 19, 20 further shows that while other augmentations perform very well for a single combination of modalities, *CMDA* performs well across the board. Apart from the transformations with CT as original modality, the results also back the formulated hypothesis.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.477	0.456	0.463	0.480	0.455	0.454
	ViT-B/16	0.600	0.466	0.640	0.441	0.719	**0.539**
AUPR-IN	ResNet-18	0.495	0.530	0.515	0.547	0.491	0.507
	ViT-B/16	0.582	0.471	0.634	0.456	0.683	**0.519**
AUPR-OUT	ResNet-18	0.538	0.541	0.512	0.551	0.537	0.522
	ViT-B/16	0.648	0.528	0.652	0.529	0.742	**0.584**
FPR95TPR	ResNet-18	0.870	0.844	0.905	0.881	0.842	**0.940**
	ViT-B/16	0.780	0.889	0.826	0.877	0.728	**0.875**

Table 12: OOD-detection metrics with multiple data augmentations. The calculated numbers represent the means over all possible combinations of original and target modality. The best values for each row are highlighted by color while *CMDA*-values that are significantly better than None-values are put in bold.

6 Discussion

6.1 Limitations and Obstacles

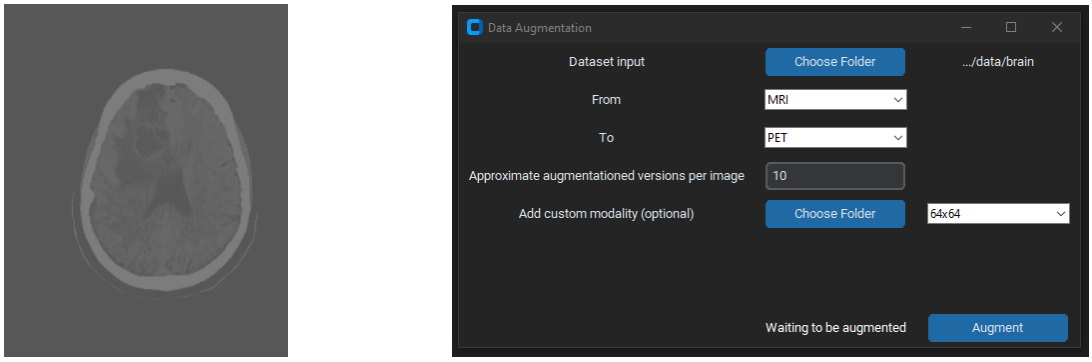
During the development of *CMDA*, many challenges arose, affecting and sometimes limiting the progress that was achieved.

Datasets As already outlined in Section 1, medical data is scarce. This was immediately noticeable in the early stages of development as there was a severe lack of suitable datasets. The challenge of finding labeled PET, MRI, and CT scans of the same anatomical structure lead to the first programming experiments being conducted with MedMNIST v2 (Yang et al., 2023) or other, modality-incomplete datasets. Eventually, the decision was made to build a custom dataset. This was achieved by combining the ADNI dataset containing PET and MRI (ADNI, 2022) with the RSNA dataset containing CT images (A. Stein, 2019). Finding a second labeled dataset with an equal anatomical structure and all three modalities included was needed to prove the generalization capabilities of *CMDA*. However, with TCGA-BLCA (Kirk et al., 2016), it was even harder to find than the first datasets.
Although the datasets used, provide the current, to best knowledge, most appropriate options, they still pose limitations. The combination of ADNI and RSNA led to the creation of a new classification task "healthy" / "not healthy" as the initial tasks did not align. This made it substantially harder for NNs to find patterns in the data. The in Section 5 presented results are also not easily reproducible because the ADNI subset cannot be shared for legal reasons. Additionally, test sets with distinct subjects had to manually be built for ADNI and TCGA-BLCA as no test sets were provided. The small size (∼300 images per modality) of all datasets also made it hard to train models, even though the size of the ADNI/RSNA dataset could be increased (∼900 images per modality) during the end of development.

Image Processing Despite DICOM being the industry standard, medical images were provided in various image formats such as .npz, .npy, .nii, .tiff, .png, or .jpg, leading to challenging formatting work. Some datasets also imposed cropping of images to get uniform dimensions across the data, or modality-specific operations like converting CT scans from HU to gray-scale.

Engineering of *CMDA* The final state of *CMDA* is the result of many tried and tested approaches, some of which were unsuccessful, especially regarding texture manipulation. The attempt of computing various textural features using related techniques such as local binary pattern (LBP) (Ojala et al., 2002), histogram of oriented gradients (HOG) (Dalal and Triggs, 2005), Gabor filter (Mehrotra et al., 1992), fast Fourier transform (FFT) (Cooley and Tukey, 1965), or Haar-Wavelet transforms (Haar, 1910) ultimately turned out to not be useful for the transformation itself. This is because the computed features could be altered but not be

reversed to images again, which was the initial idea. Their only use in this scenario was for evaluation purposes as can be seen for the GLCM features.
In the beginning, there were also doubts about whether the overall idea of *CMDA* made sense. Different modalities capture different disease patterns in different detail, which cannot be translated to other modalities, leading to augmented images lacking realism. However, it was concluded that a rough modality representation is sufficient as models can still use the augmented images to learn about the characteristics of each modality.



(a) Faulty histogram matching of a CT image.

(b) GUI for *CMDA* that was developed under wrong assumptions and later discarded.

Figure 24: Challenges during the development of *CMDA*. The left image shows the results of faulty histogram matching where the blank space around the head is transformed as well. The right image illustrates a GUI for the usage of *CMDA* that was developed and later discarded.

The augmentations themselves had to be implemented, ensuring their effectiveness across gray-scale and color images as added custom modalities might provide RGB images. The color augmentation was specifically tricky as it should only operate on the anatomical structure present in the scan, avoiding the remaining blank space around it. If ignored, this caused augmentations to look as illustrated in Figure 24a, where the color of the blank space is transformed as well.

Due to missing knowledge about existing data augmentations and their functioning, the early development also saw a graphical user interface (GUI) being built to make *CMDA*'s usage easier. This happened under to assumption that the augmentations are performed before rather than during runtime and can be seen in Figure 24b.

Similar to other data augmentations, the final version of *CMDA* works as intended but rarely still produces poor visual results. A collection of faulty augmentations can be found in Figure 25.

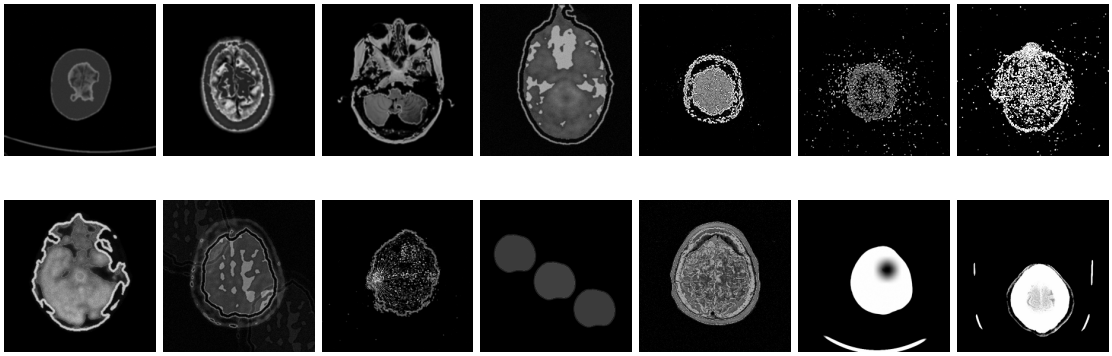


Figure 25: Various images of human brain scans with faulty augmentations by *CMDA* that were created when augmenting a large number of images. They include images from all possible modality combinations.

Experiments Both the implementation, as well as the results of the experiments conducted as part of the evaluation have partially caused problems. At one point, the final evaluation used an energy-based OOD-detector, but first setups used a self-implemented softmax-based detection that did not work correctly.
However, the biggest experimental challenge was the quantitative evaluation assessing the generalization performance of *CMDA*. Different experimental setups tried with various predefined weights, generalization tests, modality combinations, and regularization techniques, but no experiment achieved the results hypothesized to satisfaction. It was furthermore tested if a reduced size of the dataset would boost *CMDA*'s performance in comparison to no data augmentation used, but the minimal boost in augmentation metrics that was achieved could not be proven to be significant.

6.2 Analysis and Interpretation

The presented outcomes of the conducted evaluation provide valuable insights into the effectiveness of *CMDA* within the medical imaging domain. The qualitative experiments produced the expected good results across all measures implemented. This demonstrates *CMDA*'s capabilities to effectively align the image distributions from a given to a target modality. These achievements strongly suggest that it successfully preserves medical image integrity and addresses the cross-modality shift. While certain combinations of given and target modalities may yield better results than others in certain experiments, a broader analysis shows that the performances tend to balance out across all experiments. The quantitative experiments however only demonstrated satisfactory execution times but indicated poor improvements in models' generalization performance. These results suggest efficient resource utilization on the one hand. Yet, they raise questions about its utility in helping models to generalize to unseen domains and improving their understanding of certain modalities on the other.

This discrepancy between qualitative success and quantitative deficits is surprising and unexpected. The augmented images seem to look like those from the target modality but strangely enough, do not help during training. That said, these findings could be explained by the necessary combination of datasets with different classification tasks, where Alzheimer's disease has completely different visual characteristics than intracranial hemorrhage. The translation might thus not help the model to generalize across domains as it might learn about the characteristics of the target modality but not about those of the disease itself. Moreover, the small dataset size may have caused insufficient data diversity, hindering the training process of the models and leading to non-representative results. Future evaluations could therefore use more extensive datasets including a greater variety of images that all examine the same disease, potentially mitigating the issues observed in this study.
Another unforeseen finding is that the qualitative experiments conducted on the TCGA-BLCA dataset yielded worse results in comparison to the ADNI/RSNA dataset, which might imply that *CMDA* is better suited for brain imaging than for other anatomical structures. Nonetheless, the OOD-experiments still indicate its general applicability across all anatomical contexts.
Direct comparisons to related results are challenging due to *CMDA*'s novel approach of runtime cross-modality augmentation. Nevertheless, the discussed findings contribute to the broader context of demonstrating the effectiveness of modality-specific data augmentations. Furthermore, *CMDA* advances the current state of the art through a runtime data augmentation technique in the medical domain. Although it does not improve the generalization performance of deep learning algorithms as expected, it can accomplish cross-modality translation to generate new training samples that closely resemble the target modality's distribution.

7 Conclusion

This paper introduced *CMDA*, a novel cross-modality-wise data augmentation that enables real-time image translation between medical imaging modalities. It can seamlessly be integrated with existing data augmentations and works for the brain and bladder, indicating that it might be applicable to all anatomical structures. The evaluation carried out highlights substantial capabilities of *CMDA* addressing the cross-modality shift by aligning image distributions across modalities. This holds true despite observed discrepancies when it comes to generalization improvements of NNs. These results suggest that, while *CMDA* can effectively produce visually similar and realistic images of the target modality, it may not enhance a model's generalization performance across unseen domains. At least, this limitation appears to be relevant with different diseases being present.

The significance of this work lies in its potential to advance modality-specific data augmentations in the medical domain that can be used during runtime. Furthermore, it supports the creation of more diverse training data to facilitate more robust models. Still, the results show that larger and more homogeneous datasets are required to fully realize *CMDA*'s potential in improving models' abilities to generalize to novel modalities.

Building on these insights, future research could focus on refining and further evaluating the current state of *CMDA*. To be exact, implementing additional modalities, or algorithms to sample more diverse reference images could help the data augmentation to create an even greater variety of data. To enhance the performance assessment of *CMDA*, more experiments in which it is combined with other data augmentations could be conducted. Additionally, its application across various diseases could be explored. This has the potential to provide deeper insights into which medical conditions might be more sensitive to modality translation than others.

In conclusion, this work contributes to the ongoing discourse on addressing the data scarcity present in the medical imaging domain. Therefore, it provides an idea and foundation for further exploration of real-time cross-modality translation.

A Appendix

A.1 Sample augmentations

PET to MRI

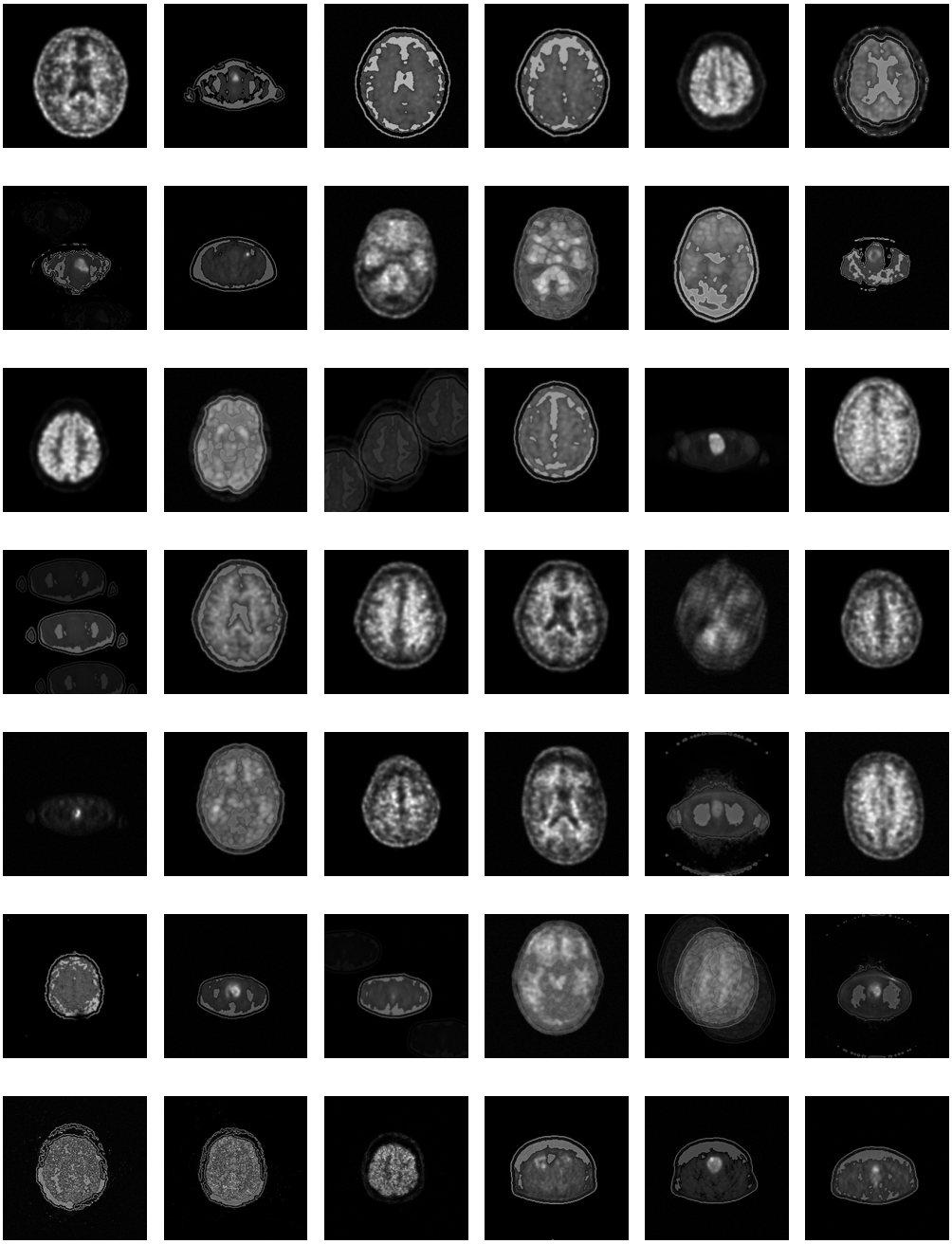


Figure 26: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy PET images of the human brain (ADNI, 2022) and bladder (Kirk et al., 2016) are transformed to target modality MRI.

PET to CT

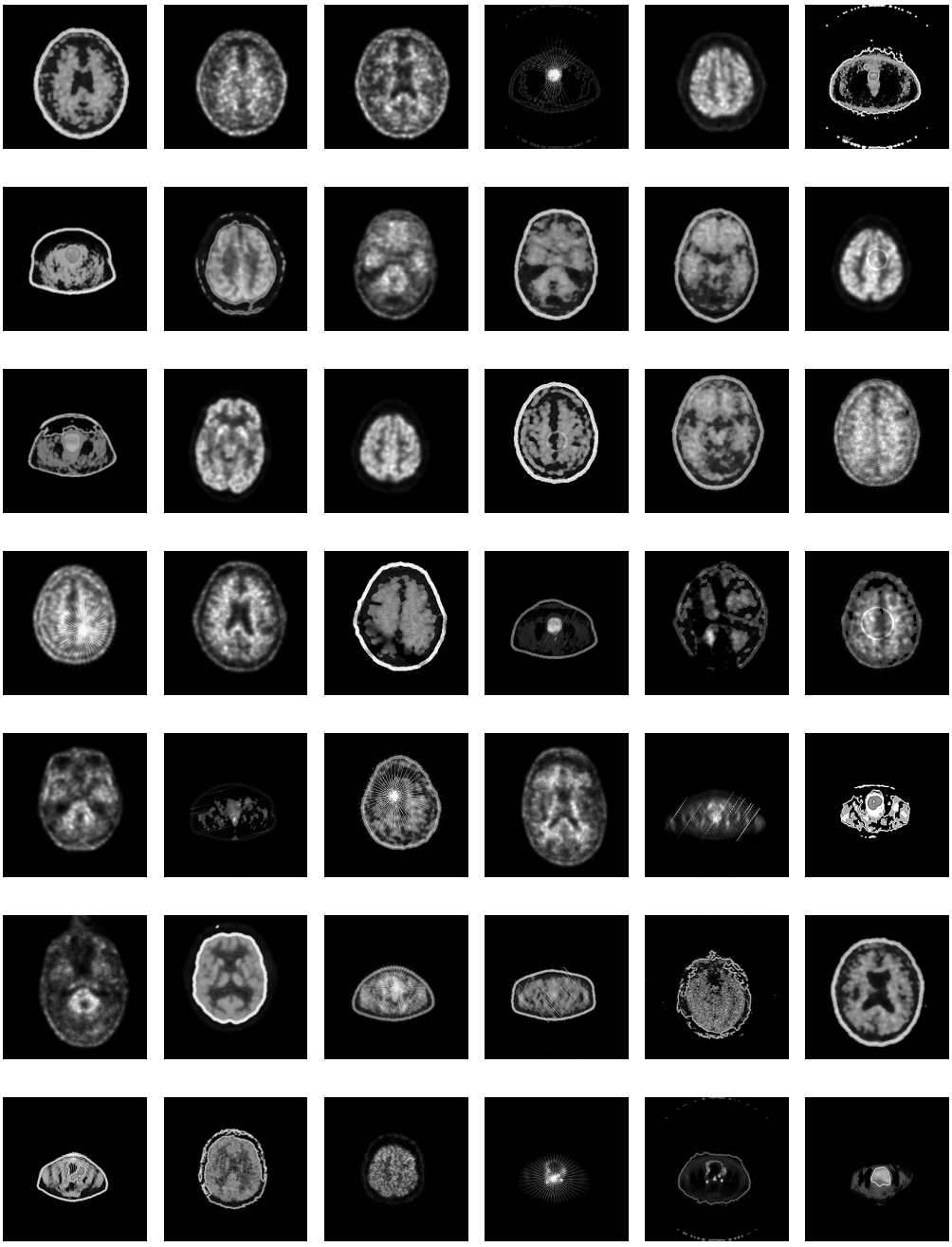


Figure 27: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy PET images of the human brain (ADNI, 2022) and bladder (Kirk et al., 2016) are transformed to target modality CT.

MRI to PET

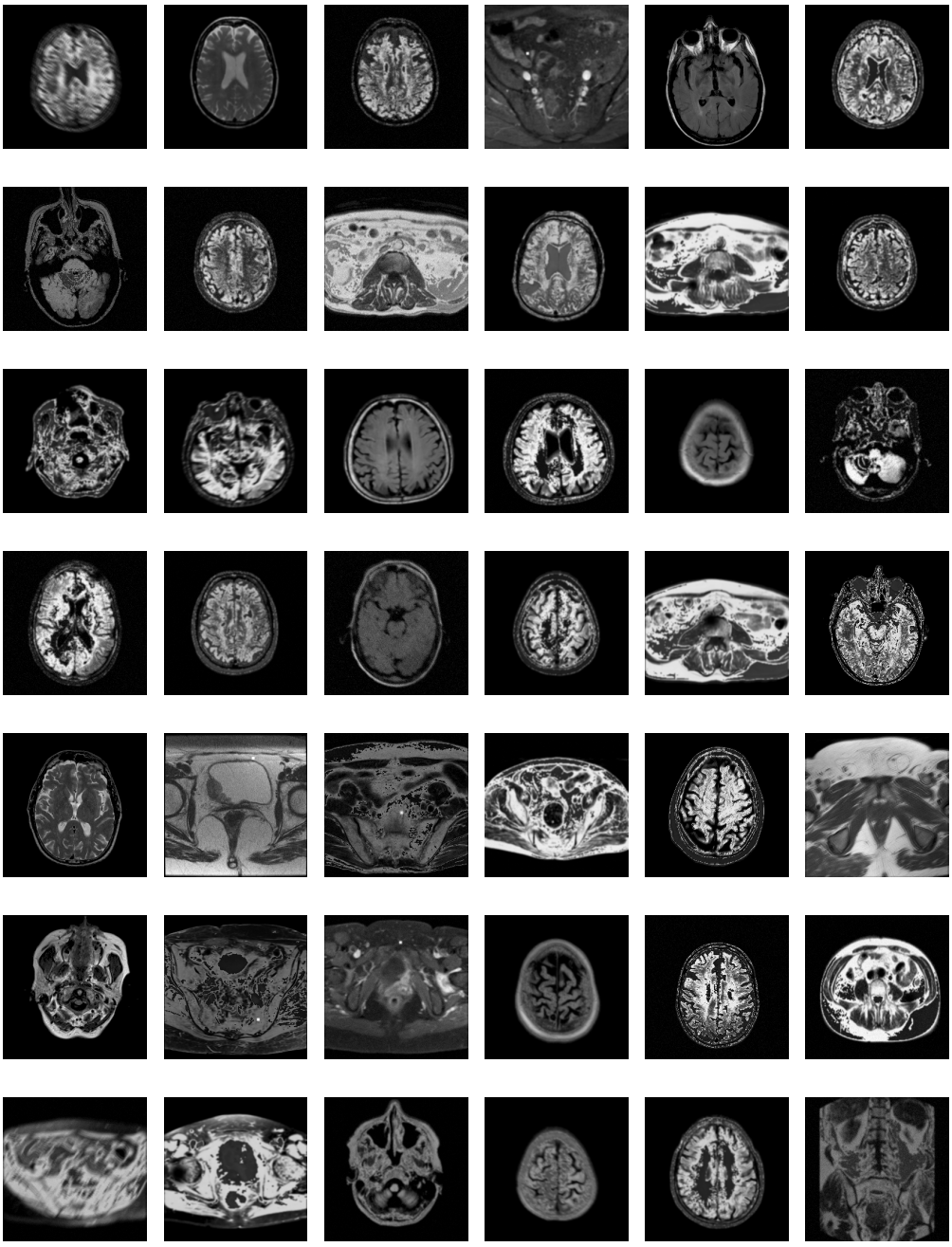


Figure 28: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy MRI images of the human brain (ADNI, 2022) and bladder (Kirk et al., 2016) are transformed to target modality PET.

MRI to CT

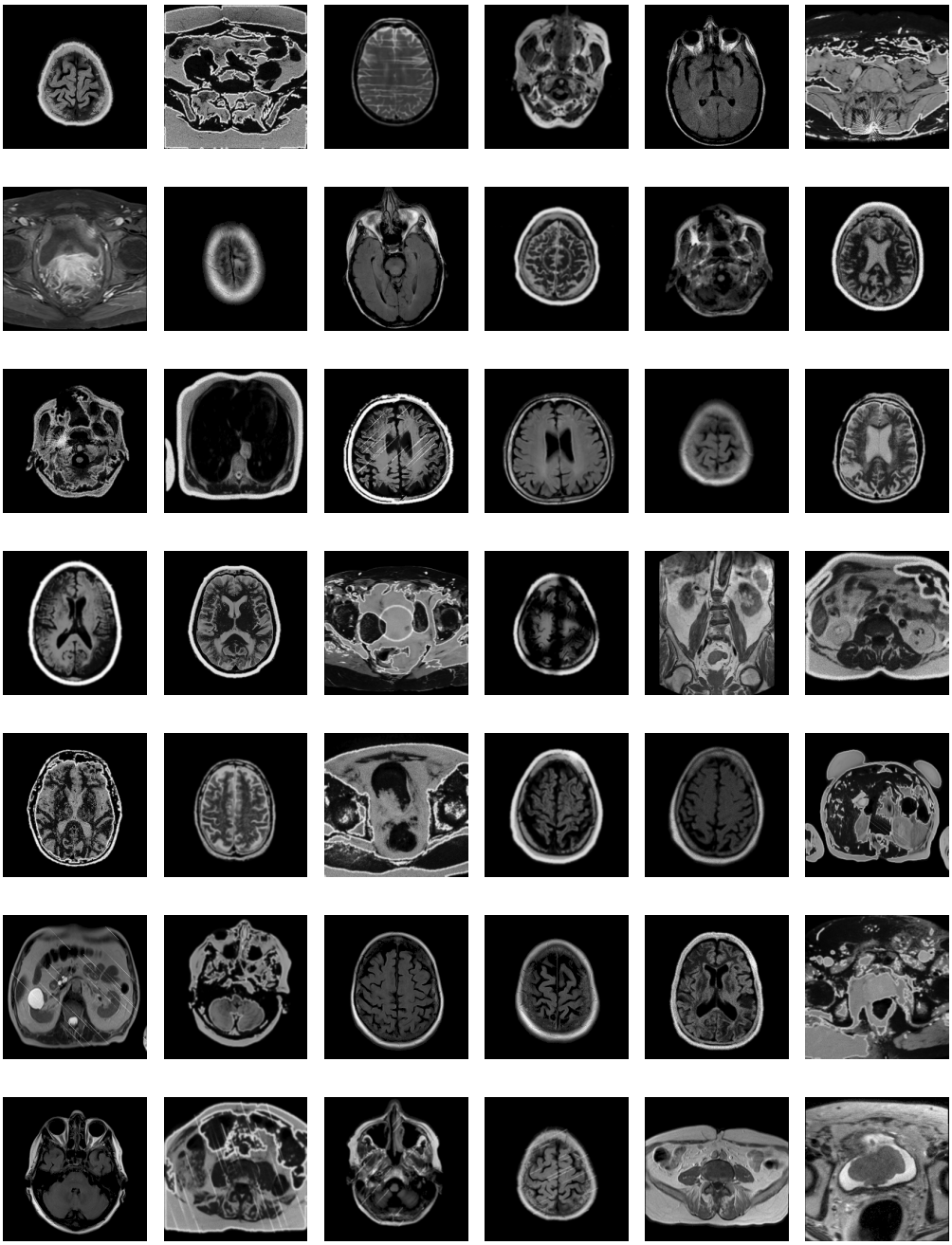


Figure 29: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy MRI images of the human brain (ADNI, 2022) and bladder (Kirk et al., 2016) are transformed to target modality CT.

CT to PET

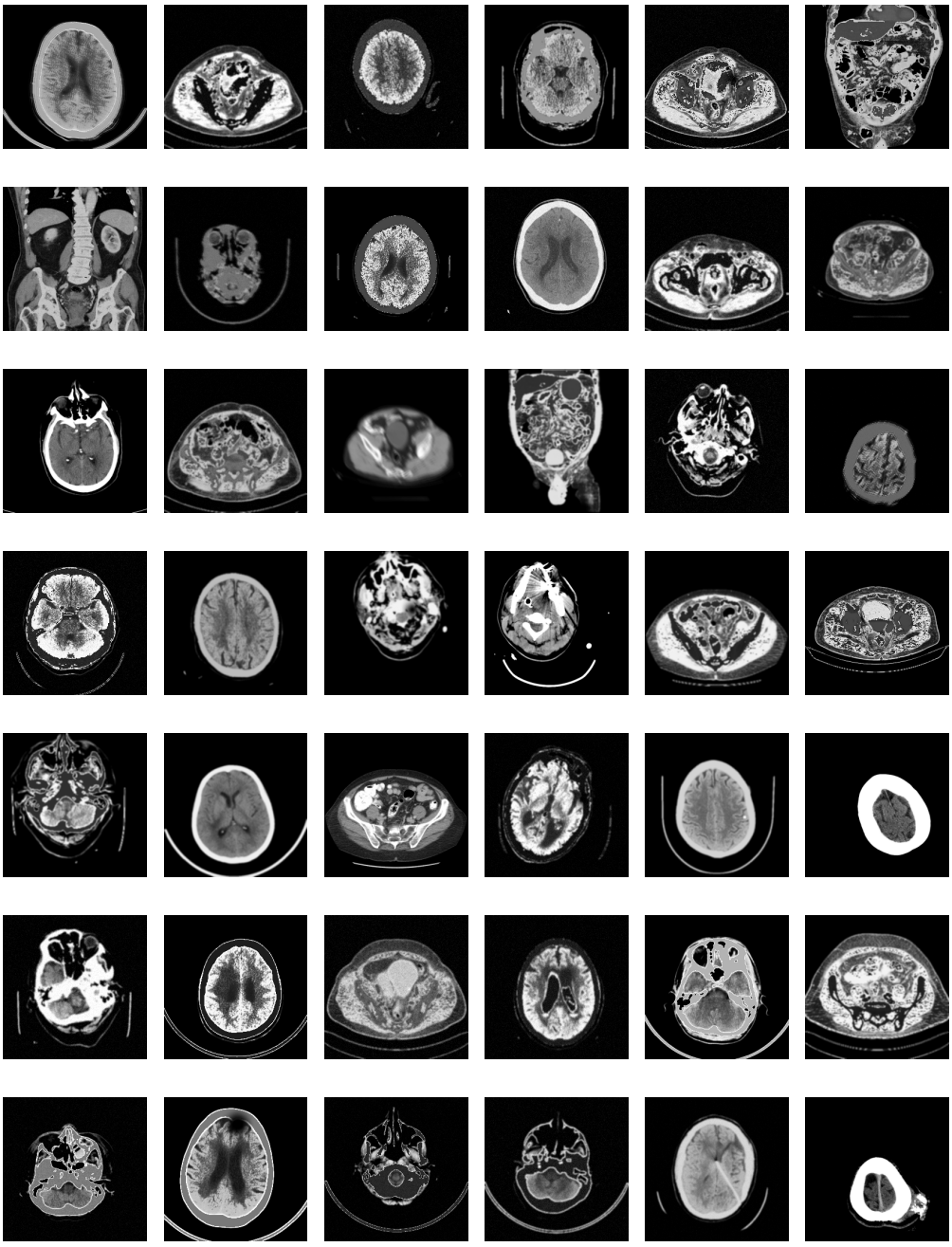


Figure 30: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy CT images of the human brain (A. Stein, 2019) and bladder (Kirk et al., 2016) are transformed to target modality PET.

CT to MRI

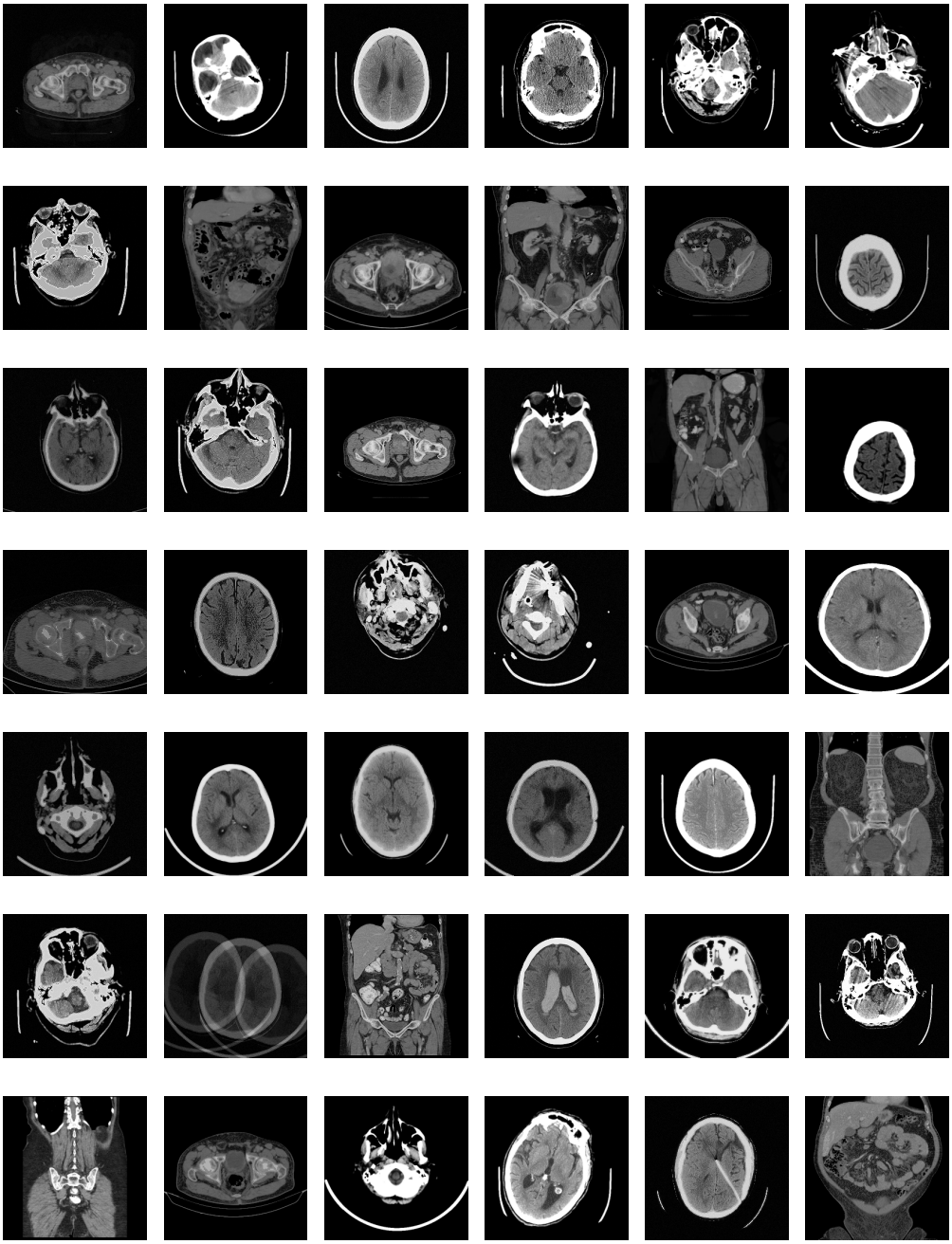


Figure 31: Randomly *CMDA*-augmented samples with zero to four possible applied augmentations, where each augmentation is equally likely to appear with an intensity between 0% and 100%. Here, healthy and unhealthy CT images of the human brain (A. Stein, 2019) and bladder (Kirk et al., 2016) are transformed to target modality MRI.

A.2 Further Qualitative Evaluation Metrics

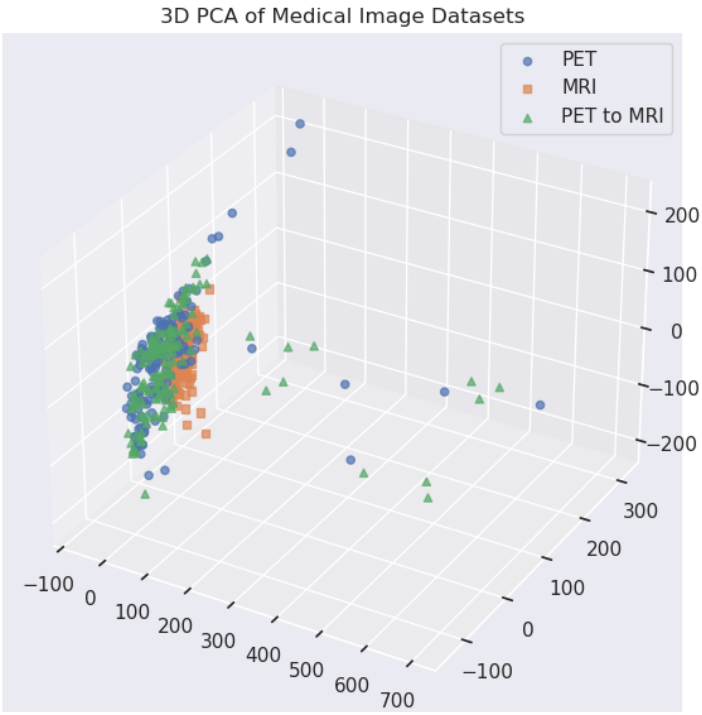
A.2.1 Brain Dataset

Modalities	Distance	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
PET to MRI	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.86	4.85	4.99	4.95	4.70	4.46	3.86	3.38	3.22	3.33	3.62
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	0.42	0.79	2.73	2.67	2.82	2.92	3.46	4.06	4.80	4.93	5.06
PET to CT	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.79	4.48	4.48	4.37	4.15	3.83	3.44	3.18	3.09	3.11	3.06
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	0.18	0.58	0.63	0.92	1.37	1.94	2.68	3.31	3.84	3.97	4.06
MRI to PET	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.96	2.98	2.98	2.95	3.37	3.74	4.23	4.52	4.73	4.88	4.85
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	1.00	4.05	4.29	4.32	3.99	3.72	3.44	3.39	3.50	3.69	3.76
MRI to CT	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.61	4.64	4.64	4.63	4.53	4.54	4.30	4.06	4.04	4.11	4.11
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	0.46	2.22	2.05	1.65	1.43	1.95	1.89	2.37	2.64	2.82	2.90
CT to PET	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.36	2.71	2.55	2.63	2.75	2.82	3.16	3.37	3.48	3.83	4.01
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	0.81	3.52	4.15	4.21	4.21	4.18	4.01	3.90	3.82	3.68	3.63
CT to MRI	$\|\|V_{D_O^A} - V_{D_T}\|\|$	4.25	3.84	3.71	3.54	3.74	3.32	3.30	3.10	2.91	2.64	2.99
	$\|\|V_{D_O^A} - V_{D_O^O}\|\|$	1.27	1.85	3.56	3.71	3.66	3.93	4.03	4.13	4.33	4.59	4.63

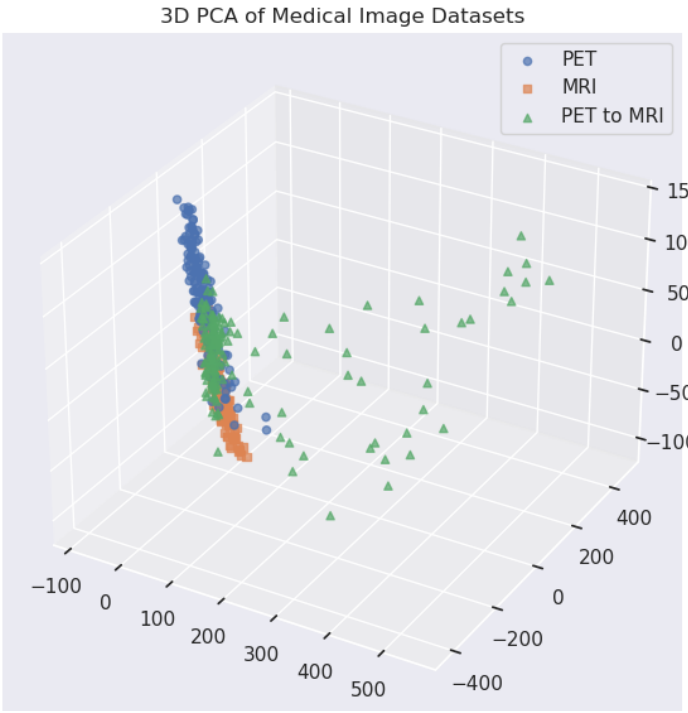
Table 13: Tabular overview of the exact Euclidean distances between GLCM features of augmented, original, and target modality datasets displayed in Figure 21. It shows that the *CMDA*-augmented dataset aligns with the target modality by an average of 22%, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

Modalities	Distance	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
PET to MRI	$d^2(f_{D_O^A}, f_{D_T})$	379.6	351.6	347.2	370.5	355.9	334.9	334.0	343.1	358.1	370.4	377.1
	$d^2(f_{D_O^A}, f_{D_O^O})$	59.2	116.1	247.5	317.7	303.0	295.1	300.8	306.8	325.2	348.1	358.1
PET to CT	$d^2(f_{D_O^A}, f_{D_T})$	422.4	367.5	328.5	321.7	340.5	362.7	350.7	357.6	369.1	352.1	349.3
	$d^2(f_{D_O^A}, f_{D_O^O})$	59.2	354.1	308.6	285.3	314.2	344.1	358.9	382.1	392.6	393.1	403.2
MRI to PET	$d^2(f_{D_O^A}, f_{D_T})$	383.3	304.3	264.7	268.1	260.1	242.1	222.8	208.8	207.7	208.4	215.0
	$d^2(f_{D_O^A}, f_{D_O^O})$	76.0	120.7	206.1	242.0	244.6	244.4	247.8	248.2	247.0	252.4	254.1
MRI to CT	$d^2(f_{D_O^A}, f_{D_T})$	366.5	354.4	352.8	350.3	364.0	361.6	362.9	358.7	354.8	355.5	351.8
	$d^2(f_{D_O^A}, f_{D_O^O})$	76.0	109.6	116.2	124.3	138.2	167.2	178.3	185.8	191.8	202.0	210.0
CT to PET	$d^2(f_{D_O^A}, f_{D_T})$	401.1	337.7	307.2	306.0	303.4	295.7	289.1	268.0	252.3	230.2	231.5
	$d^2(f_{D_O^A}, f_{D_O^O})$	162.7	192.0	226.7	244.0	259.4	269.6	271.1	276.6	285.8	291.7	297.0
CT to MRI	$d^2(f_{D_O^A}, f_{D_T})$	334.6	351.2	363.8	378.8	373.7	371.8	364.0	360.5	353.0	349.0	334.4
	$d^2(f_{D_O^A}, f_{D_O^O})$	162.7	169.6	176.9	194.5	199.9	213.0	221.0	241.6	250.1	274.9	280.5

Table 14: Tabular overview of the exact FIDs between augmented, original, and target modality datasets displayed in Figure 22. It shows that the *CMDA*-augmented dataset aligns with the target modality by an average of 25%, successfully demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

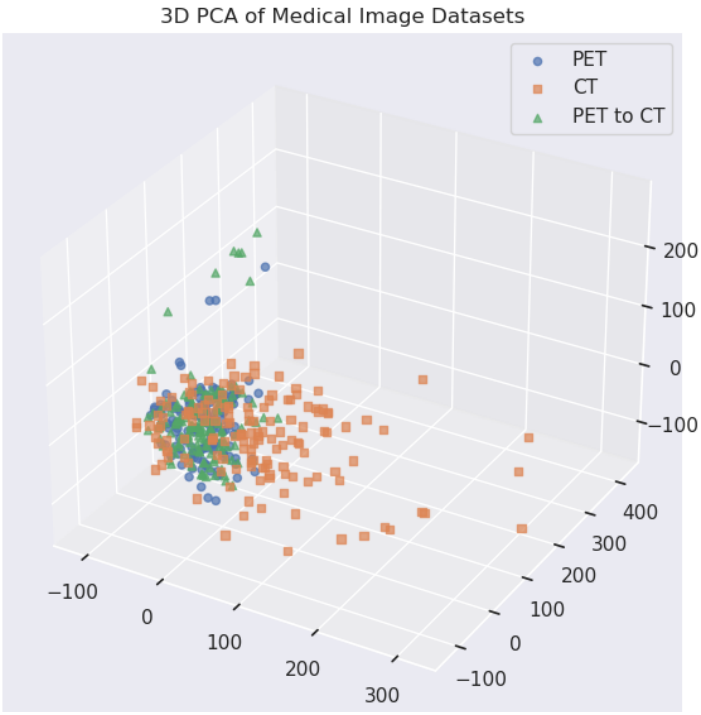


PET to MRI 0% augmented

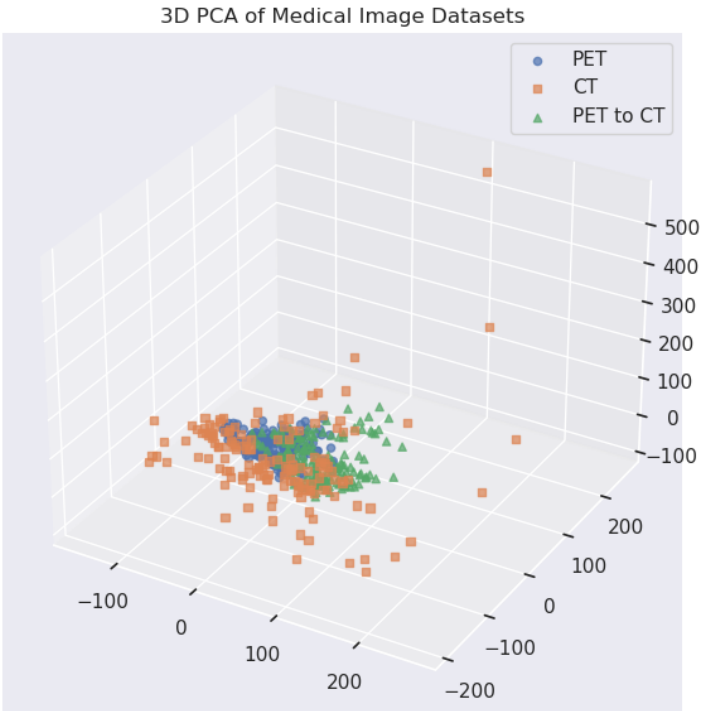


PET to MRI 100% augmented

Figure 32: Comparison of 3D-PCAs for PET dataset augmented to MRI. The top image shows PCA for PET, MRI, and PET images to be augmented. The bottom image shows PCA for the same datasets, but the PET dataset is augmented by *CMDA* with target modality MRI at 100% intensity. The augmented data points in the bottom PCA align better with the target modality, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

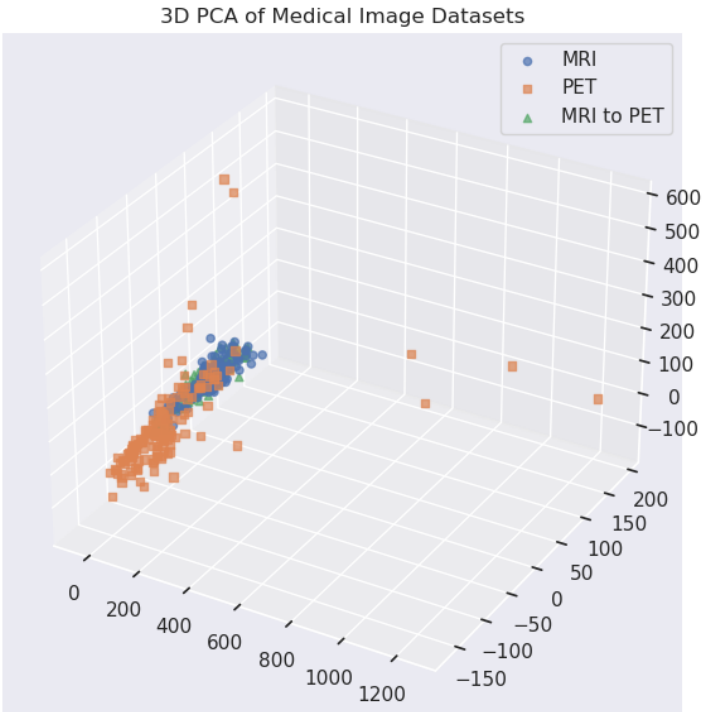


PET to CT 0% augmented

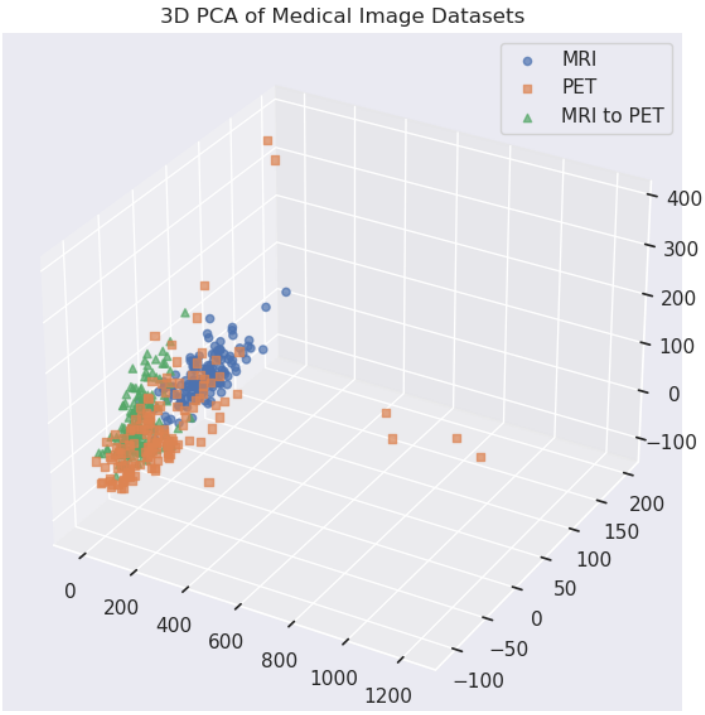


PET to CT 100% augmented

Figure 33: Comparison of 3D-PCAs for PET dataset augmented to CT. The top image shows PCA for PET, CT, and PET images to be augmented. The bottom image shows PCA for the same datasets, but the PET dataset is augmented by *CMDA* with target modality CT at 100% intensity. The augmented data points in the bottom PCA align better with the target modality, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

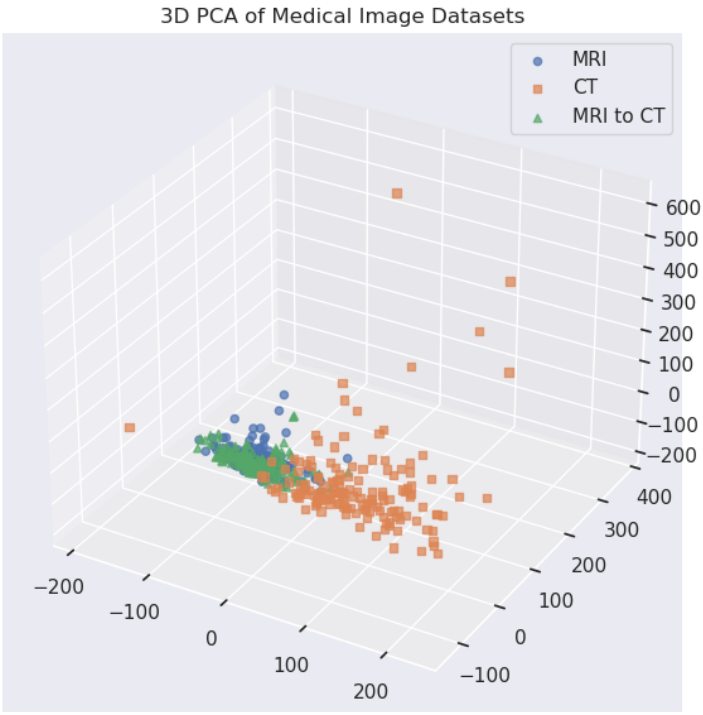


MRI to PET 0% augmented

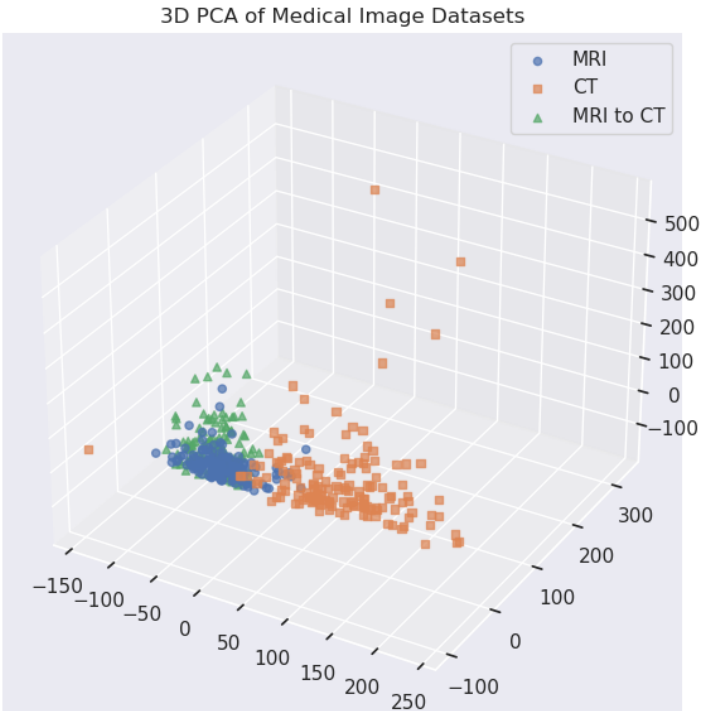


MRI to PET 100% augmented

Figure 34: Comparison of 3D-PCAs for MRI dataset augmented to PET. The top image shows PCA for MRI, PET, and MRI images to be augmented. The bottom image shows PCA for the same datasets, but the MRI dataset is augmented by *CMDA* with target modality PET at 100% intensity. The augmented data points in the bottom PCA align better with the target modality, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

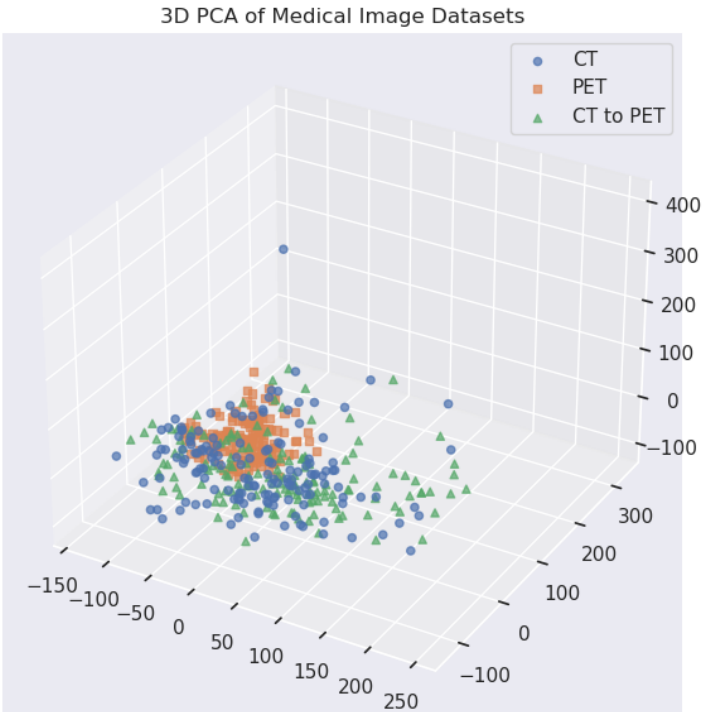


MRI to CT 0% augmented

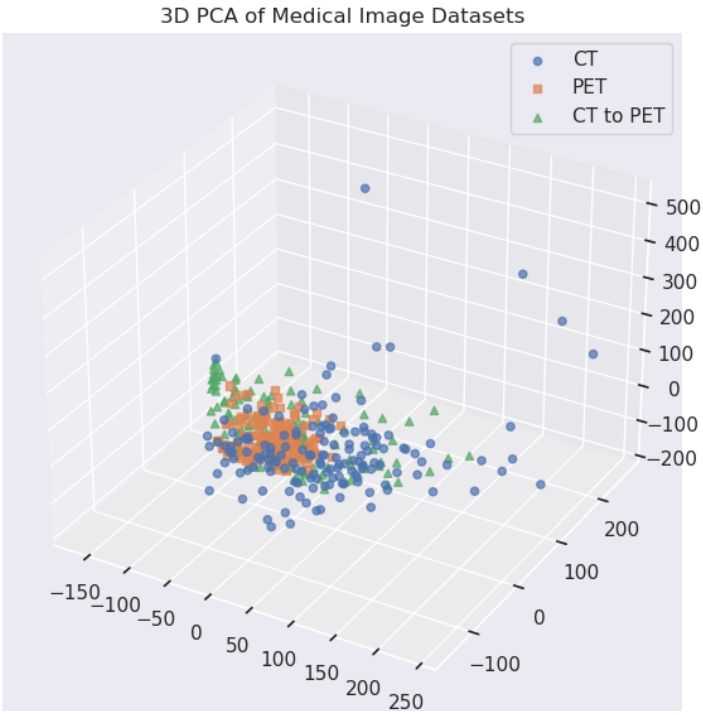


MRI to CT 100% augmented

Figure 35: Comparison of 3D-PCAs for MRI dataset augmented to CT. The top image shows PCA for MRI, CT, and MRI images to be augmented. The bottom image shows PCA for the same datasets, but the MRI dataset is augmented by *CMDA* with target modality CT at 100% intensity. The augmented data points in the bottom PCA do not align better with the target modality, not supporting *CMDA*'s statements of aligning images with the distribution of the target modality.



CT to PET 0% augmented



CT to PET 100% augmented

Figure 36: Comparison of 3D-PCAs for CT dataset augmented to PET. The top image shows PCA for CT, PET, and CT images to be augmented. The bottom image shows PCA for the same datasets, but the CT dataset is augmented by *CMDA* with target modality PET at 100% intensity. The augmented data points in the bottom PCA align better with the target modality, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

Metric	Model	no	imgaug	albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.572	0.339	0.398	0.426	0.585	0.397
	ViT-B/16	0.695	0.241	0.692	0.315	0.848	0.736
AUPR-IN	ResNet-18	0.594	0.451	0.541	0.583	0.588	0.505
	ViT-B/16	0.751	0.401	0.725	0.424	0.890	0.776
AUPR-OUT	ResNet-18	0.551	0.382	0.429	0.437	0.629	0.379
	ViT-B/16	0.695	0.347	0.679	0.390	0.807	0.624
FPR95TPR	ResNet-18	0.825	0.959	0.918	0.939	0.698	0.991
	ViT-B/16	0.652	0.948	0.705	0.921	0.641	0.934

Table 15: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for PET to MRI. The best values for each row are highlighted by color. Although imgaug performs best, *CMDA* is often close to the winning result, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.424	0.268	0.275	0.207	0.588	0.229
	ViT-B/16	0.702	0.418	0.535	0.412	0.910	0.824
AUPR-IN	ResNet-18	0.575	0.486	0.522	0.514	0.654	0.459
	ViT-B/16	0.737	0.506	0.603	0.494	0.941	0.883
AUPR-OUT	ResNet-18	0.425	0.341	0.322	0.284	0.624	0.279
	ViT-B/16	0.700	0.409	0.553	0.481	0.864	0.680
FPR95TPR	ResNet-18	0.873	0.943	0.988	1.000	0.650	0.796
	ViT-B/16	0.621	0.909	0.750	0.820	0.502	0.797

Table 16: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for PET to CT. The best values for each row are highlighted by color. Although v2 performs best, *CMDA* is often close to the winning result, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.615	0.716	0.639	0.636	0.461	0.621
	ViT-B/16	0.516	0.588	0.658	0.648	0.547	0.447
AUPR-IN	ResNet-18	0.528	0.684	0.518	0.536	0.450	0.575
	ViT-B/16	0.418	0.476	0.682	0.629	0.442	0.384
AUPR-OUT	ResNet-18	0.700	0.779	0.720	0.725	0.551	0.682
	ViT-B/16	0.608	0.646	0.618	0.683	0.659	0.604
FPR95TPR	ResNet-18	0.790	0.636	0.717	0.767	0.893	0.793
	ViT-B/16	0.919	0.887	0.955	0.858	0.816	0.839

Table 17: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for MRI to PET. The best values for each row are highlighted by color. Although RandAugment performs best, *CMDA* is often close to the winning result, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.476	0.539	0.621	0.698	0.441	0.422
	ViT-B/16	0.648	0.546	0.634	0.326	0.693	0.537
AUPR-IN	ResNet-18	0.543	0.601	0.702	0.780	0.533	0.480
	ViT-B/16	0.673	0.595	0.697	0.419	0.681	0.567
AUPR-OUT	ResNet-18	0.473	0.501	0.547	0.690	0.412	0.404
	ViT-B/16	0.604	0.517	0.579	0.381	0.672	0.484
FPR95TPR	ResNet-18	0.916	0.938	0.912	0.696	0.984	0.981
	ViT-B/16	0.859	0.888	0.862	0.949	0.790	0.942

Table 18: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for MRI to CT. The best values for each row are highlighted by color. Although v2 performs best, *CMDA* is often close to the winning result, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality.

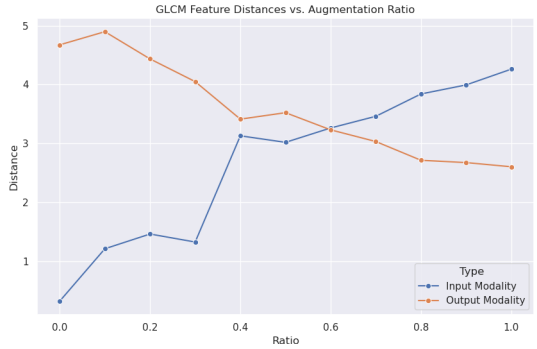
Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.311	0.450	0.410	0.412	0.287	0.534
	ViT-B/16	0.448	0.496	0.676	0.530	0.677	0.667
AUPR-IN	ResNet-18	0.314	0.424	0.384	0.381	0.348	0.456
	ViT-B/16	0.349	0.375	0.543	0.388	0.579	0.557
AUPR-OUT	ResNet-18	0.510	0.614	0.542	0.584	0.505	0.623
	ViT-B/16	0.638	0.663	0.775	0.680	0.747	0.759
FPR95TPR	ResNet-18	0.934	0.796	0.952	0.897	0.934	0.935
	ViT-B/16	0.867	0.819	0.740	0.821	0.826	0.766

Table 19: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for CT to PET. The best values for each row are highlighted by color. Overall, the non-augmented case performs the best. These results do not support *CMDA*'s statements of aligning images with the distribution of the target modality.

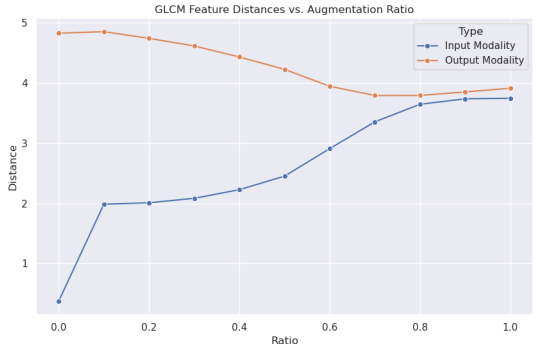
Metric	Model	None	imgaug	Albumentations	v2	RandAugment	*CMDA*
AUROC	ResNet-18	0.465	0.524	0.434	0.501	0.368	0.523
	ViT-B/16	0.593	0.508	0.649	0.412	0.637	0.664
AUPR-IN	ResNet-18	0.418	0.533	0.427	0.491	0.376	0.470
	ViT-B/16	0.562	0.472	0.653	0.382	0.571	0.691
AUPR-OUT	ResNet-18	0.572	0.630	0.511	0.586	0.520	0.567
	ViT-B/16	0.643	0.588	0.680	0.543	0.701	0.658
FPR95TPR	ResNet-18	0.883	0.791	0.944	0.889	0.887	0.943
	ViT-B/16	0.865	0.883	0.845	0.899	0.795	0.908

Table 20: OOD-detection metrics for ResNet and Vision Transformer models with different data augmentations for CT to MRI. The best values for each row are highlighted by color. Overall, v2 performs the best. These results do not support *CMDA*'s statements of aligning images with the distribution of the target modality.

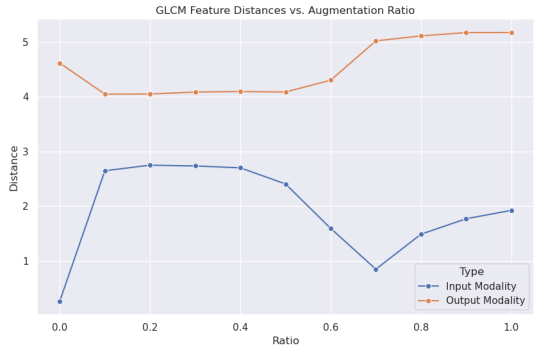
A.2.2 Bladder Dataset



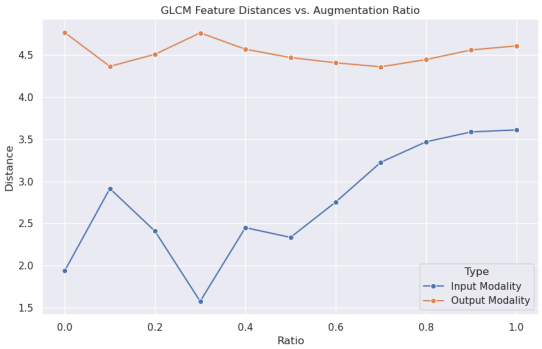
PET to MRI



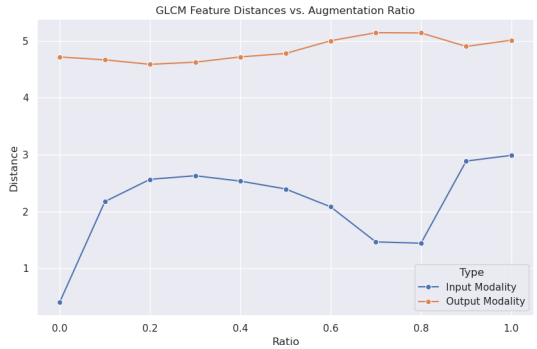
PET to CT



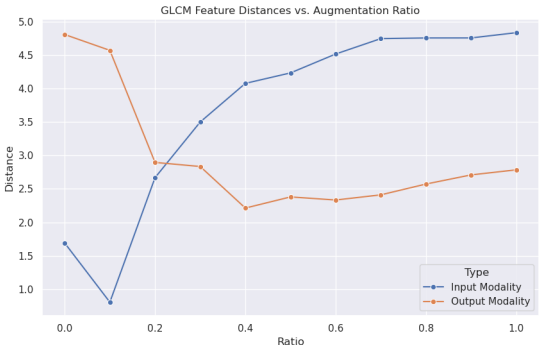
MRI to PET



MRI to CT

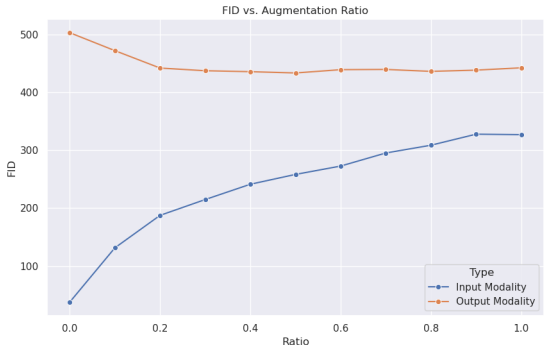


CT to PET

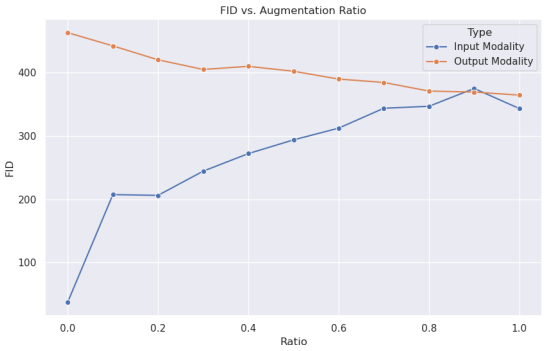


CT to MRI

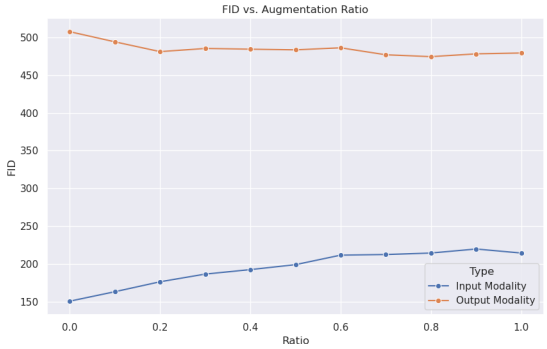
Figure 37: This experiment was conducted with the TCGA-BLCA bladder dataset. Euclidean distances between GLCM features of augmented, original, and target modality datasets. It shows that the *CMDA*-augmented dataset aligns with the target modality by an average of 14%, reasonably successfully demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality, even across other anatomies than the brain.



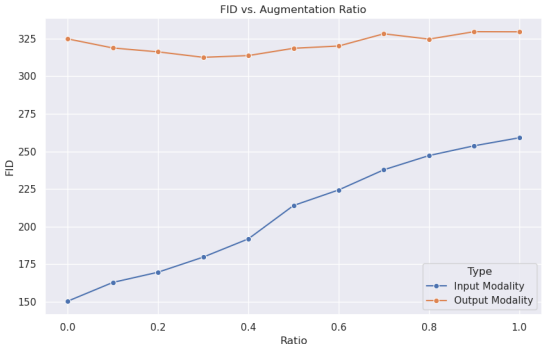
PET to MRI



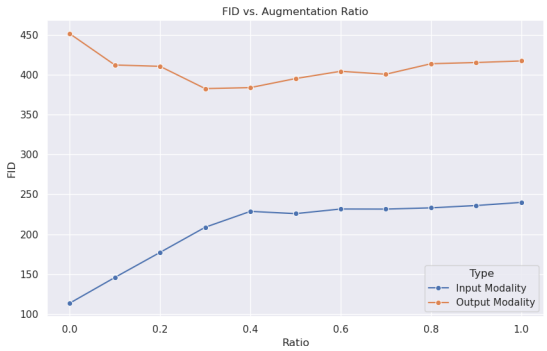
PET to CT



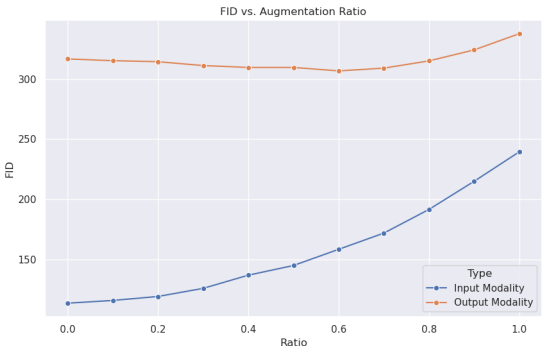
MRI to PET



MRI to CT

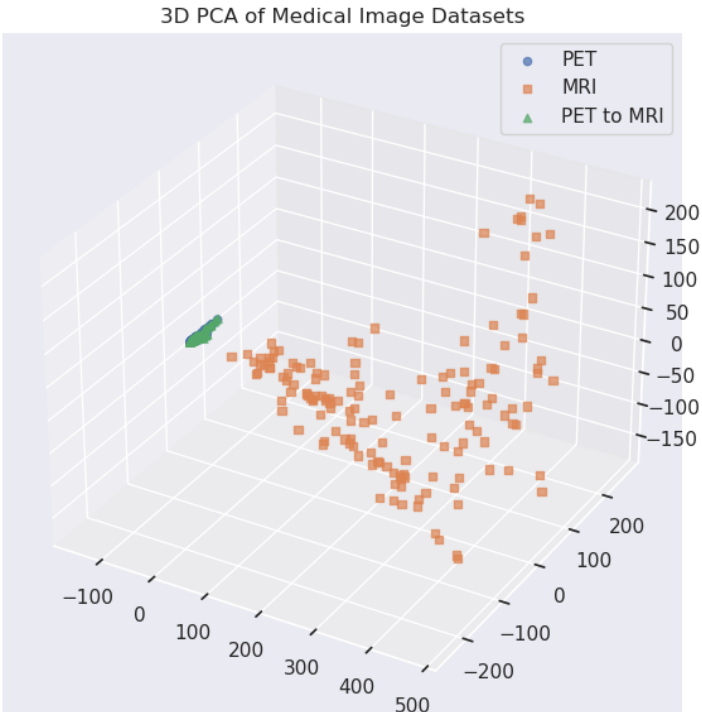


CT to PET

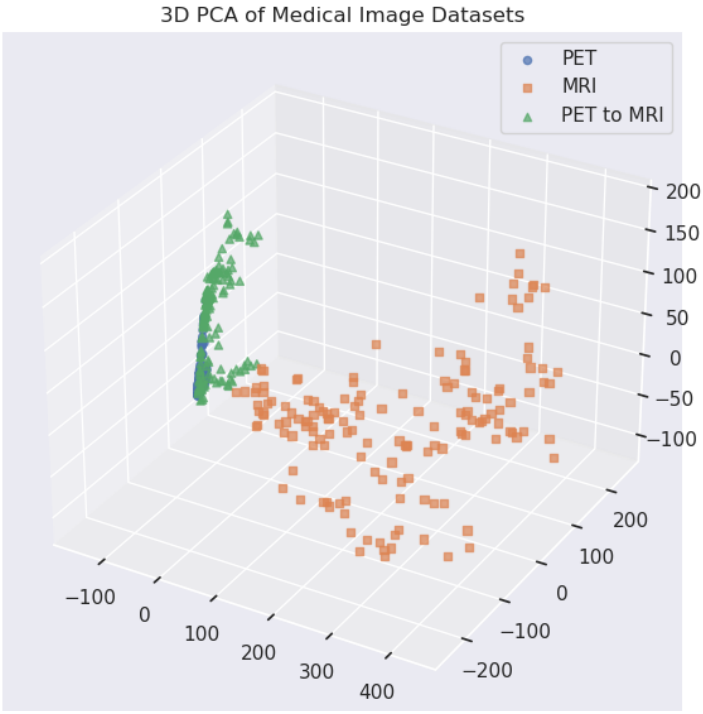


CT to MRI

Figure 38: This experiment was conducted with the TCGA-BLCA bladder dataset. FIDs of augmented, original, and target modality datasets. It shows that the *CMDA*-augmented dataset aligns with the target modality by an average of 7%, reasonably successfully demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality, even across other anatomies than the brain.

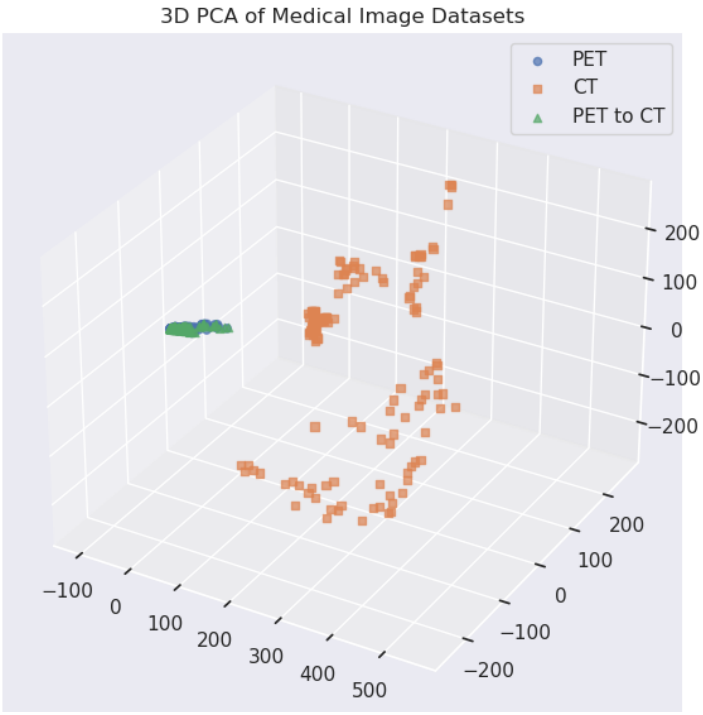


PET to MRI 0% augmented

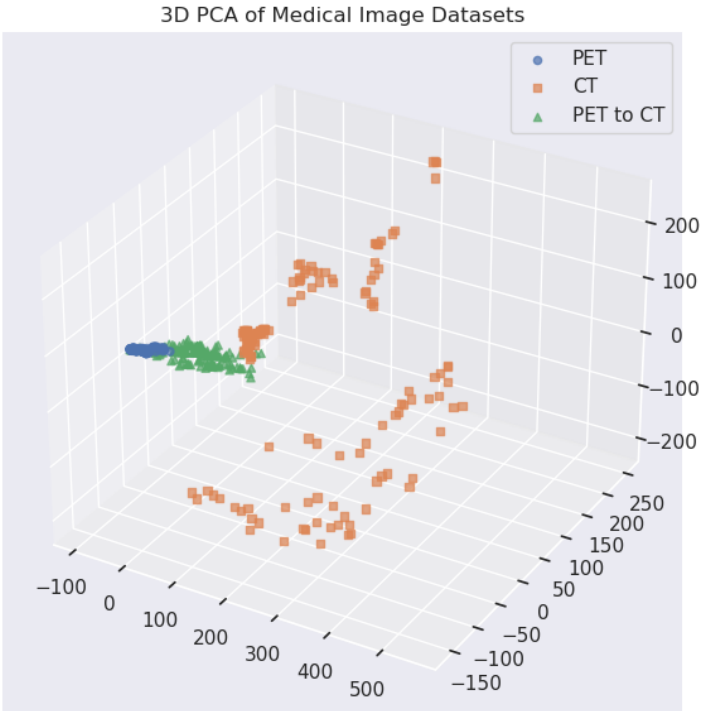


PET to MRI 100% augmented

Figure 39: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for PET dataset augmented to MRI. The top image shows PCA for PET, MRI, and PET images to be augmented. The bottom image shows PCA for the same datasets, but the PET dataset is augmented by *CMDA* with target modality MRI at 100% intensity. The augmented data points in the bottom PCA do not align better with the target modality, not supporting *CMDA*'s statements of aligning images with the distribution of the target modality, at least not across other anatomies than the brain.

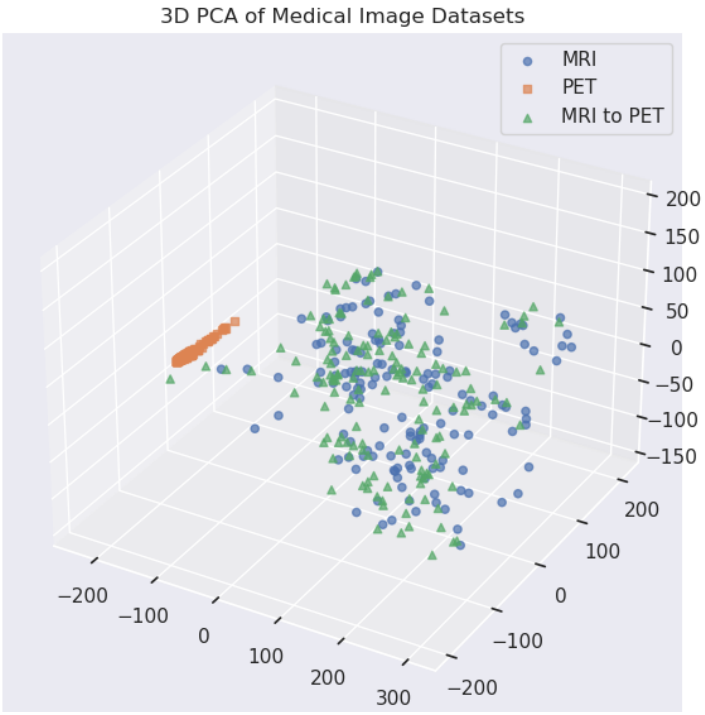


PET to CT 0% augmented

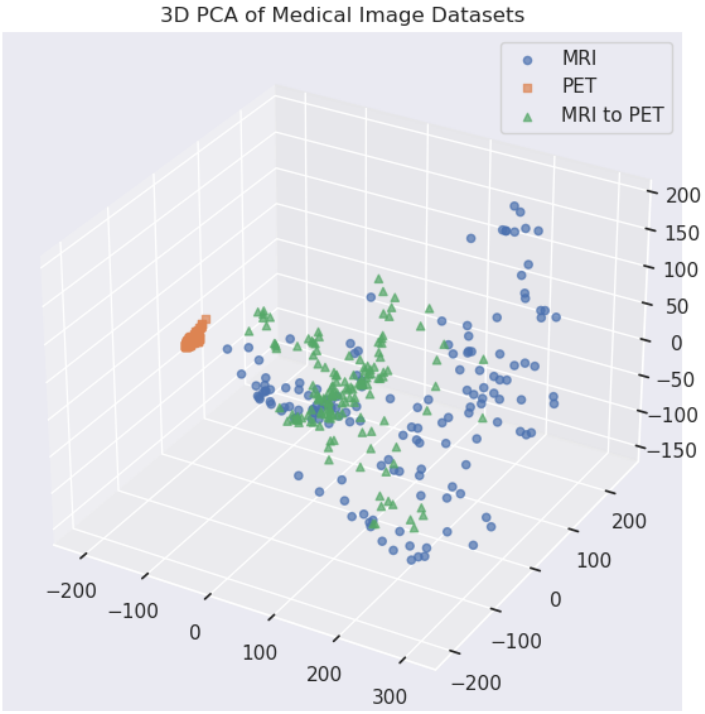


PET to CT 100% augmented

Figure 40: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for PET dataset augmented to CT. The top image shows PCA for PET, CT, and PET images to be augmented. The bottom image shows PCA for the same datasets, but the PET dataset is augmented by *CMDA* with target modality CT at 100% intensity. The augmented data points in the bottom PCA do partially align better with the target modality, reasonably successfully demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality, even across other anatomies than the brain.

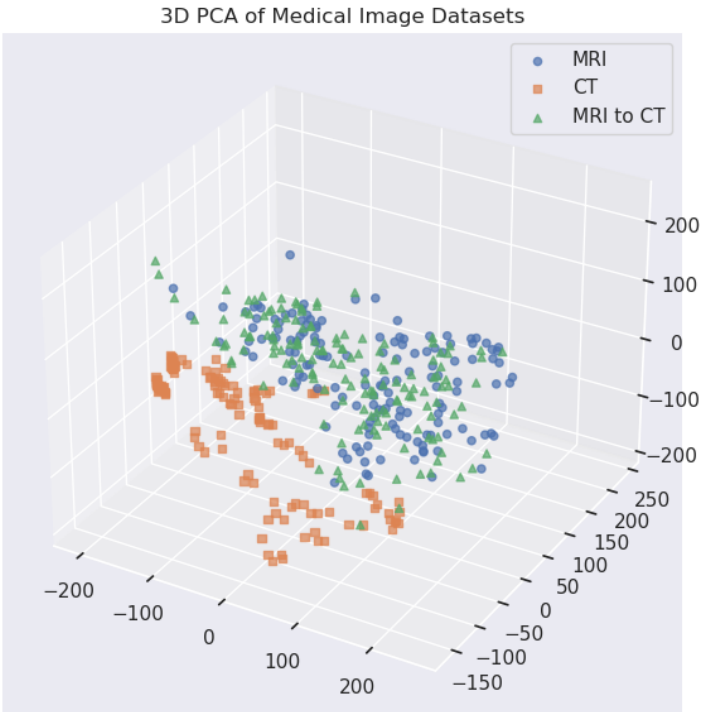


MRI to PET 0% augmented

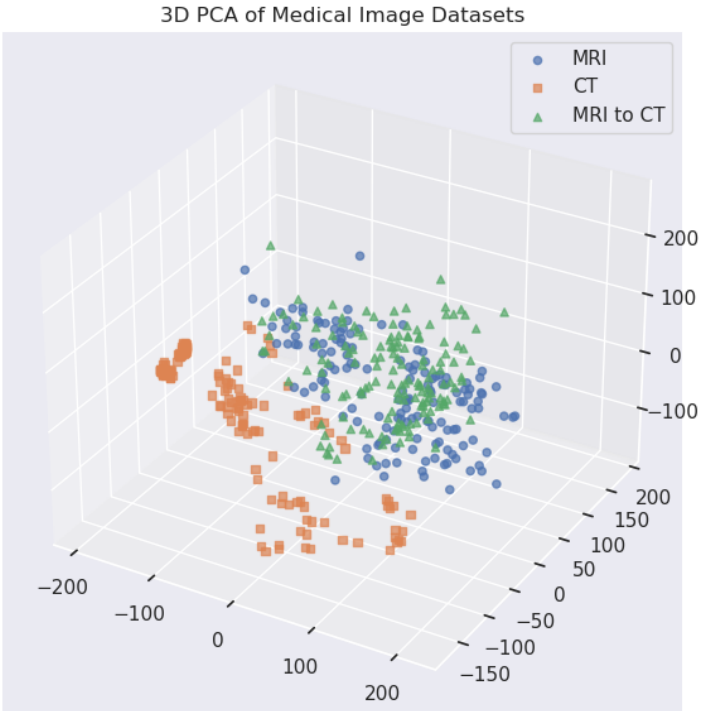


MRI to PET 100% augmented

Figure 41: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for MRI dataset augmented to PET. The top image shows PCA for MRI, PET, and MRI images to be augmented. The bottom image shows PCA for the same datasets, but the MRI dataset is augmented by *CMDA* with target modality PET at 100% intensity. The augmented data points in the bottom PCA align minimally better with the target modality, however not enough to support *CMDA*'s statements of aligning images with the distribution of the target modality, at least not across other anatomies than the brain.

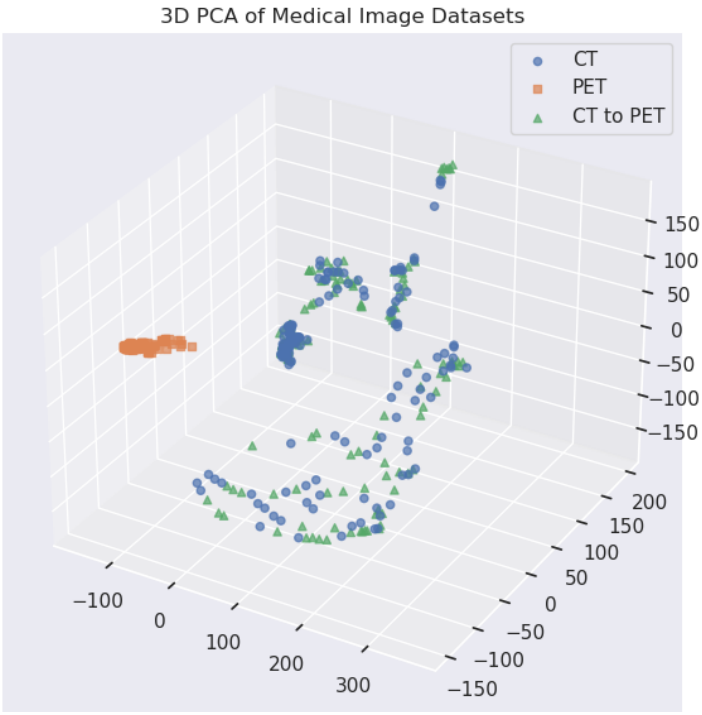


MRI to CT 0% augmented

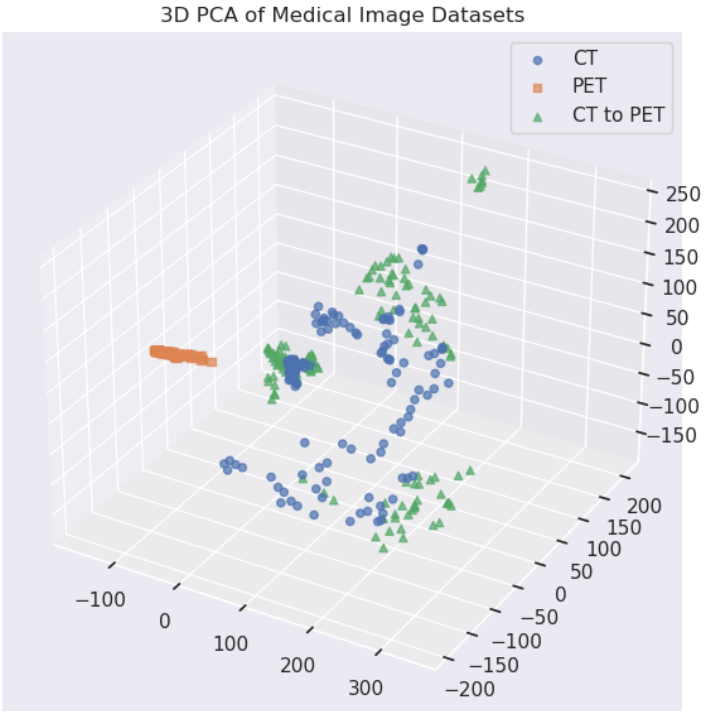


MRI to CT 100% augmented

Figure 42: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for MRI dataset augmented to CT. The top image shows PCA for MRI, CT, and MRI images to be augmented. The bottom image shows PCA for the same datasets, but the MRI dataset is augmented by *CMDA* with target modality CT at 100% intensity. The augmented data points in the bottom PCA do not align better with the target modality, not supporting *CMDA*'s statements of aligning images with the distribution of the target modality, at least not across other anatomies than the brain.

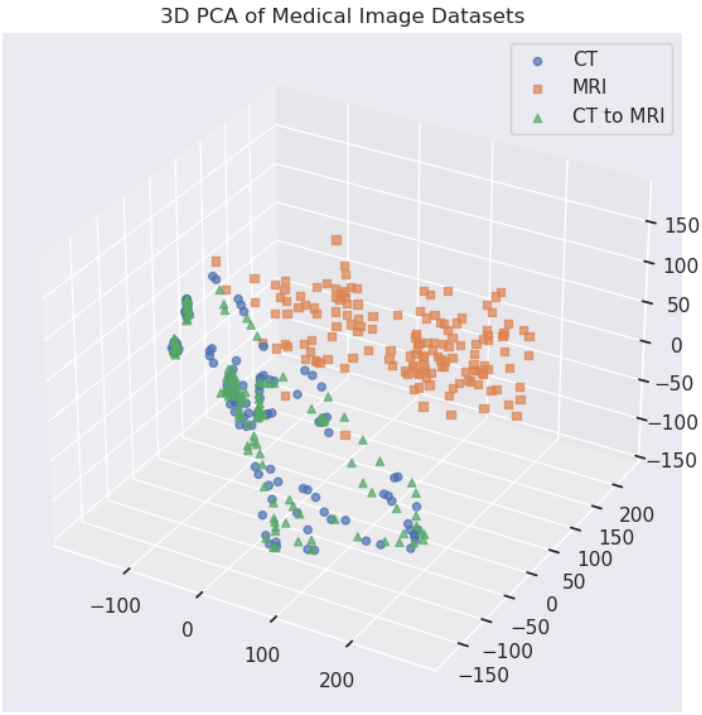


CT to PET 0% augmented

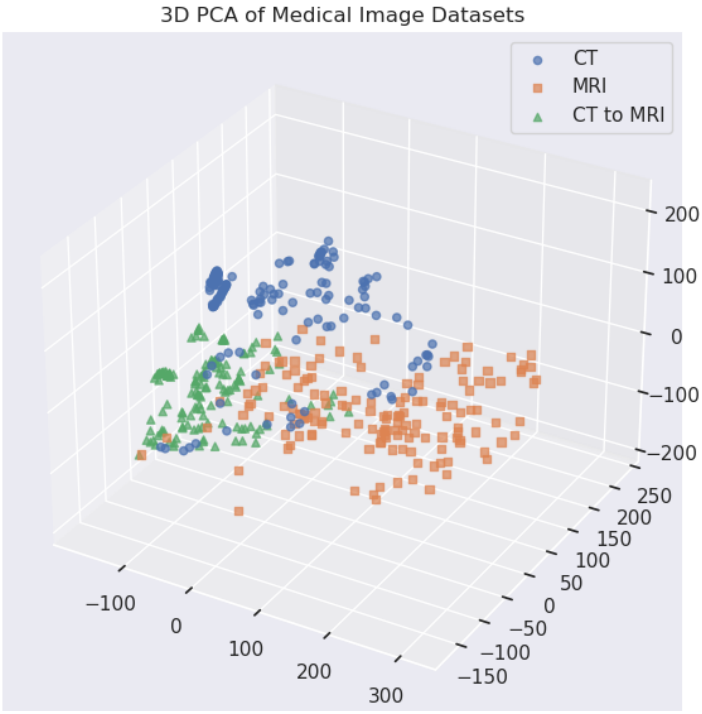


CT to PET 100% augmented

Figure 43: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for CT dataset augmented to PET. The top image shows PCA for CT, PET, and CT images to be augmented. The bottom image shows PCA for the same datasets, but the CT dataset is augmented by *CMDA* with target modality PET at 100% intensity. The augmented data points in the bottom PCA align minimally better with the target modality, however not enough to support *CMDA*'s statements of aligning images with the distribution of the target modality, at least not across other anatomies than the brain.



CT to MRI 0% augmented



CT to MRI 100% augmented

Figure 44: This experiment was conducted with the TCGA-BLCA bladder dataset. Comparison of 3D-PCAs for CT dataset augmented to MRI. The top image shows PCA for CT, MRI, and CT images to be augmented. The bottom image shows PCA for the same datasets, but the CT dataset is augmented by *CMDA* with target modality MRI at 100% intensity. The augmented data points in the bottom PCA do partially align better with the target modality, reasonably successfully demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality, even across other anatomies than the brain.

Modalities	$\boldsymbol{VAE}_O$			$\boldsymbol{VAE}_A$		
	Test Loss	MAE	RMSE	Test Loss	MAE	RMSE
PET to MRI	4847216	0.2770	0.3525	3683450	0.2816	0.3557
PET to CT	3834369	0.2124	0.3414	9156919	0.2152	0.3454
MRI to PET	801581	0.3057	0.3266	998985	0.3734	0.3904
MRI to CT	904223	0.1632	0.2386	1067878	0.1651	0.2701
CT to PET	761608	0.2882	0.3252	771650	0.2923	0.3271
CT to MRI	1395642	0.2134	0.2712	1273842	0.1953	0.2449

Table 21: This experiment was conducted with the TCGA-BLCA bladder dataset. Test Losses, MAEs, and RMSEs for the VAE experiment. The results across both VAEs are quite similar, balancing out values that differ. Therefore, they do not support *CMDA*'s statements of aligning images with the distribution of the target modality, at least not across other anatomies than the brain.

Metric	**Model**	**None**	**imgaug**	**Albumentations**	**v2**	**RandAugment**	***CMDA***
AUROC	ResNet-18	0.249	0.271	0.277	0.340	0.450	0.210
	ViT-B/16	0.979	0.966	0.977	0.961	0.969	0.729
AUPR-IN	ResNet-18	0.440	0.430	0.479	0.504	0.522	0.393
	ViT-B/16	0.982	0.966	0.977	0.960	0.961	0.729
AUPR-OUT	ResNet-18	0.504	0.485	0.509	0.503	0.571	0.421
	ViT-B/16	0.979	0.967	0.978	0.963	0.969	0.754
FPR95TPR	ResNet-18	0.876	0.866	0.824	0.785	0.757	0.947
	ViT-B/16	0.116	0.076	0.093	0.118	0.114	0.545

Table 22: This experiment was conducted with the TCGA-BLCA bladder dataset. OOD-detection metrics with multiple data augmentations. The calculated numbers represent the MAEns over all possible combinations of original and target modality. The best values for each row are highlighted by color. Here, *CMDA* noticeably performs best across all metrics, demonstrating *CMDA*'s effectiveness in aligning images with the distribution of the target modality, even across other anatomies than the brain.

A.3 Code Availability

CMDA's source code, conducted experiments, more elaborate information, permitted parameter values, examples, and a detailed explanation regarding the use of custom reference images can be taken from the corresponding GitHub repository. All code is open source and can be used for data augmentation pipelines or further research.

Bibliography

Carol Wu Chris Carr George Shih Jayashree Kalpathy-Cramer Julia Elliott kalpathy Luciano Prevedello Marc Kohli MD Matt Lungren Phil Culliton Robyn Ball Safwan Halabi MD A. Stein, MD. Rsna intracranial hemorrhage detection, 2019. URL `https://kaggle.com/competitions/rsna-intracranial-hemorrhage-detection`.

Mohei M Abouzied, Elpida S Crawford, and Hani Abdel Nabi. 18F-FDG imaging: pitfalls and artifacts. *Journal of nuclear medicine technology*, 33(3):145–155, 2005.

Alzheimer's Disease Neuroimaging Initiative ADNI. Alzheimer's disease neuroimaging initiative. `https://adni.loni.usc.edu/data-samples/adni-data/`, 2022. Accessed: 2024-08-12.

Santiago Aja-Fernández and Gonzalo Vegas-Sánchez-Ferrero. Statistical analysis of noise in mri. *Switzerland: Springer International Publishing*, 2016.

Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Küstner, Tobias Hepp, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. Medgan: Medical image translation using gans. *Computerized medical imaging and graphics*, 79:101684, 2020.

Bailey, Townsend, Valk, and Maisey. *Positron Emission Tomography: Basic Sciences*, volume 18. Springer, 2005.

Ms. Aayushi Bansal, Dr. Rewa Sharma, and Dr. Mamta Kathuria. A systematic review on data scarcity problem in deep learning: Solution and applications. *ACM Comput. Surv.*, 54(10s), sep 2022. ISSN 0360-0300. doi: 10.1145/3502287. URL `https://doi.org/10.1145/3502287`.

Julia F Barrett and Nicholas Keat. Artifacts in CT: recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004.

Markus Bayer, Marc-André Kaufhold, and Christian Reuter. A survey on data augmentation for text classification. *ACM Computing Surveys*, 55(7):1–39, 2022.

F Edward Boas, Dominik Fleischmann, et al. CT artifacts: causes and reduction techniques. *Imaging Med*, 4(2):229–240, 2012.

Christopher Bowles, Liang Chen, Ricardo Guerrero, Paul Bentley, Roger Gunn, Alexander Hammers, David Alexander Dickie, Maria Valdés Hernández, Joanna Wardlaw, and Daniel Rueckert. Gan augmentation: Augmenting training data using generative adversarial networks. *arXiv preprint arXiv:1810.10863*, 2018.

Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020.

Thorsten M Buzug. Computed tomography. In *Springer handbook of medical technology*, pages 311–342. Springer, 2011.

M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.

Prashanth Chandran, Gaspard Zoss, Paulo Gotardo, Markus Gross, and Derek Bradley. Adaptive convolutions for structure-aware style transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7972–7981, 2021.

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 865–872, 2019.

Junxiao Chen, Jia Wei, and Rui Li. Targan: target-aware generative adversarial networks for multi-modality medical image translation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 24–33. Springer, 2021.

Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.

Gary JR Cook, Eva A Wegner, and Ignac Fogelman. Pitfalls and artifacts in 18FDG PET and PET/CT oncologic imaging. In *Seminars in nuclear medicine*, volume 34, pages 122–133. Elsevier, 2004.

James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

G. Csurka. *A Comprehensive Survey on Domain Adaptation for Visual Applications*. 2017. doi: 10.1007/978-3-319-58347-1_1.

Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

Oscar Day and Taghi M Khoshgoftaar. A survey on heterogeneous transfer learning. *Journal of Big Data*, 4:1–42, 2017.

Manoj Diwakar and Manoj Kumar. A review on ct image noise and its denoising. *Biomedical Signal Processing and Control*, 42:73–88, 2018.

Jackie L Dobrovolny and Stephanie Christine G Fuentes. Quantitative versus qualitative evaluation: A tool to decide which to use. *Performance Improvement*, 47 (4):7–14, 2008.

Niklas Donges. What is transfer learning? exploring the popular deep learning approach. `https://builtin.com/data-science/transfer-learning`, 2024. Accessed: 2024-08-18.

Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

A. Farahani, S. Voghoei, K. Rasheed, and H.R. Arabnia. A brief review of domain adaptation. In *Advances in Data Science and Information Engineering. Transactions on Computational Science and Computational Intelligence*, 2021. doi: 10.1007/978-3-030-71704-9_65.

Lehrstuhl Für Mustererkennung FAU. CT Image De-Noising. `https://www5.cs.fau.de/en/our-team/balda-michael/projects/ct-image-de-noising/index.html#:~:text=The%20two%20main%20sources%20of,is%20square%2Droot%20of%20m`. Accessed: 2024-08-14.

Adam E Flanders, Luciano M Prevedello, George Shih, Safwan S Halabi, Jayashree Kalpathy-Cramer, Robyn Ball, John T Mongan, Anouk Stein, Felipe C Kitamura, Matthew P Lungren, et al. Construction of a machine learning dataset through collaboration: the rsna 2019 brain ct hemorrhage challenge. *Radiology: Artificial Intelligence*, 2(3):e190211, 2020.

S Garbarino. Quantitative and qualitative methods in impact evaluation and measuring results. *Issues Paper*, 2009.

Itai Gat, Idan Schwartz, and Alex Schwing. Perceptual score: What data modalities does your model perceive? *Advances in Neural Information Processing Systems*, 34:21630–21643, 2021.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

Mehdi Gheisari, Fereshteh Ebrahimzadeh, Mohamadtaghi Rahimi, Mahdieh Moazzamigodarzi, Yang Liu, Pijush Kanti Dutta Pramanik, Mohammad Ali Heravi, Abolfazl Mehbodniya, Mustafa Ghaderzadeh, Mohammad Reza Feylizadeh, et al. Deep learning: Applications, architectures, models, tools, and frameworks: A

comprehensive survey. *CAAI Transactions on Intelligence Technology*, 8(3):581–606, 2023.

Evgin Goceri. Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*, 56(11):12561–12605, 2023.

Lee W Goldman. Principles of ct and ct technology. *Journal of nuclear medicine technology*, 35(3):115–128, 2007.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: A survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2022. doi: 10.1109/TBME.2021.3117407.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. *arXiv preprint arXiv:2007.01434*, 2020.

Alfred Haar. On the theory of orthogonal function systems. *Mathematische Annalen*, 69(3):331–371, 1910.

Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

Ray Hashman Hashemi, William G Bradley, and Christopher J Lisanti. *MRI: the basics: The Basics*. Lippincott Williams & Wilkins, 2012.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

William R. Hendee, Gary J. Becker, James P. Borgstede, Jennifer Bosma, William J. Casarella, Beth A. Erickson, C. Douglas Maynard, James H. Thrall, and Paul E. Wallner. Addressing overutilization in medical imaging. *Radiology*, 257(1):240–245, 2010. doi: 10.1148/radiol.10100063.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.

Zeshan Hussain, Francisco Gimenez, Darvin Yi, and Daniel Rubin. Differential data augmentation techniques for medical imaging classification tasks. In *AMIA annual symposium proceedings*, volume 2017, page 979. American Medical Informatics Association, 2017.

Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019.

G.C. Kagadis, C. Kloukinas, K. Moore, J. Philbin, P. Papadimitroulas, C. Alexakos, P.G. Nagy, D. Visvikis, and W.R. Hendee. Cloud computing in medical imaging. *Med. Phys.*, 40:070901, 2013. doi: 10.1118/1.4811272.

Shizuo Kaji and Satoshi Kida. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology*, 12(3):235–248, 2019.

Hany Kasban, MAM El-Bendary, and DH Salama. A comparative study of medical imaging techniques. *International Journal of Information Science and Intelligent System*, 4(2):37–58, 2015.

Girish Katti, Syeda Arshiya Ara, and Ayesha Shireen. Magnetic resonance imaging (mri)–a review. *International journal of dental clinics*, 3(1):65–70, 2011.

Aghiles Kebaili, Jérôme Lapuyade-Lahorgue, and Su Ruan. Deep learning approaches for data augmentation in medical imaging: a review. *Journal of Imaging*, 9(4):81, 2023.

Jaehong Key and James F Leary. Nanoparticles for multimodal in vivo imaging in nanomedicine. *International journal of nanomedicine*, pages 711–726, 2014.

Ji Hye Kim, Il Jun Ahn, Woo Hyun Nam, Yongjin Chang, and Jong Beom Ra. Post-filtering of pet image based on noise characteristic and spatial sensitivity distribution. In *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*, pages 1–3. IEEE, 2013.

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Konstantin Kirchheim, Marco Filax, and Frank Ortmeier. Pytorch-ood: A library for out-of-distribution detection based on pytorch. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4351–4360, June 2022.

S. Kirk, Y. Lee, F. R. Lucchesi, N. D. Aredes, N. Gruszauskas, J. Catto, K. Garcia, R. Jarosz, V. Duddalwar, B. Varghese, K. Rieger-Christ, and J. Lemmerman. Tcga-blca, the cancer genome atlas urothelial bladder carcinoma collection. `https://doi.org/10.7937/K9/TCIA.2016.8LNG8XDR`, 2016. Accessed: 2024-06-21.

Lennart R Koetzier, Domenico Mastrodicasa, Timothy P Szczykutowicz, Niels R van der Werf, Adam S Wang, Veit Sandfort, Aart J van der Molen, Dominik

Fleischmann, and Martin J Willemink. Deep learning image reconstruction for ct: technical principles and clinical prospects. *Radiology*, 306(3):e221257, 2023.

Katarzyna Krupa and Monika Bekiesińska-Figatowska. Artifacts in magnetic resonance imaging. *Polish journal of radiology*, 80:93, 2015.

Rohit Kundu. Domain adaptation in computer vision: Everything you need to know. `https://www.v7labs.com/blog/domain-adaptation-guide`, 2022. Accessed: 2024-08-18.

David B. Larson, David C. Magnus, Matthew P. Lungren, Nigam H. Shah, and Curtis P. Langlotz. Ethics of using and sharing clinical imaging data for artificial intelligence: A proposed framework. *Radiology*, 295(3):675–682, 2020. doi: 10.1148/radiol.2020192536.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521 (7553):436–444, 2015.

Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Foundations & trends in multimodal machine learning: Principles, challenges, and open questions. *ACM Computing Surveys*, 56(10):1–42, 2024.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

Mengting Liu, Piyush Maiti, Sophia Thomopoulos, Alyssa Zhu, Yaqiong Chai, Hosung Kim, and Neda Jahanshad. Style transfer using generative adversarial networks for multi-site mri harmonization. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 313–322. Springer, 2021.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33: 21464–21475, 2020.

Bradley C Lowekamp, David T Chen, Luis Ibáñez, and Daniel Blezek. The design of simpleitk. *Frontiers in neuroinformatics*, 7:45, 2013.

Junyan Lyu, Yiqi Zhang, Yijin Huang, Li Lin, Pujin Cheng, and Xiaoying Tang. Aadg: Automatic augmentation for domain generalization on retinal image segmentation. *IEEE Transactions on Medical Imaging*, 41(12):3699–3711, 2022.

Andrzej Maćkiewicz and Waldemar Ratajczak. Principal components analysis (pca). *Computers & Geosciences*, 19(3):303–342, 1993.

Frank Mancolo. Eisen: a python package for solid deep learning. *arXiv preprint arXiv:2004.02747*, 2020.

Michalis Mazonakis and John Damilakis. Computed tomography: What and how does it measure? *European journal of radiology*, 85(8):1499–1504, 2016.

Rajiv Mehrotra, Kameswara Rao Namuduri, and Nagarajan Ranganathan. Gabor filter-based edge detection. *Pattern recognition*, 25(12):1479–1494, 1992.

Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*, pages 117–122. IEEE, 2018.

Richard Morin and Mahadevappa Mahesh. Role of noise in medical imaging. *Journal of the American College of Radiology*, 15(9):1309, 2018.

Muehllehner and Karp. Positron emission tomography. *Physics in Medicine & Biology*, 51:117–137, 2006.

Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

Alexander J Ratner, Henry Ehrenberg, Zeshan Hussain, Jared Dunnmon, and Christopher Ré. Learning to compose domain-specific transformations for data augmentation. *Advances in neural information processing systems*, 30, 2017.

Mengwei Ren, Neel Dey, James Fishbaugh, and Guido Gerig. Segmentation-renormalized deep feature modulation for unpaired image harmonization. *IEEE transactions on medical imaging*, 40(6):1519–1530, 2021.

Rennie. An introduction to the use of tracers in nutrition and metabolism. *Proceedings of the Nutrition Society*, 58:935–944, 1999.

Markus Ringnér. What is principal component analysis? *Nature biotechnology*, 26 (3):303–304, 2008.

David Roessner. Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation*, 9(2):125–132, 2000.

Saha. *Basics of PET Imaging: Physics, Chemistry, and Regulations*, volume 3. Springer, 2015.

Ludwig Schmidt, Shibani Santurkar, Dimitris Tsipras, Kunal Talwar, and Aleksander Madry. Adversarially robust generalization requires more data. *Advances in neural information processing systems*, 31, 2018.

Justus Schock and Michael Baumgartner. rising. `https://github.com/PhoenixDL/rising`, 2023. Accessed: 2024-08-28.

scikit-image. Histogram matching. `https://scikit-image.org/docs/stable/auto_examples/color_exposure/plot_histogram_matching.html`. Accessed: 2024-08-12.

Silvia Seoni, Alen Shahini, Kristen M Meiburger, Francesco Marzola, Giulia Rotunno, U Rajendra Acharya, Filippo Molinari, and Massimo Salvi. All you need is data preparation: A systematic review of image harmonization techniques in multi-center/device studies for medical support systems. *Computer Methods and Programs in Biomedicine*, page 108200, 2024.

Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19(1):221–248, 2017.

Hoo-Chang Shin, Alvin Ihsani, Swetha Mandava, Sharath Turuvekere Sreenivas, Christopher Forster, Jiook Cha, and Alzheimer's Disease Neuroimaging Initiative. Ganbert: Generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis. *arXiv preprint arXiv:2008.04393*, 2020.

Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

Akhil Singh, Vaibhav Jaiswal, Gaurav Joshi, Adith Sanjeeve, Shilpa Gite, and Ketan Kotecha. Neural style transfer: A critical review. *IEEE Access*, 9:131583–131613, 2021.

TB Smith, S Zhang, A Erkanli, D Frush, and E Samei. Variability in image quality and radiation dose within and across 97 medical facilities. *J Med Imaging (Bellingham)*, 8(5):052105, 2021. doi: 10.1117/1.JMI.8.5.052105.

Travis B Smith. MRI artifacts and correction strategies. *Imaging in Medicine*, 2(4): 445, 2010.

3D and Quantitative Imaging Laboratory Stanford. CT imaging artifacts. `https://3dqlab.stanford.edu/imaging-artifacts/`. Accessed: 2024-08-12.

Waheeda Sureshbabu and Osama Mawlawi. PET/CT imaging artifacts. *Journal of nuclear medicine technology*, 33(3):156–161, 2005.

Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, 2020.

Sara L Thrower, Karine A Al Feghali, Dershan Luo, Ian Paddick, Ping Hou, Tina Briere, Jing Li, Mary Frances McAleer, Susan L McGovern, Kristina Demas Woodhouse, et al. The effect of slice thickness on contours of brain metastases for stereotactic radiosurgery. *Advances in Radiation Oncology*, 6(4):100708, 2021.

Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

Anamaria Vizitiu, Cosmin Ioan Niţă, Andrei Puiu, Constantin Suciu, and Lucian Mihai Itu. Towards privacy-preserving deep learning based medical imaging applications. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2019. doi: 10.1109/MeMeA.2019.8802193.

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and S Yu Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022.

Jing Wang, Hongbing Lu, Zhengrong Liang, Daria Eremina, Guangxiang Zhang, Su Wang, John Chen, and James Manzione. An experimental study on the noise properties of x-ray ct sinogram data in radon space. *Physics in Medicine & Biology*, 53(12):3327, 2008.

Shengyun Wei, Shun Zou, Feifan Liao, et al. A comparison on data augmentation methods based on deep learning for audio classification. In *Journal of physics: Conference series*, volume 1453, page 012085. IOP Publishing, 2020.

Dominik Weishaupt, Victor D Köchli, and Borut Marincek. *Wie funktioniert MRI?: eine Einführung in Physik und Funktionsweise der Magnetresonanzbildgebung*. Springer, 2009.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

McKell Woodland, Austin Castelo, Mais Al Taie, Jessica Albuquerque Marques Silva, Mohamed Eltaher, Frank Mohn, Alexander Shieh, Suprateek Kundu, Joshua P. Yung, Ankit B. Patel, and Kristy K. Brock. Feature extraction for generative medical imaging evaluation: New evidence against an evolving trend, 2024. URL `https://arxiv.org/abs/2311.13717`.

Jiancheng Yang, Rui Shi, and Bingbing Ni. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 191–195, 2021.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, pages 1–28, 2024.

Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, Eric I-Chao Chang, and Yan Xu. Mri cross-modality image-to-image translation. *Scientific reports*, 10(1): 3753, 2020.

Hyeona Yim, Seogjin Seo, and Kun Na. Mri contrast agent-based multifunctional materials: Diagnosis and therapy. *Journal of Nanomaterials*, 2011(1):747196, 2011.

M Yogeshwari and G Thailambal. Automatic feature extraction and detection of plant leaf disease using glcm features and convolutional neural networks. *Materials Today: Proceedings*, 81:530–536, 2023.

Junhai Zhai, Sufang Zhang, Junfen Chen, and Qiang He. Autoencoder and its various variants. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 415–419. IEEE, 2018.

B. Zhang, B. Rahmatullah, and S.L. et al. Wang. A review of research on medical image confidentiality related technology coherent taxonomy, motivations, open challenges and recommendations. *Multimed Tools Appl*, 82(None):21867–21906, 2023. doi: 10.1007/s11042-020-09629-4.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, 2022.

A. Ziller, D. Usynin, and R. et al. Braren. Medical imaging deep learning with differential privacy. *Sci Rep*, 11(None):13524, 2021. doi: 10.1038/s41598-021-93030-0.

Declaration of Authorship

Ich erkläre hiermit gemäß §9 Abs. 12 APO, dass ich die vorstehende Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Des Weiteren erkläre ich, dass die digitale Fassung der gedruckten Ausfertigung der Abschlussarbeit ausnahmslos in Inhalt und Wortlaut entspricht und zur Kenntnis genommen wurde, dass diese digitale Fassung einer durch Software unterstützten, anonymisierten Prüfung auf Plagiate unterzogen werden kann.

Place, Date

Signature