The Lens Journal
A journal of perspectives for young researchers

Revolutionizing Medical Imaging: The Vital Role of Diffusion Models in Modern Image Augmentation

Arjun Suresh
Newsome High School Lithia, Florida, U.S
C. D'Ettorre
University College London, Gower street, London, England, UK

Abstract
AI-based medical imaging has been widely used to identify complex patterns in imaging data [1], provide a quantitative evaluation of radiographic traits and detect image modalities at various treatment stages (for example, tumor delineation).

Denoising diffusion models, a class of generative models, have garnered immense interest lately in various deep-learning problems, especially since there is not enough images to train the network [2]. The main issue in Medical Imaging is that most research groups and industry have limited data access based on small sample sizes from small geographic areas. In addition, the preparation of data is a costly and time-intensive process, the results of which are algorithms with limited utility and poor generalization [2]. Diffusion models are widely appreciated for their strong mode coverage and quality of the generated samples in spite of their known computational burdens. Capitalizing on the advances in computer vision, the field of medical imaging has also observed a growing interest in diffusion models. With the aim of helping the researcher navigate this profusion, this paper intends to provide a comprehensive overview of diffusion models, challenges, and future directions in leveraging diffusion models for refining medical image data. Ultimately, this paper underscores the pivotal role of diffusion models in shaping the present and future landscapes of medical imaging, offering unprecedented opportunities for enhanced diagnostic accuracy and improved patient care [3].

Firstly, the paper starts by exploring what exactly are diffusion models and how they are applied in the field of medical imaging. Next, it focuses on the critical analysis and review of the other documents that have been published in this field in a table format, as well as analyses the methodology used. The paper also discusses recent advancements, challenges, and future directions in leveraging diffusion models for refining medical image data. Ultimately, this exploration emphasizes the pivotal role of diffusion models in shaping the present and future landscapes of medical imaging, offering unprecedented opportunities for enhanced diagnostic accuracy and enhanced patient care.

Keywords: Diffusion Models, Medical Imaging, Data Augmentation, GANs, VAEs, Deep Learning Methods, Image Augmentation Methods.

Introduction

Medical imaging refers to several different technologies that are used to view the human body in order to diagnose, monitor or treat medical conditions [4]. The vast majority involves non-invasive visualization techniques for medical personnel to identify injuries, diagnose diseases or chronic conditions [5]. The application of AI analysis to medical imaging supports healthcare professionals in identifying problems areas or details that may be missed by the human eye. The main areas of AI application for medical purposes include imaging processing, physiological signal recognition, and neurological health issues. Researchers have applied AI to automatically recognizing complex patterns in imaging data and providing quantitative assessments of radiographic characteristics [6]. In radiation oncology, AI has been applied on different image modalities that are used at different stages of the treatment like tumor delineation and treatment assessment. Radiomics, the extraction of a large number of image features from radiation images with a high-throughput approach, is a widely researched topic today in medical imaging research [7]. AI is the essential boosting power of processing massive number of medical images and therefore uncovers disease characteristics that fail to be noticed by the naked eye.

Image augmentation is a technique that is used to artificially expand the size of a dataset [8]. This is helpful when there are datasets with very few data samples. Augmentation via generative models (models that are used to generate data similar to the data on which they are trained) like Diffusion Models improves image recognition by generating large scale image datasets from a handful of real images without human labeling effort. It is a very popular technique in computer vision and has become inseparable from deep learning applications. A diffusion probabilistic model defines a forward diffusion stage where the input data is gradually perturbed over several steps by adding Gaussian noise. The models are trained on a large dataset of billions of filtered image text pairs retrieved from the internet, enabling them to generate images of unparalleled variety. Furthermore, these models can be easily conditioned using text queries, which allows for easy control of the data generation process [9]. The usage of these models is becoming more ubiquitous as they have overcome the challenges associated with aligning posterior distributions in variational autoencoders (VAEs) and dealing with the unstable adversarial objective in Generative adversarial networks (GANs), thus outperforming these older, traditional methods in terms of visual quality and diversity [10]. Consequently, diffusion models have garnered significant attention in various domains, such as computer vision, sequence modeling, audio processing, and AI for science [11]. These models are widely appreciated for their strong mode coverage and quality of the generated samples despite their known computational burdens. Compared to traditional generative models, diffusion models have better image quality, interpretable latent space, and robustness to overfitting. Diffusion models have diverse applications across several domains, such as text-to-video synthesis, image-to-image translation, image search, and reverse image search [10].

This work aims to provide a comprehensive examination of diffusion models used in the rapidly evolving field of medical imaging and consolidate diverse studies across imaging modalities, such as MRI, CT, and PET. A meticulous analysis was conducted on the nuanced applications of diffusion models in diagnosing and understanding neurological disorders, tissue characterization, and treatment monitoring. Uniquely, this synthesis delves into the recent advancements in diffusion and other emerging techniques, comparing their strengths and limitations. By analyzing and discussing the practical implications and potential future directions, this review serves as a valuable guide for researchers and clinicians, bridging theoretical insights with real-world clinical applications in a way that previous literature review papers have not yet achieved. We have made it easier to identify patterns in how DMs are being used currently, their competitive advantages and limitations, by organizing this information in a table format.

Introduction

Diffusion models are a powerful class of probabilistic generative models that are used to learn complex data distributions. These models achieve this learning process by two principal stages: the forward and reverse diffusion process [11]. The forward diffusion process adds noise to the input data, gradually increasing the noise level until the data is altered to pure Gaussian noise. This process systematically perturbs the structure of the data distribution. The reverse diffusion process, also known as denoising, is then applied to recover the original structure of the data from the perturbed data distribution. The structure of the data distribution is systematically altered by this procedure. The outcome is a profoundly adaptable and manageable generative model that can accurately display complex data distributions from random noise [11]. Due to the theoretical basis and consistent performance of these diffusion models, they have since been applied to a variety of vision tasks, such as inpainting, super resolution, text-to-image translation, anomaly detection and segmentation [3]; they can augment as well as generate realistic images from textual input, as demonstrated in [12].

In [3], the authors have used DM for the segmentation of nuclei in histopathology images. A diffusion-based data augmentation framework that can generate histopathology images and their segmentation labels from scratch has proven to be very useful in nuclei segmentation. Manually labeling all nuclei instances for a dataset is inefficient, and obtaining a large-scale human-annotated dataset is time-consuming and labor-intensive. Therefore, augmenting a dataset with only a few labeled images to improve the segmentation performance is of great application value [3]. Denoising Diffusion Probabilistic Model (DDPM) have been used in generating 4D cardiac MR images between the diastolic and systolic phases, through the latent space provided by the parameterized Gaussian process [13]. One particular type of DDPM, Diffusion deformable models (DDM), is used to generate realistic images by learning a Markov chain process for the transformation of the simple Gaussian distribution into the data distribution [14]. These DDMs generate intermediate temporal volumes between source and target volumes. They are composed of the diffusion and the deformation modules which enables them to be able to learn spatial deformation information between the source and target volumes, so as to provide a latent code for generating intermediate frames along a geodesic path. This latent code is then simply interpolated and fed into the deformation module, which enables the DDM to generate temporal frames along the continuous trajectory all while preserving the topology of the source image. Compared to the existing deformation methods, this type of DDM achieves high performance on temporal volume generation [13]. These DDMs are a promising tool in clinical applications as it connects the frames between two images in real-time when analyzing changes in anatomical structures [15].

Another type of DM is Anomaly Detection with Denoising Diffusion Probabilistic Models (AnoDDPM) [16]. AnoDDPM is trained on healthy MRI scans of the brain to potentially detect anomalous distributions, by means of a partial diffusion process. Fig. 1 shows a sample of an MRI scan comparing a brain that is healthy and one that contains a tumor.

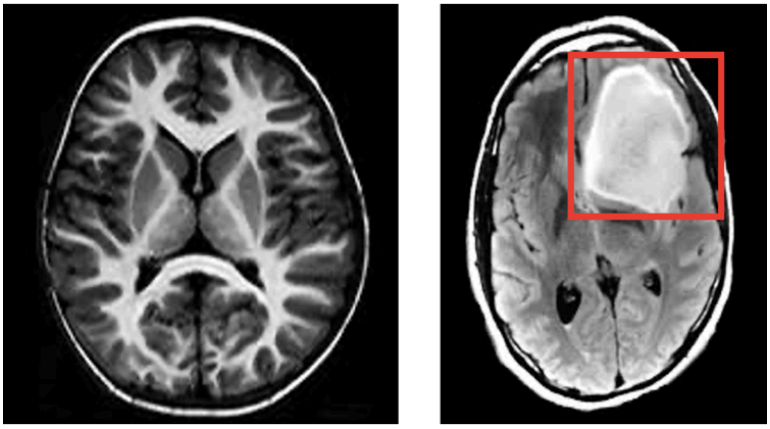

Fig. 1: On the left: MRI scan of a healthy brain image. On the right: MRI scan of a brain with tumor. Source: Brain MRI Images for Brain Tumor Detection Dataset collected by Navoneel Chakrabarty [[44].

The images are corrupted and then reconstructed to a healthy approximation. The data is subsequently compared with the original image and any anomalies in the data are detected. This process is greatly useful as it allows for better reconstruction of samples from the data distribution and scales to high resolution images well. Moreover, it also provides several advantages over adversarial training as it can better capture smaller datasets with improved sample quality and stable training. When these DDPMs are combined with simplex noise, they are able to capture large anomalous regions with stable training that does not require large datasets, avoiding the limitations common in GAN-based approaches [17]. pDDPMs, a type of Diffusion Models, apply the forward diffusion process on a small part of the input image and use the whole, partly noised image in the backward process to recover the noised patch. At test time, the trained pDDPM are then used to sequentially noise and denoise a sliding patch within the input image and then stitch the individual denoised patches to reconstruct the entire image. Thus, the Patch-based DDPMs serve as a promising approach also for brain MRI reconstruction [18].

Diffusion autoencoders can encode any image into a two-part latent code that captures both semantics and stochastic variations and allows near-exact reconstruction [19]. They allow us to visualize different grades of a given vertebra. The grading scale was proposed as follows: Degree zero corresponded to the absence of signs of degenerative disease in the lumbar spine; Grade I – presence of signs of degenerative disease up to two segments; Grade II – involvement of three or more segments; Grade III – associated deformity or signs of instability (grading scale). Fig.2 illustrates these gradings. This provides interpretability and insight into the features that contribute to automated grading [20].

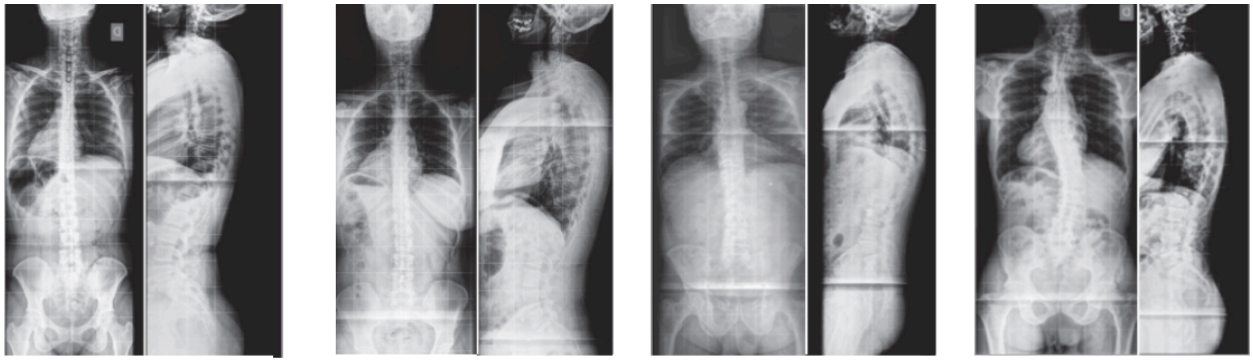

Fig. 2: Grade 0 corresponded to the absence of signs of degenerative disease in the lumbar spine; Grade I – presence of signs of degenerative disease up to two segments; Grade II – involvement of three or more segments; Grade III – associated deformity or signs of instability (grading scale). Source: image adapted from [16].

Most commonly used in the augmentation of underrepresented data, they serve to address the imbalance in computer-aided diagnosis – Even though vertebrae segmentation datasets exist, only a few graded vertebra CTs are available. The advantages of this method is that it offers inherent interpretability by enabling clinicians to visualize the inner representations of the model's decision-making process, and have the potential to scale with more unlabelled data when they are trained without any supervision [21]. Moreover, a hyperplane can be constructed in the DAE latent space that serves as an effective decision boundary for fracture detection and that the distance from this hyperplane can be leveraged to estimate fracture severity. Therefore, it is no surprise that Diffusion Models are being preferred for use instead of the likes of GANs (generative adversarial network) and CNN (Convolutional Neural Network) as they are capable of outperforming both the GAN- and CNN-based methods in terms of structural similarity index (SSIM) and peak signal-to-noise ratio (PNSR) [22]. This is one of the most advanced approaches in this field. These models have overcome the challenges associated with aligning posterior distributions in VAEs and dealing with the unstable adversarial objective in GANs [23].

These are some of the real-world examples where Diffusion Models have been applied in medical imaging. Some of the existing works like [24] focus only on a specific application, while others concentrate on a specific image modality. There have also been many review articles on recently emerging diffusion models for computer vision tasks [25] Although the reviews were published before the field was fully developed, many advances have been made in the medical field since then. On the other hand, to the best of our knowledge, none of these studies provide a comparison on the applications of diffusion models in medical imaging, which is a central aspect in advancing this research. Therefore, large gaps remain in these studies. In addition, we believe that the medical community can leverage insights from successful products of diffusion models in vision by a retrospective on the past and future research directions of diffusion models provided.

Even though there is great interest in these diffusion models, a comprehensive taxonomy and analysis of the research progress on DMs are still lacking. Due to a great influx of contributions in this field of study, a comprehensive review and comparison of the existing literature would be beneficial and timely for the community. This study aims to provide a comprehensive overview/comparison of the Applications, Image Type, Image Augmentation, Competitiveness with other models (GANs, VAEs), limitations of DMs. This paper describes the challenges and issues, and identifies common trends that raise open questions about the future development of diffusion models both as algorithms and applications in the

Methods

In order to identify and catalog all the available publications involving DMs and generative models, specifically DMs in use in medical imaging, the protocol queries two main databases: National Institutes of Health (NIH) [33], Google Scholar [34]. Some of the papers used were cited from international conferences or journals, as well as from the open-access articles stored in arXiv. Firstly, a search was conducted on Google scholar with the query 'Diffusion Models in Medical Imaging'. The research time interval was set between 2015 (when the first paper introducing DMs was published [35]) and 2023. The results were further processed and refined by removing certain papers where DMs was not actually mentioned in the Methods section of the work (which means it was just cited but not actually used in the experimental work) as used in [36] which focuses on the application of VAEs and GANs in chemical science in its methodology, but only briefly mentions DMs. This research finally generated 32 papers from 44. Out of these 32 papers, 12 of them were discarded due to them being comprehensive reviews like [37] or surveys like [38]. The aim of this paper is to provide a general overview of the current state of DMs in medical imaging. Due to time constraints, 10 of the 20 papers were randomly chosen to analyze closely. These papers focused mostly on DMs being used to augmentation images in medical imaging. These remaining 10 papers are summarized in table 1.

For analyzing the body of publications, the information in the papers were classified into 6 groupings; these were our points of interests to analyze the relevancy and use of DMs in modern image augmentation: Applications (how DMs have been applied in that study); Image Type (the images that DMs were used on), Dataset (where the images were received from); Image Augmentation (in what way were DMs used to augment/modify the images); Competitiveness with other models (GAN, VAE) [how do these DMs fair against similar other generative models, i.e. how are they more efficient/inefficient]; Limitations (focuses on the shortcomings of DMs in these applications). Therefore, the goal of this paper is to summarize the research papers to provide crucial information about these papers at a glance. For this purpose, the information is organized in a table format. Table 1 reports the proposed classifications.

TABLE 1

Applications	Image Type	Dataset	Image Augmentation	Competitiveness with other models (GAN, VAE)	Limitations
Generating high-quality image samples [26]	Facial images	• CelebAMask-HQ dataset • CIFAR-10 dataset	• Exhibits out-of-the-box generalization to downstream tasks like image super-resolution and denoising • Restore, Denoise and increase resolution of images	• Better sample quality • Impressive performance on several image synthesis • Clearer images	• Require an expensive iterative sampling procedure • Lack a low-dimensional latent representation

3D Medical Image Generation [12], [27]	• Breast MRI • Lungs CT Scan	• BraTS2020 dataset • LIDC - IDRI	Enriching synthetic images	Can synthesize high-quality medical data for MRI and CT	• Require a large amount of computational resources which limits their application to medical tasks • Only allows the processing of small 3D volumes, which impedes the processing of high-resolution MR / CT scans. • Needs excessive hardware
Nuclei Image Segmentation [3]	Histopathology images	• MoNuSeg • Kumar	We first generate a nuclei structure with an unconditional diffusion model, and then generate images conditioned on the nuclei structure	• Can generate nuclei structures with realistic nuclei shapes and spatial distribution • Can generate images of close resemblance to real histopathology images and high diversity	Requires large amounts of resources to carry out
To produce high quality samples in many domains [28], [29]	• Facial Images • Images of objects and buildings	• CIFAR10 • CelebA	DMs are trained to denoise samples corrupted by various levels of Gaussian noise. Samples are then produced by a Markov chain which, starting from white noise, progressively denoises it into an image	• Able to generate high quality samples much more efficiently • Superior sample generation quality	• They require simulating a Markov chain for many steps in order to produce a sample • They require many iterations to produce a high quality sample. This is much slower compared to GANs, which only needs one pass through a network
To augment a dataset with images sampled by prompt-based conditioning techniques [9]	Real-world images with high photorealism, variety, and detail	• LAION-5B [30] • ImageNet • Caltech256	Guidance-free diffusion model sampling using simple conditioning techniques with prompts based on the objects' class label, and personalization techniques that fine-tune the diffusion model conditioning	• Outperform GANs in terms of visual quality and diversity • Able to generate images of unparalleled variety • Can be easily conditioned	• These sophisticated models are unable to beat a simple and strong image retrieval baseline on simple downstream vision tasks.

				using text queries, thus allowing for easy control of the data generation process	
Pathology detection with reconstructions of images using score-based DMs [31]	MRI images of knees	FastMRI knee dataset	Reconstructing MRI images	• Provides a powerful way to model images using the gradient of the data distribution • Can be readily used for solving inverse problems in imaging, especially for accelerated MRI	• When reconstructing OOD data with 1D under-sampling pattern, sometimes one may observe mild aliasing-like artifacts in local edges • When pursuing extreme condition reconstruction, DMs may occasionally produce results that are unsatisfactory (e.g. sample marked with the red dotted line). • The detailed structure may also have high variance within the posterior samples, due to the high ill-posedness
To generate visually realistic images from textural descriptions [10]	Images generated by user prompts	• MS-COCO • ENTIGEN benchmark dataset • Wikiart • VIST	Generation of synthetic images by using user-prompts	• DMs have excellent inversion property • Performs well in a wide range of image editing • tasks, constituting a generalized framework	• Models trained on large scale unfiltered data may suffer from even reinforce the biases from the dataset, leading to ethical risks • Dataset bias • The large model size of DMS, prevent their deployment in efficiency-oriented environments, such as edge devices • May not always be consistent with perceptual quality • The automatic evaluation criteria is not so diverse and reliable

To generate multiple plausible annotations from a single input image [32]	• CT scan images of lungs • MRI scans • Ultrasound scans of bones	• Lung lesion segmentation (LIDC-IDRI) dataset • Bone surface segmentation (Bone-US) • Multiple sclerosis lesion segmentation (MS-MRI).	CIMD (a type of DM) uses its hierarchical structure to incorporate stochasticity at each level and doesn't require a separate network to encode prior information about the image during the inference stag	• DMs realistically model the heterogeneity of the segmentation masks without requiring any additional network to provide prior information during inference. • Better than the existing ambiguous image segmentation networks in terms of quantitative standards while producing superior qualitative results.	Since DMs trained using a single segmentation mask per input image, the models have no control over the variation and moreover, the produced masks are not necessarily diverse.

A. Applications

As reported in the table, diffusion models can be applied to a variety of purposes ranging from nuclei image segmentation to tumor identification from CT Scan images. It is important to note the different ways in which DMs are applied so as to gain a broader understanding of their place in the field of not only medical imaging but generative AI models as a whole.

CT and MRI are two of the most prevalent imaging types. CT, however, is limited in displaying images with the utmost detail for soft tissue injuries [11]. Hence, a subsequent MRI may be needed for a conclusive diagnosis after receiving the initial CT scan. Hence, in addition to being time-consuming and costly, this process may also cause misalignment between MRI and CT images. To this end, [31] took advantage of Score-based diffusion models to properly model the actual complicated data distribution of MRI scans for accelerated MRI. Fig.3 shows the method used by the authors to model data in MRI images of knees by using DMs.

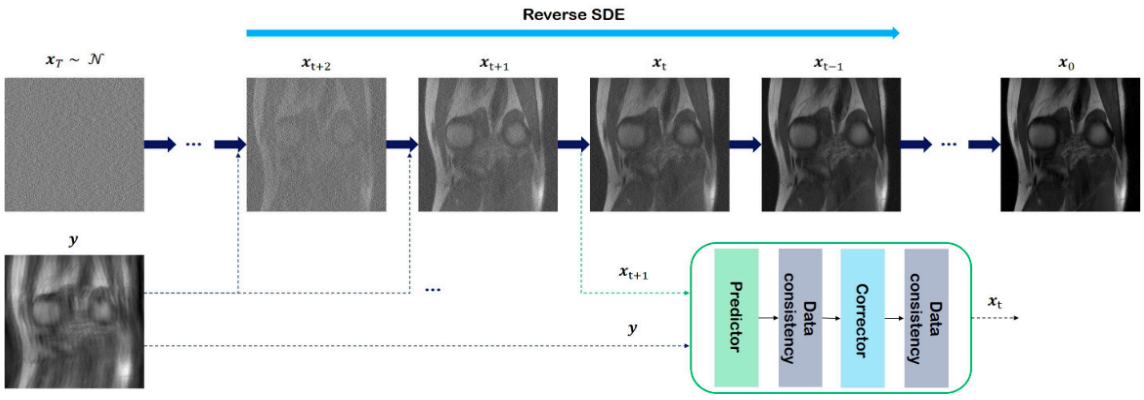

Fig. 3: Overview of the proposed method in [31] which uses DMs to model complex data in MRI scan images.

Currently, DALL-E from OpenAI [33] and Parti [34] from Google have played a huge role in exploiting large scale data for text-to-image generation. However, autoregressive nature makes these methods suffer from high computation costs and sequential error accumulation. Thus, DeepFusion has made use of DDPMs to generate 3D images from text [10].

Therefore, this column is to make known the various ways in which DMs are being used to play a pivotal role in Medical Imaging. Moreover, this also shows the wide variety of DDPMS being used for various different purposes in Medical Imaging e.g. AnoDDPM [16], TabDDPM [35].

B. Image Type

This section groups all the types of images that these DMs were used on. The most common images used were facial images as shown in Fig.4. These facial images were used as inputs to DMs in [26],[28], [29] . The other common image types were CT/MRI scans of organs e.g. lungs, breast like those [12], [31], [27] [32], which were then augmented by DMs.

Fig.4: Facial images being augmented by DMs as used in [28].

Although the most widely used images in Medical Imaging are CT and MRI scans [36], additionally, an attempt was made to show that DMs have the capacity to augment vastly diverse types of images to the utmost precision and quality. Hence, the types of images used in the study of these papers was also noted in the table.

C. Dataset

An image dataset is a collection of data compiled for a machine learning project [37]. It includes digital images curated for testing, training, and evaluating the performance of machine learning and AI generative models like DMs. From the information compiled in the table, for DMs working with facial images, the most popular and widely used data sets were the CelebA dataset (shown in Fig.5) and CIFAR-10 dataset. These were used in [26], [28], [29] – papers which focused predominantly on augmenting facial image samples. For papers that required medical data e.g. CT scans of organs, the most common choice was usage of the Lung lesion segmentation (LIDC-IDRI) dataset (Shown in Fig.6) as used in [12], [27], [32].

Fig.5: 256 x 256 samples from CelebA dataset as used in [28].

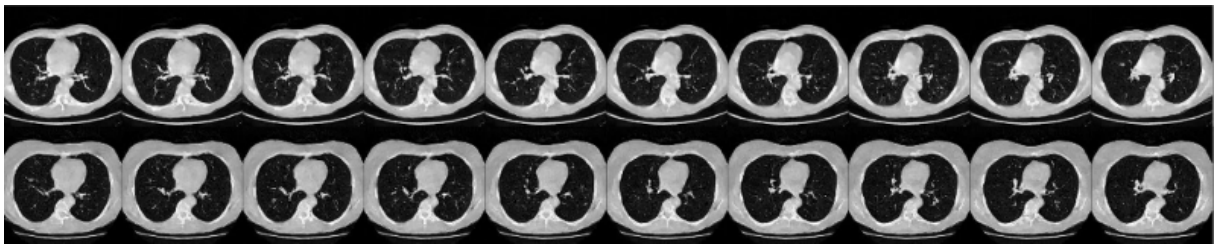

Fig.6: Images from LIDC-IDRI dataset as used in [12].

D. Image Augmentation

Data Augmentation addresses the challenge of limited data by expanding the dataset within the data space. It involves employing various methods to increase both the quantity and quality of training data, ultimately enabling the development of more effective Deep Learning models. This section was to find how the DMs were used to change the images throughout the experiment i.e. the techniques used, how they were applied to augment the images used.

E. Competitiveness with other models (GAN, VAE)

This section shows how DMs compare with other models which serve similar purposes like VAEs and GANs. GANs are capable of generating high-quality samples rapidly, but they have poor mode coverage and are prone to lack sampling diversity. Conversely, VAEs and normalizing flows suffer from the intrinsic property of low sample quality despite being witnessed in covering data modes [11]. A common concern with GANs is their training dynamics which have been recognized as being unstable, resulting in deficiencies such as mode collapse, vanishing gradients, and convergence [ref]. Furthermore, despite their remarkable achievements, the behavior of VAEs is still far from satisfactory due to some theoretical and practical challenges like balancing issues and variable collapse [38]. Despite being feasibly designed, these generative models have their specific drawbacks.

E. Limitations of DMs

This category encompasses all the concerns and limitations of DMs expressed in these papers. The most common drawbacks were that they require a large amount of resources to use which may be expensive [26], [27]; though DMs are able to produce better quality samples than other models like GANs, they require many iterations to do so. It may go through thousands of steps just to produce a single sample like in [29]. This makes it much slower when compared to GANs, which only need one pass through a network.

Results/Discussion

After analyzing the 10 papers and collecting information as reported in table 1, some common patterns and trends were identified. In this section, these patterns are discussed, along with a brief explanation of these common trends.

Firstly, a pattern was noticed in the way that DMs were being used in these papers: DMs were predominantly being used in the generation of high-quality images [10], [12], [26], [27], [28], [29] and in the augmentation of images [9], [31] whether by means of reconstruction like in [31], or via image transformations using generative adversarial networks like in [9]. In the medical field, numerous datasets may encounter significant class imbalance. DMs address this challenge by generating highly realistic images for application in the medical field. Moreover, generating synthetic images holds considerable educational merit. By constructing distinct synthetic samples for learning, DMs can meet educational needs by crafting diverse synthetic examples for teaching and training purposes. Additionally, these synthetic images can alleviate concerns surrounding data security when utilizing patient data in public environments [11]. Furthermore, these images can address a critical challenge in training deep neural networks for medical tasks. Thus, utilizing DMs for generating synthetic samples presents a promising approach to mitigating the issue of medical data scarcity considerably. The most prevalently used images were facial images and CT/MRI scans of organs like lungs. One reason why DMs are being used to augment these types of images is because they can be used to handle high-dimensional and complex data, such as medical images, which are difficult to model using conventional methods. Moreover, DMs can also provide a more efficient and accurate way to infer the underlying physical properties of the object or system and can handle uncertainty and noise in the measurements or observations [11]. The most common dataset used for facial image samples was the CelebA dataset. An integral aspect of computer vision and deep learning involves the identification of faces for various purposes. This dataset is great for training and testing models for face detection, particularly in discerning facial attributes like hair color or glasses. It serves as a robust resource for training and testing across multiple computer vision tasks, including face attribute recognition, face detection, and landmark localization within the face [39]. As for DMs being applied in medical imaging, the most common data set used was the LIDC-IDRI dataset. This dataset comprises 1018 cases, each containing clinical thoracic CT scan images accompanied by an XML file detailing the outcomes of a two-phase image annotation process conducted by four experienced thoracic radiologists [40].

The most common way DMs were used was to enrich and produce higher-quality images of higher resolution as in [26], [27]. DMs do this via a process called denoising. These models leverage a diffusion process where a noise-free image gradually transforms into the desired image through a series of intermediate noise levels. This method effectively reduces noise and enhances details in the generated images. By learning to gradually denoise from Gaussian noises with a noise prediction, diffusion models can generate high-quality clean data that fits well with real data distribution. Diffusion models trained on large-scale text-image pairs, such as Imagen and Stable Diffusion, have gained unprecedented success in text-to-image generation. Therefore, they are suitable as a powerful prior for multiple editing tasks [41]. Furthermore, augmenting existing medical images can substantially increase the sample size of the training dataset. Image translation involves altering an image's style while preserving its original content. In the context of medical image translation, this means changing the modality (e.g. CT, MRI) of the image while retaining its anatomical details and clinical representation. Thus, researchers can transform images from various modalities into the desired target modality, thereby expanding the dataset for that specific modality. Affine transformations such as flipping and scaling are the most common image augmentation operations. This method can significantly increase the number of samples. Additionally, the image stitching method, a traditional augmentation approach, merges foreground and background regions from different images through a copy-paste technique, thereby further enhancing the dataset's diversity [24].

A frequent pattern noticed from the papers gathered is that DMs have better sample quality [26], and produce images of sharper and higher quality as in [12], [28], [29], when compared to that of GANs and VAEs. According to [9], DMs have outperformed traditional generative models like GANs in terms of visual quality of images and greater diversity. Another competitive advantage of DMs is that they are able to generate highly-realistic images that bear close resemblance to the original, as observed in [3] where DMs were used to generate histopathology images with nuclei structures, and in [32]

where DMs were deployed to realistically model segmentation masks. In contrast to GANs and VAEs which directly generate data samples, DMs refine noisy data iteratively by adding noise at each step, making them well-suited for tasks like data refinement and denoising. In [42], the authors have used synthetic images generated by diffusion models and found that it improved the accuracy of skin classifiers. They further noted that models that are trained using a combination of synthetic and real data perform better than those trained using only one data source. The findings of the research suggest that synthetic images closely resemble real ones and can effectively train models for medical research. Therefore, in these cases, DMs could be a better alternative to these models as they can provide a more accurate representation of the data distribution. This leads to better reconstructions of the image when DMs are used than when other generative model types are used.

Conclusion

By reviewing various research papers, specific information was extracted and compiled into a table. These pieces of information were chosen to highlight that DMs are still being relevantly used in modern industries like medical imaging [12] and vision [43] and they have adversely different uses like in [10] which makes use of DMs to augment facial images. There was a need to show how DMs are on-par or even better than traditional generative AI models like GANs and VAEs, and thus a section was included in table 1 that compares how DMs fare against these other types of models. The datasets used and the image types that DMs have been used on were included, to compare as well as show how DMs can be used in vastly different images and image types. Lastly, we acknowledge that not every generative model is perfect and that they have their own set of limitations that still need to be improved. The most common drawbacks of DMs are that they require a large amount of resources to use, which may be expensive. Despite their capability to generate superior quality samples as compared to traditional models like GANs, DMs often demand numerous iterations to achieve desired results, due to the use of multiple denoising steps. Additionally, there's a risk of pretrained diffusion models inadvertently leaking their training data, potentially compromising evaluation processes [43]. Nevertheless, these constraints do not undermine the unique capabilities and advantages of diffusion models in producing high-quality images and their ability to function without extensive sets of labeled or unlabelled data. Instead, they present open challenges for future research and enhancement efforts. Thus, DMs remain an extremely promising tool for generative modeling.

We believe that given more time, we could have researched even more papers which would have led to a more accurate trend to be observed. Applications of DMs in medical imaging is a constantly evolving field and every week there are new DMs coming out that fulfill various purposes, so this opens up the possibility for a vast field of research. There is still much work to be done on researching DMs, their capabilities and their potential for improvement in both theoretical and empirical aspects.

References

[1] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, H. J. W. L. Aerts, and H. H. Edu, "Artificial intelligence in radiology HHS Public Access," *Nat Rev Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[2] M. P. Martin J. Willemink, "Preparing Medical Imaging Data for Machine Learning".

[3] X. Yu *et al.*, "Diffusion-based Data Augmentation for Nuclei Image Segmentation," Oct. 2023.

[4] "Medical Imaging - FDA."

[5] Luís Pinto-Coelho, "How AI-Powered Medical Imaging is Transforming Healthcare."

[6] X. Tang, "the role of artificial intelligence in medical imaging research," 2020.

[7] P. Lambin *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur J Cancer*, Mar. 2012.

[8] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, Dec. 2019.

[9] M. F. Burg *et al.*, "Image retrieval outperforms diffusion models on data augmentation," Apr. 2023.

[10] C. Zhang, C. Zhang, M. Zhang, and I. S. Kweon, "Text-to-image Diffusion Models in Generative AI: A Survey," Mar. 2023.

[11] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, I. Hacihaliloglu, and D. Merhof, "Diffusion Models in Medical Imaging: A Comprehensive Survey," 2023.

[12] F. Khader *et al.*, "Denoising diffusion probabilistic models for 3D medical image generation," 123AD.

[13] B. Kim, "Diffusion Deformable Model for 4D Temporal Medical Image Generation".

[14] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models."

[15] X. Chen *et al.*, "Recent advances and clinical applications of deep learning in medical image analysis HHS Public Access," *Med Image* , vol. 2022.

[16] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise."

[17] Z. Xiao, K. Kreis, and A. Vahdat, "Tackling the Generative Learning Trilemma with Denoising Diffusion GANs," Dec. 2021.

[18] F. Behrendt, D. Bhattacharya, J. Krüger, R. Opfer, and A. Schlaefer, "Patched Diffusion Models for Unsupervised Anomaly Detection in Brain MRI," *Proceedings of Machine Learning Research-Preprint*, 2023.

[19] S. Gong *et al.*, "Diffusion Model based Semi-supervised Learning on Brain Hemorrhage Images for Efficient Midline Shift Quantification," Jan. 2023.

[20] M. Keicher *et al.*, "Semantic Latent Space Regression of Diffusion Autoencoders for Vertebral Fracture Grading".

[21] Z. Salahuddin, H. C. Woodruff, A. Chatterjee, and P. Lambin, "Transparency of deep neural networks for medical image analysis: A review of interpretability methods," *Comput Biol Med*, 2022.

[22] C. Di Ruberto *et al.*, "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," *Journal of Imaging 2023, Vol. 9, Page 81*, Apr. 2023.

[23] D. Saxena and J. Cao, "Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions".

[24] Y. Chen *et al.*, "Generative Adversarial Networks in Medical Image augmentation: A review," *Comput Biol Med*, vol. 144, May 2022.

[25] L. Yang *et al.*, "Diffusion Models: A Comprehensive Survey of Methods and Applications," *ACM Comput Surv*, vol. 56, no. 4, Nov. 2023, doi: 10.1145/3626235.
[26] K. Pandey, A. Mukherjee, P. Rai, A. Kumar, and G. Research, "VAEs meet Diffusion Models: Efficient and High-Fidelity Generation," in *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.
[27] F. Bieder, J. Wolleb, A. Durrer, R. Sandkühler, and P. C. Cattin, "Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing," *Proceedings of Machine Learning Research-nnn*, pp. 1–16, 2023.
[28] J. Song, C. Meng, and S. Ermon, "Denoising Diffusion Implicit Models," Oct. 2020.
[29] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Jun. 2020.
[30] C. Schuhmann *et al.*, "LAION-5B: An open large-scale dataset for training next generation image-text models," Oct. 2022.
[31] J. C. Y. Hyungjin Chunga, "Score-based diffusion models for accelerated MRI".
[32] A. Rahman, J. Maria, J. Valanarasu, I. Hacihaliloglu, and V. M. Patel, "Ambiguous Medical Image Segmentation using Diffusion Models".
[33] "DALL·E: Creating images from text."
[34] "Parti: Pathways Autoregressive Text-to-Image Model."
[35] A. Kotelnikov, D. Baranchuk, I. Rubachev, and A. Babenko, "TabDDPM: Modelling Tabular Data with Diffusion Models".
[36] "The 5 Most Common Medical Imaging Techniques - Houston Physicians Hospital." Accessed: Mar. 11, 2024. [Online]. Available: https://onestepdiagnostic.com/houston-medical-imaging/
[37] "Creating an Image Dataset for Machine Learning: Quick Tutorial."
[38] A. Asperti, D. Evangelista, and E. L. Piccolomini, "A Survey on Variational Autoencoders from a Green AI Perspective," *SN Comput Sci*, 2021.
[39] "celeb_a | TensorFlow Datasets." Accessed: Mar. 18, 2024. [Online]. Available: https://www.tensorflow.org/datasets/catalog/celeb_a
[40] S. G. Armato *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans," *Med Phys*, vol. 38, no. 2, p. 915, 2011.
[41] K. Zhang, Y. Zhou, X. Xu, X. Pan, and B. Dai, "DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing".
[42] M. Akrout *et al.*, "Diffusion-based Data Augmentation for Skin Disease Classification: Impact Across Original Medical Datasets to Fully Synthetic Images".
[43] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion Models in Vision: A Survey," Sep. 2022, doi: 10.1109/TPAMI.2023.3261988.
[44] Navoneel Chakrabarty, "Brain MRI Images for Brain Tumor Detection."