A Survey of Medical Image Generation and Multimodal Generation Techniques

www.surveyx.cn

Abstract

Medical image generation is advancing healthcare by producing high-quality synthetic images that enhance diagnostic and therapeutic processes. This survey explores the interconnected fields of medical image generation, text-to-image generation, multimodal generation, and the use of Generative Adversarial Networks (GANs) and diffusion models. These technologies address data scarcity and privacy concerns, offering solutions through synthetic data augmentation. GANs, with their adversarial framework, excel in creating realistic images, while diffusion models provide high-fidelity images through iterative refinement. The survey highlights applications in medical imaging segmentation and real-world clinical settings, demonstrating significant improvements in diagnostic accuracy and treatment planning. Challenges such as computational demands and the need for robust evaluation metrics are discussed, alongside future directions focusing on model efficiency and privacy-preserving techniques. The integration of advanced generative models promises to transform healthcare delivery, supporting more informed clinical decisions and personalized medicine. Continued research is essential to optimize these technologies and expand their applicability across diverse clinical scenarios, ultimately improving patient outcomes.

1 Introduction

1.1 Significance of Medical Image Generation in Healthcare

Medical image generation is essential for advancing healthcare by producing high-quality synthetic images that improve diagnostic and therapeutic processes. This generation addresses data scarcity, a significant challenge in training deep learning models due to legal and infrastructural constraints limiting access to diverse datasets [1]. Generative models, particularly Generative Adversarial Networks (GANs), create anatomically plausible images for data augmentation and segmentation, enhancing the robustness and accuracy of AI applications in clinical settings [2].

Integrating synthetic images into medical education and training is crucial, as medical students and radiology trainees need exposure to thousands of images to detect subtle visual patterns critical for accurate diagnosis [1]. By augmenting existing datasets with synthetic images, generative models help mitigate the imbalance and scarcity of pathological findings in medical imaging, vital for developing generalizable deep learning models [2].

Moreover, medical image generation fosters novel imaging techniques and the exploration of complex multimodal data, contributing to innovative diagnostic tools and personalized treatment strategies. This is particularly relevant in MR-only radiotherapy, where synthetic CT images improve dose calculations, treatment planning, and patient outcomes [2]. The synthesis of medical reports, which generates coherent descriptions of medical images, further emphasizes the transformative impact of these technologies on diagnostic accuracy and healthcare delivery [2].

The potential of medical image generation extends beyond data scarcity; it enables exploration of internal structures and the development of new imaging modalities, essential for personalized

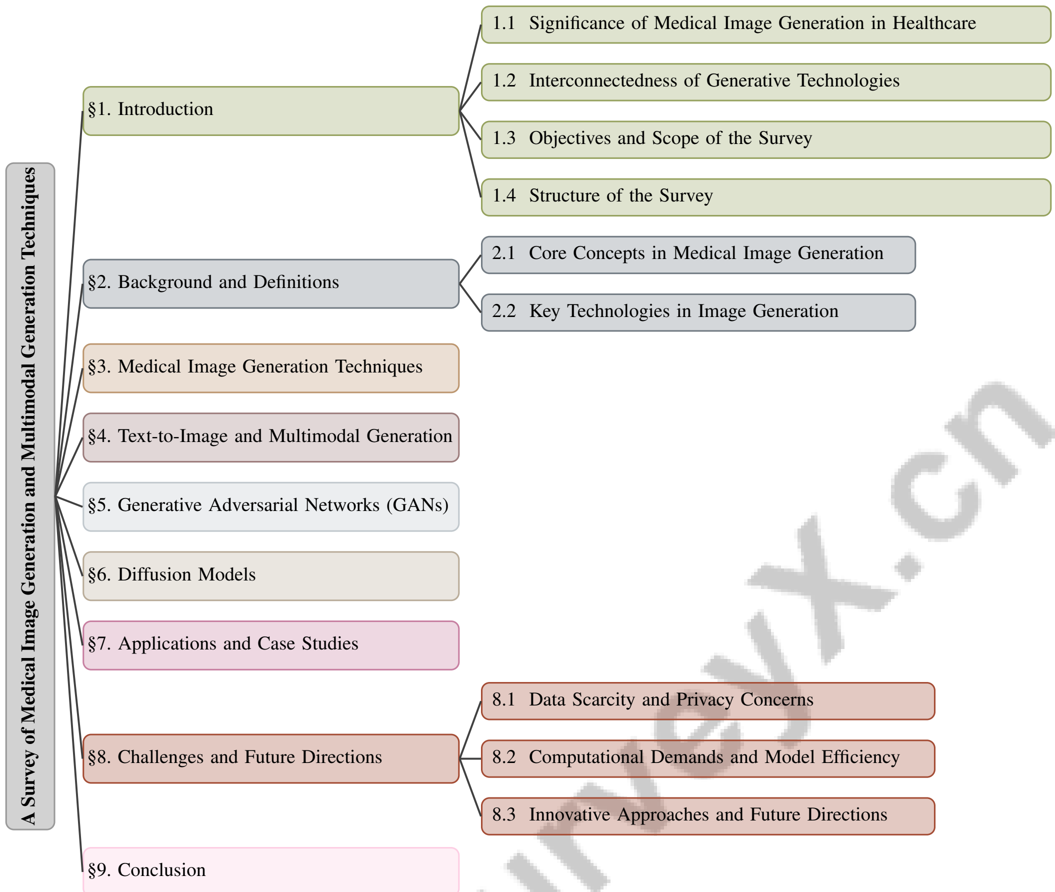

Figure 1: chapter structure

medicine and improved patient care. By leveraging generative AI and deep learning, healthcare practitioners can achieve enhanced diagnostic accuracy and efficiency through improved medical image analysis and real-time data processing. These innovations deepen the understanding of disease progression and treatment effectiveness, empowering clinicians to make informed decisions and significantly improve patient outcomes. Additionally, integrating these technologies with IoT devices facilitates smarter healthcare services and supports telemedicine, addressing computational demands and ethical considerations [3, 4].

1.2 Interconnectedness of Generative Technologies

The interconnectedness of generative technologies is pivotal in the evolution of medical imaging, where diverse methodologies converge to enhance the synthesis and interpretation of medical images. Text-to-image generation (TTI) models have progressed significantly, particularly with large models that convert textual descriptions into detailed visual representations, underscoring the synergy between natural language processing and image synthesis. This integration is essential in medical imaging, facilitating accurate conversion of intricate medical narratives into visual formats, thereby enhancing diagnostics and treatment planning. Advanced techniques such as multi-task learning frameworks and co-attention mechanisms streamline the identification of abnormalities in medical images and the generation of comprehensive reports, addressing radiologists' common challenges, including time constraints and potential errors [5, 6, 7].

Generative adversarial networks (GANs) serve as a foundational technology in this interconnected landscape, demonstrating versatility across various generative tasks. Their ability to learn nonlinear mappings from source to target images enhances the realism and applicability of generated images in clinical settings, facilitating tasks such as modality translation and data augmentation [8]. The Cycle-GAN framework exemplifies the interconnectedness of multimodal imaging techniques, enabling seamless transformation between MRI and CT images, crucial for precise treatment planning and execution [9].

Innovative approaches, such as the PCSA-GAN method, address the limitations of existing multimodal imaging methods by integrating multiple data sources, thereby expanding the scope of generative technologies in medical imaging [10]. Similarly, the PathLDM model illustrates how leveraging textual data from pathology reports can enhance image synthesis, showcasing the potential of integrating diverse data modalities to improve medical imaging outcomes [11].

Recent research emphasizes the necessity for unified models that perform both modality translation and multi-modality generation tasks simultaneously, highlighting the demand for comprehensive solutions to address medical imaging complexities [12]. Additionally, diffusion models are being explored for their potential in multi-modal generative modeling, although current models primarily focus on single data type generation, indicating an area ripe for further development [13]. The benchmark developed for standardized datasets with manual annotations of biasing artifacts also facilitates research on the debiasing effects of GAN-based data augmentation, further illustrating the interconnectedness of these technologies in refining medical imaging processes [14].

As generative technologies continue to evolve, their interconnectedness will play a crucial role in advancing medical imaging, offering new possibilities for diagnosis, treatment, and personalized medicine. By integrating textual descriptions, visual representations, and multi-modal data, technologies such as Surgical Imagen and automated medical imaging report generation are set to revolutionize healthcare delivery. Surgical Imagen utilizes advanced text-to-image generative models to create photorealistic surgical images from action triplet-based textual prompts, addressing challenges in acquiring annotated surgical data. Meanwhile, automated report generation frameworks enhance the accuracy and efficiency of medical imaging interpretations by employing multi-task learning and co-attention mechanisms to generate comprehensive reports that highlight abnormal regions in the images. Together, these innovations promise to enhance diagnostic precision and ultimately improve patient outcomes in clinical settings [6, 15].

1.3 Objectives and Scope of the Survey

This survey provides a comprehensive exploration of advancements and applications of generative models within healthcare, focusing on medical image generation, data analysis, and diagnostic processes. A primary objective is to elucidate the state-of-the-art in adversarial text-to-image synthesis models, addressing challenges such as generating high-resolution images and developing robust evaluation metrics [16]. The scope encompasses various generative techniques, including the use of GANs for stain normalization alongside traditional statistical approaches [17].

The survey extends to applications of deep learning in medical imaging, covering critical areas such as image registration, anatomical structure detection, tissue segmentation, and computer-aided diagnosis [4]. It examines the role of GANs in tasks such as image reconstruction, segmentation, detection, classification, and cross-modality synthesis [18], while highlighting the application of Convolutional Neural Networks (CNNs) and GANs in synthesizing medical images across different modalities, particularly MRI data [19].

Additionally, the survey addresses the challenge of providing diverse and realistic training scenarios in medical education, often constrained by high costs associated with standardized patients and the scarcity of comprehensive imaging datasets. This includes leveraging conditional GANs to create radiographs tailored to specific features, such as fracture status, to enhance medical training [1]. The inclusion of studies on progressively trained GANs for synthesizing high-resolution mammograms further underscores advancements in achieving unprecedented image resolutions [20].

The transformative role of Transformers in medical imaging is explored, encompassing applications in segmentation, detection, classification, reconstruction, synthesis, registration, and clinical report generation [21]. The survey also covers the use of GANs for synthetic data generation in the context of COVID-19 diagnosis, while excluding studies that do not focus on GANs or lack insights into data augmentation [22].

By defining the objectives and scope of this survey, we aim to offer an in-depth exploration of how generative models, particularly GANs, are transforming medical imaging and healthcare. This includes applications in image reconstruction, segmentation, and data augmentation, as well as

AI-generated, for reference only.

integration with advanced technologies like IoT for real-time data analysis and enhanced diagnostic capabilities. Furthermore, we will identify specific areas for future research and development, addressing challenges such as computational demands and ethical considerations associated with these innovative technologies [3, 23, 18, 24].

1.4 Structure of the Survey

This survey is meticulously structured to provide a comprehensive overview of advancements in medical image generation and multimodal generation techniques. The paper begins with an **Introduction** section, establishing the significance of medical image generation in healthcare and exploring the interconnectedness of generative technologies. It outlines the objectives and scope of the survey, concluding with an overview of the paper's organization.

The **Background and Definitions** section follows, offering insights into the core concepts and technologies underpinning medical image generation, defining key terms, and elucidating their roles and applications in healthcare.

In the survey titled **Medical Image Generation Techniques**, a comprehensive examination of various methodologies for generating medical images is presented, emphasizing the utilization of GANs and diffusion models. The survey highlights the growing interest in GANs due to their ability to create realistic medical images and annotations, crucial for applications such as image augmentation, reconstruction, and cross-modality synthesis. It discusses advancements in GAN frameworks, including Deep Convolutional GANs and CycleGANs, and their effectiveness in enhancing medical image analysis. The review also addresses challenges posed by limited labeled datasets in medical imaging and explores how these generative techniques can mitigate issues related to data scarcity, improving diagnostic accuracy and treatment efficacy [25, 18, 26, 27]. Recent advancements and methodologies are highlighted to provide a thorough understanding of the field.

The **Text-to-Image and Multimodal Generation** section examines the process of generating images from textual descriptions and multiple data sources, discussing challenges and solutions in creating realistic and accurate images while emphasizing the integration of text-to-image generation in medical imaging.

A detailed analysis of **Generative Adversarial Networks (GANs)** is provided, discussing their architecture, training processes, strengths, and limitations. This is complemented by a discussion on **Diffusion Models**, which have emerged as powerful tools for image synthesis. The study conducts a comprehensive comparison of their principles and performance against various GAN techniques, emphasizing the distinct advantages and applications of each approach, including improvements in image generation, training stability, and the ability to manipulate visual attributes, as well as the integration of advanced architectures like Capsule Networks [28, 29, 30, 31].

The survey then explores **Applications and Case Studies** of medical image generation techniques in healthcare, providing real-world examples and case studies to illustrate their clinical impact, covering enhancements in diagnostic accuracy and applications in medical imaging segmentation.

Finally, the paper addresses **Challenges and Future Directions** in the field, identifying current challenges such as data scarcity, privacy concerns, and computational demands. The section discusses innovative approaches and potential future directions for research and development.

The survey concludes with a **Conclusion** section, summarizing key findings and insights, and emphasizing the importance of continued research and development in medical image generation.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts in Medical Image Generation

Medical image generation is pivotal in healthcare, focusing on creating high-quality synthetic images to improve diagnostic accuracy and treatment planning. A key challenge is producing high-resolution images from low-resolution inputs, essential for precise medical diagnoses [20]. Techniques like super-resolution (SR) combined with image-to-image translation significantly enhance medical image quality and resolution.

AI-generated, for reference only.

Generative Adversarial Networks (GANs) are central to medical image synthesis, offering robust solutions for image generation and analysis. Traditional GANs often face challenges in generating diverse outputs, especially from text inputs, due to limitations in gradient-based optimization [32]. Conditional GANs (cGANs) transform descriptive medical narratives into detailed visuals but sometimes lack control over generated images, often relying on a single variable [33].

The scarcity of annotated data, particularly for abnormal findings, poses a significant hurdle for training effective machine learning algorithms, as large, labeled datasets are crucial [27]. This challenge is exacerbated by the high costs and time required for labeling extensive datasets, such as those for surgical scene segmentation [34]. Utilizing contextual information from pathology reports has been shown to enhance image generation quality, facilitating more accurate synthesis [11].

Cross-modality image generation, such as translating MRI to CT images, is critical for clinical decision-making, requiring accurate mapping often complicated by misaligned imaging pairs due to patient motion or organ movement [35]. Realistic 3D medical image generation from low-dimensional inputs is vital for training AI models, necessitating diverse and high-quality datasets.

Ensuring patient privacy while generating realistic synthetic data is crucial, especially where sharing real datasets is legally and ethically restricted [36]. This need is underscored by the labor-intensive nature of pixel-level annotations in medical image segmentation, highlighting the limitations of previous methods reliant on such annotations [37]. The synthesis of fluorodeoxyglucose (FDG) PET scans from structural MRI images underscores the importance of cross-modality generation in improving diagnostic accuracy and treatment planning [38].

These core concepts underscore the significance of medical image generation in producing high-quality, diverse, and privacy-preserving datasets that support various diagnostic and therapeutic applications. The automatic generation of medical imaging reports also addresses inefficiencies in reporting processes, benefiting both novice and experienced medical professionals [39]. Additionally, the field of radiomics, which involves extracting quantitative features from medical images, faces challenges such as image noise in low-dose CT images that generative models aim to mitigate, enhancing medical image analysis [40]. Data scarcity, particularly in contexts like COVID-19, hampers the training of AI models, such as Convolutional Neural Networks (CNNs), for effective diagnosis, highlighting the need for innovative generative solutions [41]. Furthermore, generating medical images that accurately reflect ordinal relationships among different severity levels is essential for effective diagnosis and treatment [42].

2.2 Key Technologies in Image Generation

Generative Adversarial Networks (GANs) and diffusion models are at the forefront of medical image generation technologies, each offering unique advantages in creating high-quality, realistic images. GANs, with their dual network architecture comprising generator and discriminator components, are crucial in addressing data scarcity and bias in medical imaging applications [41]. They have been widely used to synthesize diverse datasets, enhancing the training and performance of machine learning models in clinical diagnostics.

The versatility of GANs is demonstrated by their capacity to generate images under specific conditions, such as synthesizing pelvic radiographs conditioned on fracture status [1]. However, challenges like limited output diversity persist, especially in conditional GANs (cGANs), which often prioritize high-dimensional structured contexts over input noise vectors [43]. To address these limitations, innovative approaches integrating convolutional neural networks (CNNs) with GANs have been proposed, aiming to enhance data diversity and improve the robustness of generated outputs [44].

Diffusion models have emerged as a robust alternative to GANs, utilizing a unique denoising process to generate synthetic samples from real medical images [45]. These models employ a forward diffusion process that introduces noise into the data and a reverse process that reconstructs the original data distribution, facilitating the creation of high-resolution images. Latent Diffusion Models (LDMs) exemplify the potential of diffusion models to preserve patient privacy while generating synthetic medical images, offering significant advantages in sensitive healthcare applications [46].

The integration of diffusion models in medical imaging is further enhanced by techniques like Competence-based Multimodal Curriculum Learning (CMCL), which addresses data bias by progressively training models on instances of varying difficulty [5]. This approach improves the adaptability and robustness of diffusion models across diverse medical imaging tasks. Additionally, the application of contrastive learning methods, such as CISP, to create joint embedding spaces for 2D images and 3D shapes enhances accuracy and guidance in generating 3D shapes from images [47].

Advancements in GANs and diffusion models have significantly propelled the field of medical image generation, offering novel solutions to improve image quality, resolution, and applicability across various clinical scenarios. These technologies continue to evolve, presenting new opportunities to enhance diagnostic accuracy and treatment planning within healthcare [21].

3 Medical Image Generation Techniques

3.1 Overview of Medical Image Generation Techniques

Method Name	Technique Types	Application Scenarios	Key Innovations
MEGAN[48]	Generative Adversarial Networks	Data Augmentation	Mixture OF Experts
cGAN[1]	Conditional Gans	Training Tools	Conditional Gans
ODM[42]	Diffusion Models	Data Augmentation	Ordinal Diffusion Model
GIN[49]	Generative Invertible Networks	Virtual Patient Creation	Bidirectional Feature Mapping
PGGAN[50]	Progressive Gans	Data Augmentation	High-resolution Synthesis
CISP[47]	Diffusion Models	Shape Generation	Contrastive Image-Shape

Table 1: Comparative analysis of various medical image generation methods, highlighting their technique types, application scenarios, and key innovations. The table provides a detailed overview of the capabilities of different generative models, including Generative Adversarial Networks and diffusion models, in enhancing medical imaging through data augmentation, virtual patient creation, and high-resolution synthesis.

Medical image generation techniques significantly enhance healthcare by producing high-quality images pivotal for diagnostics and therapy. Generative Adversarial Networks (GANs) and diffusion models are prominent in this field due to their efficacy in image synthesis. GANs, with their dual network structure of generator and discriminator, address data scarcity and bias, facilitating data augmentation which improves Convolutional Neural Networks (CNNs) training and classification accuracy [48]. The Mixture of Experts GAN (MEGAN) showcases this flexibility by using multiple specialized generators to enhance image quality.

Conditional GANs (cGANs) extend traditional GAN capabilities by embedding specific conditions into the generation process, offering enhanced output control [43]. These models are particularly useful in applications like super-resolution, improving the resolution of lower-quality medical images [41]. GANs also serve as a crucial source of training images in radiology, highlighting their transformative role in medical education [1].

Diffusion models provide a compelling alternative to GANs, employing a denoising process to reconstruct high-fidelity images from noisy inputs. These models use a forward diffusion process that introduces noise, followed by a reverse process that reconstructs the original data distribution, generating high-resolution images. The Ordinal Diffusion Model (ODM) enhances image quality by incorporating ordinal relationships among classes into the noise estimation process [42].

Innovative frameworks like Generative Invertible Networks (GIN) create virtual patients by mapping between high-dimensional feature space and low-dimensional representation, preserving critical pathophysiological information [49]. Progressive Growing of GANs (PGGANs) synthesize realistic brain MR images suitable for CNN-based tumor detection [50].

Text-to-image generation methods further demonstrate versatility in medical applications. Models like DALL-E 2 synthesize realistic X-ray images [51], while the CISP method aligns 2D images with 3D shapes in a shared embedding space, enhancing the generative process through 3D characteristics [47].

Advancements in medical image generation techniques, including GANs, diffusion models, and innovative frameworks, significantly enhance diagnostic accuracy and personalized medicine. These technologies are poised to revolutionize healthcare delivery by enabling the automated generation of high-quality medical images and reports, thus improving diagnostic accuracy and expediting clinical decision-making. By integrating deep learning and generative models, these innovations streamline the interpretation of complex medical data and assist healthcare professionals in identifying

AI-generated, for reference only.

abnormalities more effectively, ultimately leading to improved patient outcomes through timely interventions [4, 3, 6, 15, 5]. Table 1 offers a comprehensive comparison of medical image generation techniques, elucidating the diverse methodologies and innovations that contribute to advancements in medical imaging.

As illustrated in Figure 2, the hierarchical categorization of medical image generation techniques highlights primary methods such as Generative Adversarial Networks (GANs), Diffusion Models, and Innovative Frameworks. Each category is further detailed with specific applications or models, showcasing their roles in enhancing image quality, resolution, and training data augmentation in medical imaging. The first example in the figure illustrates the use of GANs in the Brain Tumor Segmentation (BRATS) dataset for tasks including image-to-brain segmentation, label-to-image synthesis, and image-to-tumor segmentation. This sophisticated training process, utilizing both a generator and a discriminator, enables the creation of highly detailed and accurate medical images. The second example presents a comparative analysis of MRI images from various datasets, such as Duke, MRNet, LIDC-IDRI, and ADNI, revealing the diversity in image characteristics across different datasets, including variations in contrast and detail. Collectively, these examples provide insights into the capabilities and applications of medical image generation techniques, underscoring their role in advancing medical imaging and improving patient outcomes [52, 53].

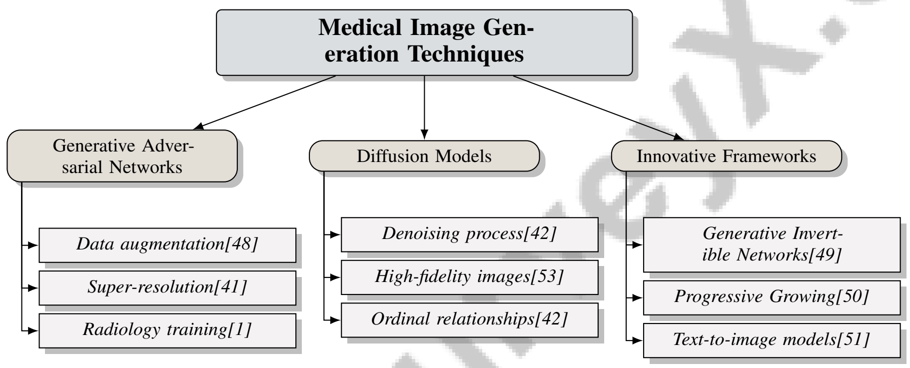

Figure 2: This figure illustrates the hierarchical categorization of medical image generation techniques, highlighting the primary methods such as Generative Adversarial Networks (GANs), Diffusion Models, and Innovative Frameworks. Each category is further detailed with specific applications or models, showcasing their roles in enhancing image quality, resolution, and training data augmentation in medical imaging.

3.2 Advanced Generative Models

Method Name	Model Types	Innovative Techniques	Clinical Applications
DDPM[54]	Ddpms	Fast Inference	Anomaly Detection
MD[55]	Diffusion Models	Class-conditional Sampling	Synthetic Ehr Generation
SegGuidedDiff[56]	Diffusion Models	Segmentation Conditioning	Rare Cases
MSGANs[57]	Conditional Gans	Mode Seeking Regularization	-
GA[25]	Gan-GA	Genetic Algorithm	Medical Image Synthesis
MEGAN[48]	Generative Adversarial Networks	Gating Networks	-

Table 2: Overview of advanced generative models in medical imaging, highlighting their respective model types, innovative techniques, and clinical applications. The table encapsulates the diversity of methodologies, from diffusion models to GAN-based architectures, and their impact on anomaly detection, synthetic data generation, and image synthesis.

Advanced generative models have transformed medical imaging by introducing innovative methodologies that enhance the synthesis of high-quality images. Diffusion models have emerged as a formidable alternative to traditional GAN-based approaches, yielding superior image fidelity and semantic alignment with textual descriptions [58]. Denoising Diffusion Probabilistic Models (DDPMs) are effective in anomaly detection and segmentation in brain imaging, leveraging training on healthy data to identify anomalies in the latent space [54].

The MedDiff model exemplifies advancements in diffusion processes by incorporating class-conditional sampling and an accelerated inference strategy to generate realistic synthetic electronic health records (EHRs), enhancing clinical data representation [55]. Similarly, the SegGuidedDiff model introduces a novel approach by conditioning the generation process on anatomical segmentation masks at each step, improving the anatomical accuracy of generated medical images [56].

Innovations in GAN architectures continue to expand the possibilities of medical image generation. Mode-seeking GANs introduce regularization terms that promote diverse output generation, addressing a critical limitation of traditional GANs [57]. The integration of genetic algorithms within the InfoGAN architecture, exemplified by the GANGA model, enhances image fidelity and diversity while accelerating the training process [25].

The MEGAN framework utilizes a gating network that dynamically selects the appropriate generator based on input conditions, enhancing performance and adaptability compared to conventional GANs [48]. This approach highlights the potential of ensemble methods in improving the quality and specificity of generated medical images.

Table 2 provides a comprehensive summary of the advanced generative models discussed, illustrating their contributions to medical imaging through various innovative techniques and clinical applications. These advancements in generative models are crucial for driving progress in medical imaging, offering novel solutions for enhancing image quality, resolution, and applicability across various clinical scenarios. As these technologies continue to evolve, they hold the potential to significantly improve healthcare delivery by facilitating high-quality medical image generation through generative AI models. This evolution not only enhances diagnostic speed and accuracy but also supports the automatic creation of detailed imaging reports, leading to more informed clinical decisions and better patient outcomes. Furthermore, the integration of these technologies with Internet of Things (IoT) systems enables real-time data analysis and predictive capabilities, enriching telemedicine and overall healthcare services while addressing challenges such as computational demands and ethical considerations [3, 59, 6, 15].

4 Text-to-Image and Multimodal Generation

4.1 Integration of Text-to-Image Generation in Medical Imaging

The integration of text-to-image generation in medical imaging represents a significant technological advancement, enhancing diagnostic processes and clinical outcomes through the use of generative models. Diffusion models, such as VQ-Diffusion, effectively generate high-quality images from textual descriptions using mask-and-replace strategies to manage contextual information [60]. The BiomedJourney framework exemplifies this by utilizing multimodal patient journey data to create counterfactual biomedical images, crucial for planning treatments [61]. The RPG framework improves generation accuracy by decomposing complex tasks into subtasks [62].

Approaches like TCIG enhance image control and quality using pre-trained segmentation models, providing actionable visual data for clinicians [63]. UNIMO-G processes interleaved text and visual inputs, showcasing multimodal approaches' potential in medical imaging [64]. These methods improve interpretability and diagnostic utility while protecting patient privacy by minimizing data replication risks [46, 7].

Techniques such as Multi-Attributed and Structured Text-to-Face Synthesis (MAST) demonstrate that structured textual descriptions can enhance GAN performance, suggesting similar benefits in medical contexts [65]. Overall, the integration of text-to-image generation offers substantial benefits, including improved image quality, diagnostic accuracy, and novel data generation capabilities, ultimately supporting personalized healthcare and clinical decision-making [66].

4.2 Multimodal Generation and Modality Translation

Multimodal generation and modality translation are pivotal in advancing medical imaging, synthesizing and interpreting complex data by leveraging multiple modalities' complementary strengths [67]. The MaxFusion framework exemplifies this by efficiently synthesizing comprehensive medical images without extensive training data, enhancing image quality and diversity [68]. Diffusion models like MT-Diffusion facilitate simultaneous generation of multiple data types, improving efficiency and image quality through structured textual descriptions [13, 65].

The WordStylist framework further integrates text-to-image generation in multimodal contexts, refining images through a diffusion process conditioned on text content and writer styles [69]. This integration is beneficial in medical imaging, where translating complex medical narratives into visual formats is crucial for diagnosis and treatment planning. These advancements enable the synthesis of high-quality, realistic images, supporting informed clinical decisions and enhancing diagnostic accuracy and treatment effectiveness [59, 6].

4.3 Challenges in Generating Complex and High-Resolution Images

Generating complex and high-resolution images in medical imaging presents significant challenges, primarily due to the extensive computational resources and time required for training diffusion models [70]. Methods like UFOGen attempt to improve generation speed without compromising quality, yet resource demands remain a barrier [71]. Fidelity and diversity are often compromised by content-level artifacts, and unidirectional bias in predicting image tokens can lead to accumulated errors [72, 60].

In text-to-image generation, DTC-GAN addresses challenges in reflecting multiple objects and spatial relationships from text, crucial for generating complex images [73]. However, controllability is often limited by computational resources and data availability [63]. Poorly initialized images can hinder refinement, complicating high-quality output achievement [74], while maintaining coherence across video frames remains challenging [75].

Frameworks like UNIMO-G face challenges such as context synthesis inaccuracies and ethical concerns related to misuse, particularly in deepfake generation [64]. Addressing these challenges is crucial for improving the quality and applicability of complex and high-resolution medical image generation, ultimately enhancing diagnostic accuracy and patient care.

Figure 3 illustrates the primary challenges in generating complex and high-resolution images, categorized into computational constraints, quality issues, and controllability challenges. Each category highlights specific methods and their associated challenges, reflecting the intricate landscape of medical image generation.

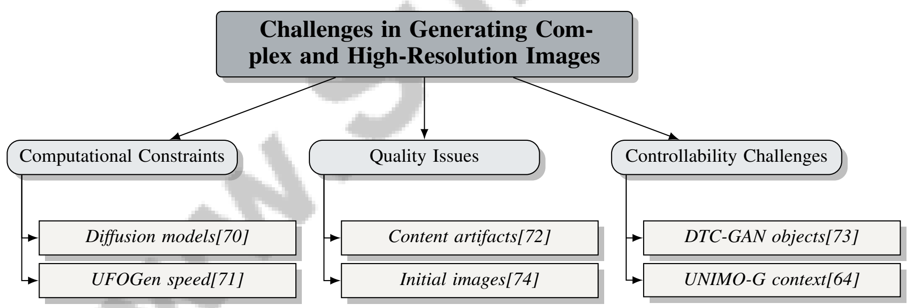

Figure 3: This figure illustrates the primary challenges in generating complex and high-resolution images, categorized into computational constraints, quality issues, and controllability challenges. Each category highlights specific methods and their associated challenges, reflecting the intricate landscape of medical image generation.

5 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) have significantly impacted medical imaging by providing sophisticated solutions for image synthesis and analysis. This section delves into the various GAN architectures and their adaptations to address specific challenges in medical imaging, highlighting their contributions and implications for clinical practice. Understanding these architectures is crucial for appreciating advancements in image generation capabilities.

5.1 Architecture and Variations of GANs

GANs have advanced medical imaging by offering diverse architectures tailored for specific image synthesis and translation tasks. The basic GAN structure, consisting of a generator and a discriminator in an adversarial setup, plays a pivotal role in generating high-quality medical images, mitigating data scarcity, and enhancing diagnostic accuracy [22]. The evolution of GANs has led to numerous network architectures, loss functions, and optimization strategies that enhance performance and adaptability.

Conditional GANs (cGANs) augment the basic GAN framework by incorporating conditioning variables, enabling controlled image generation based on specific inputs like disease presence or anatomical features [76]. ScatGAN exemplifies this by converting B-mode ultrasound images into scatter maps [77].

Deep Convolutional GANs (DCGANs) employ convolutional layers to improve image quality and resolution, making them suitable for high-resolution imaging tasks [2]. Progressive Growing of GANs (PGGANs) incrementally generate high-resolution images by adding layers to the generator and discriminator, enhancing detailed medical images [50]. PGSgan further refines synthesis quality by integrating a sketch generative adversarial network into a progressive growing scheme, improving structural details [78].

CycleGAN has emerged as a significant advancement in modality translation, enabling image transformation from one modality to another, such as MRI to CT, without paired training data [76]. CSGAN improves image transformation quality with a Cyclic-Synthesized Loss, bridging the gap between synthesized and cycled images.

Generative Invertible Networks (GIN) incorporate a GAN for generation, a CNN for reverse mapping, and a feature selection mechanism, enhancing GANs' versatility in medical imaging [49]. MEGAN uses an ensemble of specialized generators focusing on distinct modality subsets, improving performance and adaptability compared to conventional GANs [48].

These diverse architectures and variations of GANs provide robust solutions for a wide range of medical imaging challenges, including image synthesis, modality translation, feature extraction, and unsupervised learning. These advancements significantly improve diagnostic accuracy and enable personalized treatment strategies by enhancing abnormality identification and clinical workflow efficiency [4, 59, 6, 5].

5.2 Training Process and Challenges

The GAN training process involves an adversarial interaction between the generator, which creates realistic data samples, and the discriminator, which evaluates their authenticity. This framework iteratively enhances the generator's ability to produce data indistinguishable from real samples, but it presents several inherent challenges [79].

A primary challenge is instability arising from adversarial dynamics between the generator and discriminator, often leading to oscillations during training and complicating convergence to a stable solution. The complex, non-convex optimization landscape further exacerbates these issues, necessitating careful hyperparameter tuning and often resulting in prolonged training periods [80].

Mode collapse, where the generator produces limited and repetitive outputs, is another critical issue linked to imbalanced training between the generator and discriminator. Techniques such as multi-discriminator approaches have been proposed to stabilize training and mitigate mode collapse, ensuring balanced training dynamics and enhancing the diversity of generated samples [79].

Evaluating GANs poses additional challenges due to the lack of standardized metrics for assessing the quality and diversity of generated images. Visualization techniques are crucial for understanding internal representations and diagnosing issues like mode collapse and instability, aiding in refining the training process and improving output robustness [80].

Recent advancements, including convolutional and recurrent blocks for sequence generation, have shown promise in improving training efficiency and output quality. These innovations, alongside multi-discriminator systems, contribute to reducing artifacts and enhancing the realism of generated images, offering promising alternatives to traditional GAN training methods [79].

5.3 Strengths and Limitations of GANs

GANs have become integral to medical imaging, capable of generating high-quality, realistic images that enhance diagnostic precision and treatment strategies. A notable strength is their ability to create diverse datasets, particularly advantageous in medical imaging, where annotated data is often limited due to privacy concerns and high collection costs. The Mixture of Experts GAN (MEGAN) framework exemplifies this strength by enhancing image diversity and quality through specialization, despite challenges related to rapid convergence of gating networks [48].

The PGSgan framework minimizes artifacts and produces high-fidelity images with editable features, increasing the practical utility of generated images in clinical environments [78]. Furthermore, GANs have shown promise in augmenting data for AI models aimed at COVID-19 diagnosis, although issues with clinical validation and generalization persist [41]. The PLAN model illustrates GANs' capability to generate privacy-preserving synthetic samples, maintaining performance levels comparable to real data while safeguarding privacy [81].

Despite these strengths, GANs face limitations affecting their broader application in medical imaging. A primary challenge is instability during training, often leading to mode collapse and limiting the diversity of generated outputs [79]. Traditional GANs also struggle with ensuring one-to-one mappings between image domains, resulting in biased translations and reduced fidelity in cross-modality applications.

Evaluating GAN performance remains a significant hurdle, as existing metrics may not fully capture the visual quality and clinical relevance of generated images. Limitations observed in models like DALL-E 2 highlight these challenges, as it produces realistic X-ray images but struggles with pathological images and cross-sectional modalities [51]. Additionally, GANs typically require substantial computational resources and large datasets for effective training, posing a barrier in medical imaging contexts where data is scarce.

6 Diffusion Models

6.1 Emergence and Advantages of Diffusion Models

Diffusion models have revolutionized image synthesis in medical imaging, offering a robust alternative to generative adversarial networks (GANs) by progressively denoising data to produce high-fidelity images [45]. Their iterative refinement process preserves intricate details crucial for medical applications [82]. These models enhance stability and diversity in generated datasets, with innovations like the Wavelet Diffusion GAN improving inference times and image quality through wavelet transforms [82]. Furthermore, the Vector Quantized Diffusion Model (VQ-Diffusion) enhances text-to-image generation quality using a vector quantized variational autoencoder [60].

Diffusion models' versatility is highlighted by their ability to synthesize multi-modal data. The Segmentation-Guided Diffusion (SegGuidedDiff) method, for instance, generates images conditioned on anatomical segmentation masks, allowing precise anatomical control [56]. The Ordinal Diffusion Model (ODM) introduces a novel loss function to maintain ordinal relationships, advancing standard diffusion techniques [42]. Innovative benchmarks have been established to evaluate these models, focusing on unique artifacts and detection methods [83]. Tutorials on diffusion models aim to bridge knowledge gaps in generative tasks, including text-to-image and text-to-video generation [66].

The advent of diffusion models marks a significant advancement in image synthesis, improving stability, diversity, and efficiency. These developments in deep learning models continue to enhance medical imaging, addressing challenges like accurate interpretation of complex data and driving AI advancements [4, 59, 6].

6.2 Working Principles of Diffusion Models

Diffusion models are powerful tools in image synthesis, particularly in medical imaging, due to their ability to produce high-quality, realistic images. The core of these models is the denoising process, which iteratively refines noisy data to reconstruct the original image distribution. This involves a forward diffusion phase, where noise is added, and a reverse phase, where noise is removed to yield a clear image [45]. Their effectiveness in medical image generation is enhanced by low-dimensional latent spaces, ensuring computational efficiency and robust control over the generation process [84]. Models like SegGuidedDiff integrate multi-class anatomical segmentation masks at each denoising step, ensuring anatomical constraints are met [56].

Wavelet transforms incorporated into the diffusion process, as shown by the Wavelet Conditional Diffusion GAN (WCDGAN), enhance image super-resolution by reducing training and inference times while improving quality [82]. The nested diffusion process further refines image generation by employing outer and inner diffusion processes, enhancing intermediate predictions and overall quality [85]. Beyond static images, diffusion models have been adapted for video generation, ensuring temporal coherence and high-quality output [75]. These working principles, characterized by iterative denoising and advanced techniques, establish diffusion models as formidable tools in medical image generation, significantly enhancing diagnostic accuracy and treatment planning [66].

6.3 Comparison with Generative Adversarial Networks (GANs)

Comparing Generative Adversarial Networks (GANs) and diffusion models reveals distinct advantages and applications for each in medical image generation. GANs, known for their adversarial training process, produce highly realistic images that closely mimic real medical data [86]. This makes them effective where authenticity is critical. Conversely, diffusion models achieve high-fidelity image synthesis through iterative refinement, as exemplified by the Vector Quantized Diffusion Model (VQ-Diffusion), which addresses unidirectional bias and error accumulation in text-to-image generation [60]. This approach allows diffusion models to produce images with superior quality and detail, suitable for applications demanding high-resolution outputs.

A notable advantage of diffusion models is their ability to generate high-quality intermediate images, as demonstrated by the Nested Diffusion Process, which allows for anytime generation, enhancing flexibility and image quality at various stages [85]. This is particularly beneficial in medical imaging, where iterative refinement and precision are paramount. Tutorials on diffusion models emphasize their efficacy in generating high-fidelity images through iterative refinement, positioning them as formidable competitors to GANs [66]. While GANs excel in generating realistic images via adversarial training, diffusion models offer enhanced stability and detail through structured denoising processes.

In recent years, advancements in medical imaging have significantly transformed diagnostic practices and patient care. A comprehensive understanding of these developments requires an exploration of the hierarchical structure of applications and case studies within this field. Figure 4 illustrates this structure, highlighting key enhancements in diagnostic accuracy and the diverse applications of medical imaging segmentation. The figure also emphasizes the real-world clinical impacts of these advancements, showcasing the pivotal role of generative models, such as Generative Adversarial Networks (GANs) and diffusion models, in improving both diagnostic precision and segmentation accuracy. Moreover, these innovations contribute to medical education, ultimately enhancing healthcare delivery and patient outcomes.

7 Applications and Case Studies

7.1 Enhancements in Diagnostic Accuracy

Generative models, notably GANs and diffusion models, have markedly improved diagnostic accuracy by generating synthetic images that enhance clinical tasks. GANs, for instance, have been instrumental in producing synthetic MRI images, which augment datasets and significantly improve the segmentation performance of models like U-Net [87]. The integration of synthetic data into medical imaging workflows not only bolsters model robustness but also mitigates data scarcity and privacy concerns, offering a viable solution for data augmentation and anonymization [88]. This is crucial in environments with limited annotated datasets and stringent privacy regulations.

Generative models address mode collapse issues in conditional GANs (cGANs) through mode-seeking regularization, thereby enhancing image diversity without sacrificing quality, which is vital for diagnostic applications [57]. The automation of image analysis, supported by synthetic data, has advanced diagnostic accuracy by reducing medical professionals' workload and improving patient outcomes through more precise diagnostic tools [44]. The ability of generative models to

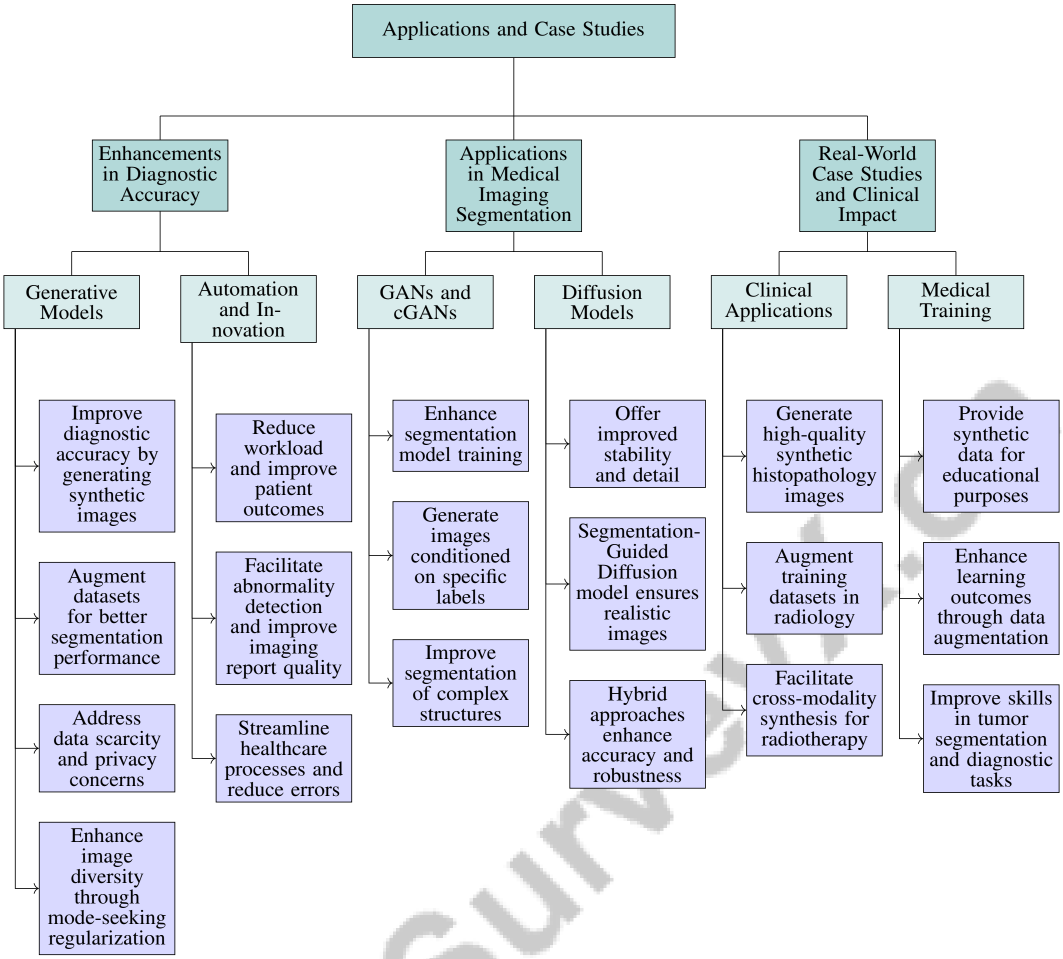

Figure 4: This figure illustrates the hierarchical structure of applications and case studies in medical imaging, highlighting enhancements in diagnostic accuracy, applications in medical imaging segmentation, and real-world clinical impacts. It showcases the role of generative models like GANs and diffusion models in improving diagnostic precision, segmentation accuracy, and medical education, ultimately enhancing healthcare delivery and patient outcomes.

produce comprehensive and accurate medical image representations highlights their potential to refine diagnostic precision and clinical decision-making.

Recent innovations in medical image generation, particularly through GANs and automatic report generation frameworks, are transforming diagnostic accuracy. These advancements facilitate efficient abnormality detection, improve imaging report quality, and enable better integration of complex data, ultimately enhancing healthcare delivery and patient care by streamlining processes and reducing errors associated with manual interpretations [18, 6].

7.2 Applications in Medical Imaging Segmentation

Image generation techniques have revolutionized medical imaging segmentation, improving the delineation of anatomical structures with precision and efficiency. GANs provide a robust framework for generating synthetic images that enhance segmentation model training, particularly by augmenting limited datasets [2]. Conditional GANs (cGANs) extend these capabilities by generating images conditioned on specific labels, facilitating more accurate segmentation of complex structures, as seen in MRI tumor segmentation tasks [76, 50].

Diffusion models have emerged as powerful tools in segmentation, offering improved stability and detail in generated images. The Segmentation-Guided Diffusion (SegGuidedDiff) model, for example, conditions the generation process on anatomical segmentation masks, ensuring realistic and clinically relevant synthetic images [56]. This is crucial for tasks like organ segmentation in CT scans, where precision is essential.

Hybrid approaches combining GANs' adversarial training with diffusion models' denoising capabilities have further advanced segmentation accuracy and robustness [45]. The application of these techniques significantly improves the ability to delineate anatomical structures, supporting informed clinical decisions and enhancing patient outcomes. As deep learning and AI continue to reshape medical imaging, these technologies promise to enhance diagnostic accuracy and efficiency, paving the way for personalized medicine and precision healthcare [4, 59, 6].

7.3 Real-World Case Studies and Clinical Impact

Advanced generative models have demonstrated significant potential in real-world clinical settings, enhancing diagnostic and therapeutic processes. Diffusion probabilistic models, for instance, generate high-quality synthetic histopathology images that outperform traditional GANs, providing detailed images crucial for accurate pathological analysis [89]. This advancement addresses data scarcity and improves diagnostic algorithm training, enhancing clinical decision-making.

In radiology, GAN-generated synthetic data augments training datasets, particularly in modalities where acquiring large annotated datasets is challenging. For example, synthetic mammography images generated by GANs enhance breast cancer detection models, improving diagnostic accuracy and reducing false positives [2]. This underscores the clinical impact of generative models in providing reliable early cancer detection tools, enhancing patient outcomes.

The generation of synthetic CT and MRI images facilitates cross-modality synthesis, crucial for radiotherapy treatment planning. These models provide accurate synthetic images, enabling precise dose calculations and treatment simulations, thereby optimizing therapeutic interventions and minimizing adverse effects [50].

Generative models also enhance medical training by providing synthetic data that serves as a valuable resource for students and trainees. This approach mitigates data scarcity challenges and enhances learning outcomes through data augmentation and anonymization. Synthetic abnormal MRI images, for instance, allow students to practice realistic scenarios, improving skills in tumor segmentation and other diagnostic tasks [88, 5, 52, 24].

The real-world applications of generative models in medical imaging demonstrate their transformative potential in improving diagnostic accuracy, optimizing treatment strategies, and enhancing medical education. As these technologies evolve, they promise to enhance healthcare delivery by leveraging advanced AI, enabling faster and more accurate diagnoses, facilitating real-time data analysis, and supporting medical report automation. These innovations pave the way for personalized medicine and improved patient care, addressing critical healthcare challenges like cost reduction and efficiency enhancement [3, 59, 6, 4].

8 Challenges and Future Directions

8.1 Data Scarcity and Privacy Concerns

Data scarcity and privacy concerns present significant obstacles in the realm of medical image generation, impacting the advancement of generative models. The availability of large, annotated datasets is crucial for training deep learning models to interpret complex medical imaging data, yet such datasets are scarce due to the intricate nature of medical images and the expertise required for accurate annotation [44]. Generative models like GANs and diffusion models offer solutions by creating synthetic datasets to augment limited real-world data, circumventing logistical and legal constraints [1]. However, the quality of synthetic data often relies on expert evaluations, introducing subjectivity and affecting consistency [2].

Privacy concerns further complicate the use of real patient data for model training due to ethical and legal issues. Balancing patient privacy with the utility of generated datasets requires careful strategies, such as differential privacy, to maintain data quality while addressing ethical concerns [1]. The generalizability of generative models is limited by insufficient validation and dataset transparency, restricting their applicability in diverse clinical contexts [41]. Additionally, the computational demands of training generative models, which sometimes produce low-quality outputs, pose practical challenges in clinical settings [44].

Addressing these challenges is crucial for the progression of medical image generation. Developing resource-efficient and ethically responsible approaches can overcome data scarcity and privacy issues, enhancing healthcare solutions. Advanced generative techniques that mitigate data scarcity and improve image quality highlight the transformative potential of these models in medical image analysis [1].

8.2 Computational Demands and Model Efficiency

The computational demands and efficiency of generative models are critical barriers to their widespread clinical adoption. While powerful, GANs and diffusion models often require substantial computational resources, limiting their practical use, particularly in resource-constrained environments [32]. These models involve complex computations and extensive training data, contributing to significant computational overhead. Diffusion models, for instance, require numerous steps in the diffusion process, extending the time needed to generate high-quality images and limiting their scalability [82]. Traditional diffusion models often produce intermediate images that fail to align with the learned image manifold, complicating their integration into clinical workflows [85].

Memory-efficient GAN-based methods have been explored to reduce the memory footprint in processing large images, particularly in high-resolution 3D medical imaging [32]. These approaches optimize architecture and training processes to enhance computational efficiency without compromising image quality. However, optimizing multiple loss functions and fine-tuning dual generators in advanced GAN frameworks remain computationally challenging.

Innovations in model architecture and optimization strategies are essential to address these computational demands. Integrating wavelet transforms in diffusion models has shown promise in reducing training and inference times, improving overall efficiency [82]. Exploring alternative optimization methods, despite potentially increasing computational resource requirements, could enhance model efficiency and performance [32].

8.3 Innovative Approaches and Future Directions

Innovative approaches in medical image generation aim to address existing challenges and expand the capabilities of generative models, particularly through advancements in GANs and diffusion models. Future research will focus on enhancing the explainability of GAN methods, increasing dataset sizes, and ensuring clinical validation to facilitate real-world applications [41]. Developing robust theoretical frameworks to elucidate GAN behaviors, along with exploring novel architectures and loss functions, remains a priority [43].

In diffusion models, future research will aim to improve ordinal relationship loss for stricter adherence to class differences, exploring applications of generated images in medical analysis, particularly for data augmentation [42]. Efforts to enhance image quality, incorporate image-level class guidance, and extend the model's capabilities for segmentation-guided image translation are anticipated [56].

Refining methodologies for privacy-preserving generative models, including developing hybrid metrics that incorporate memorization assessments, is crucial for enhancing privacy and augmentation properties. Optimizing load-balancing regularization and extending the Mixture of Experts GAN (MEGAN) to other generative models are promising future research directions [48].

Addressing data scarcity remains a significant focus, with developing unsupervised and semi-supervised learning techniques critical for improving model interpretability and gaining trust from healthcare professionals [44]. Expanding benchmarks to include additional modalities and enhancing the robustness of evaluation metrics will be essential for advancing the quality and applicability of generated images [2].

Future research could also improve the handling of imperfect segmentation masks to enhance the quality of generated images, particularly relevant for segmentation-guided image translation [63]. Additionally, developing new metrics for evaluating generated faces and experimenting with different GAN architectures could further improve text-to-face synthesis [65].

9 Conclusion

Generative models hold immense promise for revolutionizing healthcare by enhancing diagnostic precision, elevating the quality of medical services, and enabling efficient real-time data analysis. Techniques such as Generative Adversarial Networks (GANs) and diffusion models have advanced the field of medical image generation, tackling issues of data scarcity and privacy. The creation of synthetic images, such as synthetic CMR images, provides effective solutions to the constraints of annotated medical datasets. Innovations like the Diffusion Deformable Model (DDM) illustrate the potential of diffusion models in clinical settings, particularly for tracking anatomical changes. Furthermore, the transformation of ultrasound images into mammogram-like images through sophisticated pipelines underscores the capability of generative models to improve diagnostic processes, notably in breast cancer detection.

Despite these advancements, challenges persist, including the computational intensity of generative models and the need for comprehensive evaluation metrics. The role of Transformers in various applications highlights their potential to address existing limitations and enhance model efficacy. Continued research and development are essential to surmount these obstacles, enhance the efficiency of generative models, and expand their use across diverse clinical environments.

References

[1] Samuel G. Finlayson, Hyunkwang Lee, Isaac S. Kohane, and Luke Oakden-Rayner. Towards generative adversarial networks as a new paradigm for radiology education, 2018.

[2] Basel Alyafi, Oliver Diaz, Joan C Vilanova, Javier del Riego, and Robert Marti. Quality analysis of dcgan-generated mammography lesions, 2020.

[3] Yuan Sun and Jorge Ortiz. Rapid review of generative ai in smart medical applications, 2024.

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

[5] Fenglin Liu, Shen Ge, Yuexian Zou, and Xian Wu. Competence-based multimodal curriculum learning for medical report generation. *arXiv preprint arXiv:2206.14579*, 2022.

[6] Baoyu Jing, Pengtao Xie, and Eric Xing. On the automatic generation of medical imaging reports. *arXiv preprint arXiv:1711.08195*, 2017.

[7] Bram de Wilde, Anindo Saha, Maarten de Rooij, Henkjan Huisman, and Geert Litjens. Medical diffusion on a budget: Textual inversion for medical image generation, 2024.

[8] Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12):2720–2730, 2018.

[9] Leonardo Crespi, Samuele Camnasio, Damiano Dei, Nicola Lambri, Pietro Mancosu, Marta Scorsetti, and Daniele Loiacono. Leveraging multimodal cyclegan for the generation of anatomically accurate synthetic ct scans from mris, 2024.

[10] Fuyou Mao, Lixin Lin, Ming Jiang, Dong Dai, Chao Yang, Hao Zhang, and Yan Tang. Cross-modal medical image generation based on pyramid convolutional attention network, 2024.

[11] Srikar Yellapragada, Alexandros Graikos, Prateek Prasanna, Tahsin Kurc, Joel Saltz, and Dimitris Samaras. Pathldm: Text conditioned latent diffusion model for histopathology, 2023.

[12] Minghui Hu, Chuanxia Zheng, Heliang Zheng, Tat-Jen Cham, Chaoyue Wang, Zuopeng Yang, Dacheng Tao, and Ponnuthurai N. Suganthan. Unified discrete diffusion for simultaneous vision-language generation, 2022.

[13] Changyou Chen, Han Ding, Bunyamin Sisman, Yi Xu, Ouye Xie, Benjamin Z. Yao, Son Dinh Tran, and Belinda Zeng. Diffusion models for multi-modal generative modeling, 2024.

[14] Agnieszka Mikołajczyk, Sylwia Majchrowska, and Sandra Carrasco Limeros. The (de)biasing effect of gan-based augmentation methods on skin lesion images, 2022.

[15] Chinedu Innocent Nwoye, Rupak Bose, Kareem Elgohary, Lorenzo Arboit, Giorgio Carlino, Joël L. Lavanchy, Pietro Mascagni, and Nicolas Padoy. Surgical text-to-image generation, 2024.

[16] Stanislav Frolov, Tobias Hinz, Federico Raue, Jörn Hees, and Andreas Dengel. Adversarial text-to-image synthesis: A review, 2021.

[17] Jack Breen, Kieran Zucker, Katie Allen, Nishant Ravikumar, and Nicolas M. Orsi. Generative adversarial networks for stain normalisation in histopathology, 2024.

[18] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.

[19] Azin Shokraei Fard, David C. Reutens, and Viktor Vegh. Cnns and gans in mri-based cross-modality medical image estimation, 2021.

[20] Dimitrios Korkinof, Tobias Rijken, Michael O'Neill, Joseph Yearsley, Hugh Harvey, and Ben Glocker. High-resolution mammogram synthesis using progressive generative adversarial networks, 2019.

[21] Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. Transformers in medical imaging: A survey. *Medical image analysis*, 88:102802, 2023.

[22] Hazrat Ali, Christer Gronlund, and Zubair Shah. Leveraging gans for data scarcity of covid-19: Beyond the hype, 2023.

[23] Yuda Bi. Exploring the power of generative deep learning for image-to-image translation and mri reconstruction: A cross-domain review, 2023.

[24] Meng Li and Brian Lovell. End to end generative meta curriculum learning for medical data augmentation, 2022.

[25] M. AbdulRazek, G. Khoriba, and M. Belal. Gan-ga: A generative model based on genetic algorithm for medical image generation, 2023.

[26] Nripendra Kumar Singh and Khalid Raza. Medical image generation using generative adversarial networks, 2020.

[27] 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, 2018.

[28] Mathijs Pieters and Marco Wiering. Comparing generative adversarial network techniques for image creation and modification, 2018.

[29] Hamed Alqahtani, Manolya Kavakli-Thorne, and Gulshan Kumar. Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, 28:525–552, 2021.

[30] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, 35(4):3313–3332, 2021.

[31] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks, 2018.

[32] Victor Costa, Nuno Lourenço, João Correia, and Penousal Machado. Exploring generative adversarial networks for text-to-image generation with evolution strategies, 2022.

[33] Haowei Yang, Yuxiang Hu, Shuyao He, Ting Xu, Jiajie Yuan, and Xingxin Gu. Applying conditional generative adversarial networks for imaging diagnosis, 2024.

[34] Yihang Zhou, Rebecca Towning, Zaid Awad, and Stamatia Giannarou. Image synthesis with class-aware semantic diffusion models for surgical scene segmentation, 2024.

[35] Bowen Xin, Tony Young, Claire E Wainwright, Tamara Blake, Leo Lebrat, Thomas Gaass, Thomas Benkert, Alto Stemmer, David Coman, and Jason Dowling. Deformation-aware gan for medical image synthesis with substantially misaligned pairs, 2024.

[36] August DuMont Schütte, Jürgen Hetzel, Sergios Gatidis, Tobias Hepp, Benedikt Dietz, Stefan Bauer, and Patrick Schwab. Overcoming barriers to data sharing with medical image generation: A comprehensive evaluation, 2021.

[37] Chun-Mei Feng. Enhancing label-efficient medical image segmentation with text-guided diffusion models, 2024.

[38] Apoorva Sikka, Skand Peri, Jitender Singh Virk, Usma Niyaz, and Deepti R. Bathula. Mri to pet cross-modality translation using globally and locally aware gan (gla-gan) for multi-modal diagnosis of alzheimer's disease, 2024.

[39] Cristiana Tiago, Sten Roar Snare, Jurica Sprem, and Kristin McLeod. A domain translation framework with an adversarial denoising diffusion model to generate synthetic datasets of echocardiography images, 2024.

[40] Junhua Chen, Chong Zhang, Alberto Traverso, Ivan Zhovannik, Andre Dekker, Leonard Wee, and Inigo Bermejo. Generative models improve radiomics reproducibility in low dose cts: A simulation study, 2021.

[41] Hazrat Ali and Zubair Shah. Combating covid-19 using generative adversarial networks and artificial intelligence for medical images: A scoping review, 2022.

[42] Shumpei Takezaki and Seiichi Uchida. An ordinal diffusion model for generating medical images with different severity levels, 2024.

[43] P Manisha and Sujit Gujar. Generative adversarial networks (gans): What it can generate and what it cannot?, 2019.

[44] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of decision making*, pages 323–350, 2017.

[45] Abdullah al Nomaan Nafi, Md. Alamgir Hossain, Rakib Hossain Rifat, Md Mahabub Uz Zaman, Md Manjurul Ahsan, and Shivakumar Raman. Diffusion-based approaches in medical image generation and analysis, 2024.

[46] Salman Ul Hassan Dar, Marvin Seyfarth, Isabelle Ayx, Theano Papavassiliu, Stefan O. Schoenberg, Robert Malte Siepmann, Fabian Christopher Laqua, Jannik Kahmann, Norbert Frey, Bettina Baeßler, Sebastian Foersch, Daniel Truhn, Jakob Nikolas Kather, and Sandy Engelhardt. Unconditional latent diffusion models memorize patient imaging data: Implications for openly sharing synthetic data, 2025.

[47] Cristian Sbrolli, Paolo Cudrano, and Matteo Matteucci. Can shape-infused joint embeddings improve image-conditioned 3d diffusion?, 2024.

[48] David Keetae Park, Seungjoo Yoo, Hyojin Bahng, Jaegul Choo, and Noseong Park. Megan: Mixture of experts of generative adversarial networks for multimodal image generation, 2018.

[49] Jialei Chen, Yujia Xie, Kan Wang, Zih Huei Wang, Geet Lahoti, Chuck Zhang, Mani A Vannan, Ben Wang, and Zhen Qian. Generative invertible networks (gin): Pathophysiology-interpretable feature mapping and virtual patient generation, 2018.

[50] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. Infinite brain mr images: Pggan-based data augmentation for tumor detection, 2019.

[51] Lisa C. Adams, Felix Busch, Daniel Truhn, Marcus R. Makowski, Hugo JWL. Aerts, and Keno K. Bressem. What does dall-e 2 know about radiology?, 2022.

[52] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks, 2018.

[53] Firas Khader, Gustav Mueller-Franzes, Soroosh Tayebi Arasteh, Tianyu Han, Christoph Haarburger, Maximilian Schulze-Hagen, Philipp Schad, Sandy Engelhardt, Bettina Baessler, Sebastian Foersch, Johannes Stegmaier, Christiane Kuhl, Sven Nebelung, Jakob Nikolas Kather, and Daniel Truhn. Medical diffusion: Denoising diffusion probabilistic models for 3d medical image generation, 2023.

[54] Walter H. L. Pinaya, Mark S. Graham, Robert Gray, Pedro F Da Costa, Petru-Daniel Tudosiu, Paul Wright, Yee H. Mah, Andrew D. MacKinnon, James T. Teo, Rolf Jager, David Werring, Geraint Rees, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Fast unsupervised brain anomaly detection and segmentation with diffusion models, 2022.

AI-generated, for reference only.

[55] Huan He, Shifan Zhao, Yuanzhe Xi, and Joyce C Ho. Meddiff: Generating electronic health records using accelerated denoising diffusion model, 2023.

[56] Nicholas Konz, Yuwen Chen, Haoyu Dong, and Maciej A. Mazurowski. Anatomically-controllable medical image generation with segmentation-guided diffusion models, 2024.

[57] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis, 2019.

[58] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, and Junmo Kim. Text-to-image diffusion models in generative ai: A survey, 2024.

[59] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820–838, 2021.

[60] Shuyang Gu, Dong Chen, Jianmin Bao, Fang Wen, Bo Zhang, Dongdong Chen, Lu Yuan, and Baining Guo. Vector quantized diffusion model for text-to-image synthesis, 2022.

[61] Yu Gu, Jianwei Yang, Naoto Usuyama, Chunyuan Li, Sheng Zhang, Matthew P. Lungren, Jianfeng Gao, and Hoifung Poon. Biomedjourney: Counterfactual biomedical image generation by instruction-learning from multimodal patient journeys, 2023.

[62] Ling Yang, Zhaochen Yu, Chenlin Meng, Minkai Xu, Stefano Ermon, and Bin Cui. Mastering text-to-image diffusion: Recaptioning, planning, and generating with multimodal llms, 2024.

[63] Salaheldin Mohamed. Tcig: Two-stage controlled image generation with quality enhancement through diffusion, 2024.

[64] Wei Li, Xue Xu, Jiachen Liu, and Xinyan Xiao. Unimo-g: Unified image generation through multimodal conditional diffusion, 2024.

[65] Rohan Wadhawan, Tanuj Drall, Shubham Singh, and Shampa Chakraverty. Multi-attributed and structured text-to-face synthesis, 2021.

[66] Stanley H. Chan. Tutorial on diffusion models for imaging and vision, 2025.

[67] Debin Meng, Christos Tzelepis, Ioannis Patras, and Georgios Tzimiropoulos. Mm2latent: Text-to-facial image generation and editing in gans with multimodal assistance, 2024.

[68] Nithin Gopalakrishnan Nair, Jeya Maria Jose Valanarasu, and Vishal M Patel. Maxfusion: Plugplay multi-modal generation in text-to-image diffusion models, 2024.

[69] Konstantina Nikolaidou, George Retsinas, Vincent Christlein, Mathias Seuret, Giorgos Sfikas, Elisa Barney Smith, Hamam Mokayed, and Marcus Liwicki. Wordstylist: Styled verbatim handwritten text generation with latent diffusion models, 2023.

[70] Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. Multidiffusion: Fusing diffusion paths for controlled image generation, 2023.

[71] Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. Ufogen: You forward once large scale text-to-image generation via diffusion gans, 2023.

[72] Anton Razzhigaev, Arseniy Shakhmatov, Anastasia Maltseva, Vladimir Arkhipkin, Igor Pavlov, Ilya Ryabov, Angelina Kuts, Alexander Panchenko, Andrey Kuznetsov, and Denis Dimitrov. Kandinsky: an improved text-to-image synthesis with image prior and latent diffusion, 2023.

[73] Stanislav Frolov, Prateek Bansal, Jörn Hees, and Andreas Dengel. Dt2i: Dense text-to-image generation from region descriptions, 2022.

[74] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis, 2019.

[75] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022.

[76] Sahar Almahfouz Nasser, Ashutosh Sharma, Anmol Saraf, Amruta Mahendra Parulekar, Purvi Haria, and Amit Sethi. Transforming breast cancer diagnosis: Towards real-time ultrasound to mammogram conversion for cost-effective diagnosis, 2023.

[77] Andrawes Al Bahou, Christine Tanner, and Orcun Goksel. Scatgan for reconstruction of ultrasound scatterers using generative adversarial networks, 2019.

[78] Jiamin Liang, Xin Yang, Haoming Li, Yi Wang, Manh The Van, Haoran Dou, Chaoyu Chen, Jinghui Fang, Xiaowen Liang, Zixin Mai, Guowen Zhu, Zhiyi Chen, and Dong Ni. Synthesis and edition of ultrasound images via sketch guided progressive growing gans, 2020.

[79] Isabela Albuquerque, João Monteiro, and Tiago H. Falk. Learning to navigate image manifolds induced by generative adversarial networks for unsupervised video generation, 2019.

[80] Moritz Platscher, Jonathan Zopes, and Christian Federau. Image translation for medical image generation – ischemic stroke lesions, 2021.

[81] Matteo Pennisi, Federica Proietto Salanitri, Giovanni Bellitto, Simone Palazzo, Ulas Bagci, and Concetto Spampinato. A privacy-preserving walk in the latent space of generative models for medical applications, 2023.

[82] Lorenzo Aloisi, Luigi Sigillo, Aurelio Uncini, and Danilo Comminiello. A wavelet diffusion gan for image super-resolution, 2024.

[83] Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. On the detection of synthetic images generated by diffusion models, 2022.

[84] Pranav Aggarwal, Hareesh Ravi, Naveen Marri, Sachin Kelkar, Fengbin Chen, Vinh Khuc, Midhun Harikumar, Ritiz Tambi, Sudharshan Reddy Kakumanu, Purvak Lapsiya, Alvin Ghouas, Sarah Saber, Malavika Ramprasad, Baldo Faieta, and Ajinkya Kale. Controlled and conditional text to image generation with diffusion prior, 2023.

[85] Noam Elata, Bahjat Kawar, Tomer Michaeli, and Michael Elad. Nested diffusion processes for anytime image generation, 2023.

[86] Marija Habijan and Irena Galic. Generation of artificial ct images using patch-based conditional generative adversarial networks, 2022.

[87] Eric Carver, Zhenzhen Dai, Evan Liang, James Snyder, and Ning Wen. Improvement of multiparametric mr image segmentation by augmenting the data with generative adversarial networks for glioma patients, 2019.

[88] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3*, pages 1–11. Springer, 2018.

[89] Puria Azadi Moghadam, Sanne Van Dalen, Karina C. Martin, Jochen Lennerz, Stephen Yip, Hossein Farahani, and Ali Bashashati. A morphology focused diffusion probabilistic model for synthesis of histopathology images, 2022.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.