

DiffusionTBAD: Rendering CTA images for type B aortic dissection diagnosis[☆]

Ayman Abaid^a, Muhammad Ali Farooq^b, Niamh Hynes^c, Peter Corcoran^b,
Ihsan Ullah^a,*

^a School of Computer Science, University of Galway and Data Science Institute, Ireland

^b C3I Group, School of Engineering, College of Science and Engineering, University of Galway, Ireland

^c School of Medicine, University of Galway, Galway, Ireland

ARTICLE INFO

Dataset link: <https://github.com/AymanAbaid/DiffusionTBAD>

Keywords:

Aortic dissection
Data synthesis
Stable diffusion
Text-to-image
Computed tomography angiography

ABSTRACT

The success of diffusion models in medical imaging highlights their potential to generate high-quality synthetic datasets that closely resemble real clinical data, addressing limited dataset availability and patient privacy concerns. We present *DiffusionTBAD*, a novel text-to-image diffusion-based pipeline for synthesizing diagnostically accurate computed tomography angiography (CTA) images of type B aortic dissection (TBAD). Using few-shot learning, *DiffusionTBAD* fine-tunes a diffusion model guided by textual prompts to capture the distinct features and variability of TBAD cases. The synthetic data are evaluated using quantitative diversity and similarity metrics, as well as downstream task performance. Augmenting real TBAD datasets with synthetic images improved supervised classification accuracy from 67% to 76%, and pre-training on synthetic images increased segmentation DICE scores from 66% to 70%. Additionally, qualitative assessment by eight healthcare professionals confirmed the high visual realism and diagnostic plausibility of the generated images. These results demonstrate that *DiffusionTBAD* can enhance model performance while reducing reliance on real patient data, enabling privacy-preserving development of medical imaging models.

1. Introduction

Deep learning (DL) techniques have significantly advanced various visual tasks in medical imaging, encompassing both diagnostic and prognostic applications (Anaya-Isaza et al., 2021). These methods excel at automatically modeling complex, high-level features in data, leading to improvements in accuracy, automation, and overall efficiency. To ensure that DL models generalize well and capture a wide range of potential variations rather than simply memorizing the training data, it is crucial to have data with significant variability. However, acquiring sufficient high-quality data, particularly for rare diseases, is a time-consuming task prone to human error. This challenge is further compounded by stringent privacy regulations, the requirement for patient consent, and the sensitive nature of medical information. To address the challenges of generalization and adaptation in DL models with limited data, the generation of synthetic images using diffusion models has recently emerged as a promising approach (Güven and Talu, 2023; Chuquicusma et al., 2018; Ali et al., 2022; Abaid et al., 2024a). By employing conditional synthesis, large, diverse datasets of medical images can be generated, enhancing model robustness while addressing data privacy issues.

Aortic dissection (AD) is a severe cardiovascular disorder marked by a tear in the inner layer of the aorta. This tear causes a separation between the inner and middle layers, forming a false lumen (FL) alongside the existing true lumen (TL), as shown in Fig. 1. The presence of FL disrupts normal blood flow, leading to reduced blood supply to vital organs and limbs. In some cases, a blood clot may develop within the false lumen, known as false lumen thrombosis (FLT), resulting in very low, non-pulsatile blood pressure in the affected area (Pepe et al., 2020). The Stanford Classification system categorizes AD into two types based on the location of the tear: Type A and Type B. Stanford Type B aortic dissection (TBAD) occurs when the dissection takes place in the descending part of the aorta. Although TBAD is rare, it demands immediate medical attention as it constitutes a medical emergency. Rapid diagnosis and treatment significantly enhance the chances of survival, even in the presence of the condition's potential severity. A key factor in predicting the prognosis of TBAD is the rapid and precise quantification of anatomical features, especially the lumen size (Sailer et al., 2017). These characteristics aid in recognizing patients who may be prone to experiencing adverse events as the disease progresses. Precise segmentation of the aortic lumen is an essential initial stage

[☆] This article is part of a Special issue entitled: 'GenAI for Medical Imaging' published in Computerized Medical Imaging and Graphics.

* Corresponding author.

E-mail addresses: a.abaid1@universityofgalway.ie (A. Abaid), muhammadali.farooq@universityofgalway.ie (M.A. Farooq), niamh.hynes@universityofgalway.ie (N. Hynes), peter.corcoran@universityofgalway.ie (P. Corcoran), ihsan.ullah@universityofgalway.ie (I. Ullah).

<https://doi.org/10.1016/j.compmedimag.2026.102740>

Received 21 July 2025; Received in revised form 8 January 2026; Accepted 27 February 2026

Available online 2 March 2026

0895-6111/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

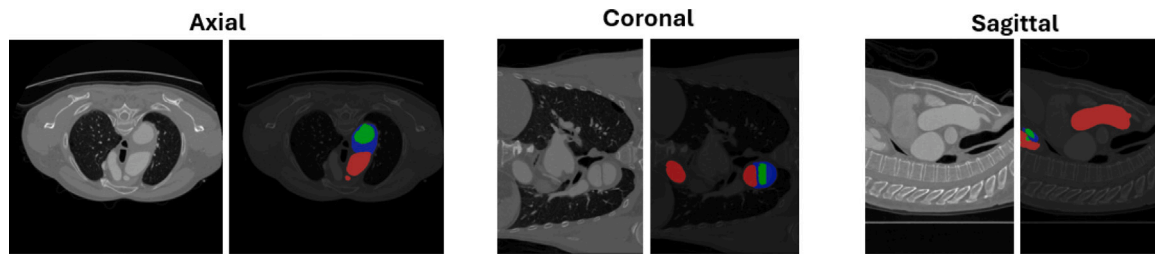


Fig. 1. Visualization of TBAD CTA slices and corresponding masks, highlighting FLT (blue), TL (green), and FL (red) across axial, coronal, and sagittal views.

in assessing these anatomical traits. CTA is commonly used to diagnose TBAD because it can rapidly generate high-quality images, is minimally invasive, and provides clearer visualization of relevant features. Recent advancements in DL provide a unique opportunity for automated and precise segmentation of TL, FL, and FLT. However, as discussed earlier, DL models require large amounts of data, which poses challenges in effectively training them for accurate classification and segmentation tasks, particularly when data availability is limited. To address this issue, we explore the potential of utilizing text-to-image (TtI) diffusion models such as Stable Diffusion (Rombach et al., 2022) and DALL-E (Ramesh et al., 2021) for generating synthetic 2D CTA axial slices. While this design aligns with current data availability and privacy constraints, it inherently operates on independent 2D slices and therefore does not enforce inter-slice or volumetric anatomical consistency.

Few-shot learning methods, such as Textual Inversion (Gal et al., 2022) and DreamBooth (Ruiz et al., 2023), enable targeted adaptation of pre-trained TtI diffusion models to new concepts using only a small number of examples. This is particularly advantageous in medical imaging, where annotated datasets are limited and capturing subtle anatomical and pathological features is critical. Full network fine-tuning would require substantially more data and carries a higher risk of over-fitting, which can distort fine-grained structures. To further improve training efficiency while preserving the generalization capabilities of the pre-trained model, we employ the parameter-efficient fine-tuning method Low-Rank Adaptation (LoRA) (Hu et al., 2021).

In this study, we employ a pre-trained TtI diffusion model, Stable Diffusion, and fine-tune it on TBAD CTA images using the few-shot fine-tuning approach DreamBooth (Ruiz et al., 2023), along with the PEFT method Low-Rank Adaptation (LoRA) (Hu et al., 2021). We systematically compare few-shot adaptation, full fine-tuning, and other PEFT methods, providing a balanced analysis of their relative strengths and limitations. Our results indicate that synthetic data generated through DiffusionTBAD produce images that accurately capture TBAD-specific vascular patterns and align closely with real patient data. Experimental evaluations demonstrate that augmenting real datasets with these synthetic images significantly improves the performance of machine learning models compared to using real data alone. To promote transparency and collaboration within the research community, we are releasing the generated TBAD CTA images as an [open-source dataset](#). By making this resource publicly available, we aim to advance cardiovascular imaging research and inspire innovation in related domains.

The core contributions of this paper are as follows:

- We propose **DiffusionTBAD**, a customized text-to-image (TtI) diffusion imaging pipeline. This pipeline fine-tunes a pre-trained diffusion model using few-shot learning and leverages LoRA to generate synthetic images while optimizing for reduced computational overhead.
- We generate and open-source new synthetic TBAD data using DiffusionTBAD, covering five distinct class variations: True Lumen (TL), False Lumen (FL), False Lumen Thrombus (FLT), a combination of TL + FL, and a Background class (lacking TL, FL, or FLT).
- The reliability of the synthetic data is validated through extensive quantitative evaluations using standard metrics and downstream machine learning tasks, providing insights into the effectiveness of the generation process and its potential utility.
- Furthermore, an in-depth qualitative evaluation was conducted by involving eight healthcare professionals to assess the clinical plausibility and visual realism of the synthetic images. This assessment included both paired image comparisons and class-specific realism evaluations.

2. Related works

2.1. Generation of synthetic medical images

The medical research community has demonstrated growing interest in generating synthetic images to address challenges such as data scarcity and model training limitations. Generative models, including Generative Adversarial Networks (GANs) (Goodfellow et al., 2020) and Diffusion Models (Dhariwal and Nichol, 2021), have shown exceptional capabilities in synthesizing realistic images (Güven and Talu, 2023; Chuquicusma et al., 2018; Ali et al., 2022; Abaid et al., 2024a). For instance, Ghorbani et al. (2020) successfully produced high-resolution skin lesion images using pix2pix GANs, which were indistinguishable from real ones by experts. Similarly, Güven and Talu (2023) employed GigaGAN, a GAN variant, for TtI synthesis to generate brain MRIs that matched the quality of real images. Additionally, Chuquicusma et al. (2018) demonstrated that GAN-generated lung cancer nodule images were nearly indistinguishable from real images, even to trained radiologists.

While GANs have achieved state-of-the-art (SoTA) performance in various medical imaging tasks, they exhibit notable limitations. These include poor mode coverage and limited sampling diversity, which can impact the variability and comprehensiveness of generated images (Kazerouni et al., 2023). Furthermore, recent advancements in TtI synthesis using GANs, as explored in Kang et al. (2023), still face challenges in text alignment and visual quality, indicating that these models are not yet on par with production-grade frameworks like Stable Diffusion. These limitations highlight the need for exploring alternative approaches, to overcome the constraints associated with GANs.

Diffusion models, on the other hand, have shown superior performance in this area. Their ability to capture fine-grained details and complexities, along with their capacity to generate diverse outputs, makes them particularly well-suited for conditional image generation tasks. For instance, Pan et al. (2023) explored the use of a diffusion model combined with a transformer architecture instead of the traditional U-Net for generating chest X-rays, heart MRIs, pelvic CTs, and abdominal CTs. A Visual Turing assessment conducted by three medical physicists confirmed that the synthetic images produced were highly realistic and could be valuable in scenarios with limited data availability. Similarly, Ali et al. (2022) employed diffusion models alongside GANs to generate brain MRI and chest X-ray images, facilitating the recognition of thoracic abnormalities in chest radiographs. Pre-trained diffusion models, such as Stable Diffusion and DALL-E, are trained on

Table 1
Comparison of Various Algorithms for Analyzing CTA and 3D Volume Images.

Paper	Data	Algorithms	TL	FL	FLT
Hahn et al. (2020)	2D CTA images	2 CNNs	87	89	–
Cheng et al. (2020)	2D CTA images	Enhanced U-Net + aortic circularity analysis	91	–	–
Abaid et al. (2024b)	2D CTA images	2D U-Net	83	85	30
Cao et al. (2019)	3D CTA volumes	CNN trained using serial multi-task models	93	91	–
Wobben et al. (2021)	3D CTA volumes	Sequential multi-task 3D residual U-Net	86	86	50
Yao et al. (2021)	3D CTA volumes	3D U-Net	79	68	52
Jung et al. (2024)	3D CTA volumes	3D transformer + 3D U-Net	92	88	63

extensive datasets containing billions of images. This training enables them to learn diverse representations, facilitating the generation of novel images that exceed the variability present in their training data, even when prompted with highly improbable inputs. The scale of their training datasets suggests that these models likely include medical image data, potentially allowing them to comprehend the composition and structure of CT scans, X-rays, and ultrasounds (Adams et al., 2023). This opens up promising applications, particularly as domain-specific fine-tuning holds potential for enhancing the generation of accurate and clinically relevant medical data.

Recent research in this domain has focused on two main strategies: training generative models from scratch and fine-tuning pre-trained diffusion models. While these methods have proven effective, they are computationally intensive and often require large-scale datasets, limiting their applicability to smaller medical datasets. Additionally, fine-tuning large models for each downstream task poses a significant challenge, as it typically involves full fine-tuning of all dense layers. To address these challenges, parameter-efficient fine-tuning techniques, such as LoRA, and few-shot learning approaches, like Textual Inversion and DreamBooth, have gained increasing attention.

This work builds upon the study conducted in Abaid et al. (2024a), which demonstrated that diffusion models trained on natural images can be adapted to generate CTA images in scenarios with limited training data. In that study, the authors generated and evaluated a limited set of CTA samples using few-shot learning. In this study, we compare full fine-tuning, Textual Inversion, and DiffusionTBAD in terms of the quality of the generated images and the training time required. We also perform a comprehensive evaluation of the utility of synthetic images. Specifically, we assess their impact on downstream tasks, including classification and segmentation, providing deeper insights into their potential to enhance model performance in medical imaging applications.

2.2. Type-B aortic dissection segmentation using deep learning

DL approaches have been widely utilized in the management of TBAD, with several investigations leveraging CTA images to segment TL, FL and FLT. Although many studies have focused on 3D volumes, the potential of DL-driven segmentation has been showcased in both 2D and 3D contexts. Table 1 summarizes the characteristics of various investigations on DL-based TBAD segmentation.

Hahn et al. (2020) proposed a 2D network for axial CT images to automate the segmentation of the entire aorta, TL, and FL, yielding promising outcomes. Similarly, Cheng et al. (2020) employed contrast-enhanced CT images within a U-Net framework integrated with Squeeze and Excitation (SE) blocks post-encoder blocks. They further enhanced the U-Net decoder by incorporating a deformable feature decoder module, replacing bilinear interpolation and deconvolution with deformable convolution to improve image quality. However, neither of these approaches accounted for the segmentation of FLT. Abaid et al. (2024b) employed a 2D U-Net and its variants, incorporating atrous convolutions and custom layers, to segment TL, FL, and FLT from axial CTA images. While their method achieved favorable results for TL and FL segmentation, the performance on FLT segmentation remained suboptimal.

Expanding on these efforts, authors in Cao et al. (2019) utilized a dataset containing 276 volumes, focusing on the segmentation of the aorta, TL, and FL. The authors implemented three different 3D convolutional neural networks (CNNs) with varying configurations: CNN1, CNN2, and CNN3. CNN1 is a single-task network that individually segments the whole aorta, TL, and FL. In contrast, CNN2 employs a multi-task network framework to simultaneously segment the whole aorta, TL, and FL. CNN3 utilizes a serial multi-task network framework with two 3D U-Net models, achieving the best results with Dice Similarity Coefficient (DICE) of 0.93 for TL and 0.91 for FL. Similarly, Wobben et al. (2021) used a dataset consisting of 164 CTA scans with ground truth masks for TL, FL, and FLT. They proposed three models: (1) a single-step multi-task model that segments TL, FL, and the Background; (2) a sequential multi-task model that first segments the whole aorta, TL, and FL, and subsequently segments FLT from FL; and (3) a single-step single-task model dedicated to segmenting FLT. The second model achieved the highest performance, with DICE scores of 0.86, 0.86, and 0.50 for TL, FL, and FLT, respectively (Wobben et al., 2021). Additionally, Yao et al. (2021) employed a dataset containing 100 CTA volumes and proposed a two-step pipeline for segmentation. The first step involves region of interest (ROI) extraction, followed by ROI segmentation in the second step. For the segmentation, a 3D U-Net is applied to the extracted ROI regions, achieving DICE scores of 0.79, 0.68, and 0.50 for TL, FL, and FLT, respectively. Lastly, recent work (Jung et al., 2024), a novel architecture named ZOZI-Seg was introduced. This cascaded network combines the features of a transformer and U-Net while integrating a Zoom out Zoom in (ZOZI) mechanism. This study utilized a dataset comprising 253 CTA images, achieving DICE of 0.91, 0.88, and 0.63 for TL, FL, and FLT, respectively.

In prior research, significant efforts have been directed towards enhancing segmentation accuracy of DL models by training them on real-world data. Approximately half of the studies on TBAD have focused on segmenting commonly available substructures, such as the TL and FL. This is due to the rarity of FLT in patients, which substantially contributes to the lower segmentation accuracy for FLT. The limited diversity in the available data poses a challenge for DL models in effectively learning the underlying image representations.

This paper introduces a novel methodology for synthesizing CTA images to support TBAD diagnosis, representing, to the best of our knowledge, the first attempt in this domain. Our approach leverages T1 diffusion models that are fine-tuned on medical or out-of-distribution data to generate synthetic datasets without training generative models from scratch, which is both time consuming and computationally expensive. We adopt a 2D slice-based strategy, converting 3D CTA volumes into axial slices for training and inference. This choice aligns the pipeline with the inherently 2D design of diffusion models, enhancing adaptability across modalities while maintaining diagnostic fidelity. Although this reduces volumetric context, prior studies (Hahn et al., 2020; Cheng et al., 2020; Abaid et al., 2024b) have demonstrated that high-resolution 2D slices can capture sufficient local anatomical detail for accurate segmentation and classification while addressing challenges associated with limited training data and the high computational demands of 3D modeling. Importantly, this framework enables the generation of synthetic datasets that improve downstream performance, particularly for rare substructures such as FLT.

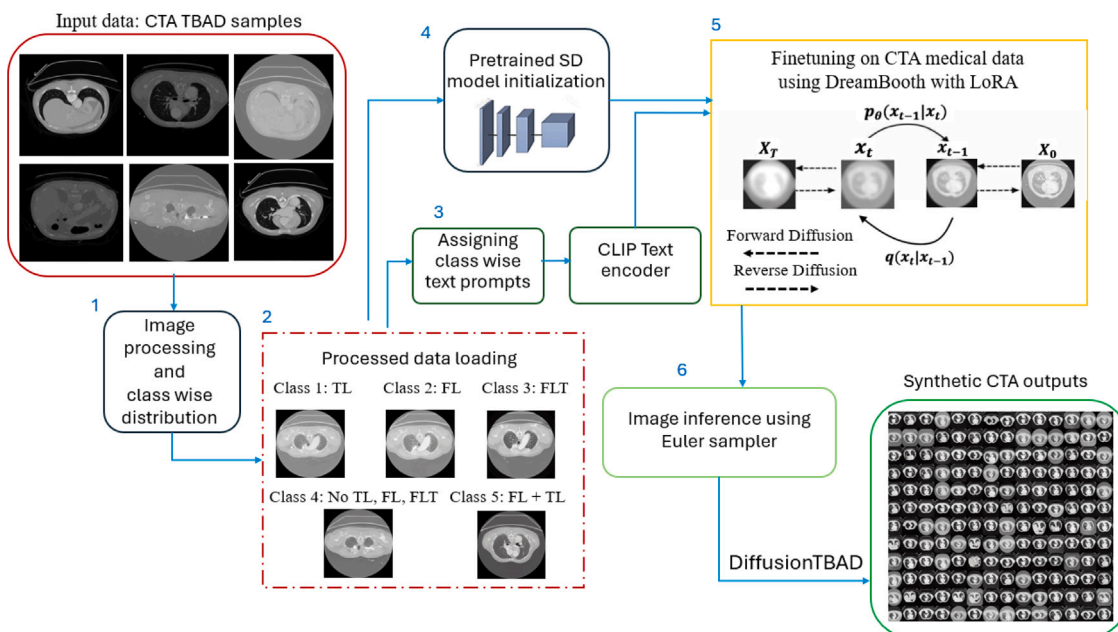


Fig. 2. Block diagram of the proposed methodology. Pre-processed input images are assigned class-specific text prompts, which are then converted into embedding. Both the image data and text embedding are utilized to fine-tune the diffusion model. The numerical labels on each block denote the sequential order of steps in the pipeline.

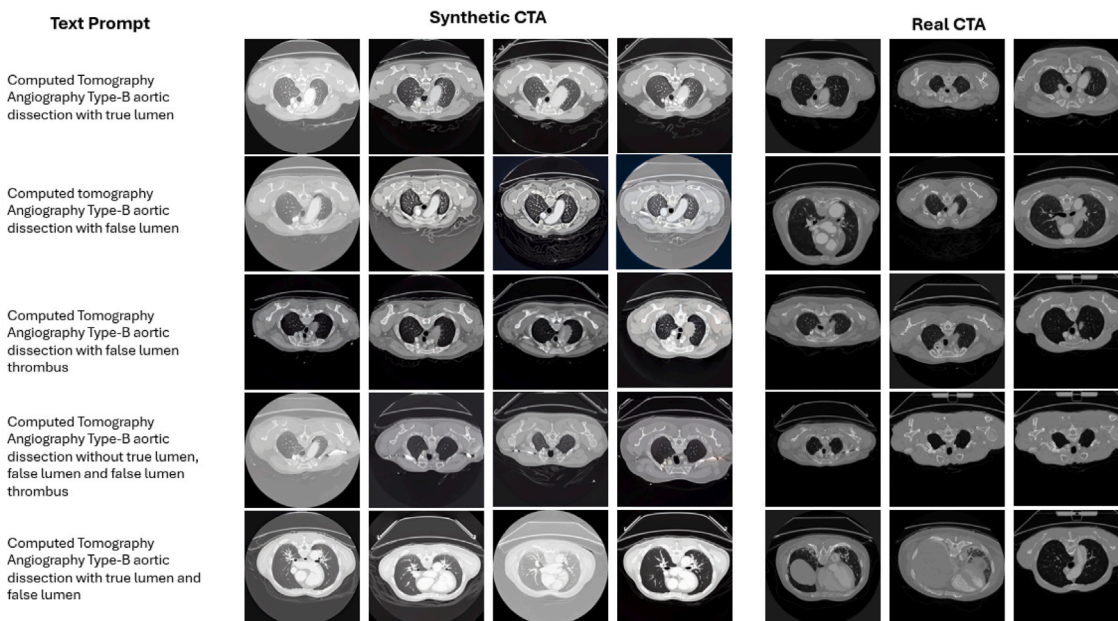


Fig. 3. Synthetic CTA images of Type-B aortic dissection generated from text prompts, illustrating variations in the true lumen, false lumen, and false lumen thrombus, shown in comparison with the ground truth.

3. Proposed methodology

This section provides a detailed overview of the adapted methodology and the validation process employed in this study. Fig. 2 presents a comprehensive block diagram illustrating the adapted methodology used for generating CTA TBAD data. The methodology is organized into subsections: 3.1 covers data and preprocessing, 3.2 describes the diffusion models and fine-tuning, 3.3 explains image sampling and textual prompt construction, and 3.4 discusses the validation and evaluation of generated samples.

3.1. Seed data sourcing and image processing

The seed data plays a crucial role in tuning diffusion models especially for medical imaging application to ensure optimal quality and accuracy of the generated synthetic images. As depicted in block 1–2 of Fig. 2 the first step includes acquiring seed data samples from publicly available TBAD datasets. For this purpose we have used the ImageTBAD dataset (Yao et al., 2021). The ImageTBAD dataset, comprises 100 CTA volumes and corresponding labels from 100 patients. Each CTA volume provides axial, sagittal, and coronal views alongside their respective

labels as shown in Fig. 1. Segmentation is performed manually by two experienced cardiologists and the process of labeling each image took between 1 to 1.5 h. The segmentation mask encompasses three distinct substructures: TL, FL, and FLT. Among the 100 patients, 68 include FLT within their images, while the remaining 32 images do not feature this substructure. For this study, we transformed the volumes into 2D images, focusing solely on axial images and their corresponding masks. Each axial image and its label are sized at 512×512 pixels, resulting in a total of 34,380 images. These images are divided into three sets: training, testing, and validation. The training set contains 22,248 images, while the testing and validation sets consist of 4256 and 4493 images, respectively.

3.2. Fine-tuning diffusion models via few-shot learning

Diffusion models (Ho et al., 2020) are a class of generative models used to generate realistic samples from a given data distribution by leveraging a forward diffusion process and a reverse denoising process. The forward diffusion process is a Markov chain with T steps, starting with a data sample x_0 drawn from a real data distribution $q(x)$ and gradually adding Gaussian noise over T discrete time steps, transforming it into a sequence of noise samples x_1, x_2, \dots, x_T . The forward process at time step t can be defined as follows:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I). \quad (1)$$

where x_t is the noisy version of x_0 at time step t , β_t is the variance schedule parameter at time step t that controls the noise level, I is the identity matrix, and \mathcal{N} denotes a Gaussian distribution. The overall forward process is described by:

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}).$$

The reverse denoising process aims to reconstruct x_0 from x_T by learning a reverse Markov chain parameterized by a neural network θ , expressed as

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \quad (2)$$

where μ_θ and Σ_θ are the mean and variance predicted by the neural network. A visual representation of the diffusion process is shown in Block 5 of Fig. 2.

Diffusion models, as demonstrated in various studies (Dhariwal and Nichol, 2021; Nichol et al., 2021; Saharia et al., 2022), can be conditioned on attributes such as class labels, text descriptions, or even images. Classifier-free guidance (Ho and Salimans, 2022) represents a novel technique within conditional diffusion models aimed at enhancing the richness and diversity of generated samples without the necessity of explicit classifier input. Throughout the training, the model intermittently receives conditioning information. Joint training is conducted to refine both an unconditional denoising diffusion model $p_\theta(z)$, governed by a score estimator $\epsilon_\theta(z|\lambda)$, and a conditional model $p_\theta(z|c)$, driven by $\epsilon_\theta(z|\lambda, c)$, leveraging a single neural network to parameterize both models. Sampling is then facilitated through a linear combination of the conditional and unconditional score estimates, described by following:

$$\tilde{\epsilon}_\theta(z|\lambda, c) = (1 + w)\epsilon_\theta(z|\lambda, c) - w\epsilon_\theta(z|\lambda). \quad (3)$$

where c is the class identifier and w regulates the influence of the classifier's guidance. Throughout the training process, the diffusion model optimizes the following squared error loss:

$$\mathbb{E}_x, c, \epsilon, t, w, |\hat{x}(\alpha_t x + \sigma_t \epsilon, c) - x|^2 \quad (4)$$

This loss function evaluates the model's prediction \hat{x} by comparing it to the ground-truth image x . By minimizing this loss, the model is trained to progressively denoise noisy inputs $\alpha_t x + \sigma_t \epsilon$, while maintaining alignment with the conditioning vector c .

DreamBooth is a fine-tuning technique designed to adapt diffusion models with limited samples, specifically addressing the problem of language drift, where the model overfits and loses diversity in its outputs. To mitigate this, DreamBooth introduces a class-specific prior preservation loss, as shown in Eq. (4), alongside the original loss in Eq. (5). This approach ensures the model retains the diversity of the original data distribution while smoothly integrating new samples. It preserves the prior by associating the class name with the specific instance during fine-tuning. By leveraging the semantic prior embedded in the model, DreamBooth generates diverse, semantically consistent outputs that reflect both the original class and the newly introduced subject's characteristics.

$$\lambda W_{t'} \left\| \hat{x}_\theta(\alpha_{t'} x_{pr} + \sigma_{t'} \epsilon', c_{pr}) - x_{pr} \right\|^2 \quad (5)$$

Rather than fully fine-tuning all dense layers of the diffusion model, Low-Rank Adaptation (LoRA) offers a more computationally efficient alternative. LoRA trains a small set of parameters, Δ , which significantly reduces computational overhead. This method involves adding a low-rank weight matrix, ΔW , to the existing trained weight matrix, leading to faster training speed and reduced VRAM requirements. Suppose W_0 represents the weights of the pre-trained large model that remain frozen and do not receive gradient updates. In this context, A and B are the trainable parameters and low rank matrices. The update to W_0 is constrained by representing it with a low-rank decomposition:

$$W_0 + \Delta W = W_0 + BA. \quad (6)$$

Both W_0 and ΔW are multiplied by the same input x , and their outputs are summed, resulting in the modified forward pass:

$$W_0 x + \Delta W x = W_0 x + BA x. \quad (7)$$

Here $B \in \mathbb{R}^{d \times r}$ and $A \in \mathbb{R}^{r \times k}$, with $\text{rank } r \ll \min(d, k)$. As demonstrated in Step 5 of Fig. 2, LoRA has been utilized to expedite the fine-tuning process by reducing the number of trainable parameters. This reduction enhances model fine-tuning efficiency and makes it more feasible in environments with limited computational resources, such as academic research settings.

3.3. Prompt engineering and text encoding for rendering CTA data

The input text prompt is crucial during the fine-tuning of TtI diffusion models, as it guides the model in generating images that align with specific textual descriptions. In the case of DreamBooth, a unique identifier is introduced to bind the new subject during the fine-tuning phase. Prompts are constructed in the form of “[class noun] [identifier]” enhancing the model's vocabulary with unique identifier-subject pairs. The [identifier] uniquely identifies the subject, while the [class noun] serves as a general class descriptor. This technique leverages the relationship between the specific subject and its class, utilizing the model's pre-existing knowledge about the class. In our study, the [class noun] refers to “computed tomography angiography type B aortic dissection,” while the [identifier] denotes the various lumens depicted in the images, such as ‘true lumen,’ ‘false lumen,’ or combinations like ‘true lumen and false lumen.’ This method was employed to generate text prompts, which were instrumental in fine-tuning the model for each class, as illustrated in block 3 of Fig. 2. The prompts are first tokenized using Byte Pair Encoding (BPE) (Sennrich, 2015), which represents the input text as a sequence of subword tokens. The tokenized text is input into CLIP's text encoder (Radford et al., 2021) to generate embeddings, which are then used to condition the diffusion model. Fig. 3 showcases the generation of CTA images with varying TBAD characteristics alongside the textual prompts used for training.

Table 2

Quantitative evaluation of synthetic data using various metrics with a sample size of 10000. Metrics marked with \uparrow indicate that higher values are better, while those marked with \downarrow indicate that lower values are preferred.

Method	Clean-fid \downarrow	KID \downarrow	MS-SSIM \uparrow	BiomedCLIPScore \uparrow	RadFID \downarrow	Training Time (h)
Full fine-tuning	67.16	0.05	0.39	26.2	7.05	15.30
Full fine-tuning + LoRA	119.24	0.29	0.29	31.1	5.87	2.70
Textual Inversion	148.12	0.13	0.20	21.1	12.14	11.50
DiffusionTBAD	50.14	0.03	0.40	26.5	9.15	6.30
DiffusionTBAD+	91.95	0.10	0.31	31.1	6.99	2.90

Table 3

Comparison of CLIP scores across different training strategies and the proposed approach with a sample size of 2000 per class. Results are reported for each class (TL, FL, FLT, Background, and TL+FL). Higher values indicate better alignment.

Method	TL	FL	FLT	Background	TL+FL
Full fine-tuning	29.5	29.4	29.1	27.8	28.2
Full fine-tuning + LoRA	28.2	28.7	28.0	27.9	27.4
Textual Inversion	29.4	30.3	25.8	27.1	27.9
DiffusionTBAD	29.3	29.6	28.9	28.6	28.0
DiffusionTBAD+	28.3	28.6	27.7	28.0	27.4

3.4. Validation of synthetic data

Our evaluation and validation pipeline assesses the quality of synthetic images based on three essential criteria: fidelity, diversity, and utility. Fidelity is measured using established metrics such as the Fréchet Inception Distance (FID) (Heusel et al., 2017). FID leverages pre-trained CNNs, such as the Inception network trained on ImageNet, to compare feature activations between real and synthetic images. It calculates the Fréchet distance between the distributions of features from real and generated data, where lower FID scores indicate higher image quality. Despite its widespread use, recent studies have identified several limitations of FID, including biases arising from normalization assumptions, sensitivity to low-level image preprocessing, and inefficiency with small sample sizes (Jayasumana et al., 2024). To address these shortcomings, we employed alternative metrics: clean-fid (Parmar et al., 2022), which is robust to low-level preprocessing artifacts, and Kernel Inception Distance (KID) (Bińkowski et al., 2018), which does not rely on normalization assumptions, is sensitive to both mode collapse and quality issues, and performs effectively with smaller sample sizes. These metrics were utilized to comprehensively evaluate the fidelity of the generated images. To further validate semantic consistency, we also computed the standard CLIP score (Hessel et al., 2021) for each class, which quantifies how closely the synthetic images align with their respective textual class prompts. This class-wise CLIP evaluation provides fine-grained insights into whether generated samples correctly represent the targeted diagnostic categories. In addition, to account for the unique properties of medical imaging, we incorporated Rad-FID (Osuala et al., 2023), a domain-adapted variant of FID computed using radiology pre-trained backbones, and the BiomedCLIP (Zhang et al., 2025) score, which measures cross-modal alignment between generated images and biomedical text embeddings.

Diversity assessment incorporates visual inspection and quantitative analysis, employing Multi-Scale Structural Similarity (MSSIM) (Wang et al., 2003) to capture differences in the distributions of real and synthetic data. MSSIM measures the structural similarity between images, where higher values denote greater similarity in structure and texture. Finally, utility is evaluated through comprehensive testing across a spectrum of classification and segmentation tasks, ensuring the practical effectiveness and applicability of the synthesized dataset.

4. Experimental setup and results

This section outlines the experimental setup and provides a comprehensive analysis of the results obtained during this research. The

overall experiments were conducted on a server-grade system equipped with NVIDIA RTX A6000 Ada Generation GPU with 48 GB of VRAM. In this study, we utilized Stable Diffusion Model version 1.5 as the base architecture for fine-tuning on TBAD images. The complete fine-tuning process is illustrated in Block 5 of Fig. 2. The fine-tuning procedure involved training the model for 15,000 optimization steps with a batch size of 1 and base learning rate of $1e-4$. We trained two variants of DiffusionTBAD: the standard version and a version enhanced with LoRA, referred to as DiffusionTBAD+. To optimize performance, we employed the 8-bit AdamW optimizer in conjunction with LoRA, which effectively reduced computational overhead. The Discrete Denoising Scheduler (DDS) was selected as the noise scheduler. CTA images were generated with the Euler sampler due to its superior performance (Farooq et al., 2024), with 50 sampling steps. We evaluated the performance of DiffusionTBAD against two widely adopted methods: Textual Inversion and the full fine-tuning technique. Additionally, we analyzed the impact of incorporating LoRA into the full fine-tuning process. In the subsequent phase, 2000 CTA images were rendered for each class using unique text prompts, as illustrated in Fig. 3. Since raw diffusion outputs occasionally contained unrealistic or non-medical samples, we introduced an automated filtering step prior to evaluation. Specifically, we trained a ResNet50 classifier to distinguish CTA images (medical) from ImageNet samples (non-medical). All generated images were passed through this classifier, and only those identified as medical were retained. This step reduced the impact of common failure modes, such as hallucinated anatomical structures, noisy textures, or non-vascular patterns, thereby ensuring that downstream evaluation was conducted on clinically meaningful images. Representative erroneous images, exhibiting unrealistic features and abnormal coloring not found in CTA scans, are shown in Fig. 5. Training time analysis revealed that DiffusionTBAD required approximately 6.3 h, while DiffusionTBAD+ reduced this to 2.9 h. In comparison, Textual Inversion required 11.5 h, and full fine-tuning without and with LoRA took 15.3 and 2.70 h, respectively. The results indicate that DiffusionTBAD+ effectively balances computational efficiency and performance, establishing it as a robust and practical approach for generating realistic and clinically meaningful medical images.

5. Evaluation of synthetic data

The following sections provide information about the evaluation of synthetic data and discuss the corresponding results. In Section 5.1, synthetic images are evaluated using quality and diversity metrics. Sections 5.2 and 5.3 assess the utility of synthetic images in downstream machine learning tasks which includes classification and segmentation tasks. Section 5.4 examines the generalizability of synthetic data, while Section 5.5 details validation by clinician.

5.1. Quantitative metrics

The quantitative evaluation results, summarized in Table 2, demonstrate the superior performance of the proposed DiffusionTBAD model. It achieved a clean-FID score of 50.14, a KID of 0.03, and an MS-SSIM of 0.40, indicating strong fidelity and structural consistency. In contrast, the LoRA variant (DiffusionTBAD+) obtained a clean-FID of 91.95, a KID of 0.10, and an MS-SSIM of 0.31, reflecting a clear performance degradation. Fine-tuning with LoRA also resulted in higher clean-FID

Table 4

Performance comparison of the classifier using real data versus a hybrid dataset in terms of overall accuracy and class-wise accuracy. Blue and red colors indicate an increase and decrease in values, respectively, compared to the baseline using only real images for training.

Setup	No. of Training Images		Overall Accuracy	Class-Wise Accuracy				
	Real	Synthetic		TL	FL	FLT	Background	FL+TL
Real	18233	-	0.6750.02	0.4550.07	0.8550.05	0.2050.00	0.8950.02	0.9850.02
Synthetic	Full fine-tuning	18000	0.3250.01 (10.35)	0.0050.01 (10.45)	0.0050.00 (10.85)	0.4050.00 (10.20)	0.9150.04 (10.02)	0.2950.10 (10.69)
	DiffusionTBAD	18000	0.2450.02 (10.43)	0.0150.01 (10.44)	0.0150.02 (10.84)	0.5350.00 (10.33)	0.3550.05 (10.54)	0.2950.07 (10.69)
	DiffusionTBAD+	18000	0.3850.03 (10.29)	0.4450.11 (10.01)	0.3250.10 (10.53)	0.0050.00 (10.20)	0.5750.07 (10.32)	0.5950.06 (10.39)
Hybrid	Full fine-tuning	18233	0.7250.01 (10.05)	0.6550.07 (10.20)	0.8750.02 (10.20)	0.2050.00 (-)	0.9450.04 (10.05)	0.9450.04 (10.04)
	DiffusionTBAD	18233	0.7450.02 (10.07)	0.7450.09 (10.29)	0.8950.01 (10.04)	0.2050.00 (-)	0.9350.04 (10.04)	0.9350.04 (10.05)
	DiffusionTBAD+	18233	0.7650.00 (10.09)	0.8050.06 (10.35)	0.8750.02 (10.02)	0.2050.00 (-)	0.9650.04 (10.07)	0.9550.03 (10.03)

Table 5

Structure-aware evaluation results (DICE scores) for synthetic images across each class (TL, FL, FLT, Background, and TL+FL) for all methods.

Method	TL	FL	FLT	Background	TL+FL
Full fine-tuning	0.84	0.99	0.63	0.25	0.40
Full fine-tuning + LoRA	0.38	0.42	0.43	0.25	0.33
Textual Inversion	0.25	0.25	0.25	0.25	0.25
DiffusionTBAD	0.75	0.73	0.48	0.49	0.37
DiffusionTBAD+	0.38	0.44	0.51	0.50	0.35

scores, suggesting reduced fidelity, while full fine-tuning achieved a score of 67.16, representing an improvement over LoRA but still falling short of DiffusionTBAD. In both cases, LoRA-based variants consistently exhibited lower fidelity and diversity than their non-LoRA counterparts.

Among all methods, Textual Inversion showed the weakest performance, with the highest clean-FID (148.12), the lowest MS-SSIM (0.20), and the highest KID, underscoring its limited capacity to generate high-quality and diverse images. These findings confirm the effectiveness of DiffusionTBAD in producing diagnostically accurate and structurally consistent medical images, outperforming alternative approaches across multiple evaluation metrics. Domain-specific evaluations further reinforce these results. DiffusionTBAD and full fine-tuning achieved comparable Rad-FID scores (26.5 and 26.2, respectively), with DiffusionTBAD also attaining a moderate BiomedCLIP score of 9.15. Interestingly, Textual Inversion reached a low Rad-FID (21.1) and a high BiomedCLIP score of 12.14, yet its structural fidelity and diversity remained limited, highlighting the necessity of complementary evaluation metrics.

Table 3 presents class-wise CLIP scores for TBAD categories TL, FL, FLT, Background, and FL+TL. DiffusionTBAD achieved competitive per-class scores across all categories, with a mean CLIP score of 28.9, the highest among all methods. These results indicate that DiffusionTBAD generates images that are semantically consistent with their intended classes while outperforming alternative training strategies. In addition to image-level metrics, we incorporated a structure-aware evaluation to assess anatomical consistency of the generated images. Following prior work (Zhang et al., 2018), we computed a proxy metric based on predicted segmentation masks. For each synthetic–real pair, the synthetic image was segmented using a pre-trained 2D UNet, and multi-class DICE scores was computed against the ground-truth mask of its nearest real neighbor. As shown in Table 5, DiffusionTBAD demonstrates strong structure preservation across TL, FL, and Background regions, outperforming all LoRA-based variants and substantially exceeding Textual Inversion. Notably, the very high TL and FL DICE scores for the full fine-tuning baseline likely reflect over-fitting or partial memorization, as full-model tuning is known to reduce sample diversity and increase the risk of instance-level leakage on small medical datasets. In contrast, DiffusionTBAD achieves competitive structure-aware performance while avoiding collapse towards training examples, indicating better generalization and lower memorization risk. Finally, training efficiency is a key consideration. DiffusionTBAD required 6.3 h of training, offering a favorable balance between performance and computational cost. DiffusionTBAD+ reduced training time to 2.9 h but at the expense of fidelity. Full fine-tuning required substantially longer training, up to 15.3 h, without surpassing DiffusionTBAD in overall image quality.

5.2. Effectiveness of TBAD data in classification tasks

In order to evaluate the efficacy of synthetic data in conveying pertinent class information during classifier training, we carried out experiments utilizing 3 distinct sets: real, synthetic, and hybrid. The real set used only actual patient images, while the synthetic set used only generated images. The hybrid set incorporated all samples from both the real training images and both the synthetic set. For the test set, 30 images were selected for each class from real data, to fairly evaluate the classifier’s performance. In the classification task, our goal is to categorize each slice into one of the following categories: TL, FL, FLT, Background, and FL+TL. We utilize a pre-trained EfficientNet-B4 model for classification, with only the parameters of the classification head randomly initialized and updated during fine-tuning. We fine-tune the classifiers using categorical cross-entropy loss and the Adam optimizer with a learning rate of 0.01. Model fine-tuning continues until early stopping criteria are met based on validation set performance, with a patience threshold of three epochs. Each classifier is subsequently evaluated using the same real testing set. The evaluation metrics include overall accuracy and class-wise accuracy. The results in Table 4 indicate that incorporating synthetic data, whether from Full fine-tuning, DiffusionTBAD, or DiffusionTBAD+, together with real data, improves classification performance. However, the proposed DiffusionTBAD and DiffusionTBAD+ approaches achieve higher overall and class-wise accuracy while requiring significantly less training time.

5.3. Effectiveness of TBAD data in segmentation tasks

In the segmentation task, the objective is to accurately delineate regions corresponding to TL, FL, and FLT. The U-Net model was trained under two settings: (1) a pre-trained configuration, in which the model was first pre-trained on synthetic data generated by various approaches and subsequently fine-tuned on real data; and (2) a non-pretrained configuration, in which the model was trained exclusively on 100% real data. While synthetic data from most generative approaches increased the overall mean Intersection over Union (IoU) and DICE scores, it caused a decline in the class-wise mean for FL. However, synthetic data generated by DiffusionTBAD+ proved to be an exception. The results, presented in Table 6, demonstrate that pre-training the U-Net on synthetic data from DiffusionTBAD+, followed by fine-tuning with real data, improved the mean IoU by 4% and the mean DICE score by 3%. Notable improvements were observed in segmentation performance for the TL and FLT classes, with DICE score increases of 3% and 10%, respectively, compared to training without pre-training. Similarly, pre-training the U-Net on synthetic data from DiffusionTBAD, followed by fine-tuning with real data, resulted in a 2% improvement in mean IoU and 3% increase in the mean DICE score. This approach also yielded DICE score improvements for the TL and FLT classes by 3% and 5%, respectively, compared to the non-pretrained setting. However, a slight 1% decrease in the DICE score for the FL class was observed. Overall, the significant improvement in segmentation accuracy for the FLT class, despite the limited availability of real data images, highlights the utility of synthetic data in addressing dataset imbalances and enhancing overall model performance.

Table 6

Performance comparison of the U-Net model with and without synthetic data pre-training, reporting metrics such as mean IoU, mean DICE scores, and class-wise IoU and DICE scores.

Pre-training	IoU				DICE			
	Mean	TL	FL	FLT	Mean	TL	FL	FLT
Full fine tuning	0.68(\uparrow 0.02)	0.74(\uparrow 0.03)	0.70(\downarrow 0.04)	0.27(\uparrow 0.09)	0.77(\uparrow 0.02)	0.85(\uparrow 0.02)	0.82(\downarrow 0.03)	0.40(\uparrow 0.10)
Full fine tuning + LoRA	0.68 (\uparrow 0.02)	0.76 (\uparrow 0.05)	0.73 (\downarrow 0.01)	0.23 (\uparrow 0.04)	0.76 (\uparrow 0.01)	0.86 (\uparrow 0.03)	0.84 (\downarrow 0.01)	0.33 (\uparrow 0.03)
Textual Inversion	0.68 (\uparrow 0.02)	0.75 (\uparrow 0.04)	0.74 (-)	0.22 (\uparrow 0.03)	0.76 (\uparrow 0.01)	0.86 (\uparrow 0.03)	0.85 (-)	0.33 (\uparrow 0.02)
DiffusionTBAD(ours)	0.68 (\uparrow 0.02)	0.76 (\uparrow 0.05)	0.72 (\downarrow 0.02)	0.23 (\uparrow 0.04)	0.76 (\uparrow 0.01)	0.86 (\uparrow 0.03)	0.84 (\downarrow 0.01)	0.35 (\uparrow 0.05)
DiffusionTBAD+(ours)	0.70 (\uparrow 0.04)	0.77 (\uparrow 0.06)	0.75 (\uparrow 0.01)	0.28 (\uparrow 0.09)	0.78 (\uparrow 0.03)	0.87 (\uparrow 0.04)	0.85 (-)	0.40 (\uparrow 0.10)
No Pre-training	0.66	0.71	0.74	0.19	0.75	0.83	0.85	0.30

Table 7

Confusion matrices summarizing participants' classifications of real and synthetic images. Part 1 involved paired evaluation of real and synthetic images (n = 8 participants), while Part 2 involved class-wise realism assessment (n = 3 participants). Rows indicate predicted labels and columns indicate the actual image type.

Part		Real(Actual)	Synthetic(Actual)
1	Predicted Real	31	35
	Predicted Synthetic	9	5
2	Predicted Real	6	8
	Predicted Synthetic	5	4

5.4. Assessment of generalization and adaptation of synthetic TBAD data

To assess the generalization and robustness of the proposed pipeline on external data with heterogeneous imaging conditions, we conducted experiments using an open-source TBAD dataset (Mayer et al., 2024), which includes 40 CT scans without false lumen thrombosis (FLT). Three separate EfficientNet-B4 models were trained under distinct data augmentation strategies: (i) no augmentation, (ii) conventional augmentation, and (iii) augmentation with synthetic data. Among these approaches, the highest performance was achieved using a hybrid strategy that combined synthetic and real data. The resulting classification accuracies were 70% without augmentation, 76% with conventional augmentation, and 79% with synthetic data augmentation, representing a 9% absolute improvement over the baseline. These results suggest that the generated synthetic data can effectively complement or even substitute real data in scenarios where data sharing is restricted.

5.5. Subjective image evaluation through clinical expertise

A qualitative assessment was conducted with eight healthcare professionals representing diverse clinical specialties and experience levels. Among the participants, two had less than 3 years of clinical experience, while the remaining six had 11 or more years of practice. The cohort comprised four cardiologists, one vascular surgeon, one radiographer, one medical trainee, and one radiologist. Each participant was presented with a series of synthetic images and asked to evaluate their clinical plausibility, visual realism, and potential diagnostic utility. The evaluation consisted of two parts: the first part involved paired and independent image assessments, where participants compared images side-by-side and also rated images individually. The second part focused on class-wise realism, where participants assessed the visual realism of images grouped by diagnostic category. Data collection was performed via a structured online questionnaire.

In the first section, all eight participants reviewed ten images (five real and five synthetic). As shown in Table 7, the results indicate a high recall of 0.75 for identifying real images, suggesting that most real images were correctly recognized. However, the precision was only 0.47, meaning that just 47% of the images labeled as real were actually real. In other words, 53% of the time, participants mistook synthetic images for real ones. This high false-positive rate indicates that the synthetic images exhibited a high degree of visual realism and were able to convincingly mimic real anatomical features. In the second section, three participants completed an extended survey consisting of

ten class-specific image pairs. Each pair contained one real and one synthetic image depicting a distinct anatomical feature (e.g., a true lumen). Participants were asked to identify the real image and assess the plausibility of the synthetic one. As summarized in Table 7, the recall dropped to 0.54, and precision further declined to 0.42. This means that only 42% of the images labeled as real were actually real, with 58% being synthetic. These results reinforce the high visual plausibility of the generated images, as even experienced clinical observers were frequently misled and struggled to reliably distinguish real from synthetic content.

The results suggest that the synthetic images achieved a noteworthy level of realism. However, several limitations should be noted. The small number of participants in the second part (n=3) limits generalizability. The lower accuracy in identifying real images (59%) in class-specific comparisons also suggests that context and image complexity may affect perception. Overall, the results are promising, showing that synthetic images can approach real-world realism. Still, further validation with larger cohorts and clinically grounded tasks is necessary.

6. Discussion

In this study, we show that the proposed DiffusionTBAD and DiffusionTBAD+ approaches can generate diagnostically accurate synthetic TBAD CTA images, even when trained on a small and highly imbalanced dataset. Our dataset spanned from as few as 121 samples in the FLT class to nearly 14,000 samples in the FL+TL class. Despite this significant class imbalance, incorporating synthetic images led to notable improvements in downstream model performance, with classification accuracy increasing by 9% and segmentation accuracy by 4%. The synthetic images were evaluated using both qualitative assessments and quantitative metrics, providing complementary perspectives on their realism and utility. Furthermore, our approach reduces computational costs compared to full network fine-tuning, making it particularly suitable for low-data scenarios and enabling scalable generation of high-quality medical images on consumer-grade GPUs. While the diffusion-based pipeline shows strong potential for producing realistic TBAD data, several limitations and challenges must be carefully considered in interpreting these results.

6.1. Metric limitations

Prior studies (Stein et al., 2024; Xing et al., 2023) have highlighted that widely used metrics such as FID are often ill-suited for medical imaging due to unrealistic assumptions of normality, prompting the adoption of alternatives like clean-FID, KID, and RadFID. In our analysis (Fig. 4(a), Fig. 4(b)), clean-FID and KID demonstrated greater stability than FID; however, their scores did not consistently correlate with downstream task performance. Similarly, domain-specific metrics showed limited alignment with human judgment (Konz et al., 2024; Woodland et al., 2024), underscoring challenges in evaluating synthetic medical images using conventional quantitative measures.

LoRA-based variants generally exhibited lower fidelity and diversity metrics compared to non-LoRA models. Nevertheless, DiffusionTBAD+ achieved superior segmentation performance and comparable classification results (Tables 4 and 6), demonstrating that lower perceptual

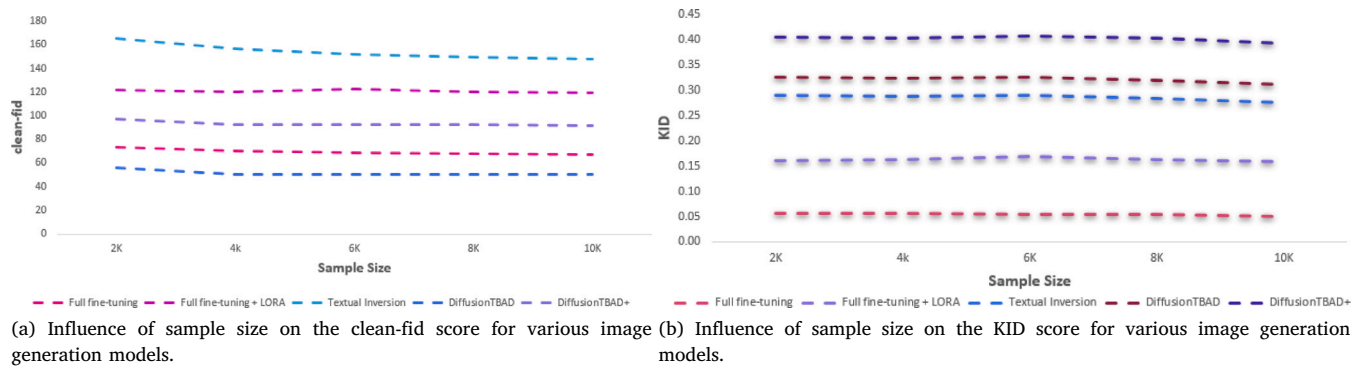


Fig. 4. Comparison of clean-fid and KID scores for various image generation models.

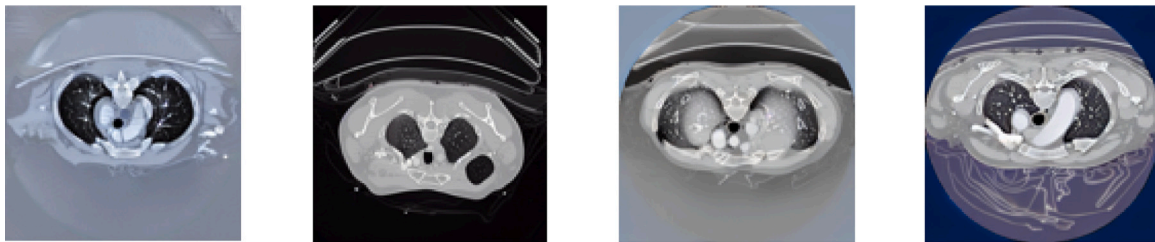


Fig. 5. Examples of synthetic CTA images exhibiting failure modes in unfiltered diffusion outputs. These images illustrate typical issues, including unrealistic vascular morphology, anatomical discontinuities, abnormal coloring, and texture artifacts.

scores do not necessarily compromise downstream utility. This observation emphasizes the limitations of current quantitative metrics in capturing task-relevant features and aligns with prior work (Xing et al., 2023) suggesting that high fidelity and diversity are not strictly required for clinically useful synthetic medical images. Collectively, these findings reveal a persistent gap between numerical evaluation scores and practical clinical utility. Although this study focused on the usability of synthetic images for classification and segmentation, the identification of reliable evaluation metrics remains an unresolved challenge (Stein et al., 2023). Advancing this area will be essential for the safe and effective integration of synthetic data into medical imaging workflows.

6.2. Expert opinion analysis

Subjective evaluations offered complementary insights. Experts frequently rated the synthetic images as realistic, with some even misclassify them as real. Nevertheless, occasional failure cases were observed, including non-anatomical structures, noisy textures, and implausible vascular formations. To address these issues, we incorporated an automated filtering step (Section 4), which improved dataset reliability. Even so, these artifacts highlight the need for additional strategies to detect and mitigate residual errors in generative pipelines. These results highlight both the promise and the limitations of diffusion-based synthesis: on one hand, the models can generate realistic medical images that experts may mistake for real; on the other hand, occasional artifacts and errors demonstrate that synthetic data cannot fully replace real clinical data. Importantly, expert-based evaluations remain constrained by the limited availability of qualified reviewers, introducing uncertainty into the assessment of medical realism.

6.3. Spatial constraints and deployment considerations

While the use of 2D images has proven beneficial for TBAD diagnosis (Abaid et al., 2024a), reliance on 2D data inherently limits the representation of spatial complexity present in 3D volumetric CTA. This

limitation may affect the accuracy of prognostic models for intricate anatomical structures. Moreover, although generative models demonstrate considerable promise, their clinical deployment must proceed cautiously. Careful implementation of bias detection and mitigation strategies is essential to ensure fairness, reliability, and safe application across diverse patient populations.

7. Conclusion and future work

In conclusion, our proposed framework demonstrates a robust capability to generate realistic synthetic CTA data, offering substantial value for augmenting small medical datasets. The synthetic TBAD CTA images were rigorously evaluated through various metrics and clinician assessments, yielding promising results. This study shows that adapting pre-trained diffusion models via few-shot learning can effectively generate synthetic medical images, which can be utilized in self-supervised pre-training, significantly enhancing TBAD classification and segmentation performance in data-limited environments. Additionally, the methodology successfully addressed domain shift issues, adapting to diverse imaging conditions with up to a 9% improvement when the model was tested on data from different scanner. By making our synthetic dataset publicly available, we aim to address the challenge of acquiring large imaging datasets while mitigating privacy concerns.

As a direction for future work, extending the proposed framework towards volumetric or slice-conditioned diffusion models represents a natural progression beyond the current 2D, slice-independent setting. Such extensions could enable explicit modeling of inter-slice anatomical coherence and more accurate representation of complex spatial relationships, better aligning synthetic data generation with clinical workflows. Furthermore, incorporating clinically relevant factors such as patient demographics, comorbidities, and disease stage, which are known to influence TBAD outcomes (Onitsuka et al., 2004), may improve fairness and reduce bias in downstream learning tasks. Finally, integrating temporal modeling to simulate longitudinal disease progression could further enhance the clinical utility of diffusion-based synthetic data for prognosis and treatment planning.

CRediT authorship contribution statement

Ayman Abaid: Writing – original draft, Visualization, Validation, Resources, Formal analysis, Data curation, Conceptualization. **Muhammad Ali Farooq:** Writing – review & editing, Formal analysis, Data curation. **Niamh Hynes:** Writing – review & editing, Validation. **Peter Corcoran:** Writing – review & editing, Supervision, Funding acquisition. **Ihsan Ullah:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

This work is supported with the financial support of Science Foundation Ireland, Ireland under Grant Agreement No SFI/12/RC/2289 P2 for the Insight SFI Research Centre for Data Analytics, University of Galway, Ireland and by ADAPT - Centre for Digital Content Technology, Enterprise Ireland. This publication has also emanated from research supported by Taighde Éireann – Research Ireland under Grant No. 18/CRT/6223.

Data availability

The synthetic images generated in this study are publicly available at <https://github.com/AymanAbaid/DiffusionTBAD>.

References

- Abaid, A., Farooq, M.A., Hynes, N., Corcoran, P., Ullah, I., 2024a. Synthesizing CTA image data for Type-B Aortic Dissection using stable diffusion models. arXiv preprint arXiv:2402.06969.
- Abaid, A., Ilancheran, S., Iqbal, T., Hynes, N., Ullah, I., 2024b. Exploratory analysis of Type B Aortic Dissection (TBAD) segmentation in 2D CTA images using various kernels. *Comput. Med. Imaging Graph.* 118, 102460.
- Adams, L.C., Busch, F., Truhn, D., Makowski, M.R., Aerts, H.J., Bresslem, K.K., 2023. What does DALL-E 2 know about radiology? *J. Med. Internet Res.* 25, e43110.
- Ali, H., Murad, S., Shah, Z., 2022. Spot the fake lungs: Generating synthetic medical images using neural diffusion models. In: *Irish Conference on Artificial Intelligence and Cognitive Science*. Springer, pp. 32–39.
- Anaya-Isaza, A., Mera-Jiménez, L., Zequera-Diaz, M., 2021. An overview of deep learning in medical imaging. *Informatics Med. Unlocked* 26, 100723.
- Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A., 2018. Demystifying mmd gans. arXiv preprint arXiv:1801.01401.
- Cao, L., Shi, R., Ge, Y., Xing, L., Zuo, P., Jia, Y., Liu, J., He, Y., Wang, X., Luan, S., 2019. Fully automatic segmentation of Type B Aortic Dissection from CTA images enabled by deep learning. *Eur. J. Radiol.* 121, 108713.
- Cheng, J., Tian, S., Yu, L., Ma, X., Xing, Y., 2020. A deep learning algorithm using contrast-enhanced computed tomography (CT) images for segmentation and rapid automatic detection of aortic dissection. *Biomed. Signal Process. Control.* 62, 102145.
- Chuquicusma, M.J., Hussein, S., Burt, J., Bagci, U., 2018. How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis. In: 2018 IEEE 15th International Symposium on Biomedical Imaging. ISBI 2018, IEEE, pp. 240–244.
- Dhariwal, P., Nichol, A., 2021. Diffusion models beat gans on image synthesis. *Adv. Neural Inf. Process. Syst.* 34, 8780–8794.
- Farooq, M.A., Abaid, A., Ullah, I., Corcoran, P., 2024. A comparative study on diffusion sampling methods across diverse medical imaging modalities. In: *Proceedings of the Asian Conference on Computer Vision*. pp. 193–206.
- Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G., Cohen-Or, D., 2022. An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618.
- Ghorbani, A., Natarajan, V., Coz, D., Liu, Y., 2020. Dermgan: Synthetic generation of clinical skin images with pathology. In: *Machine Learning for Health Workshop*. PMLR, pp. 155–170.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. *Commun. ACM* 63 (11), 139–144.
- Güven, S.A., Talu, M.F., 2023. Brain MRI high resolution image creation and segmentation with the new GAN method. *Biomed. Signal Process. Control.* 80, 104246.
- Hahn, L.D., Mistelbauer, G., Higashigaito, K., Koci, M., Willemink, M.J., Sailer, A.M., Fischbein, M., Fleischmann, D., 2020. CT-based true-and false-lumen segmentation in Type B Aortic Dissection using machine learning. *Radiol.: Cardiothorac. Imaging* 2 (3), e190179.
- Hessel, J., Holtzman, A., Forbes, M., Bras, R.L., Choi, Y., 2021. Clipscore: A reference-free evaluation metric for image captioning. arXiv preprint arXiv:2104.08718.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* 30.
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* 33, 6840–6851.
- Ho, J., Salimans, T., 2022. Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598.
- Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 2021. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.
- Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., Kumar, S., 2024. Rethinking fid: Towards a better evaluation metric for image generation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9307–9315.
- Jung, J., Oh, H.M., Jeong, G., Kim, T., Koo, H.J., Lee, J., Yang, D.H., 2024. ZOZI-seg: A transformer and unet cascade network with zoom-out and zoom-in scheme for aortic dissection segmentation in enhanced CT images. *Comput. Biol. Med.* 108494.
- Kang, M., Zhu, J., Zhang, R., Park, J., Shechtman, E., Paris, S., Park, T., 2023. Scaling up gans for text-to-image synthesis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10124–10134.
- Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hachailoglu, I., Merhof, D., 2023. Diffusion models in medical imaging: A comprehensive survey. *Med. Image Anal.* 102846.
- Konz, N., Osuala, R., Verma, P., Chen, Y., Gu, H., Dong, H., Chen, Y., Marshall, A., Garrucho, L., Kushibar, K., et al., 2024. Fréchet Radiomic Distance (FRD): A versatile metric for comparing medical imaging datasets. arXiv preprint arXiv:2412.01496.
- Mayer, C., Pepe, A., Hossain, S., Karner, B., Arnreiter, M., Kleesiek, J., Schmid, J., Janisch, M., Hannes, D., Fuchsjäger, M., 2024. Type B Aortic Dissection CTA collection with true and False Lumen Expert annotations for the development of AI-based algorithms. *Sci. Data* 11 (1), 596.
- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M., 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint arXiv:2112.10741.
- Onitsuka, S., Akashi, H., Tayama, K., Okazaki, T., Ishihara, K., Hiromatsu, S., Aoyagi, S., 2004. Long-term outcome and prognostic predictors of medically treated acute Type B Aortic Dissections. *Ann. Thorac. Surg.* 78 (4), 1268–1273.
- Osuala, R., Skorupko, G., Lazrak, N., Garrucho, L., Garcia, E., Joshi, S., Jouide, S., Rutherford, M., Prior, F., Kushibar, K., et al., 2023. medigan: a Python library of pre-trained generative models for medical image synthesis. *J. Med. Imaging* 10 (6), 061403.
- Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., 2023. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Phys. Med. Biol.* 68 (10), 105004.
- Parmar, G., Zhang, R., Zhu, J., 2022. On aliased resizing and surprising subtleties in gan evaluation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11410–11420.
- Pepe, A., Li, J., Rolf-Pissarczyk, M., Gsaxner, C., Chen, X., Holzapfel, G.A., Egger, J., 2020. Detection, segmentation, simulation and visualization of aortic dissections: a review. *Med. Image Anal.* 65, 101773.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*. PMLR, pp. 8748–8763.
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I., 2021. Zero-shot text-to-image generation. In: *International Conference on Machine Learning*. Pmlr, pp. 8821–8831.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. CVPR, pp. 10684–10695.

- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K., 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M., 2022. Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10.
- Sailer, A.M., van Kuijk, S.M., Nelemans, P.J., Chin, A.S., Kino, A., Huininga, M., Schmidt, J., Mistelbauer, G., Baeumler, K., Chiu, P., 2017. Computed tomography imaging features in acute uncomplicated stanford Type-B Aortic Dissection predict late adverse events. *Circ.: Cardiovasc. Imaging* 10 (4), e005709.
- Sennrich, R., 2015. Neural machine translation of rare words with subword units. arXiv preprint [arXiv:1508.07909](https://arxiv.org/abs/1508.07909).
- Stein, G., Cresswell, J., Hosseinzadeh, R., Sui, Y., Ross, B., Villecroze, V., Liu, Z., Caterini, A.L., Taylor, E., Loaiza-Ganem, G., 2023. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (Eds.), In: *Advances in Neural Information Processing Systems*, vol. 36, Curran Associates, Inc., pp. 3732–3784, URL https://proceedings.neurips.cc/paper_files/paper/2023/file/0bc795afae289ed465a65a3b4b1f4eb7-Paper-Conference.pdf.
- Stein, G., Cresswell, J., Hosseinzadeh, R., Sui, Y., Ross, B., Villecroze, V., Liu, Z., Caterini, A.L., Taylor, E., Loaiza-Ganem, G., 2024. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. *Adv. Neural Inf. Process. Syst.* 36.
- Wang, Z., Simoncelli, E.P., Bovik, A.C., 2003. Multiscale structural similarity for image quality assessment. In: *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, vol. 2, Ieee, pp. 1398–1402.
- Wobben, L.D., Codari, M., Mistelbauer, G., Pepe, A., Higashigaito, K., Hahn, L.D., Mastrodicasa, D., Turner, V.L., Hinostroza, V., Bäuml, K., 2021. Deep learning-based 3D segmentation of true lumen, false lumen, and false lumen thrombosis in Type-B Aortic Dissection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, pp. 3912–3915.
- Woodland, M., Castelo, A., Al Taie, M., Albuquerque Marques Silva, J., Eltaher, M., Mohn, F., Shieh, A., Kundu, S., Yung, J.P., Patel, A.B., et al., 2024. Feature extraction for generative medical imaging evaluation: New evidence against an evolving trend. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 87–97.
- Xing, X., Felder, F., Nan, Y., Papanastasiou, G., Walsh, S., Yang, G., 2023. You don't have to be perfect to be amazing: Unveil the utility of synthetic images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 13–22.
- Yao, Z., Xie, W., Zhang, J., Dong, Y., Qiu, H., Yuan, H., Jia, Q., Wang, T., Shi, Y., Zhuang, J., 2021. Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection. *Front. Physiol.* 12, 732711.
- Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., et al., 2025. A multimodal biomedical foundation model trained from fifteen million image–text pairs. *NEJM AI* 2 (1), A1oa2400640.
- Zhang, Z., Yang, L., Zheng, Y., 2018. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 9242–9251.