Article

Progressive Disease Image Generation with Ordinal-Aware Diffusion Models

Meryem Mine Kurt [*] , Ümit Mert Çağlar , Alptekin Temizel

Posted Date: 5 September 2025

doi: 10.20944/preprints202509.0426.v1

Keywords: medical image synthesis; diffusion models; ulcerative colitis; disease progression modeling; ordinal classification; endoscopy; computer-aided diagnosis; Mayo Endoscopic Score

Article

Progressive Disease Image Generation with Ordinal-Aware Diffusion Models

Meryem Mine Kurt [1,2,*], Ümit Mert Çağlar [2] and Alptekin Temizel [2,]

[1] ASELSAN Inc., Ankara 06200, Turkey
[2] Graduate School of Informatics, Middle East Technical University, Ankara 06800, Turkey
[*] Correspondence: megundogan@aselsan.com

Abstract

Ulcerative Colitis (UC) lacks longitudinal visual data, which limits both disease progression modeling and the effectiveness of computer-aided diagnosis systems. These systems are further constrained by sparse intermediate disease stages and the discrete nature of the Mayo Endoscopic Score (MES). Meanwhile, synthetic image generation has made significant advances. In this paper, we propose novel ordinal embedding architectures for conditional diffusion models to generate realistic UC progression sequences from cross-sectional endoscopic images. By adapting Stable Diffusion v1.4 with two specialized ordinal embeddings, Basic Ordinal Embedder using linear interpolation and Additive Ordinal Embedder modeling cumulative pathological features, our framework converts discrete MES categories into continuous progression representations. The Additive Ordinal Embedder outperforms alternatives, achieving superior distributional alignment (CMMD 0.4137, recall 0.6331) and disease consistency comparable to real data (Quadratic Weighted Kappa 0.8425, UMAP Silhouette Score 0.0571). The generated sequences exhibit smooth transitions between severity levels while maintaining anatomical fidelity. This work establishes a foundation for transforming static medical datasets into dynamic progression models and demonstrates that ordinal-aware embeddings can effectively capture disease severity relationships, enabling synthesis of underrepresented intermediate stages. These advances support applications in medical education, diagnosis, and synthetic data generation.

Keywords: medical image synthesis; diffusion models; ulcerative colitis; disease progression modeling; ordinal classification; endoscopy; computer-aided diagnosis; Mayo Endoscopic Score

1. Introduction

Computer-aided diagnosis (CAD) systems have revolutionized endoscopic practice by enhancing the accuracy of lesion detection, standardizing diagnostic criteria, and aiding in clinical decision-making [1]. However, developing robust AI models for endoscopic applications remains fundamentally constrained by limited comprehensive datasets that adequately represent the complete spectrum of disease progression. This limitation becomes particularly pronounced when considering the emerging paradigm of personalized disease trajectory prediction, which represents a transformative frontier in medical AI [2–4]. Such predictive capabilities, successfully demonstrated across various medical domains through longitudinal symptom monitoring [5–8], remain largely unexplored in endoscopic applications. This challenge is particularly acute in inflammatory bowel diseases such as ulcerative colitis (UC), where current scoring systems inadequately represent the continuous nature of pathological evolution.

Ulcerative colitis is a chronic inflammatory bowel disease characterized by inflammation and ulceration of the colonic mucosa. The severity of UC is generally assessed using the Mayo Endoscopic Score (MES), which ranges from 0 to 3 [9]. While this scoring system provides a useful standardized clinical assessment, it imposes artificial discrete categories on what is inherently a continuous pathological process [10–12]. Although UC severity exists along a continuum, medical image datasets commonly consist of isolated samples annotated with discrete MES levels and controlled image sets capturing the same anatomical region as it progresses through different severity stages are often lacking. In addition, current computer-aided endoscopic systems primarily focus on static image analysis, largely due to this lack of longitudinal data. While the generation of longitudinal disease progression has been explored in other medical domains [8,13,14], there remains a research gap regarding endoscopic applications of disease progression. Such longitudinal data would be valuable for developing comprehensive visual training materials and simulation tools, helping clinicians better recognize subtle transitions between MES stages and gain deeper insight into disease dynamics on a patient-specific level [15].

Recent advances in generative artificial intelligence, particularly diffusion models, have demonstrated remarkable capabilities in medical image synthesis [16,17]. Unlike Generative Adversarial Networks (GANs), diffusion models offer better training stability, sample diversity, and generation quality, making them particularly suitable for medical applications requiring high accuracy and reliability [18]. On the other hand, since only discrete MES data are available, current generative models can only synthesize discrete UC classes, leaving a significant gap in data generation.

This study addresses these fundamental limitations by introducing novel ordinal class embedding architectures specifically designed for medical image generation. Our approach converts cross-sectional endoscopic datasets into continuous progression models, enabling the synthesis of realistic intermediate disease stages that are often underrepresented in clinical data. The key innovation lies in developing specialized embeddings that capture the cumulative nature of pathological features, recognizing that higher severity levels encompass the characteristics of lower levels while also introducing additional pathological manifestations.

The primary contributions are:

- Development of ordinality-aware embedding strategies that model disease progression relationships, rather than treating severity levels as independent categories.
- Adaptation of state-of-the-art diffusion models for medical image synthesis with domain-specific modifications.
- Generation of realistic synthetic longitudinal datasets to unlock new possibilities for image-based UC trajectory analysis.
- Exploration of integration with generative data augmentation techniques to improve deep learning model training using synthetic data.

2. Related Work

2.1. Computer-Aided Diagnosis in Endoscopy

Computer-aided diagnosis has emerged as a transformative technology in endoscopic practice, with deep learning models demonstrating exceptional performance in detecting and classifying gastrointestinal lesions [19]. The Mayo Endoscopic Score (MES) serves as the standard index for evaluating UC disease activity, grading inflammation severity on a scale from 0 to 3: 0 indicates normal mucosa; 1 corresponds to mild disease with erythema and decreased vascular pattern; 2 reflects moderate disease with marked erythema and friability; and 3 denotes severe disease characterized by spontaneous bleeding and ulceration [20].

Existing CAD systems predominantly rely on static image analysis and require extensive datasets covering the full pathological spectrum [21]. Recent advances in edge computing have enabled real-time AI deployment for endoscopic applications [22], yet these systems remain focused on static classification tasks, limiting their ability to model temporal disease evolution.

2.2. Evolution of Generative Models and Medical Image Synthesis

The progression from traditional generative approaches to modern diffusion models represents a significant advancement in medical image synthesis. The early frameworks included Variational Autoencoders (VAEs) [23] and autoregressive models such as PixelRNN [24], which established foundational probabilistic approaches. Generative Adversarial Networks (GANs) [25] subsequently revolutionized the field through adversarial training between generator and discriminator networks, leading to sophisticated architectures including StyleGAN variants [26,27].

The comparative analysis by Dhariwal and Nichol [28] revealed that diffusion models outperform GANs in image quality and diversity while providing superior training stability. Given these advantages, the research on diffusion-based approaches for medical imaging applications is accelerated, leveraging models reliability and consistency.

GAN-based medical image synthesis has been extensively applied across multiple imaging modalities, including radiography, computed tomography, magnetic resonance imaging, and histopathology [18]. Çağlar et al. [29] specifically addressed the classification of ulcerative colitis by combining StyleGAN2-ADA with active learning, demonstrating that class-specific synthetic data augmentation enhances classification accuracy when training data is limited, with separate models trained for each Mayo Endoscopic Score category.

Despite these achievements, GANs face substantial limitations in clinical applications, including mode collapse, training instability, and limited sample diversity [18]. These issues are particularly problematic in medical contexts where subtle image features carry significant clinical implications.

Recent diffusion model applications in medical imaging have demonstrated superior capabilities for preserving anatomical details and maintaining clinical validity. Kazerouni et al. [17] provide a comprehensive survey categorizing diffusion models across medical applications including image translation, reconstruction, and generation. Zhang et al. [30] introduced texture-preserving diffusion models for cross-modal synthesis, while Müller-Franzes et al. [18] developed Medfusion, a conditional latent diffusion model specifically designed for medical image generation that outperforms state-of-the-art GANs across multiple medical domains.

2.3. Ordinal Relationships in Medical Image Generation

A critical limitation of existing generative approaches is their treatment of disease severity as independent categorical variables rather than ordinal progressions. Takezaki and Uchida [31] introduced an Ordinal Diffusion Model by incorporating ordinal relationship loss functions to maintain severity class relationships during generation. This framework improves distribution estimation through interpolation/extrapolation capabilities, making it particularly applicable to medical conditions with well-defined severity scales.

However, current ordinal approaches primarily focus on discrete classification improvements rather than generating smooth progression sequences that could support medical education and training applications. This represents a significant gap where comprehensive disease progression modeling remains underexplored.

2.4. Disease Progression Modeling

Traditional disease progression modeling has relied primarily on longitudinal clinical data and statistical approaches that fail to capture visual manifestations of disease evolution [8]. Recent innovations have explored temporal medical data generation through video synthesis frameworks. Cao et al. [8] introduced Medical Video Generation (MVG), a zero-shot framework for disease progression simulation enabling manipulation of disease-specific characteristics without requiring pre-existing video datasets.

Li et al. [32] developed Endora, combining spatial-temporal video transformers with latent diffusion models for high-resolution endoscopy video synthesis. While these approaches demonstrate promise for temporal content generation, they typically require substantial computational resources and complex training procedures that may limit clinical applicability.

In particular, for UC disease progression, the discrete nature of current scoring systems fails to capture the continuous pathological evolution characteristics [33]. The scarcity of intermediate disease states in clinical datasets restricts the development of comprehensive educational resources and training materials for medical practitioners.

Current approaches to UC image generation have primarily focused on classification tasks rather than progression modeling. While GAN-based methods [29] and ordinal diffusion approaches [31] have shown promise for discrete severity classification, the generation of smooth, clinically meaningful progression sequences remains an unsolved challenge that could significantly benefit medical education and computer-aided diagnosis systems.

3. Materials and Methods

This work presents a conditional diffusion framework for synthesizing realistic disease progression sequences from cross-sectional endoscopic data. The approach employs specialized ordinal class embeddings that capture the progressive nature of disease severity, enabling generation of smooth transitions between discrete MES levels while maintaining anatomical consistency. Two novel embedding strategies are introduced: the Basic Ordinal Embedder, which facilitates smooth transitions via linear interpolation between discrete severity classes, and the Additive Ordinal Embedder, which explicitly models the cumulative nature of pathological features. By adapting the Stable Diffusion v1.4 architecture with medical specific modifications and replacing general purpose text conditioning with domain specific ordinal embeddings, the framework enables precise control over disease severity while maintaining high image quality and training stability.

3.1. Model Architecture

The proposed framework adapts Stable Diffusion v1.4 [34] for medical image generation by replacing the standard CLIP text encoder with domain-specific ordinal embeddings. The architecture comprises three key components: (1) a frozen Variational Autoencoder (VAE) for efficient latent space operations, (2) a fine-tuned U-Net backbone for the denoising process, and (3) specialized ordinal class embeddings for medical conditioning. The training pipeline is demonstrated in Figure 1 and inference pipeline is presented in Figure 2.

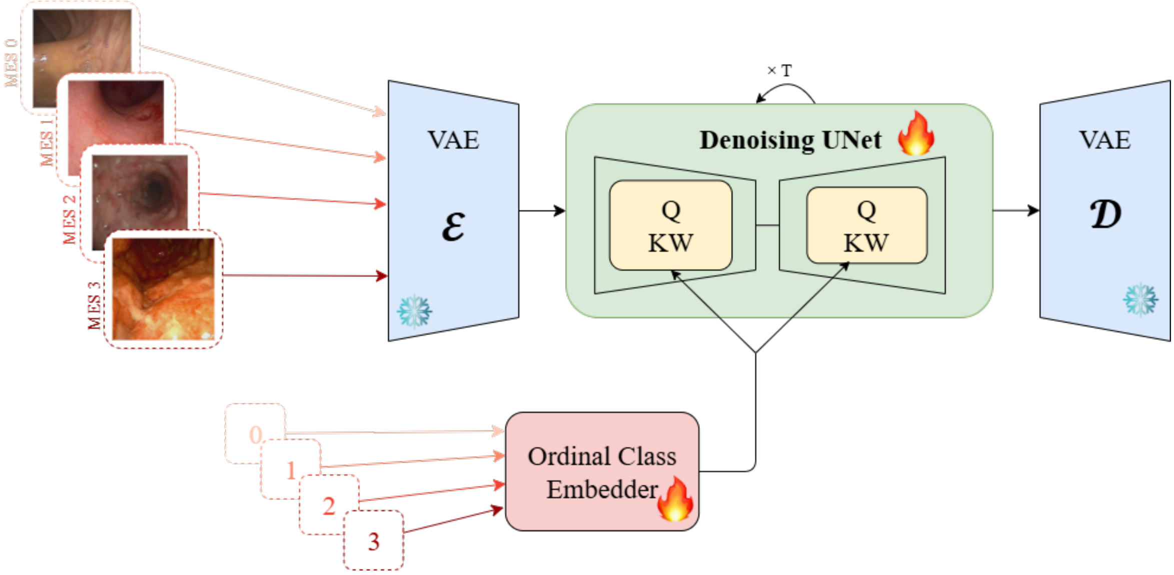


Figure 1. Training pipeline showing the encoding of endoscopic images with Mayo Endoscopic Scores having discrete values as in LIMUC dataset. The denoising U-Net is conditioned on ordinal class embeddings that capture disease severity relationships.

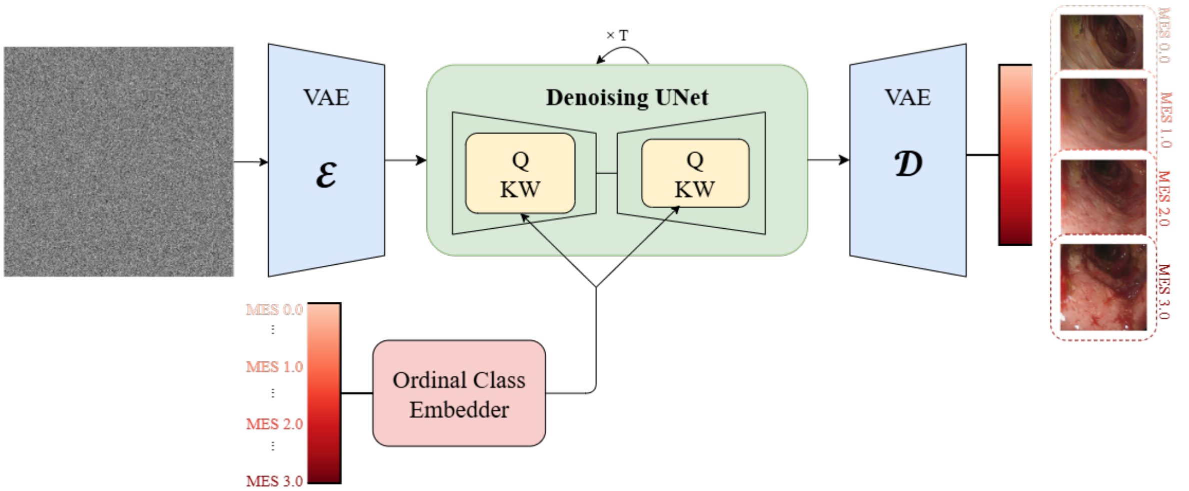


Figure 2. Generation pipeline demonstrating the synthesis process that is able to condition on a continuous values of MES using Ordinal Class Embedder, allowing generation of intermediate disease stages.

The denoising process follows the standard diffusion objective with ordinal class conditioning:

$$\mathcal{L} = \mathbb{E}_{\mathbf{z}_0,\epsilon,t}\left[||\epsilon - \epsilon_\theta(\mathbf{z}_t, t, c)||^2\right] \tag{1}$$

where c represents ordinal class conditioning, $\mathbf{z}_t$ is the noisy latent at timestep t, ϵ is added noise, and ϵ_θ is the predicted noise.

The **Basic Ordinal Embedder (BOE)** creates learnable embeddings for each integer Mayo Endoscopic Score of 0 to 3 and performs linear interpolation for fractional values:

$$\mathbf{E}(y) = \begin{cases} (1-\alpha)\mathbf{E}[y] + \alpha\mathbf{E}[y+1] & \text{if } y \notin \mathbb{Z} \\ \mathbf{E}[y] & \text{if } y \in \mathbb{Z} \end{cases} \tag{2}$$

where $y \in [0, K-1]$ denotes a class label as K is the total number of ordinal classes, $\alpha = y - \lfloor y \rfloor$ and $\mathbf{E}_i \in \mathbb{R}^{K\times 768}$ are learnable embeddings.

The **Additive Ordinal Embedder (AOE)** explicitly models the cumulative nature of pathological features by initializing embeddings monotonically:

$$\mathbf{E}_i = \mathbf{E}_0 + \sum_{j=1}^{i} \mathbf{\Delta}_j, \quad i = 0, 1, 2, 3 \tag{3}$$

where $\mathbf{E}_0$ represents normal mucosa and $\mathbf{\Delta}_j \in \mathbb{R}^{768}$ represents additive pathological features at severity level j.

3.2. Dataset and Preprocessing

The study used the Labeled Images for Ulcerative Colitis (LIMUC) dataset [35], comprising 11,276 endoscopic images from 564 patients across 1,043 colonoscopy procedures, which is the largest publicly available labeled dataset for UC research. Expert gastroenterologists assigned Mayo Endoscopic Scores (MES) using majority voting for discordant cases. The dataset exhibits significant class imbalance where high severity classes are scare: MES 0 (54.14%), MES 1 (27.70%), MES 2 (11.12%), and MES 3 (7.67%).

Following established protocols [29], 992 images containing medical instruments or annotations were excluded to prevent model bias, resulting in 10,284 images. Images were cropped from 352 × 288 to 224 × 224 pixels to remove metadata while preserving the endoscopic view, then resized to 256 × 256 using bicubic interpolation. Patient-level data splitting ensured no leakage between training (70%), validation (15%), and test (15%) sets.

3.3. Training Protocol

Training was conducted on NVIDIA V100 GPU having 16GB VRAM with batch size 32. The learning rate was set to 1×10^{-5} for 21,000 optimization steps, corresponding to approximately 672,000 image-conditioning pairs. Three key stabilization strategies were implemented.

Exponential Moving Average (EMA) was applied with a decay rate $\beta = 0.999$ to reduce training volatility and improve sample quality [36]:

$$\theta_{\text{EMA}} \leftarrow \beta \cdot \theta_{\text{EMA}} + (1 - \beta) \cdot \theta \tag{4}$$

Min-SNR-γ Weighting was utilized with $\gamma = 1.0$ to address gradient conflicts across timesteps and accelerate convergence [37].

Class-Balanced Sampling with inverse class frequencies was implemented to address dataset imbalance. The 30% synthetic data augmentation strategy was implemented to address severe class imbalance, particularly for MES 3 consisting only 7.67% of dataset. Synthetic images were generated using the GAN-based method from [29], which demonstrated improved minority class representation without compromising model performance. This proportion was empirically determined to maximize training stability while preventing overfitting to synthetic data patterns.

3.4. Evaluation Framework

Four complementary metrics were employed to assess generation quality: Fréchet Inception Distance (FID) [38], CLIP Maximum Mean Discrepancy (CMMD) [39], and precision/recall metrics [40]. This multi-metric approach addresses known limitations of individual metrics and provides comprehensive quality assessment.

Ordinal consistency was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), classification accuracy, and Quadratic Weighted Kappa (QWK) metrics. Additionally, Uniform Manifold Approximation and Projection (UMAP) [41] analysis compared manifold structures between real and generated datasets using ResNet-18 feature representations.

3.5. Experimental Design

Comprehensive comparisons were conducted between proposed ordinal embedding approaches and CLIP text embedder baseline under identical experimental conditions. All methods employed the same diffusion architecture, training procedures, and evaluation protocols to ensure fair comparison. Progressive generation evaluation assessed interpolation capabilities using fine-grained MES increments to examine transition smoothness and anatomical consistency preservation.

Guidance scale of 2.0 provides optimal balance between conditioning strength and natural image appearance. All experiments employed DDPM scheduler for training and inference with 50 denoising steps.

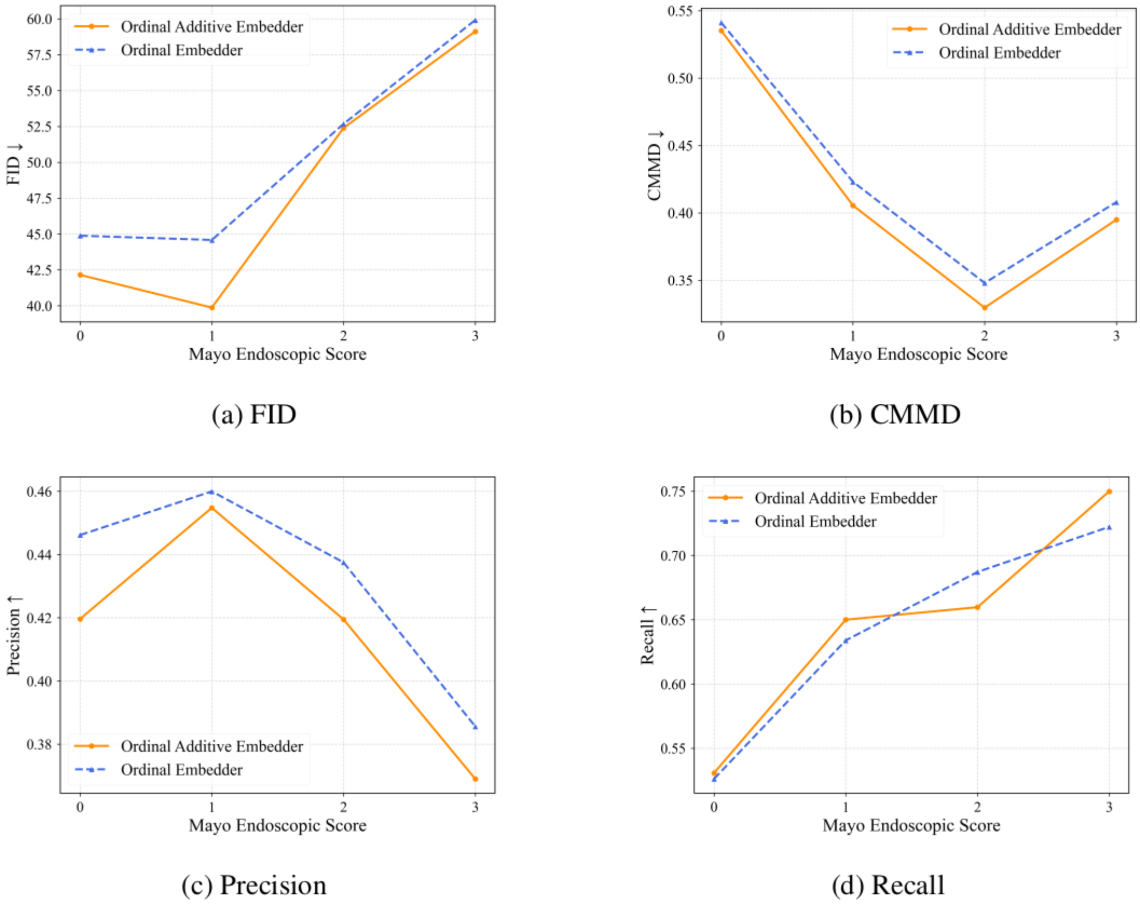
4. Results

4.1. Quantitative Performance Evaluation

The comparative results across embedding strategies with four complementary image quality metrics are presented in Table 1. AOE achieves better performance in terms of CMMD and Recall metrics, indicating better distributional alignment and improved coverage of the data manifold. Although FID scores are marginally higher for ordinal methods, CLIP-based metrics provide more reliable assessment than Inception-based metrics for medical content evaluation. The class-wise analysis in Figure 3 reveals that the AOE consistently outperforms alternatives across all Mayo Endoscopic Score severity levels, demonstrating robust performance throughout the entire spectrum of the disease.

Table 1. Quantitative comparison of Basic Ordinal Embedder (BOE) and Additive Ordinal Embedder (AOE) methods with the CLIP baseline for ulcerative colitis progression generation.

Method	CMMD (↓)	FID (↓)	Precision (↑)	Recall (↑)
CLIP	0.4246	**30.506**	**0.4942**	0.5954
BOE	0.4227	36.072	0.4616	0.6207
AOE	**0.4137**	34.675	0.4614	**0.6331**



(a) FID

(b) CMMD

(c) Precision

(d) Recall

Figure 3. Metric-wise comparison of embedder types across ordinal severity classes.

4.2. Disease Progression Synthesis Results

The framework's capability to generate clinically meaningful disease progression sequences was evaluated through systematic analysis of synthetic progression trajectories. Figure 4 demonstrates smooth transitions from normal mucosa (MES 0) to severe ulcerative colitis (MES 3) in increments of 0.20.

The generated sequences exhibit key characteristics essential for clinical validity, such as maintaining anatomical consistency in the progression stages, the gradual introduction of pathological characteristics corresponding to the Mayo Endoscopic Scoring criteria and realistic intermediate stages. Notably, the framework successfully interpolates between discrete training classes to create plausible intermediate disease states typically underrepresented in clinical datasets.

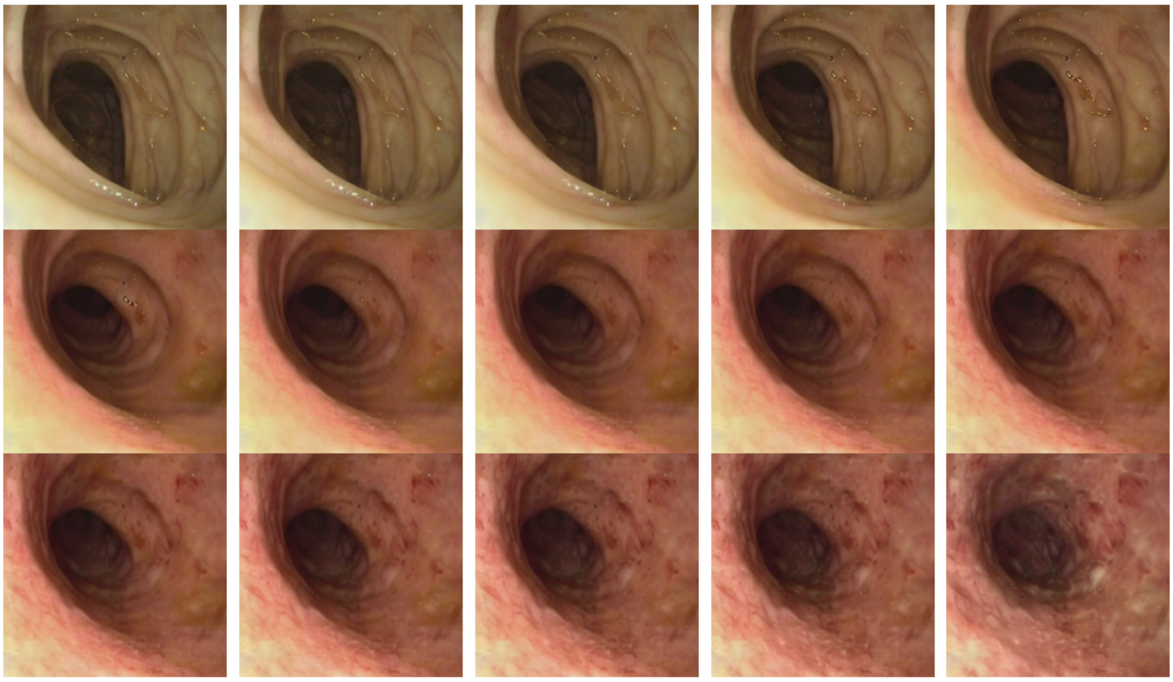


Figure 4. Disease progression sequence generated using the Additive Ordinal Embedder, showing smooth transitions from MES 0 to 3 in increments of 0.20. The progression demonstrates gradual introduction of pathological features while maintaining anatomical consistency.

4.3. Validation and Consistency Analysis

Consistency was assessed using a downstream task which is a ResNet-18 regression model trained on the LIMUC dataset to predict continuous MES from generated progression sequences. A total of 1,100 progression sequences with 0.1 increments were generated from MES 0.0 to 3.0. Due to the stochastic nature of the diffusion model, which occasionally produces suboptimal images, the best performing 80% of sequences were selected for evaluation, resulting in a total of 27,280 synthetic images. This selection process effectively eliminates stochastic outliers while maintaining sufficient data volume for robust evaluation. The Oracle baseline represents the upper bound performance achievable using real test data from the LIMUC dataset, comprising 1,443 images, and serves as the reference standard for assessing the validity of synthetically generated progression sequences.

Table 2. Ordinal consistency assessment of Basic Ordinal Embedder (BOE) and Additive Ordinal Embedder (AOE) with the CLIP baseline using ResNet-18 regression model.

Dataset	#Images	RMSE	MAE	Accuracy	QWK
Oracle	1,443	0.454	0.333	0.7651	0.8591
CLIP	27,280	0.9507	0.7448	0.3896	0.4625
BOE	27,280	0.5171	0.4374	0.6239	0.8420
AOE	27,280	**0.5112**	**0.4238**	**0.6374**	**0.8425**

The results demonstrate that ordinal embedding approaches achieve ordinal consistency comparable to real data, with Quadratic Weighted Kappa (QWK) scores of 0.8420 and 0.8425 closely matching the test data performance. The CLIP baseline exhibits poor ordinal consistency with QWK of 0.4625, highlighting the critical importance of task-specific embedding approaches for medical severity assessment. More results on various regression models are provided in Appendix A.

Figure 5 illustrates regression predictions for the specific progression sequence shown in Figure 4, demonstrating smooth ordinal relationships maintained across continuous MES values with MAE of 0.45 and RMSE of 0.22.

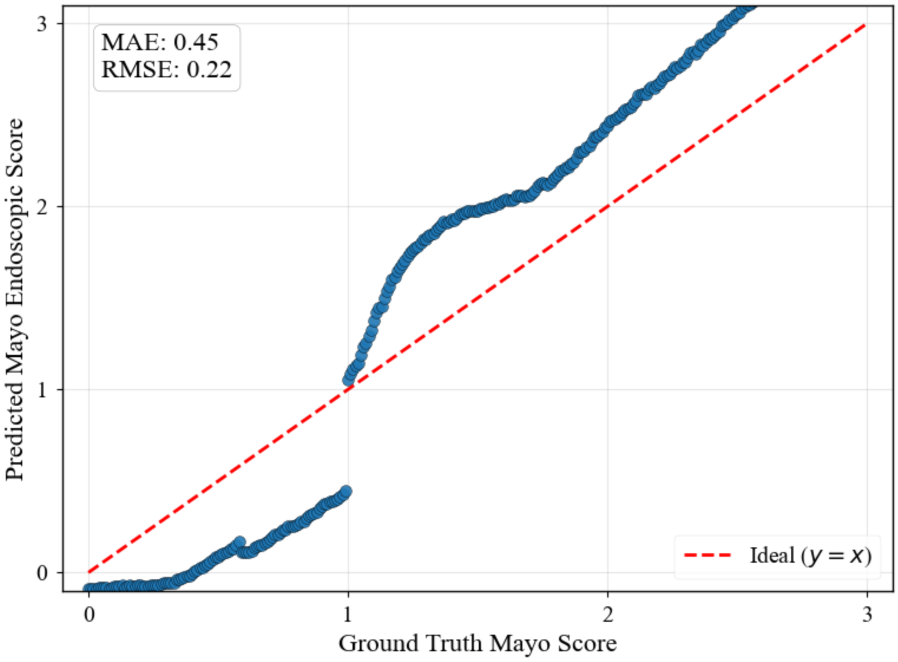


Figure 5. ResNet-18 regression predictions on synthetically generated progression sequences using the Additive Ordinal Embedder with 0.01 increments between MES 0-3. The smooth progression along the ideal diagonal demonstrates successful interpolation between discrete training classes.

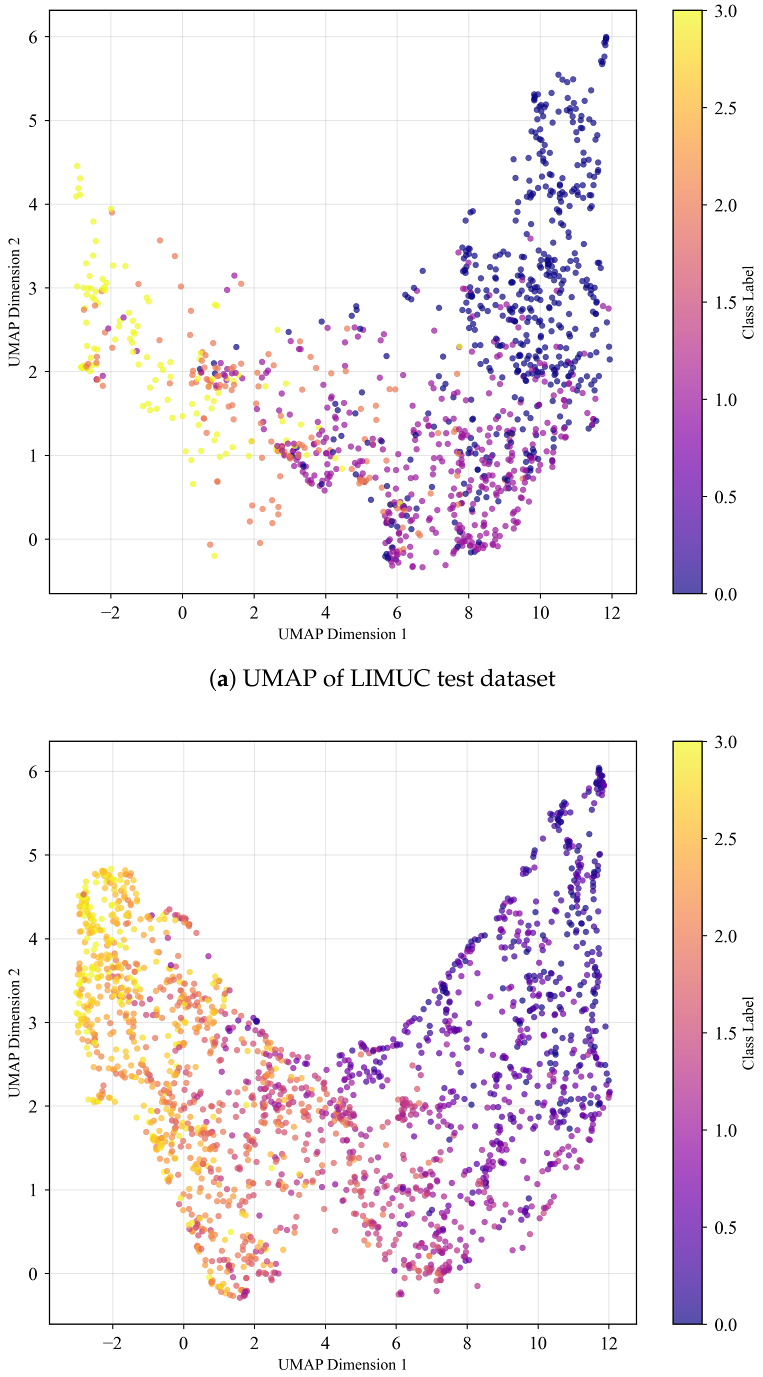
4.4. Ablation Studies and Design Validation

Comprehensive ablation studies confirm the necessity of ordinal aware embedding design. The CLIP baseline fails to maintain clinical consistency across MES levels, producing discontinuous progressions with inadequate interpolation capabilities.

UMAP manifold analysis in Figure 6 demonstrates that generated progression sequences occupy relevant regions of the feature space while providing enhanced coverage of the disease progression continuum. The Silhouette Score of 0.0571 indicates substantial manifold overlap between real and synthetic data, where a score approaching 0 signifies that the manifolds overlap significantly. The synthetic data maintains distributional alignment with real clinical data and extends into intermediate stages, effectively addressing class-imbalance problems often encountered in CAD systems by populating underrepresented classes while preserving the real feature space.

Figure 7 illustrates the fundamental limitation of CLIP embeddings for medical progression modeling. The regression predictions show significant scatter across all MES levels, with predictions failing to follow the expected diagonal progression pattern. This poor ordinal consistency arises because CLIP's text-based training objective optimizes for semantic similarity between textual descriptions and visual content rather than capturing numerical relationships or ordinal progressions. When MES are converted to text strings like “0.0”, “1.0”, “2.0”, “3.0”, the embedding space interprets these as distinct semantic categories instead of points along a continuous severity spectrum, preventing meaningful interpolation between severity levels and resulting in discontinuous progression sequences unsuitable for clinical applications requiring smooth disease evolution modeling.

doi:10.20944/preprints202509.0426.v1



(**a**) UMAP of LIMUC test dataset

(**b**) UMAP manifold of generated progression dataset by Ordinal Additive Embedder

Figure 6. UMAP manifold comparison between real test dataset and generated dataset. The Silhouette Score between these datasets is 0.0571.

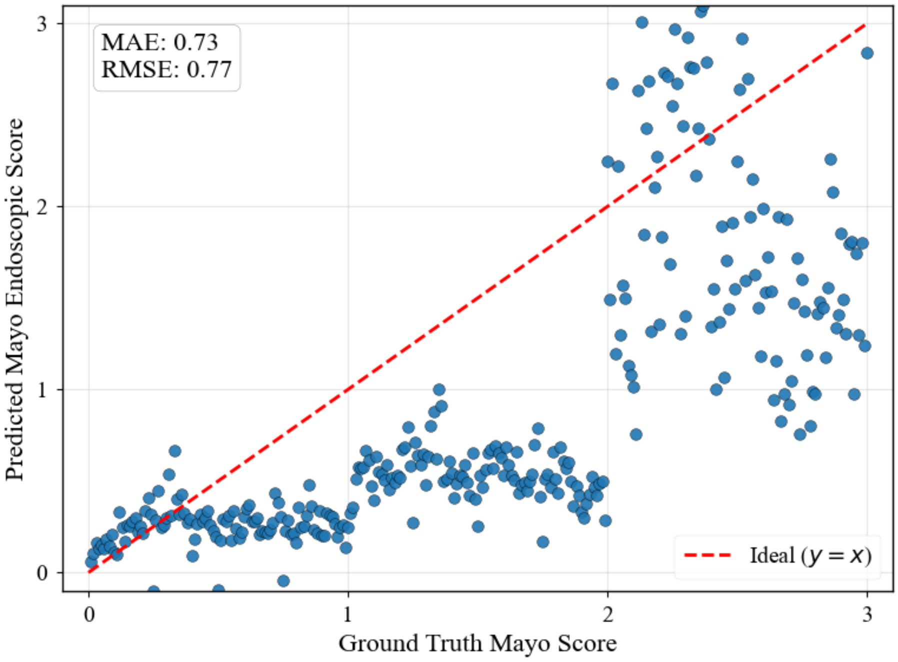


Figure 7. ResNet-18 regression predictions on synthetically generated progression sequences using the CLIP Embedder with 0.01 increments between MES 0-3. The failure in progression indicates that the CLIP Embedder's embedding space treats classes as distinct semantic categories.

5. Discussion

5.1. Educational Impact and Applications

The proposed framework shows promise for endoscopic training by generating realistic ulcerative colitis progression sequences covering the full disease spectrum. The sequences demonstrate strong ordinal consistency, with a Quadratic Weighted Kappa of 0.84, suggesting suitability for integration into computer-aided diagnosis systems requiring accurate severity assessment. By synthesizing intermediate disease stages, the framework may help address gaps in medical education where examples of disease progression are limited. Additionally, its ability to produce anatomically consistent sequences with gradual pathological changes supports development of automated assessment and disease trajectory estimation tools for UC severity grading. These results indicate potential to enhance training programs and clinical decision-making through synthetic data augmentation of underrepresented disease states.

5.2. Technical Contributions and Methodological Advances

The superior performance of ordinal embedding architectures supports the hypothesis that specialized conditioning mechanisms, rather than general-purpose text embeddings, are required for effectively modeling disease progression. The Additive Ordinal Embedder's design, which models higher severity levels as cumulative accumulations of pathological features, aligns with clinical understanding of UC progression and demonstrates measurable improvements across all evaluation metrics.

Ablation studies reveal that CLIP embeddings fail to capture ordinal relationships essential for progression modeling in medical applications. When MES are converted to text strings, the embedding space interprets these as distinct semantic categories rather than points along a continuous severity spectrum, which prevents meaningful interpolation between severity levels. This categorical interpretation results in poor clinical consistency, demonstrating why domain-specific embedding design is crucial for medical image generation.

The successful adaptation of diffusion models for medical applications through ordinal conditioning represents a methodological advancement with broader implications. Unlike previous approaches treating disease severity as independent categorical variables, the framework explicitly models progressive pathological evolution, enabling generation of clinically meaningful intermediate stages supporting both educational and research applications. The proposed approach enables the augmentation of underrepresented classes, particularly those corresponding to higher disease severity levels, by generating synthetic data conditioned on healthy samples. This synthetic data effectively estimates disease progression from the current severity state, allowing simulation of how conditions such as Ulcerative Colitis may worsen or improve over time. Furthermore, our continuous progression framework addresses common issues in previous categorical data generation methods, such as mode collapse and lack of diversity, thereby enhancing the robustness and realism of the generated samples.

5.3. Computer-Aided Diagnosis and Clinical Integration

The computational validation results indicate the potential for integrating synthetic progression data into computer-aided diagnosis systems. Generated sequences maintain ordinal relationships which could enhance the robustness of automated assessment systems by providing additional training examples and improving classification accuracy across all severity levels.

For potential real-time endoscopic applications, synthetic progression sequences could serve as reference standards during colonoscopy procedures. This would enable clinicians to compare observed pathology against expected progression patterns. Furthermore, the framework's ability to generate fine-grained severity increments offers unprecedented granularity for severity assessment when combined with regressive models, potentially supporting more nuanced diagnostic decisions than traditional discrete scoring systems like MES.

UMAP manifold analysis reveals that generated sequences occupy relevant feature space regions while extending coverage to intermediate stages underrepresented in clinical datasets. This enhanced coverage is particularly valuable for training robust disease progression models requiring comprehensive pathological spectrum representation, addressing limitations of datasets biased toward discrete MES assessments.

5.4. Limitations and Methodological Considerations

Several limitations warrant consideration despite the framework's demonstrated advantages. The jump in MES 0 shown in Figure 3 indicates that the dataset does not contain smooth transition images from healthy to mild UC, which may suggest that the framework's reliance on discrete training classes introduces artificial boundaries that do not fully capture the continuous nature of biological processes.

Furthermore, the current approach concentrates on visual progression modeling, omitting temporal dynamics and patient-specific factors that also influence disease evolution. Addressing these limitations in future studies by incorporating longitudinal data and personalized progression modeling approaches would be beneficial. While current computational validation demonstrates promising results, clinical validation by medical professionals is essential before any practical implementation in clinical settings.

6. Conclusions

This study proposes ordinal class embedding architectures for conditional diffusion models to generate ulcerative colitis progression sequences from cross-sectional endoscopic data. By transforming discrete Mayo Endoscopic Score classifications into continuous progression models, the framework enables synthesis of intermediate disease stages typically underrepresented in traditional datasets. Experimental results show that the Additive Ordinal Embedder performs favorably across metrics.

ResNet-18 regression analysis further suggests that generated sequences preserve medically relevant ordinal relationships and outperform general-purpose CLIP embeddings. This capacity to generate smooth transitions between severity levels addresses limitations of classification-based datasets and offers potential tools for endoscopic training and computer-aided diagnosis. The synthesized intermediate stages could augment training datasets, enhance automated assessment systems, and support clinical decision-making across the UC severity spectrum. Additionally, generating more data for underrepresented disease classes may help mitigate data imbalance issues, while the progressive nature of the sequences could facilitate disease trajectory forecasting.

Generating synthetic longitudinal datasets opens new avenues for disease trajectory prediction research. These sequences can enable predictive models to forecast disease evolution from current severity assessments, supporting personalized treatment planning and monitoring. Integration with real-time endoscopic systems offers promising clinical translation potential by enhancing quality assessment and providing intra-procedural feedback. Cross-modal applications linking endoscopic progression with histopathology or clinical biomarkers could further deepen disease understanding and individualized care. Future work should explore advanced data augmentation and interpolation methods to address discrete transitions, standardize evaluation protocols, and develop benchmark datasets to accelerate research and adoption.

Despite these advances, clinical validity must be confirmed by experienced gastrointestinal physicians to ensure practical utility and safety. Future research should also focus on preserving anatomical consistency, capturing temporal dynamics for longitudinal modeling, incorporating additional clinical data, and generalizing the framework to other progressive conditions with ordinal staging.

Author Contributions: Conceptualization, Ü.M.Ç. and A.T.; methodology, M.M.K. and Ü.M.Ç.; software, M.M.K.; validation, M.M.K. and Ü.M.Ç.; formal analysis, M.M.K.; investigation, M.M.K.; resources, A.T.; data curation, Ü.M.Ç.; writing—original draft preparation, M.M.K.; writing—review and editing, M.M.K., Ü.M.Ç. and A.T.; visualization, M.M.K.; supervision, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by ASELSAN Inc.

Informed Consent Statement: Not applicable

Data Availability Statement: LIMUC (Labeled Images for Ulcerative Colitis) dataset used in this study is publicly available.

Acknowledgments: This work has been supported by Middle East Technical University Scientific Research Projects Coordination Unit under grant number ADEP-704-2024-11486. The experiments reported in this work were fully performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA)

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CAD	Computer-aided diagnosis
CLIP	Contrastive Language-Image Pre-training
CMMD	CLIP Maximum Mean Discrepancy
EMA	Exponential Moving Average
FID	Fréchet Inception Distance
GANs	Generative Adversarial Networks
LIMUC	Labeled Images for Ulcerative Colitis
MAE	Mean Absolute Error
MES	Mayo Endoscopic Score
MVG	Medical Video Generation
QWK	Quadratic Weighted Kappa
RMSE	Root Mean Square Error
UC	Ulcerative colitis
UMAP	Uniform Manifold Approximation and Projection
VAE	Variational Autoencoder

Appendix A

To ensure the robustness and generalizability of our findings regarding ordinal consistency, we conducted comprehensive validation using multiple regression architectures. This appendix presents validation results using InceptionV3, ResNet-50, and Vision Transformer (ViT) regression models, all trained on the LIMUC dataset following identical protocols.

Table A1 presents the InceptionV3 regression validation results. The Additive Ordinal Embedder (AOE) closely approaches the Oracle performance in ordinal consistency. The InceptionV3 architecture's multi-scale feature extraction capabilities appear particularly well-suited for capturing the visual characteristics relevant to UC severity assessment.

Table A1. Validation results using InceptionV3 regression model for ordinal consistency assessment comparing Basic Ordinal Embedder (BOE) and Additive Ordinal Embedder (AOE) with CLIP Baseline.

Dataset	#Images	RMSE	MAE	Accuracy	QWK
Oracle	1,443	0.4447	0.3010	0.7762	0.8647
CLIP	27,280	0.9775	0.7637	0.3781	0.4328
BOE	27,280	0.4593	0.3683	0.6449	0.8451
AOE	27,280	**0.4494**	**0.3568**	**0.6594**	**0.8490**

Table A2 shows the ResNet-50 validation results. These results validate that the benefits of ordinal embeddings scale effectively with increased model capacity.

Table A2. Validation results using ResNet-50 regression model for ordinal consistency assessment.

Dataset	#Images	RMSE	MAE	Accuracy	QWK
Oracle	1,443	0.4726	0.3360	0.7505	0.8437
CLIP	27,280	0.9698	0.7614	0.3723	0.4474
BOE	27,280	0.5204	0.4259	0.6164	0.8370
AOE	27,280	**0.5086**	**0.4116**	**0.6362**	**0.8431**

Table A3 presents the Vision Transformer validation results, extending our evaluation to modern attention-based architectures. While showing slightly lower absolute performance compared to convolutional networks, the relative ordering between methods remains consistent. The performance decrease observed in ViT compared to convolutional networks may be attributed to the small and imbalanced nature of the LIMUC dataset, as transformer architectures typically require larger datasets for optimal performance.

Table A3. Validation results using ViT regression model for ordinal consistency assessment.

Dataset	#Images	RMSE	MAE	Accuracy	QWK
Oracle	1,443	0.5421	0.4058	0.7159	0.8022
CLIP	27,280	1.2682	1.0299	0.3011	0.3124
BOE	27,280	0.5577	0.4567	0.5848	**0.8222**
AOE	27,280	**0.5508**	**0.4547**	**0.5885**	0.8146

Across all four regression architectures, the results consistently validate the superiority of ordinal embedding strategies over general-purpose CLIP embeddings. This architectural independence demonstrates that the observed benefits are fundamental to the ordinal embedding approach rather than artifacts of specific model designs, providing strong evidence for robustness of our framework across diverse computer vision architectures commonly used in medical imaging applications.

References

1. Ochiai, K.; Ozawa, T.; Shibata, J.; Ishihara, S.; Tada, T. Current status of artificial intelligence-based computer-assisted diagnosis systems for gastric cancer in endoscopy. *Diagnostics* **2022**, *12*, 3153.
2. Li, L.; Qiu, J.; Saha, A.; Li, L.; Li, P.; He, M.; Guo, Z.; Yuan, W. Artificial intelligence for biomedical video generation. *arXiv preprint arXiv:2411.07619* **2024**.
3. Sun, T.; He, X.; Li, Z. Digital twin in healthcare: Recent updates and challenges. *Digital health* **2023**, *9*, 20552076221149651.
4. Chen, J.; Shi, Y.; Yi, C.; Du, H.; Kang, J.; Niyato, D. Generative AI-driven human digital twin in IoT-healthcare: A comprehensive survey. *IEEE Internet of Things Journal* **2024**.
5. Schulam, P.; Arora, R. Disease trajectory maps. *Advances in neural information processing systems* **2016**, *29*.
6. Lim, B.; van der Schaar, M. Disease-atlas: Navigating disease trajectories using deep learning. In Proceedings of the Machine Learning for Healthcare Conference. PMLR, 2018, pp. 137–160.
7. Bhagwat, N.; Viviano, J.D.; Voineskos, A.N.; Chakravarty, M.M.; Initiative, A.D.N.; et al. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS computational biology* **2018**, *14*, e1006376.
8. Cao, X.; Liang, K.; Liao, K.D.; Gao, T.; Ye, W.; Chen, J.; Ding, Z.; Cao, J.; Rehg, J.M.; Sun, J. Medical video generation for disease progression simulation. *arXiv preprint arXiv:2411.11943* **2024**.
9. Satsangi, J.; Silverberg, M.; Vermeire, S.; Colombel, J. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **2006**, *55*, 749–753.
10. Fuerstein, J.; Moss, A.C.; Farraye, F. Ulcerative Colitis. *Mayo ClinProc* **2019**, *94*, 1357–1373.
11. Xie, T.; Zhang, T.; Ding, C.; Dai, X.; Li, Y.; Guo, Z.; Wei, Y.; Gong, J.; Zhu, W.; Li, J. Ulcerative Colitis Endoscopic Index of Severity (UCEIS) versus Mayo Endoscopic Score (MES) in guiding the need for colectomy in patients with acute severe colitis. *Gastroenterology report* **2018**, *6*, 38–44.
12. Kellermann, L.; Riis, L.B. A close view on histopathological changes in inflammatory bowel disease, a narrative review. *Digestive Medicine Research* **2021**, *4*.
13. Jung, E.; Luna, M.; Park, S.H. Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression. *Pattern Recognition* **2023**, *133*, 109061.
14. Ravi, D.; Blumberg, S.B.; Ingala, S.; Barkhof, F.; Alexander, D.C.; Oxtoby, N.P.; Initiative, A.D.N.; et al. Degenerative adversarial neuroimage nets for brain scan simulations: Application in ageing and dementia. *Medical Image Analysis* **2022**, *75*, 102257.
15. Cooke, M.; Irby, D.M.; O'Brien, B.C. *Educating physicians: A call for reform of medical school and residency*; John Wiley & Sons, 2010.
16. Ktena, I.; Wiles, O.; Albuquerque, I.; Rebuffi, S.A.; Tanno, R.; Roy, A.G.; Azizi, S.; Belgrave, D.; Kohli, P.; Cemgil, T.; et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nature Medicine* **2024**, *30*, 1166–1173.
17. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Medical image analysis* **2023**, *88*, 102846.
18. Müller-Franzes, G.; Niehues, J.M.; Khader, F.; Arasteh, S.T.; Haarburger, C.; Kuhl, C.; Wang, T.; Han, T.; Nolte, T.; Nebelung, S.; et al. A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports* **2023**, *13*, 12098.
19. Nie, B.; Zhang, G. Ulcerative Severity Estimation Based on Advanced CNN–Transformer Hybrid Models. *Applied Sciences* **2025**, *15*, 7484.
20. Matsuoka, K.; Kobayashi, T.; Ueno, F.; Matsui, T.; Hirai, F.; Inoue, N.; Kato, J.; Kobayashi, K.; Kobayashi, K.; Koganei, K.; et al. Evidence-based clinical practice guidelines for inflammatory bowel disease. *Journal of gastroenterology* **2018**, *53*, 305–353.
21. Kawamoto, A.; Takenaka, K.; Okamoto, R.; Watanabe, M.; Ohtsuka, K. Systematic review of artificial intelligence-based image diagnosis for inflammatory bowel disease. *Digestive Endoscopy* **2022**, *34*, 1311–1319.
22. Gong, E.J.; Bang, C.S. Edge Artificial Intelligence Device in Real-Time Endoscopy for the Classification of Colonic Neoplasms. *Diagnostics* **2025**, *15*, 1478.
23. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv (Cornell University)* **2013**. https://doi.org/10.48550/arXiv.1312.6114.
24. van den Oord, A.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel Recurrent Neural Networks. *arXiv (Cornell University)* **2016**.

25. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems* **2014**, *27*, 73–76. https://doi.org/10.1007/978-3-658-40442-0_9.
26. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4401–4410.
27. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 01 2020, pp. 8110–8119. https://doi.org/10.1109/cvpr42600.2020.00813.
28. Dhariwal, P.; Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **2021**, *34*, 8780–8794. https://doi.org/10.48550/arxiv.2105.05233.
29. Çağlar, Ü.M.; İnci, A.; Hanoğlu, O.; Polat, G.; Temizel, A. Ulcerative Colitis Mayo Endoscopic Scoring Classification with Active Learning and Generative Data Augmentation. In Proceedings of the 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2023, pp. 462–467.
30. Zhang, Y.; Li, L.; Wang, J.; Yang, X.; Zhou, H.; He, J.; Xie, Y.; Jiang, Y.; Sun, W.; Zhang, X.; et al. Texture-preserving diffusion model for CBCT-to-CT synthesis. *Medical Image Analysis* **2025**, *99*, 103362.
31. Takezaki, S.; Uchida, S. An Ordinal Diffusion Model for Generating Medical Images with Different Severity Levels. *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* **2024**, pp. 1–5.
32. Li, C.; Liu, H.; Liu, Y.; Feng, B.Y.; Li, W.; Liu, X.; Chen, Z.; Shao, J.; Yuan, Y. Endora: Video generation models as endoscopy simulators. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2024, pp. 230–240.
33. Gajendran, M.; Loganathan, P.; Jimenez, G.; Catinella, A.P.; Ng, N.; Umapathy, C.; Ziade, N.; Hashash, J.G. A comprehensive review and update on ulcerative colitis. *Disease-a-month* **2019**, *65*, 100851.
34. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695.
35. Polat, G.; Kani, H.; Ergenc, I.; Alahdab, Y.; Temizel, A.; Atug, O. Labeled images for ulcerative colitis (limuc) dataset. *Accessed March* **2022**.
36. Karras, T.; Aittala, M.; Lehtinen, J.; Hellsten, J.; Aila, T.; Laine, S. Analyzing and improving the training dynamics of diffusion models. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 24174–24184.
37. Hang, T.; Gu, S.; Li, C.; Bao, J.; Chen, D.; Hu, H.; Geng, X.; Guo, B. Efficient diffusion training via min-snr weighting strategy. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 7441–7451.
38. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **2017**, *30*. https://doi.org/10.48550/arxiv.1706.08500.
39. Jayasumana, S.; Ramalingam, S.; Veit, A.; Glasner, D.; Chakrabarti, A.; Kumar, S. Rethinking fid: Towards a better evaluation metric for image generation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 01 2024, pp. 9307–9315. https://doi.org/10.1109/cvpr52733.2024.00889.
40. Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; Aila, T. Improved precision and recall metric for assessing generative models. *Advances in neural information processing systems* **2019**, *32*. https://doi.org/10.48550/arxiv.1904.06991.
41. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* **2018**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.