Few-Shot Flow Matching with Stochastic Barycentric Sampling for Image Synthesis

Aleksandra Kulbaka[1], John Bonnici[1], Ben Glocker[1⋆], and Bernhard Kainz[1,2⋆]

[1] Departament of Computing, Imperial College London, UK
a.kulbaka24@imperial.ac.uk
[2] Friedrich–Alexander University Erlangen–Nürnberg, DE

Abstract. The integration of automated diagnostic support into clinical practice is often hindered by data scarcity and strict privacy constraints in local centers. While generative models offer a potential solution via data augmentation, standard architectures suffer from mode collapse when conditioned on few-shot local samples. Therefore, we propose Few-Shot Flow Matching (FSFM). Our core contribution, Stochastic Barycentric Sampling, treats retrieved latent neighbors as vertices of a simplex, dynamically sampling conditioning vectors from the continuous convex hull to prevent mode collapse and preserve structural variance. We apply FSFM to Diabetic Retinopathy (DR) fundus image synthesis, adapting from a large unlabelled source to a constrained target clinic under three protocol-stratified evaluation cohorts. FSFM provides consistent improvements in 5-class DR grading over real-only baselines in the extreme-scarcity regime ($N \leq 100$), with the largest absolute gains on the homogeneous-protocol dilated cohort (QWK 0.10 $\rightarrow$ 0.20 at N=20, 0.29 $\rightarrow$ 0.37 at N=50). Stochastic Barycentric Sampling produces 26-39% higher synthetic diversity than deterministic centroid conditioning. Code is available at https://anonymous.4open.science/r/retina_flow_matching-4860.

Keywords: Image Synthesis · Flow Matching · Data Augmentation.

1 Introduction

The deployment of automated diagnostic classifiers at smaller clinical sites is routinely undermined by extreme data scarcity. Training or fine-tuning deep networks demands substantial local data to overcome domain shifts introduced by varying scanner hardware, imaging protocols, and patient demographics [11], yet peripheral institutions rarely possess more than a few dozen labelled images per pathology. Compounding this bottleneck, the vast majority of archived clinical imaging remains unlabelled, with diagnostic ground truths trapped in unstructured electronic health records whose abstraction requires costly expert effort [19]. Generative data augmentation can be a potential solution: a model

⋆ Joint supervision.

trained on a large source distribution is able to, in principle, synthesize target-domain images that restore the statistical diversity needed for robust classifier training [18,17]. In practice, however, standard generative architectures conditioned on only a handful of local samples suffer from severe mode collapse, producing synthetic images that merely replicate the anchor set and fail to provide the structural variance required to overcome domain shift [10,12].

To address this, we propose **Few-Shot Flow Matching (FSFM)**, a privacy-preserving generative framework built on Continuous Normalizing Flows [14] that learns the geometry of a medical image manifold entirely from unlabelled source data and adapts to a new target domain using as few as 50 local samples. Our core methodological contribution is **Stochastic Barycentric Sampling**: rather than conditioning the generative vector field on the deterministic centroid of retrieved latent neighbors, we treat these neighbors as vertices of a simplex, and at every forward pass, draw a fresh conditioning vector uniformly from the continuous convex hull they span. By continuously sweeping this k-simplex via Dirichlet distribution, the model is forced to learn a dense mapping from sparse feature anchors to image space, producing diverse synthetic outputs without any additional training or fine-tuning at the target site. Crucially, inference is fully decoupled from the source data, satisfying zero-knowledge privacy constraints suitable for federated clinical deployment.

We apply it to the clinically relevant task of synthesizing Diabetic Retinopathy fundus images, adapting from a large unlabelled source (EyePACS [4,8]) to constrained target dataset (MESSIDOR-2 [1,6]) using small number of local anchors. Classifiers augmented with FSFM synthetic data yield consistent accuracy improvement in the low-data regimes.

Our contributions are therefore:

- *FSFM*, a privacy-preserving generative framework that adapts to target medical domains using label-free pre-training, decoupling inference entirely from the source training data.
- *Stochastic Barycentric Sampling*, a novel conditioning mechanism that interpolates within the continuous convex hull of latent neighbors, producing more diverse synthetic outputs than deterministic centroid conditioning.
- To the best of our knowledge, we present *the first application of Flow Matching for few-shot retinal image synthesis*, achieving downstream classification gains using only as little as 20 local anchors.

Related Work. Diffusion models have been used for data augmentation in medical settings, such as pelvis radiographs [13], and skin disease [18,20]. In ophthalmology, they were used for fundus synthesis for classification [17] and label-preserving latent manipulation for DR recognition [21]. Notably, at MICCAI 2024, Yuan et al. [20] successfully adapted pre-trained generative models for fundus imaging using several thousand target samples. In few-shot settings, Few-Shot Diffusion Models aggregate patch features via a Vision Transformer [9], an approach extended to cerebral aneurysm geometries [7]. Furthermore, [3] demonstrated LoRA-based Flow Matching for cross-spectral translation. While highly

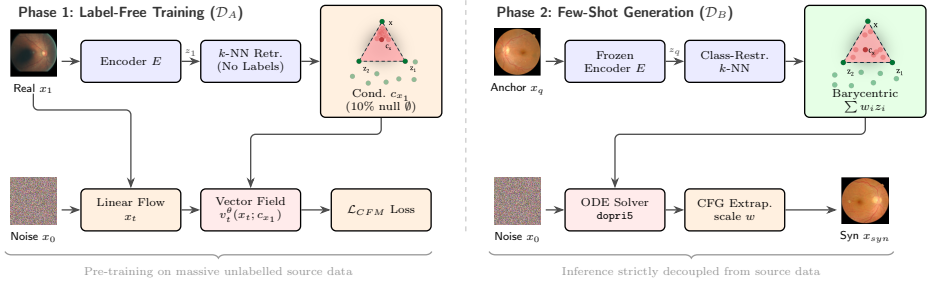

Fig. 1: **Overview of FSFM. Left:** In Phase 1, the model learns a continuous vector field entirely from unlabelled source data. During pre-training, the barycentric conditioning (c_{x_1}) is heavily weighted toward the original anchor, aligning the conditioning signal with the ground-truth image targeted by the flow matching loss. **Right:** In Phase 2, the frozen prior adapts to the local target domain using Stochastic Barycentric Sampling to draw dynamic conditioning vectors from a local, class-restricted neighbor simplex.

effective, these approaches face practical barriers in constrained clinical environments and typically necessitate iterative target-site fine-tuning [3,20], assume access to large annotated pre-training corpora [13], or focus primarily on perceptual image metrics rather than downstream diagnostic tasks [7]. We combine retrieval-augmented Flow Matching [14,15] generation, where nearest-neighbor features decouple memory from synthesis [2], with latent-space oversampling, conceptually similar to DeepSMOTE [5]. Stochastic Barycentric Sampling extends these paradigms by sampling uniformly over the convex hull instead of relying on deterministic centroids or simple pairwise interpolation. Our latent conditioning leverages RETFound [22], pre-trained on 1.6M retinal images, utilizing DINOv2 features [16] that preserve complex DR topology and heavily outperform generic ImageNet embeddings.

2 Method

Our objective is to adapt a generative model to a data-scarce target clinical domain without violating patient privacy constraints. Let $\mathcal{D}_A = \{x_i^A\}_{i=1}^M$ denote a large, unlabelled source dataset from a central institution, Hospital A, where M is large and covers a highly diverse data distribution. $\mathcal{D}_B = \{(x_i^B, y_i^B)\}_{i=1}^N$ represents a constrained target dataset from a local clinic, Hospital B, where N is extremely small and y_i^B denotes the corresponding disease grade.

Crucially, strict "zero-knowledge" privacy protocols prohibit querying the source dataset $\mathcal{D}_A$ during the inference process. The generative model must rely entirely on the limited local anchors in $\mathcal{D}_B$ to synthesize a target-specific augmented dataset $\mathcal{D}_{syn}$ that accurately reflects the local scanner characteristics and preserves the complex topology of the disease grades.

Conditional Flow Matching. We adopt the Conditional Flow Matching (CFM) framework [14,15], which learns a continuous vector field $v_t^\theta(x_t; c_{x_t})$ pushing samples from a Gaussian prior $p_0 = \mathcal{N}(0, I)$ to the data distribution p_1. Using the linear interpolant $x_t = (1-t)x_0 + tx_1$ with target field $u_t(x_t|x_0, x_1) = x_1 - x_0$, the network is trained via

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1], x_1\sim p_1, x_0\sim p_0}\left[\|v_t^\theta(x_t; c_{x_t}) - (x_1 - x_0)\|^2\right]. \tag{1}$$

The straight-line trajectories enable fast ODE integration during inference compared to the curved, stochastic paths typical of standard diffusion models.

Latent Retrieval and Stochastic Barycentric Sampling. To condition the continuous vector field without relying on explicit labels during training, we map the available images into a semantic feature space. We experiment with different pre-trained feature extractors E to map images into a latent space $\mathcal{Z}$. Standard few-shot conditioning paradigms typically compute the deterministic centroid of anchor and its retrieved k neighbors, $c_{avg} = \frac{1}{k+1}\sum_{i=1}^{k+1} z_i$. To increase the data diversity, we propose *Stochastic Barycentric Sampling*. Rather than utilizing a single deterministic point, we assume that valid, high-fidelity representations of the target domain lie continuously within the convex hull (simplex) spanned by the local neighbor set $\mathcal{N}(x)$. At each forward pass, we sample the conditioning vector c as a convex combination of the anchor and its k-nearest neighbors:

$$c = \sum_{i=1}^{k+1} w_i z_i, \quad \text{subject to} \quad \sum_{i=1}^{k+1} w_i = 1, \quad w_i \geq 0 \tag{2}$$

To sweep the continuous volume of this k-simplex uniformly, the barycentric weights $\mathbf{w} = [w_1, \ldots, w_{k+1}]$ are drawn from a Dirichlet distribution, $\mathbf{w} \sim \mathrm{Dir}(\mathbf{1})$.

By conditioning the vector field v_t^θ on these dynamically sampled, continuously varying barycentric coordinates, the model is forced to learn a dense mapping from the discrete feature anchors to the image space. This acts like a topological augmentation, allowing FSFM to synthesize infinite variations from a highly constrained local dataset.

Phase 1: Label-Free Source Pre-Training (Fig. 1, Left). We train the generative vector field v_t^θ on the unlabelled source dataset $\mathcal{D}_A$. At each step, we sample real data $x_1 \sim \mathcal{D}_A$, noise $x_0 \sim \mathcal{N}(0, I)$, and time $t \sim \mathcal{U}[0, 1]$. We extract the embedding $z_1 = E(x_1)$ and retrieve its 2-nearest neighbors across $\mathcal{D}_A$ using cosine similarity. This retrieval is entirely label-free, hence forcing the generative model to learn the natural, unsupervised geometry of the medical manifold. To construct the conditioning vector c_{x_1} while ensuring the generative vector field aligns with the ground-truth target x_1, barycentric sampling is anchor-biased: the primary embedding z_1 is assigned a random dominant weight $(0.6 - 0.9)$, while the remaining weight is distributed randomly among the 2 neighbors. The network θ is optimized via $\mathcal{L}_{CFM}$. To enable Classifier-Free Guidance (CFG) during inference, c_{x_1} is dropped to a null embedding $\emptyset$ with 10% probability.

Phase 2: Few-Shot Target Generation (Fig. 1, Right). To synthesize the augmented dataset $\mathcal{D}_{syn}$ at Hospital B, we deploy the frozen pre-trained model.

For a target class y^B, we select a random anchor $x_q \in \mathcal{D}_B$ and retrieve its 2 class-restricted neighbors to preserve diagnostic integrity. Because no ground-truth image exists to restrict generation during inference, the barycentric weights $\mathbf{w}$ are drawn uniformly from a symmetric Dirichlet distribution $\mathrm{Dir}(\mathbf{1})$, allowing c_{x_q} to fully explore the continuous volume of the simplex. Starting from pure noise $x_0 \sim \mathcal{N}(0, I)$, the final image x_1 is generated by integrating the ODE using a `dopri5` solver. At each step, CFG extrapolates the vector field:

$$\tilde{v}_t(x_t; c_{x_q}) = v_t^\theta(x_t; \emptyset) + w \cdot \big(v_t^\theta(x_t; c_{x_q}) - v_t^\theta(x_t; \emptyset)\big) \tag{3}$$

where w is the guidance scale.

3 Evaluation

We applied FSFM to synthesizing DR fundus images for adaptation across clinical centres. The generative model is trained on the unlabelled EyePACS [4,8] dataset (M = 35,126); the target clinic is simulated using MESSIDOR-2 [1,6] (1,744 fundus images from 874 patients, three French ophthalmology departments, Topcon TRC NW6 camera). Figure 2 shows representative FSFM generations paired with visually similar real anchors across DR grades: the synthetic samples reproduce overall retinal appearance, macula/optic disc placement, and grade-consistent features. Vascular patterns differ between paired samples: exact reproduction would indicate memorisation of individual patients.

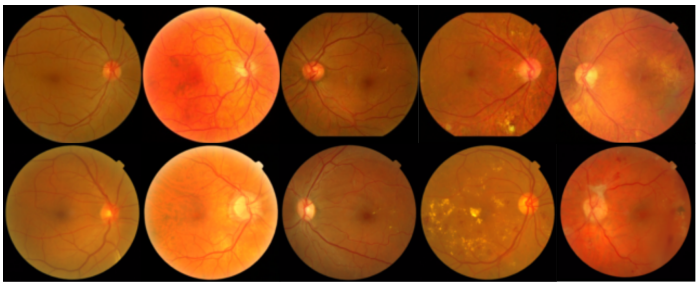

Fig. 2: FSFM generations of dilated cohort with $GS = 1.5$ (top row) paired with the real anchor most visually similar to each generation (bottom row), across DR grades G0-G4 (columns, left to right). Synthetic samples preserve the local retinal structure of their conditioning neighborhood while introducing variation.

Cohort design. We use MESSIDOR-2 comprising of Messidor-Original (1,058 PNG, three centres, mixed dilation) and Messidor-Extension (690 JPG, single centre, no dilation). Since FSFM conditions on the convex hull of retrieved neighbours, it implicitly assumes local consistency in feature space; mixed protocols can induce cross-protocol retrievals across visually distinct modes. We

stratify along pupil dilation, which is the clinical confounder most directly affecting retinal appearance and yield three cohorts: **Dilated** (385 patients, 769 images; Lariboisière + Saint-Étienne), **Non-dilated** (489 patients, 975 images; Brest CHRU), and **All MESSIDOR-2** (874 patients, 1,744 images) as a stress test under protocol violation. Severe DR is extremely scarce on Non-dilated (4 Class 3, 7 Class 4 patients in the reference pool); 5-class results there probe FSFM under extreme scarcity.

Splits, anchor budgets, and variance. Each cohort is partitioned via a single patient-level stratified split (on worst-eye DR grade), with both eyes of every patient placed in the same split: 50% reference pool, 40% test, 10% validation. We draw nested subsets at $N \in \{20, 50, 100, 200, 300\}$ from each reference pool, balanced where data permits. To quantify variance, we draw three independent anchor orderings (Seeds A, B, C); for each (N, anchor seed) we train three classifier seeds, yielding nine runs per condition.

Table 1: Five-class DR grading results at N ∈ {20, 50, 100} real anchors. Real: trained on N anchors with augmentation. FSFM-aug.: trained on N anchors + 1000 FSFM-generated synthetic samples with the same augmentation pipeline.

Cohort	N	Condition	QWK		BalAcc	Class 0	Class 4
Dilated	20	Real	$0.10_{\pm 0.10}$		$0.25_{\pm 0.05}$	$0.16_{\pm 0.12}$	$\mathbf{0.50}_{\pm 0.31}$
		FSFM-aug.	$\mathbf{0.20}_{\pm 0.08}$	(+0.10↑)	$\mathbf{0.29}_{\pm 0.03}$	$\mathbf{0.22}_{\pm 0.13}$	$0.28_{\pm 0.19}$
	50	Real	$0.29_{\pm 0.10}$		$0.33_{\pm 0.04}$	$0.26_{\pm 0.10}$	$0.28_{\pm 0.17}$
		FSFM-aug.	$\mathbf{0.37}_{\pm 0.04}$	(+0.09↑)	$\mathbf{0.35}_{\pm 0.04}$	$\mathbf{0.34}_{\pm 0.11}$	$\mathbf{0.39}_{\pm 0.12}$
	100	Real	$0.40_{\pm 0.11}$		$0.34_{\pm 0.04}$	$0.49_{\pm 0.09}$	$0.12_{\pm 0.08}$
		FSFM-aug.	$\mathbf{0.44}_{\pm 0.04}$	(+0.04↑)	$\mathbf{0.36}_{\pm 0.04}$	$\mathbf{0.55}_{\pm 0.09}$	$\mathbf{0.28}_{\pm 0.14}$
Non-dilated	20	Real	$0.01_{\pm 0.03}$		$0.23_{\pm 0.05}$	$0.17_{\pm 0.13}$	$0.33_{\pm 0.47}$
		FSFM-aug.	$\mathbf{0.07}_{\pm 0.04}$	(+0.06↑)	$\mathbf{0.27}_{\pm 0.05}$	$\mathbf{0.38}_{\pm 0.11}$	$\mathbf{0.52}_{\pm 0.21}$
	50	Real	$0.05_{\pm 0.07}$		$0.24_{\pm 0.06}$	$0.43_{\pm 0.32}$	$0.38_{\pm 0.45}$
		FSFM-aug.	$\mathbf{0.13}_{\pm 0.05}$	(+0.08↑)	$\mathbf{0.32}_{\pm 0.04}$	$\mathbf{0.59}_{\pm 0.08}$	$\mathbf{0.63}_{\pm 0.18}$
	100	Real	$\mathbf{0.16}_{\pm 0.06}$		$0.26_{\pm 0.03}$	$\mathbf{0.68}_{\pm 0.13}$	$0.05_{\pm 0.07}$
		FSFM-aug.	$0.13_{\pm 0.06}$	(−0.02↓)	$\mathbf{0.27}_{\pm 0.03}$	$0.61_{\pm 0.12}$	$\mathbf{0.32}_{\pm 0.15}$
All	20	Real	$0.03_{\pm 0.03}$		$0.24_{\pm 0.04}$	$0.23_{\pm 0.17}$	$\mathbf{0.41}_{\pm 0.38}$
		FSFM-aug.	$\mathbf{0.09}_{\pm 0.07}$	(+0.06↑)	$\mathbf{0.26}_{\pm 0.04}$	$\mathbf{0.23}_{\pm 0.17}$	$0.23_{\pm 0.15}$
	50	Real	$0.09_{\pm 0.06}$		$\mathbf{0.29}_{\pm 0.03}$	$0.29_{\pm 0.14}$	$\mathbf{0.32}_{\pm 0.14}$
		FSFM-aug.	$\mathbf{0.15}_{\pm 0.05}$	(+0.06↑)	$0.28_{\pm 0.04}$	$\mathbf{0.33}_{\pm 0.18}$	$0.23_{\pm 0.20}$
	100	Real	$0.20_{\pm 0.07}$		$\mathbf{0.32}_{\pm 0.03}$	$\mathbf{0.40}_{\pm 0.12}$	$\mathbf{0.27}_{\pm 0.13}$
		FSFM-aug.	$\mathbf{0.25}_{\pm 0.05}$	(+0.05↑)	$\mathbf{0.32}_{\pm 0.02}$	$0.37_{\pm 0.06}$	$\mathbf{0.27}_{\pm 0.15}$

Training. The generative vector field v_t^θ is a U-Net (channel multipliers $1, 2, 4, 8$; attention resolutions $32, 16, 8$; base channels 128), trained for 400 epochs with Adam (learning rate 10^{-4}) using the `torchcfm` library. We use frozen RETFound

DINOv2 embeddings as E, motivated by linear probing on EyePACS (DINOv2 QWK 0.50 vs ImageNet 0.31). ResNet-18 classifiers are trained for 200 epochs with AdamW (lr 10^{-4}, weight decay 10^{-3}), cosine annealing with warm restarts ($T_0 = 10$, $T_{\text{mult}} = 2$), and weighted cross-entropy with label smoothing 0.1 (class weights: inverse square-root of training-set class frequencies); we select the highest validation ROC-AUC checkpoint. Augmentations applied per batch to both real-only and FSFM-augmented sets: flips, rotation in $[-180°, 180°]$, perspective warping, jitter, and random sharpness, each $p = 0.5$. Each real anchor is augmented differently at every step, providing implicit oversampling; classical anchor duplication would therefore add negligible signal beyond our real-only baseline.

3.1 Scarcity impact

Table 1 reports 5-class DR grading results at $N \in \{20, 50, 100\}$ across the three cohorts. FSFM augmentation provides consistent QWK improvements over real-only baselines in the extreme-scarcity regime, with the largest absolute gain on the dilated cohort at N=20 (+0.10 QWK) and the largest relative gain on the non-dilated cohort at N=20 and N=50. Class-4 (proliferative DR) recall, the rarest grade, shows the most substantial improvements on the non-dilated cohort where FSFM-augmented classifiers detect proliferative cases at rates exceeding 50% versus 33-38% for real-only baselines. Figure 3 extends this comparison to higher anchor budgets (N up to 300). The FSFM advantage attenuates as N grows, with real-only classifiers reaching parity by N=200 on the dilated cohort. This means that synthetic augmentation provides the largest benefit in the extreme few-shot regime where real-only baselines have insufficient signal to learn discriminative features.

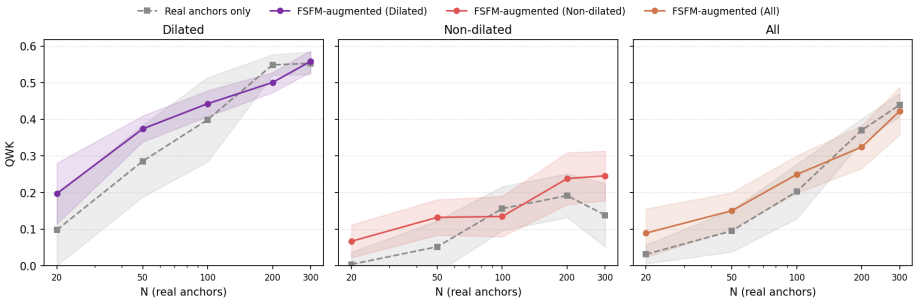

Fig. 3: 5-class DR grading QWK across three protocol-stratified cohorts. Solid lines: FSFM-augmented classifier (N real anchors + 1000 synthetic). Dashed lines: real-only baseline. Shaded regions: ±1 standard deviation across nine runs per condition (3 anchor seeds ×3 classifier seeds). Largest improvements are in the extreme few-shot regime ($N \leq 100$), with gains attenuating as N grows.

3.2 Diversity gain

To isolate the contribution of Stochastic Barycentric Sampling at inference time, we compare against deterministic centroid conditioning (uniform 1/3 weights on anchor and its two neighbours) using the same SBS-pre-trained generator. As shown in Figure 4, SBS produces consistently higher LPIPS diversity (26-39% increase across all cohort-N combinations) at comparable classifier QWK. We hypothesise that classifier gains are similar between conditions because the generator, pre-trained with stochastic conditioning, has internalised the diversity-promoting behaviour into its vector field, partially closing the gap when sampling deterministically at inference. The persistent diversity gap in synthetic outputs nonetheless confirms SBS's role as a topological augmentation mechanism.

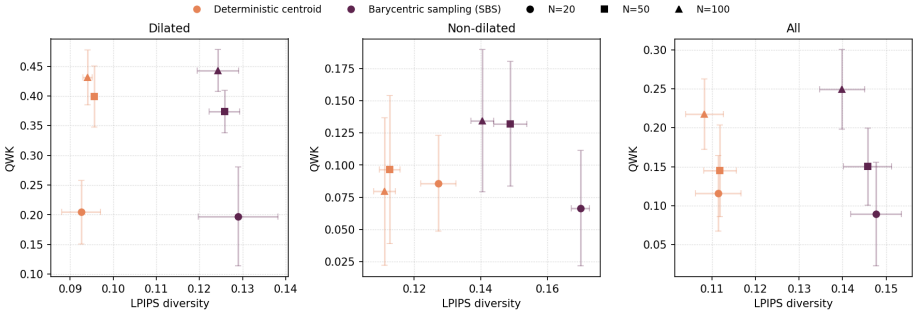

Fig. 4: Deterministic centroid versus Stochastic Barycentric Sampling, evaluated using the same SBS-pre-trained generator. Each panel shows LPIPS diversity (x-axis) versus downstream classifier QWK (y-axis) for $N \in \{20, 50, 100\}$ (marker shapes). SBS consistently produces higher-diversity synthetic samples than deterministic centroid conditioning at comparable classifier performance. Error bars: ± 1 standard deviation.

Discussion. We present this work as a proof of concept for retrieval-augmented few-shot flow matching in clinical synthesis. Our generative backbone is trained on 35k EyePACS images, which are approximately 2% of the data used by state-of-the-art retinal foundation models such as RETFound [22], and we expect the observed classifier gains at $N \leq 100$ to scale with stronger generative priors. Stochastic Barycentric Sampling produces measurably more diverse outputs than deterministic centroid conditioning, supporting its role as a topological augmentation mechanism. However, our sampling assumes approximate convexity of the local latent neighborhood; for extremely rare pathologies where anchor embeddings are sparse, interpolation may traverse low-density regions and produce anatomically implausible samples. We validated utility only through downstream classifier gains; qualitative assessment by expert ophthalmologists remains necessary to confirm visual fidelity. Finally, while designed to be privacy-preserving,

generative models can still memorise specific training samples; formal privacy audits of FSFM-generated samples are a necessary step before any clinical deployment.

4 Conclusion

We presented Few-Shot Flow Matching (FSFM), a privacy-preserving framework that adapts generative priors to data-scarce clinical domains without accessing source data at inference. Stochastic Barycentric Sampling produces measurably more diverse synthetic outputs than deterministic centroid conditioning, with downstream classifier gains concentrated in the extreme-scarcity regime ($N \leq 100$). Future work will investigate stronger generative priors, density-adaptive neighborhood scaling, and latent feature arithmetic to isolate pathology-specific vectors, extending FSFM toward patient-specific counterfactual generation.

Acknowledgments. AK is supported by a UKRI DTP award. JB is supported by UK Research and Innovation [UKRI AI Centre for Doctoral Training in Digital Healthcare number EP/Y0309741] and Ultromics Ltd. BK is supported by the ERC (101083647, 101246053) and DFG (513220538, 512819079). BG is supported by the Royal Academy of Engineering as part of the Kheiron/RAEng Research Chair.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abràmoff, M.D., Folk, J.C., Han, D.P., Walker, J.D., Williams, D.F., Russell, S.R., Massin, P., Cochener, B., Gain, P., Tang, L., et al.: Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA ophthalmology **131**(3), 351–357 (2013). https://doi.org/10.1001/jamaophthalmol.2013.1743
2. Blattmann, A., Rombach, R., Oktay, K., Müller, J., Ommer, B.: Semi-Parametric Neural Image Synthesis (2022), https://arxiv.org/abs/2204.11824
3. Clouser, M., Khezeli, K., Kalantari, J.: Few-shot LoRA adaptation of a flow-matching foundation model for cross-spectral object detection (2026), https://arxiv.org/abs/2601.04381
4. Cuadros, J., Bresnick, G.: EyePACS: An adaptable telemedicine system for diabetic retinopathy screening. Journal of Diabetes Science and Technology **3**(3), 509–516 (2009). https://doi.org/10.1177/193229680900300315
5. Dablain, D., Krawczyk, B., Chawla, N.V.: DeepSMOTE: Fusing deep learning and smote for imbalanced data (2021), https://arxiv.org/abs/2105.02340
6. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J.R., Massin, P., Erginay, A., et al.: Feedback on a publicly distributed image database: the messidor database. Image Analysis & Stereology **33**(3), 231–234 (2014). https://doi.org/10.5566/ias.1155
7. Deo, Y., Lin, F., Dou, H., Cheng, N., Ravikumar, N., Frangi, A.F., Lassila, T.: Few-shot learning in diffusion models for generating cerebral aneurysm geometries. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2024). https://doi.org/10.1109/ISBI56570.2024.10635313

8. Dugas, E., Jared, Jorge, Cukierski, W.: Diabetic retinopathy detection. https://kaggle.com/competitions/diabetic-retinopathy-detection (2015), kaggle
9. Giannone, G., Nielsen, D., Winther, O.: Few-shot diffusion models (2022), https://arxiv.org/abs/2205.15463
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks (2014), https://arxiv.org/abs/1406.2661
11. Guan, H., Liu, M.: Domain adaptation for medical image analysis: A survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2022). https://doi.org/10.1109/TBME.2021.3117407
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020), https://arxiv.org/abs/2006.11239
13. Khosravi, B., Rouzrokh, P., Mickley, J.P., Faghani, S., Mulford, K., Yang, L., Larson, A.N., Howe, B.M., Erickson, B.J., Taunton, M.J., Wyles, C.C.: Few-shot biomedical image segmentation using diffusion models: Beyond image generation. Computer Methods and Programs in Biomedicine **242**, 107832 (Dec 2023). https://doi.org/10.1016/j.cmpb.2023.107832
14. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling (2023), https://arxiv.org/abs/2210.02747
15. Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R.T.Q., Lopez-Paz, D., Ben-Hamu, H., Gat, I.: Flow matching guide and code (2024), https://arxiv.org/abs/2412.06264
16. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: DINOv2: Learning robust visual features without supervision (2024), https://arxiv.org/abs/2304.07193
17. Pandey, P.U., Micieli, J.A., Ong Tone, S., Eng, K.T., Kertes, P.J., Wong, J.C.Y.: Realistic fundus photograph generation for improving automated disease classification. British Journal of Ophthalmology **109**(7), 791–798 (2025). https://doi.org/10.1136/bjo-2024-326122
18. Sagers, L.W., Diao, J.A., Melas-Kyriazi, L., Groh, M., Rajpurkar, P., Adamson, A.S., Rotemberg, V., Daneshjou, R., Manrai, A.K.: Augmenting medical image classifiers with synthetic data from latent diffusion models (Aug 2023). https://doi.org/10.48550/arXiv.2308.12453
19. Yang, B., Yang, Y., Li, Q., Lin, D., Li, Y., Zheng, J., Cai, Y.: Classification of medical image notes for image labeling by using minbert. Tsinghua Science and Technology **28**(4), 613–627 (2023). https://doi.org/10.26599/TST.2022.9010012
20. Yuan, Z., Fang, Z., Huang, Z., Wu, F., Yao, Y.F., Li, Y.: Adapting Pre-trained Generative Model to Medical Image for Data Augmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 79–89 (Sep 2024), https://link.springer.com/chapter/10.1007/978-3-031-72086-4_8
21. Zhao, Z., Yang, J., Faghihroohi, S., Huang, K., Maier, M., Navab, N., Nasseri, M.A.: Label-Preserving Data Augmentation in Latent Space for Diabetic Retinopathy Recognition. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 284–294. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-43898-1_28

22. Zhou, Y., Chia, M.A., Wagner, S.K., Ayber, M.S., Williamson, D.J., Jasber, R.R., Gilbert, T., Lord, J., Patel, P.J., Ravindran, S.S., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (2023). https://doi.org/10.1038/s41586-023-06555-x