Inverse Flow Matching with Applications to Medical Image Classification

Mshari Alsirhni[1,2], Senjian An[1], Ling Li[1], and Qilin Li[1]

[1] School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, Western Australia 6102, Australia
[2] Information Technology Department, College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia

Abstract. Flow Matching (FM) models generate images by integrating an ordinary differential equation (ODE) that transports Gaussian noise to a target data distribution. Because the same ODE can be integrated backward, a conditional FM model also provides a natural inverse map from an image to a label-conditioned latent variable. This paper studies this inverse map as an interpretable signal for medical image classification. We show that, when an image is inverted using its correct class label, the inverse output retains little residual image structure and the recovered latent is closer to Gaussian noise; when an incorrect label is used, the inverse output retains structured image remnants. We quantify this behavior using spectral descriptors—spectral flatness, low-frequency energy, anisotropy, and radial spectral slope—together with cosine similarity between the input image and inverse output. MNIST is used as a proof-of-concept benchmark, where inverse-flow statistics achieve 95.19% test accuracy using only cosine similarity and spectral slope. We then evaluate the method on CAMUS echocardiography view classification, a medical image task distinguishing two-chamber and four-chamber views. Reliable echocardiographic view classification supports downstream cardiac ultrasound tasks, including segmentation and functional assessment. On CAMUS, inverse-flow features achieve 92.50% test accuracy and test average precision of 0.884 for correct-versus-incorrect label inversion. These results support inverse-flow analysis as a compact, interpretable bridge between conditional generation and medical image classification.

Keywords: Flow Matching · Inverse Flow · Medical Image Classification · Echocardiography · Gaussianity.

1 Introduction

Generative models that map simple priors to complex data distributions underpin advances in vision, graphics, and machine learning. In medical imaging, such models are increasingly attractive because they can capture data distributions, support synthesis, and potentially reveal uncertainty or structure not exposed by standard discriminative classifiers. GANs [6], VAEs [12], and diffusion models [8,20] have led this progress, but require adversarial optimization, approximate inference, or lengthy stochastic sampling procedures.

Flow Matching (FM) [16,1] offers a deterministic ODE-based formulation that avoids some of these limitations. By directly regressing a time-dependent velocity field that transports Gaussian noise to the data, FM supports stable training and efficient sampling using standard ODE solvers. A distinctive advantage of FM lies in its *exact invertibility*: since the generative process is defined by an ODE, reversing the learned flow field yields a mathematically well-defined map from data back to noise.

Despite this theoretical property, the inverse flow of FM has received little attention. Existing work focuses mainly on the quality of forward generation, architectural design, or theoretical links to diffusion models and optimal transport [1,17,4] . The inverse process, which does not require encoders or optimization-based inversion, remains underexplored.

The general idea of using a conditional generative model as a classifier by scoring candidate labels has been explored previously in diffusion models. In particular, diffusion-as-classifier methods evaluate candidate labels, or their corresponding text prompts, using diffusion likelihood, ELBO, or denoising-error proxy scores [15,3]. Our work builds on this broader generative label-scoring perspective, but studies a different setting: inverse conditional Flow Matching. Instead of relying on diffusion denoising or diffusion likelihood proxies, we invert the conditional flow and score the resulting label-dependent latent or residual structure. Thus, our novelty is not the general idea of converting a conditional generative model into a classifier, but rather the formulation and analysis of label scoring through inverse conditional Flow Matching, using Gaussianity-inspired spectral and structural inverse-flow descriptors.

This paper investigates Inverse Flow Matching in the conditional setting, with an emphasis on medical image classification. We uncover an interesting phenomenon: when a real image is inverted using its *correct* class label, the recovered latent variable closely resembles Gaussian noise; when an incorrect label is used, the inverse output retains structured, semantic remnants of the image. These label-dependent differences are strong enough to enable accurate classification.

We introduce spectral and structural measures to quantify deviations from Gaussianity, demonstrate their empirical effectiveness, and provide theoretical insights into their behavior. The resulting framework shows that inverse-flow analysis is not merely a byproduct of FM but a powerful tool for understanding and exploiting continuous-time generative models.

This work makes the following major contributions:

- **Inverse-flow characterization.** We provide a systematic study of inverse flows in Conditional Flow Matching (CFM) models for label scoring, positioning this setting as a Flow-Matching counterpart to diffusion-as-classifier methods while focusing on label-dependent inverse-flow statistics.

- **Spectral and structural descriptors.** We introduce a suite of interpretable descriptors for inverse-flow output, including spectral flatness, low-frequency energy, anisotropy, radial spectral slope, and cosine similarity. We further demonstrate that these statistics reliably distinguish correct from incorrect label inversions.
- **Generative-model-based classification.** We show that these descriptors enable accurate classification using only a lightweight linear head (logistic regression) on a few interpretable features, rather than a dedicated deep discriminative backbone, using MNIST as a proof of concept and CAMUS echocardiography view classification as the main medical image application.
- **Theoretical insight.** We use straight-path interpolants and vector-field perturbation analysis to explain why correct-label inverse flows recover Gaussian noise, while mismatched labels produce structured residuals.

In summary, we reposition flow matching as a classification-capable and analytically interpretable framework. Using inverse flow matching and inverse-flow statistics, we demonstrate that class information emerges naturally from flow dynamics, opening new applications of flow matching beyond image generation and toward health-oriented image analysis.

2 Related Work

2.1 Flow Matching and Continuous-Time Generative Models

Flow Matching [16] learns a time-dependent velocity field $\mathbf{u}_\theta(t, \mathbf{x})$ that defines an ODE that transports a simple prior (e.g., Gaussian) to the data distribution. The velocity is trained via a supervised regression objective using analytically known target velocities derived from interpolants between noise and data.

FM is closely related to Neural ODEs [2] and Continuous Normalizing Flows (CNFs) [7]. Unlike CNFs, which require log-determinant Jacobians for likelihood evaluation, FM focuses purely on velocity regression and can be used without explicit density computation. Stochastic and rectified variants [1,17,4] connect FM to diffusion models and optimal transport.

2.2 Conditional Flow Matching and Classifier-Free Guidance

Conditional Flow Matching (CFM) extends FM by conditioning the velocity field on labels:

$$\mathbf{u}_\theta(t, \mathbf{x} \mid k).$$

Practical implementations often use U-Net [19] backbones with FiLM-based conditioning [18]: time and label embeddings modulate intermediate feature maps via learned affine transformations. Attention layers are often used to aggregate global context information.

Classifier-Free Guidance (CFG) [9], originally proposed for diffusion models, interpolates between unconditional and conditional predictions. In the FM setting, one defines:

$$\mathbf{u}_{\mathrm{CFG}}(t, \mathbf{x}) = \mathbf{u}_\theta(t, \mathbf{x}|\varnothing) + \gamma \left(\mathbf{u}_\theta(t, \mathbf{x}|k) - \mathbf{u}_\theta(t, \mathbf{x}|\varnothing)\right),$$

where $\varnothing$ is the empty set that implies unconditional generation, $\gamma \geq 1$ controls the trade-off between fidelity and diversity. The larger γ increases the semantic alignment but also amplifies the curvature in the learned flow field.

2.3 Inversion in Generative Models

In GANs and diffusion models, mapping from data to latent space typically requires encoders or optimization-based inversion, which can be costly or yield approximate results. In contrast, FM defines a deterministic ODE and the inverse mapping corresponds to the solution of the reverse-time ODE. This work leverages this property to study the statistics of inverse flows and their applications.

3 Inverse of Flow Matching Models

Given an FM model defined by:

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}_\theta(t, \mathbf{x}(t)), \quad t \in [0, 1], \tag{1}$$

with $\mathbf{x}(0) \sim \mathcal{N}(0, I)$ and $\mathbf{x}(1) \sim p_{\mathrm{data}}$, we can define the inverse process by $\mathbf{z}(t) = \mathbf{x}(1-t)$:

$$\begin{aligned} \frac{d\mathbf{z}}{dt} &= \frac{d\mathbf{x}}{ds}\frac{ds}{dt} \quad (s = 1 - t) \\ &= -\mathbf{u}_\theta(s, \mathbf{x}(s)) \\ &= -\mathbf{u}_\theta(1-t, \mathbf{z}(t)). \end{aligned} \tag{2}$$

Thus, starting from $\mathbf{z}(0) = \mathbf{x}(1)$ (a data sample), integrating Eq. (2) from $t = 0$ to $t = 1$ yields $\mathbf{z}(1)$, which ideally lies in the Gaussian prior.

In the conditional case, the velocity field depends on a label k:

$$\frac{d\mathbf{z}}{dt} = -\mathbf{u}_\theta(1-t, \mathbf{z}(t) \,|\, k). \tag{3}$$

We denote the output noise for a real image I_{real} under label k by V_k. If k is the correct label, the reversed output is expected to lie in the Gaussian prior, otherwise, the reversed output should be far from the Gaussian noise.

Figure 1 provides a diagram for the overview of the forward and reverse flow matching.

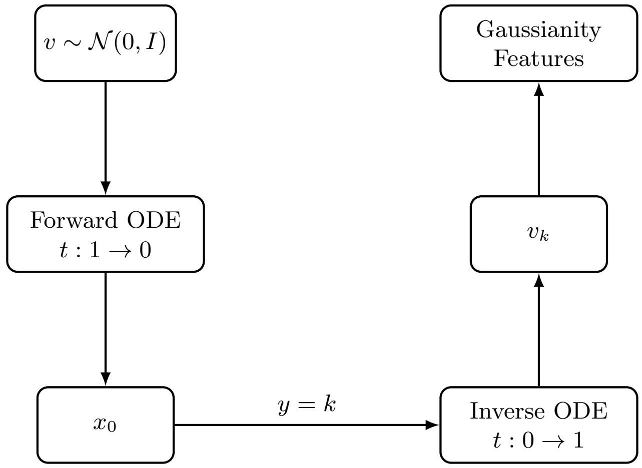

Fig. 1. Overview of forward and inverse flow matching. Forward integration map Gaussian noise to images; inverse integration maps images back to label-conditioned latents. Gaussianity of v_k varies with condition correctness.

3.1 Theoretical Analysis of Gaussianity in Inverse CFM

We now provide intuition for why inverse flows under the correct label approximate Gaussian noise, whereas mismatched labels produce structured deviations.

Straight-Path Interpolants and Ideal Velocities A common FM setup uses straight-path interpolants:

$$\mathbf{x}(t) = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1,$$

where $\mathbf{x}_0$ is Gaussian noise and $\mathbf{x}_1$ is a data sample. The ideal velocity field is:

$$\mathbf{u}^\star(t, \mathbf{x}) = \mathbf{x}_1 - \mathbf{x}_0,$$

constant along the path. When the CFM model is trained on correctly labeled pairs $(\mathbf{x}_0, \mathbf{x}_1, k_*)$, the learned conditional field $\mathbf{u}_\theta(t, \mathbf{x}|k_*)$ approximates $\mathbf{u}^\star$ along these paths.

For the inverse process with correct label k_*, we then have:

$$\frac{d\mathbf{z}}{dt} \approx -\mathbf{u}^\star(1 - t, \mathbf{z}(t)),$$

which closely follows a straight reverse path from $\mathbf{x}_1$ to $\mathbf{x}_0$. Since $\mathbf{x}_0$ is Gaussian, the endpoint $\mathbf{z}(1)$ is approximately Gaussian.

Label Mismatch as Vector Field Perturbation If an incorrect label $k \neq k_*$ is used, the learned field can be decomposed as:

$$\mathbf{u}_\theta(t, \mathbf{x}|k) = \mathbf{u}^\star(t, \mathbf{x}|k_*) + \Delta_k(t, \mathbf{x}),$$

where Δ_k captures label-dependent deviations from the ideal velocity. The inverse ODE becomes:

$$\frac{d\mathbf{z}}{dt} = -\mathbf{u}^\star(1-t, \mathbf{z}(t)) - \Delta_k(1-t, \mathbf{z}(t)).$$

Even if $\mathbf{u}^\star$ dominates, integrating this perturbed system accumulates structured errors correlated with class mismatch, leading to non-Gaussian endpoints V_k that retain semantic remnants of the image.

Effect of Classifier-Free Guidance CFG modifies the conditional velocity as:

$$\mathbf{u}_{\mathrm{CFG}}(t, \mathbf{x}) = \mathbf{u}_\theta(t, \mathbf{x}|\varnothing) + \gamma\big(\mathbf{u}_\theta(t, \mathbf{x}|k) - \mathbf{u}_\theta(t, \mathbf{x}|\varnothing)\big).$$

The perturbation term becomes:

$$\Delta_k^{(\gamma)} = \gamma\big(\mathbf{u}_\theta(t, \mathbf{x}|k) - \mathbf{u}_\theta(t, \mathbf{x}|\varnothing)\big),$$

which scales with γ. For inverse flows, higher guidance amplifies differences between correct and incorrect labels: correct-label inverse flows better recover the Gaussian prior; incorrect labels produce more pronounced non-Gaussian structure. This motivates using Gaussianity tests and similarity metrics as discriminative features.

4 Spectral and Structural Features of Inverse Flows

For each real image I_{real} with ground-truth label k_*, we compute inverse-flow outputs under all labels:

$$V_k = \mathrm{InverseFLOW}(I_{\mathrm{real}}; k), \quad k = 0, \ldots, m.$$

Where $m+1$ is the number of classes. When $k = k_*$, V_k should resemble Gaussian white noise; otherwise, it should contain structured remnants. We quantify this using the following features.

4.1 Spectral Features from the 2D Power Spectrum

Let $V_k \in \mathbb{R}^{H \times W}$ denote the inverse-flow output image under condition k, and let $\hat{V}_k(f_x, f_y)$ be its two-dimensional discrete Fourier transform (DFT) [5,10]. For pixel coordinates (x, y) and frequency indices (f_x, f_y), the DFT is defined as

$$\hat{V}_k(f_x, f_y) = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} V_k(x, y) \exp\left(-2\pi i \left(\frac{f_x x}{W} + \frac{f_y y}{H}\right)\right),$$

where $f_x = 0, \ldots, W-1$ and $f_y = 0, \ldots, H-1$.

The Fourier coefficients $\hat{V}_k(f_x, f_y) \in \mathbb{C}$ represent the contribution of spatial frequencies along the horizontal and vertical directions. The power spectrum is given by

$$P_k(f_x, f_y) = |\hat{V}_k(f_x, f_y)|^2.$$

To obtain an orientation-invariant description, we convert (f_x, f_y) to polar coordinates and define the radial frequency $r = \sqrt{f_x^2 + f_y^2}$. The radially averaged power spectrum $P_k(r)$ is computed by averaging $P_k(f_x, f_y)$ over all frequency bins with similar radius [21]. Low values of r correspond to large-scale, smooth image structures, whereas large r capture fine-scale details and noise.

Spectral flatness. The spectral flatness (SF) of V_k is defined as:

$$\mathrm{SF}(V_k) = \frac{\exp(\mathbb{E}[\log P_k])}{\mathbb{E}[P_k]}, \tag{4}$$

where expectations are taken over frequency bins. White Gaussian noise yields SF close to 1; structured signals yield SF < 1.

Low-frequency energy fraction. Let r_0 be a threshold radius. The low-frequency energy fraction is:

$$\mathrm{LF}(V_k) = \frac{\sum_{r<r_0} P_k(r)}{\sum_r P_k(r)}. \tag{5}$$

Digits and structured images tend to have larger low-frequency energy; Gaussian noise distributes energy more uniformly.

Anisotropy. We convert (f_x, f_y) to polar coordinates (r, θ) and define the angular energy $A_k(\theta)$ by integrating over a radial band. Normalizing $A_k(\theta)$ to sum to 1 yields a discrete distribution $p(\theta)$. The anisotropy is defined as:

$$\mathrm{ANISO}(V_k) = 1 - \frac{H(p)}{\log M}, \tag{6}$$

where $H(p)$ is the entropy and M is the number of angular bins. White noise yields near-uniform $p(\theta)$ and low anisotropy; structured images with oriented strokes yield higher anisotropy.

Spectral slope. We fit a line to $\log P_k(r)$ versus $\log r$:

$$\log P_k(r) \approx a_k \log r + b_k,$$

and use a_k as the spectral slope. For white noise, $a_k \approx 0$ (flat spectrum); for structured images, a_k is typically negative (more low-frequency power).

4.2 Cosine Similarity between Image and Inverse Output

In addition to spectral features, we introduce a structural feature that captures alignment between the real image and its inverse output:

$$\mathrm{CosSim}(I_{\mathrm{real}}, V_k) = \frac{\langle I_{\mathrm{real}}, V_k \rangle}{\|I_{\mathrm{real}}\|_2 \, \|V_k\|_2}. \tag{7}$$

When $k = k_*$, V_k ideally contains no image structure and CosSim is close to zero. For incorrect labels, V_k often contains residual patterns of the input image, yielding higher cosine similarity. Empirically, this feature provides a strong discriminative signal and significantly improves classification performance when combined with spectral features.

5 Classification Using Inverse-Flow Features

For each image I_{real} with label k_*, and each label k, we compute V_k and the feature vector:

$$\phi_k = \big[\mathrm{SF}(V_k), \mathrm{LF}(V_k), \mathrm{ANISO}(V_k), a_k, \mathrm{CosSim}(I_{\mathrm{real}}, V_k)\big].$$

We assign binary labels:

$$y_k = \begin{cases} 1, & \text{if } k = k_* \\ 0, & \text{if } k \neq k_*. \end{cases}$$

A logistic regression classifier trained on $\{\phi_k, y_k\}$ distinguishes Gaussian (true label) vs. non-Gaussian (incorrect label) cases.

Aggregating the probabilities per-label $\hat{p}_k$ and selecting $\arg\max_k \hat{p}_k$ produce a predicted class.

Interpretability. We call the proposed classifier interpretable in a feature-level and inversion-level sense. The final decision is based on a small set of named, physically meaningful statistics rather than a high-dimensional hidden representation: spectral flatness and low-frequency energy measure departure from white-noise Gaussianity, radial spectral slope measures residual natural-image structure, anisotropy measures directional structure, and cosine similarity measures how much of the input image remains aligned with the inverse output. For each candidate label k, the method also exposes the intermediate inverse image V_k, so a prediction can be inspected by comparing whether the selected label produces the most noise-like latent and whether rejected labels preserve anatomical remnants. Thus, the explanation for a decision is given by the per-label inverse outputs and their feature values; for example, a correct CAMUS view label is supported when its inverse output has lower residual similarity and more Gaussian-like spectral statistics than the competing view.

6 Experiments

The experiments have two roles. First, MNIST is used only as a controlled proof of concept to verify the inverse-flow Gaussianity effect and to identify compact features. Second, CAMUS echocardiography is used as the example of applications of the inverse flow matching to medical image classification.

6.1 Datasets and Evaluation Protocol

MNIST proof of concept. MNIST contains 60,000 training images and 10,000 test images from ten digit classes. Images are normalized to $[-1, 1]$ and resized to 32×32. For each image, inverse-flow outputs are computed under all ten label conditions. This benchmark is not used to claim medical relevance; it is used to demonstrate the mechanism in a familiar setting where the correct label is unambiguous.

CAMUS medical image task. The main medical imaging experiment uses the CAMUS echocardiography dataset [14]. We formulate an image-level view-classification task between two standard apical views: two-chamber (2CH) and four-chamber (4CH). Reliable discrimination between 2CH and 4CH acquisitions is clinically relevant because these views are routinely used for cardiac assessment and left ventricular volume quantification [13]. Correct view identification is also an important preprocessing step in automated echocardiography pipelines before downstream segmentation, quantitative measurement, and functional analysis [11,23]. For each patient, we use the End-Diastole (ED) and End-Systole (ES) frames and exclude segmentation masks and temporal sequence files. We follow the official patient-disjoint split provided with the CAMUS dataset files, using 400 patients for training, 50 patients for validation, and 50 patients for testing. This gives 1600 training images, 200 validation images, and 200 test images, where each patient contributes four images: 2CH-ED, 2CH-ES, 4CH-ED, and 4CH-ES. The task is balanced because each patient contributes the same number of 2CH and 4CH images.

Model and inverse-feature classifier. For both datasets, the velocity field is implemented by a compact conditional U-Net with FiLM-based time and label conditioning, Group Normalization, SiLU activations, and an intermediate self-attention block. The model is trained using straight-path Flow Matching. ODE sampling and inversion use the Dopri5 solver with tolerances $\texttt{rtol} = 10^{-5}$ and $\texttt{atol} = 10^{-5}$; classifier-free guidance uses $\gamma = 2$ unless otherwise stated. For each image I and each candidate label k, we compute $V_k = \text{InverseFLOW}(I; k)$ and extract spectral flatness (F), low-frequency energy fraction (L), anisotropy (A), radial spectral slope (S), and cosine similarity (C). A logistic regression classifier is trained on these inverse-flow features to distinguish correct-label inversions from incorrect-label inversions. The final class prediction is obtained by selecting the label with the largest correctness probability, $\hat{k} = \arg\max_k \hat{p}_k$.

6.2 MNIST Proof-of-Concept Results

Figure 2 illustrates the central phenomenon on MNIST. When the condition matches the true digit label, the inverse output is visually closer to unstructured noise; incorrect labels often leave digit-like residuals. This supports the use of Gaussianity and residual-structure statistics as classification features.

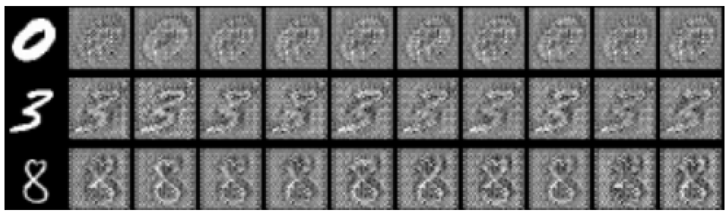

Fig. 2. MNIST inverse-flow outputs under all candidate label conditions. Correct-label inversions are more noise-like, while incorrect labels preserve stronger digit remnants.

Table 1. MNIST digit classification accuracy using inverse-flow statistics. CosSim, LowFreq, Aniso, Flatness, and Slope are denoted by C, L, A, F, S, and $SPs = \{S, F, L, A\}$ denotes all four spectral features.

	F	L	A	S	C	C+L	C+S	C+SPs
Training	0.5336	0.6923	0.2530	0.6546	0.9100	0.9450	**0.9531**	0.9518
Testing	0.5443	0.7022	0.2534	0.6601	0.9031	0.9463	**0.9519**	0.9516

Table 1 shows that cosine similarity is the strongest single feature, achieving 90.31% test accuracy, while spectral flatness and anisotropy are weaker as standalone signals. This suggests that the dominant cue is not a direct standalone test of latent Gaussianity, but rather a residual-direction or residual-structure consistency signal. Among the spectral features, low-frequency energy and radial spectral slope remain the most informative complementary descriptors. The compact pair $C + S$ achieves the best MNIST test accuracy of 95.19%, essentially matching the full feature set. We therefore interpret the proposed features as Gaussianity-inspired inverse-flow descriptors, with cosine similarity providing the dominant classification signal, and reserve the main experimental emphasis for the medical task below.

6.3 CAMUS Medical Image Classification Results

Figure 3 shows samples generated by the conditional FM model for 2CH and 4CH conditions. These samples are included to confirm that the model learns classconditioned echocardiographic image distributions; classification is performed using inverse-flow statistics rather than direct image features.

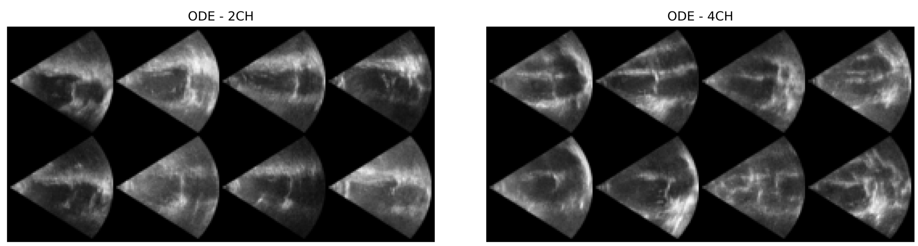

Fig. 3. Generated CAMUS echocardiography images using Flow Matching. Left: 2CH samples. Right: 4CH samples.

Figure 4 shows representative CAMUS inversions. For 4CH and 2CH inputs, the correct label produces a more noise-like inverse output, whereas the wrong label retains visible anatomical structure. This behavior mirrors the MNIST proof-of-concept but is more clinically relevant because the residual structure corresponds to view-dependent cardiac anatomy.

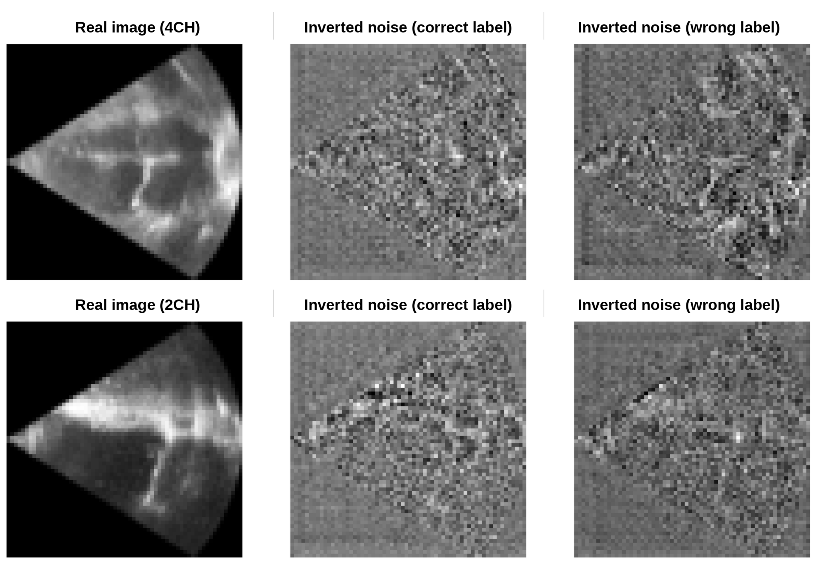

Fig. 4. Label-dependent inversion on CAMUS. Top: a 4CH example. Bottom: a 2CH example. In each row, the input image is compared with inverse outputs under correct and incorrect labels. Correct-label inversion is more Gaussian-like, while incorrect-label inversion preserves anatomical remnants.

Table 2. Indicative comparison with previously reported downstream CAMUS image-level classification results.

Model / Method	Input / Representation	Setting	ACC
ResNet18 [22]	Real images	Direct image classifier	0.89
ResNet18 [22]	DDPM synthetic images	50 sampling steps	0.78
ResNet18 [22]	MOTFM synthetic images	10 inference/ODE steps	0.82
ResNet50 [22]	Real images	Direct image classifier	0.88
ResNet50 [22]	DDPM synthetic images	50 sampling steps	0.78
ResNet50 [22]	MOTFM synthetic images	10 inference/ODE steps	0.81
Ours	Real images	Inverse-flow, C only	0.905
Ours	**Real images**	**Inverse-flow, All features**	**0.925**

Note. The ResNet18, ResNet50, DDPM, and MOTFM results are reported by Yazdani et al. [22]. DDPM denotes Denoising Diffusion Probabilistic Models [8]; DDPM (50) indicates 50 diffusion sampling steps. MOTFM denotes Medical Optimal Transport Flow Matching [22]; MOTFM (10) indicates 10 inference/ODE sampling steps. For our method, real CAMUS images are transformed into inverse-flow outputs through image-to-noise inversion. Classification is then performed using statistics extracted from these inverse-flow outputs. Here, C denotes cosine similarity, and All features denotes $\{F, L, A, S, C\}$, where F is spectral flatness, L is low-frequency fraction, A is anisotropy, and S is spectral slope. These previously reported values are included as contextual reference points, since they were obtained under the original setting of Yazdani et al. [22]. Therefore, Table 2 should be interpreted as an indicative comparison rather than a controlled same-split benchmark. Accordingly, while the reported reference values are useful for context, the absence of a fully matched same-split discriminative baseline remains a limitation for direct head-to-head interpretation.

Table 2 compares our CAMUS image-level view-classification results with the downstream classification results reported by Yazdani et al. [22]. Their best real-image classifier achieved an ACC of 0.89, while classifiers trained on synthetic images generated by DDPM and MOTFM achieved lower performance. Using only cosine similarity (C), our inverse-flow-based method achieved 0.905 ACC on the CAMUS test set, and using all five inverse-flow features improved the result to 0.925 ACC. This suggests that inverse-flow statistics extracted from image-to-noise inverse outputs provide discriminative information for CAMUS view classification.

Figure 5 further evaluates the underlying binary decision of whether a candidate inverse output corresponds to the correct label. The test average precision of 0.884 indicates that inverse-flow features are not only useful for final view classification but also provide a calibrated correctness signal for label-conditioned inversion. Together, the MNIST and CAMUS experiments strengthen the empirical case for inverse FM as an interpretable generative-model-based classifier for health imaging applications.

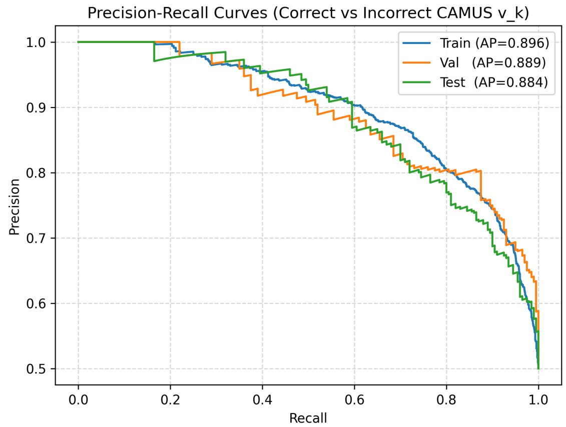

Fig. 5. Precision–Recall curves for correct-versus-incorrect CAMUS inversions. The average precision is 0.896 on train, 0.889 on validation, and 0.884 on test.

7 Conclusion

This paper presented inverse-flow analysis for Conditional Flow Matching models and applied it to interpretable medical image classification. We showed that inverse flows under the correct label condition map real images back to more Gaussian-like noise, while incorrect label conditions produce structured, non-Gaussian outputs. We introduced spectral and structural features to quantify these differences.

Experiments on MNIST demonstrate the concept in a controlled setting, while the CAMUS echocardiography results show that the same inverse-flow statistics transfer to a real medical image classification task. The best CAMUS setting achieves 92.50% test accuracy for 2CH-vs.-4CH view classification, and the correct-versus-incorrect inversion classifier achieves 0.884 test average precision.

These findings suggest that inverse-flow analysis is a useful tool for understanding and exploiting continuous-time generative models in health imaging. Future work includes extending the method to disease classification, larger multi-center datasets, multimodal conditioning, and stochastic flow variants.

References

1. Albergo, M.S., Boffi, N.M., Vanden-Eijnden, E.: Stochastic interpolants: A unifying framework for flows and diffusions. arXiv preprint arXiv:2303.08797 (2023)
2. Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. Advances in neural information processing systems **31** (2018)
3. Clark, K., Jaini, P.: Text-to-image diffusion models are zero shot classifiers. Advances in Neural Information Processing Systems **36**, 58921–58937 (2023)

4. Geng, Z., Deng, M., Bai, X., Kolter, J.Z., He, K.: Mean flows for one-step generative modeling. arXiv preprint arXiv:2505.13447 (2025)
5. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson, 4 edn. (2018)
6. 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 **27** (2014)
7. Grathwohl, W., Chen, R.T., Bettencourt, J., Sutskever, I., Duvenaud, D.: Ffjord: Free-form continuous dynamics for scalable reversible generative models. arXiv preprint arXiv:1810.01367 (2018)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
9. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
10. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice Hall (1989)
11. Khamis, H., Zurakhov, G., Azar, V., Raz, A., Friedman, Z., Adam, D.: Automatic apical view classification of echocardiograms using a discriminative learning dictionary. Medical image analysis **36**, 15–21 (2017)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
13. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., et al.: Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. European Heart Journal-Cardiovascular Imaging **16**(3), 233–271 (2015)
14. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., Lartizien, C., D'hooge, J., Lovstakken, L., Bernard, O.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging **38**(9), 2198–2210 (2019). `https://doi.org/10.1109/TMI.2019.2900516`
15. Li, A.C., Prabhudesai, M., Duggal, S., Brown, E., Pathak, D.: Your diffusion model is secretly a zero-shot classifier. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2206–2217 (2023)
16. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
17. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
18. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
20. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
21. Unser, M.: Sum and difference histograms for texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8**(1), 118–125 (1986). `https://doi.org/10.1109/TPAMI.1986.4767760`

22. Yazdani, M., Medghalchi, Y., Ashrafian, P., Hacihaliloglu, I., Shahriari, D.: Flow matching for medical image synthesis: Bridging the gap between speed and quality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2025)
23. Zhang, J., Gajjala, S., Agrawal, P., Tison, G.H., Hallock, L.A., Beussink-Nelson, L., Lassen, M.H., Fan, E., Aras, M.A., Jordan, C., et al.: Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation **138**(16), 1623–1635 (2018)