



PUNCH: Physics-informed uncertainty-aware network for coronary hemodynamics

Sukirt Thakur^{a,*,}, Marcus Roper^{b,}, Yang Zhou^{a,}, Dmitry Yu. Isaev^{a,}, Reza Akbarian Bafghi^{c,}, Brahmajee K. Nallamothu^{d,}, C. Alberto Figueroa^{e,f,}, Srinivas Paruchuri^{a,}, Scott Burger^{a,}, Carlos Collet^{g,}, Maziar Raissi^{h,}

^a AngioInsight, Inc., USA

^b Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, 90095, USA

^c University of Colorado Boulder, Boulder, CO, 80309, USA

^d Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA

^e Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, 48109, USA

^f Department of Surgery, University of Michigan, Ann Arbor, MI, 48109, USA

^g Clinical Trial Center, Cardiovascular Research Foundation, New York, NY, USA

^h University of California, Riverside, Riverside, CA, 92521, USA

ARTICLE INFO

Keywords:

Coronary flow reserve
Coronary microvascular dysfunction
Coronary angiography
Physics-informed neural networks
Uncertainty quantification
Variational inference

ABSTRACT

More than 10 million coronary angiograms are performed globally each year, providing a gold standard for detecting obstructive coronary artery disease. Yet, no obstructive lesions are identified in 70% of patients evaluated for ischemic heart disease. Up to half of these patients have undiagnosed, life-limiting coronary microvascular dysfunction (CMD), which remains under-detected due to the limited availability of invasive wire-based tools (Doppler-wire or thermodilution) used to measure coronary flow reserve (CFR). Here, we introduce PUNCH, a wire-free, uncertainty-aware framework for estimating CFR from paired resting and hyperemic coronary angiographic acquisitions. PUNCH integrates physics-informed neural networks with variational inference to infer coronary blood flow from first-principles models of contrast transport, without requiring ground-truth flow measurements or population-level training. The pipeline runs in approximately three minutes per patient on a single GPU. Evaluated on 1000 synthetic kymographs and a feasibility-scale, single-center cohort of 20 patients with matched invasive bolus thermodilution CFR, PUNCH produces CFR point estimates that correlate strongly with the invasive reference and uncertainty intervals that widen under image degradation; the raw posterior intervals are however under-dispersed relative to nominal coverage, and we discuss the recalibration step that would be required before threshold-based clinical use. As a proof-of-concept feasibility study, this work illustrates how physics-informed inference could expand the physiological information extractable from existing angiographic imaging, pending validation in larger, multi-center, multi-territory cohorts.

1. Introduction

Coronary microvascular dysfunction (CMD) is a prevalent yet underdiagnosed condition, present in up to 70% of patients with angina and non-obstructive coronary arteries undergoing invasive angiography (Kunadian et al., 2021). This diagnostic oversight disproportionately impacts female patients (Ong et al., 2018). Unlike epicardial stenosis, CMD reflects abnormalities in flow and structure that are invisible on standard angiography but critically influence myocardial perfusion and patient outcomes (Murthy et al., 2012; Taqueti and Di Carli, 2018).

The gold standard for diagnosing coronary microvascular function is the measurement of coronary flow reserve (CFR) using invasive catheter-based techniques such as Doppler flow wires or thermodilution (Taqueti and Di Carli, 2018; Barbato et al., 2004). In thermodilution, a known volume of cold saline is injected into the coronary artery, and downstream temperature changes are recorded by a thermistor-tipped catheter. The resulting temperature–time curve — an indicator-dilution profile — allows for estimation of blood flow based on the area under the curve (Fearon et al., 2003). Although reliable, thermodilution

* Corresponding author.

E-mail address: sukirt@angioinsight.com (S. Thakur).

<https://doi.org/10.1016/j.media.2026.104273>

Received 29 April 2026; Received in revised form 14 August 2026; Accepted 16 August 2026

Available online 3 September 2026

1361-8415/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

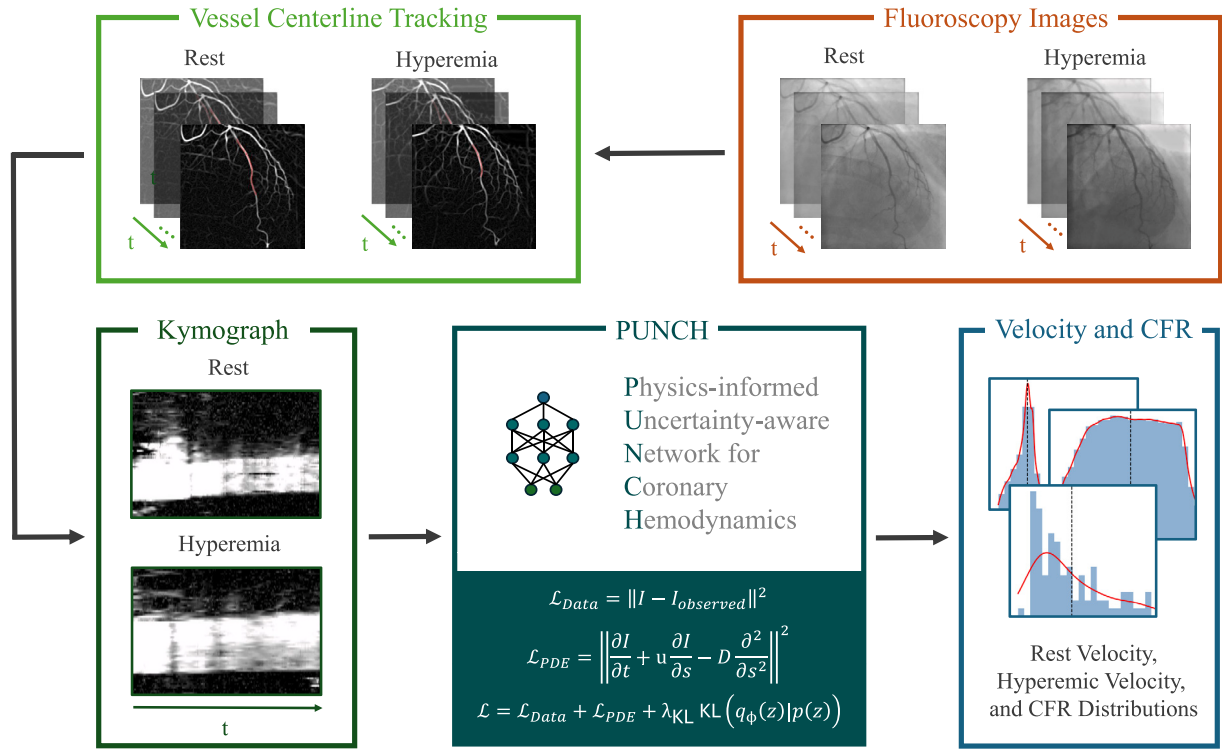


Fig. 1. PUNCH enables uncertainty-aware estimation of CFR from standard angiography. Fluoroscopy images acquired at rest and hyperemia undergo vessel centerline tracking to generate spatiotemporal intensity maps (kymographs). These kymographs serve as input to a physics-informed neural network that jointly optimizes data fidelity ($\mathcal{L}_{Data}$), advection–diffusion physics constraints ($\mathcal{L}_{PDE}$), and variational inference regularization ($\mathcal{L}_{KL}$). The framework outputs posterior distributions of rest velocity, hyperemic velocity, and CFR, providing probabilistic estimates with quantified uncertainty. The dual-branch architecture models both physiological states while sharing a common latent variable z to capture measurement uncertainty from noise, motion artifacts, and incomplete contrast filling.

requires precise timing, is sensitive to injection technique and temperature variability, and carries procedural risks inherent to placement of a coronary guide-wire in the distal coronary artery (De Bruyne et al., 1994; Fearon et al., 2003). These practical constraints have limited the widespread invasive evaluation of CFR, leaving most patients without a definitive functional diagnosis.

In principle, contrast-enhanced coronary angiography encodes flow information analogous to thermodilution: the transport and dispersion of contrast dye through the coronary lumen reflect local hemodynamics (Zierler, 2000; Ball et al., 2018). Recently, there has been an interest in using coronary angiography as a diagnostic tool for coronary microvascular dysfunction using data-driven approaches (Yang et al., 2026). However, extracting quantitative flow metrics from these time-sequenced images remains challenging due to image noise, low frame rates, cardiac motion, and vessel overlap, all of which create conditions that make it difficult to directly track the contrast front advancing through the coronary arterial tree.

Recent advances in physics-informed machine learning offer a promising pathway for addressing the inverse problem of inferring ‘hidden’ hemodynamic variables, such as blood velocity, from observed contrast dye transport dynamics. In contrast to conventional data-driven approaches, physics-informed neural networks (PINNs) incorporate known physical laws directly into the learning process, allowing the estimation of hidden physiological quantities from indirect measurements, such as medical images (Raissi et al., 2019). However, most existing PINN-based approaches yield only deterministic point estimates and provide no measure of confidence, which is an important limitation in clinical settings where image quality, motion, and contrast delivery can vary substantially (Yang and Perdikaris, 2019; Zou et al., 2025; Linka et al., 2022).

Here, we introduce an uncertainty-aware, physics-informed framework to estimate CFR from paired resting and hyperemic coronary angiographic acquisitions. Because coronary angiography is itself an invasive catheter-based procedure, the practical claim we make is not that the assessment is non-invasive, but that it is *wire-free*: it avoids the additional intravascular pressure-wire or Doppler-wire instrumentation otherwise required for invasive CFR measurement, and it relies on imaging that is acquired during a procedure already being performed. The key insight is simple: contrast material injected during angiography behaves as an intravascular tracer. As it travels through the coronary artery, it is transported by blood flow and simultaneously spreads due to velocity heterogeneity and mixing. Transport observations encode information about coronary blood flow and microvascular resistance.

Our approach leverages this physical behavior to infer flow dynamics from angiographic image sequences while explicitly quantifying uncertainty. Rather than relying on population-level training or labeled flow data, each patient is independently analyzed using first-principles contrast transport physics. Importantly, the framework produces not only a CFR estimate, but also a credible interval that reflects the uncertainty related to image quality, motion artifacts, and incomplete contrast filling. This uncertainty-aware formulation enables a safer interpretation of angiography-derived physiology and provides a pathway toward robust, wire-free assessment of coronary microvascular function from imaging acquired during invasive angiographic procedures already used in clinical care. We position the present study as a proof-of-concept feasibility evaluation.

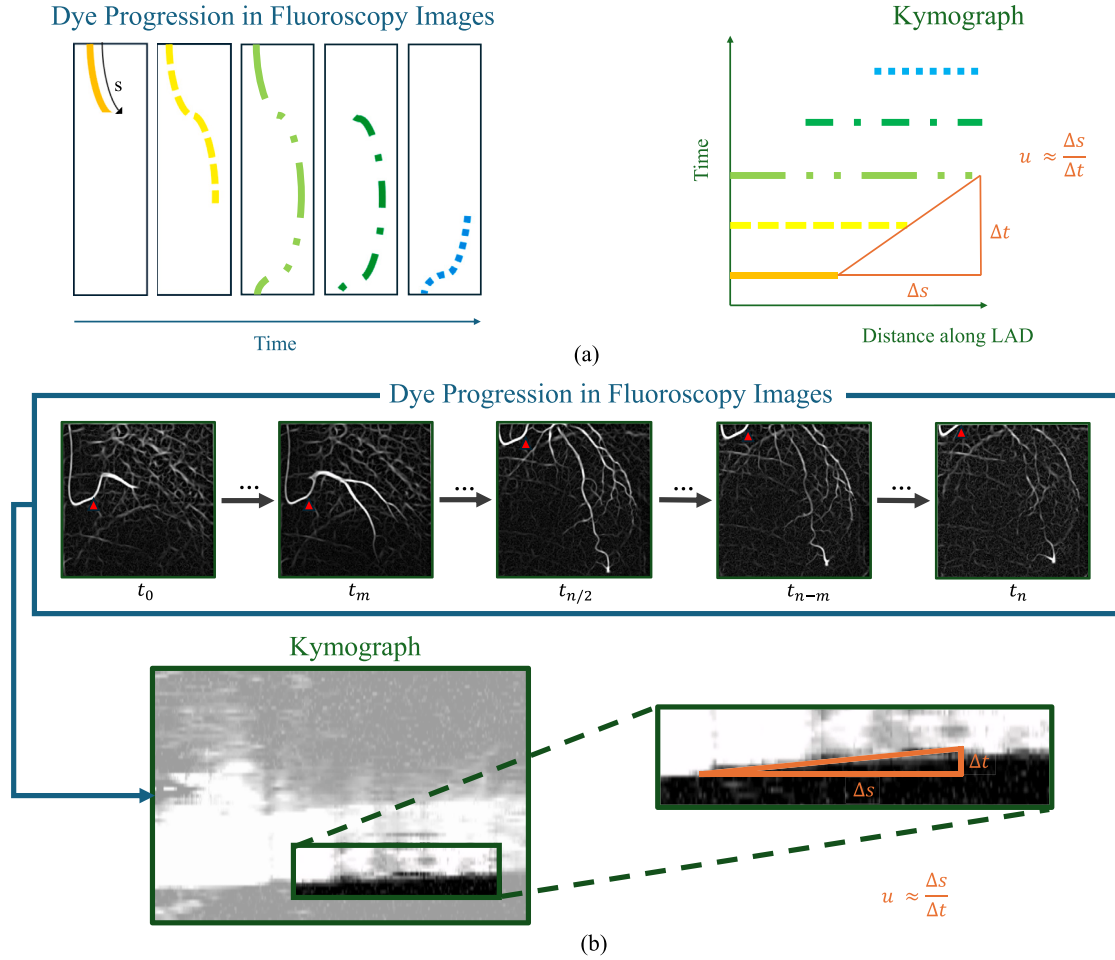


Fig. 2. Kymograph construction from coronary angiography. Spatiotemporal representation of contrast transport enables velocity inference. (a) Sequential fluoroscopy frames show progressive contrast dye advancement through the left anterior descending (LAD) coronary artery. Vessel centerline tracking identifies corresponding anatomical positions across frames. (b) Stacking intensity profiles along the centerline over time constructs a kymograph where the x -axis represents distance along the vessel (s) and the y -axis represents time (t). Diagonal patterns in the kymograph reflect contrast propagation, with slope $\Delta s / \Delta t$ approximating local blood velocity u . This representation transforms the inverse problem of velocity estimation from fluoroscopy into a tractable physics-informed learning task governed by advection–diffusion dynamics.

2. Materials and methods

2.1. Overview of the flow estimation framework

The goal of the proposed framework is to estimate the CFR from paired resting and hyperemic coronary angiographic acquisitions while providing an explicit measure of confidence in each prediction. CFR is defined clinically as the ratio of blood flow during pharmacologically induced hyperemia to that at rest and is a key marker of coronary microvascular function.

To achieve this, we model how injected contrast material moves and spreads along the coronary artery over time. We focus on the left anterior descending (LAD) artery as this is the most common artery studied in coronary microvascular dysfunction testing in routine clinical practice. Because coronary flow is predominantly longitudinal over the imaged segment of the LAD, contrast transport can be approximated as a one-dimensional process along the centerline of the vessel (Fig. 1). This approximation is widely used in indicator-dilution theory and enables tractable inference from standard fluoroscopic images without requiring three-dimensional vessel reconstruction (Zierler, 1962, 2000).

Clinical angiograms often exhibit noise, motion artifacts, variable contrast injection, and incomplete vessel opacification. To account for these factors, the framework explicitly models uncertainty during inference. A shared latent variable of dimension Z_{dim} , denoted $\mathbf{z}$, is

introduced to represent unobserved sources of data degradation and physiological variability. Conceptually, this latent variable allows the model to express lower confidence when the observed image data are ambiguous or degraded, and higher confidence when the data are clear and consistent with physical expectations (Kingma and Welling, 2014; Rezende et al., 2014).

From each angiographic sequence, the model estimates three physically significant quantities:

1. contrast intensity along the vessel as a function of space coordinate (s) and time (t) conditioned on the latent vector ($\mathbf{z}$), $I(s, t; \mathbf{z})$;
2. cross-sectionally averaged and time-averaged axial blood velocity conditioned on the latent vector, $u(s; \mathbf{z})$; and
3. an effective dispersion coefficient conditioned on the latent vector, $D(s, t; \mathbf{z})$, which captures contrast spread due to velocity variations, mixing, and unresolved three-dimensional effects.

Together, these quantities describe how contrast propagates through the coronary artery and form the basis for estimating flow reserve.

The three quantities $I(s, t; \mathbf{z})$, $u(s; \mathbf{z})$, and $D(s, t; \mathbf{z})$ are represented using neural networks that are trained jointly under physical constraints. These networks are optimized in such a way that the predicted contrast dynamics satisfy the governing laws of contrast transport while remaining consistent with the observed angiographic intensities.

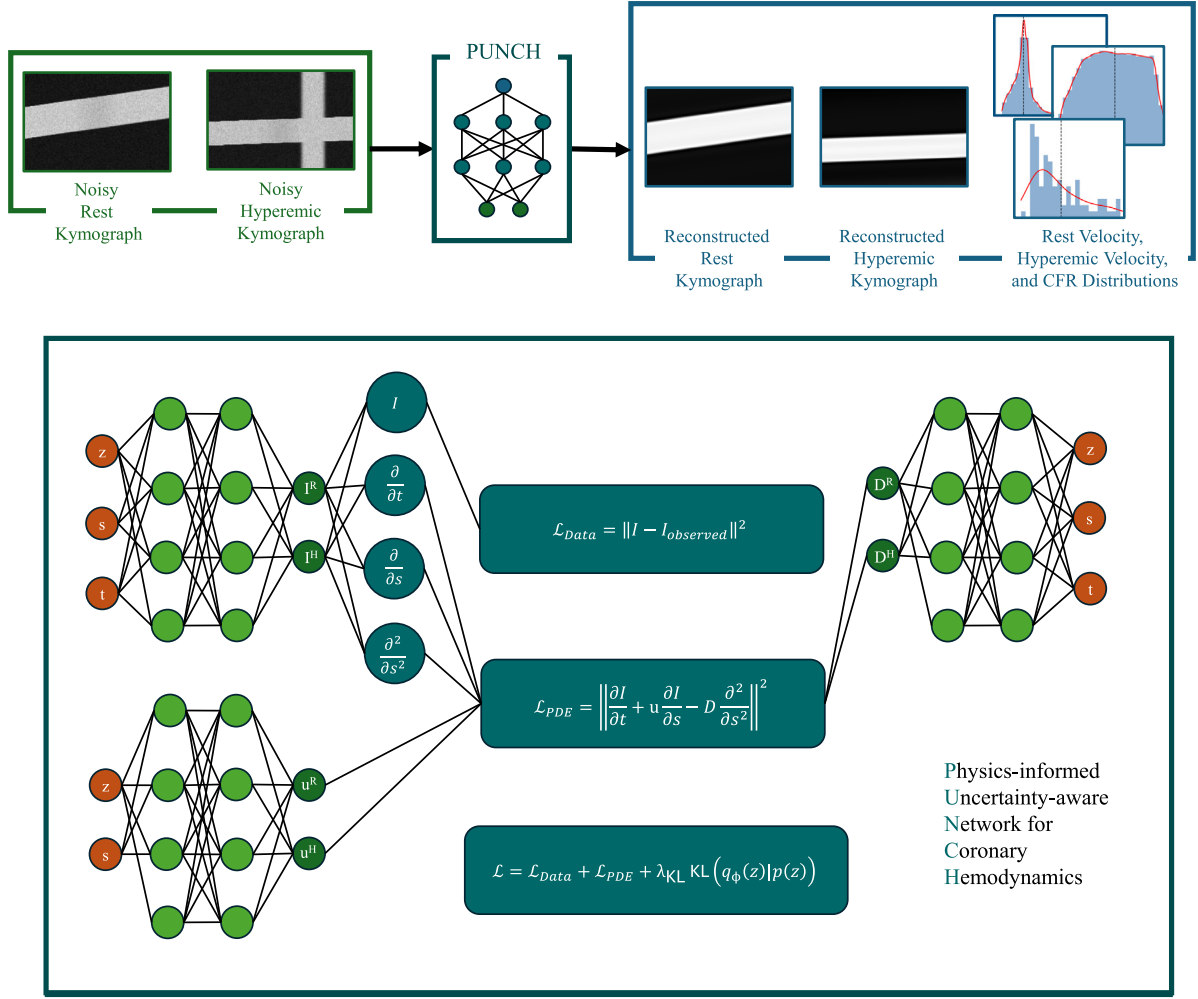


Fig. 3. Physics-informed neural networks with variational inference enable joint rest-hyperemia modeling. The framework consists of three subnetworks per physiological state: intensity predictor $I(s, t)$, velocity predictor $u(s)$, and dispersion predictor $D(s, t)$. Rest (subscript R) and hyperemia (subscript H) branches share a common latent variable $z \in \mathbb{R}^d$ sampled from a learned variational posterior $q_\phi(z)$, ensuring coherent uncertainty propagation between states. Training optimizes a composite loss combining data fidelity to observed noisy kymographs ($\mathcal{L}_{Data}$), physics-based residuals enforcing the advection–diffusion PDE ($\mathcal{L}_{PDE}$), and Kullback–Leibler regularization ($\lambda_{KL} \text{KL}$). Automatic differentiation computes spatial and temporal derivatives for physics enforcement. At inference, Monte Carlo sampling from the posterior generates distributions over velocity fields and CFR, yielding probabilistic estimates with model-predicted confidence intervals.

To derive the law of contrast transport, we introduce another variable $A(s, t)$ to represent the cross-sectional area of the LAD. We use simplified notation in the following text for readability. By conserving the mass of blood and contrast fluid in an infinitesimal section of LAD, we arrive at the equations:

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial s}(uA) = 0, \quad \frac{\partial}{\partial t}(IA) + \frac{\partial}{\partial s}(uIA) = \frac{\partial}{\partial s}\left(AD \frac{\partial I}{\partial s}\right). \quad (1)$$

Combining the two equations leads to a simplified statement of contrast agent conservation:

$$\frac{\partial I}{\partial t} + u \frac{\partial I}{\partial s} = \frac{1}{A} \frac{\partial(AD)}{\partial s} \frac{\partial I}{\partial s} + D \frac{\partial^2 I}{\partial s^2}. \quad (2)$$

The effective Péclet number for contrast agent transported within the LAD is high: in a typical fit of our model to experimental data, we find $\text{Pe} \equiv uL/D \sim 4000$, based on a typical vessel length $L \sim 0.1$ m, dispersion coefficient $D \sim 10^{-6}$ m²/s and blood flow velocity $u \sim 0.04$ m/s. This means that the advective terms on the left-hand side of the equation dominate over diffusion, which is consistent with imaging data showing a uniform bolus of contrast agent advancing through the vessel. The only place where diffusion is strong enough to balance advection is in short regions of length $\sim d/\text{Pe}^{1/2}$ around the leading and trailing fronts of the bolus, where d is the vessel diameter. Because

gradients in I are stronger than gradients of A or D , within these short regions we may at leading order neglect the first term on the right-hand side, whereas away from the fronts both diffusion terms are negligible. Hence we arrive at a uniformly valid approximation for the one-dimensional advection–diffusion of contrast agent through the vessel:

$$\frac{\partial I}{\partial t} + u \frac{\partial I}{\partial s} - D \frac{\partial^2 I}{\partial s^2} = 0. \quad (3)$$

Here, the diffusion term provides *viscosity regularization* (Evans, 2010) of sharp fronts of transported contrast agent.

Model training is formulated as a multi-objective optimization problem combining three terms:

$$\mathcal{L} = \mathcal{L}_{Data} + \mathcal{L}_{PDE} + \lambda_{KL} \text{KL}(q_\phi(z) \| p(z)), \quad (4)$$

where $\mathcal{L}_{Data}$ enforces agreement with observed angiographic intensities, $\mathcal{L}_{PDE}$ enforces physical consistency, and the final term regularizes the uncertainty representation. This regularization is a Kullback–Leibler regularization term which prevents overfitting while allowing uncertainty to increase adaptively in regions of poor data quality.

At inference time, uncertainty is propagated by sampling z from the learned posterior ($N = 100$ samples), producing a distribution of predicted velocity and dispersion fields (Gal and Ghahramani, 2016;

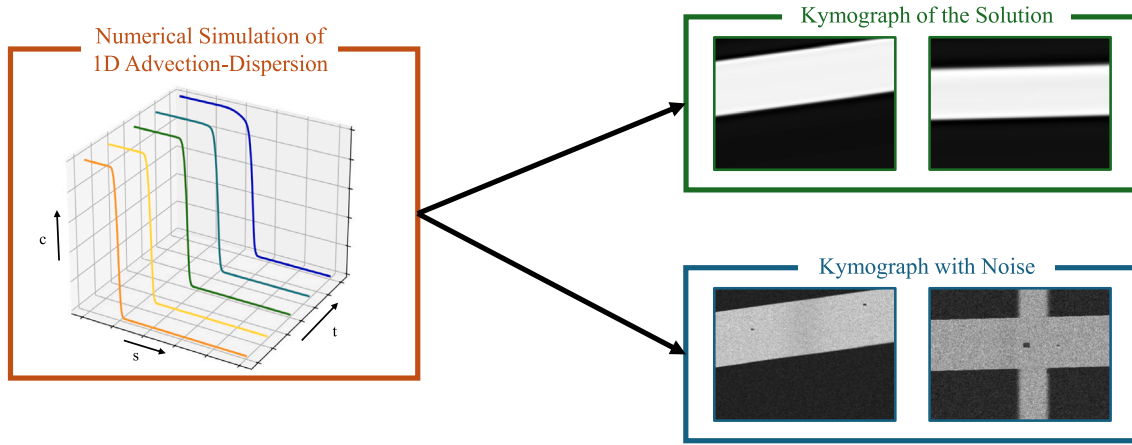


Fig. 4. Numerical simulation enables systematic evaluation under known ground truth. The one-dimensional advection–dispersion equation is solved numerically with randomly generated velocity profiles $u(s)$ and dispersion coefficients $D(s, t)$ to produce noise-free synthetic kymographs. Realistic angiographic noise and artifacts are then applied, including Poisson shot noise, spatial blur, low-frequency intensity variations, vertical banding, localized intensity dropout, and temporal pixelation. This corruption pipeline generates 1000 synthetic cases spanning mild to severe degradation, enabling quantitative assessment of model accuracy, uncertainty calibration, and robustness to image quality deterioration under controlled conditions where true flow parameters are known.

Kingma and Welling, 2014). CFR is computed for each sample as the ratio of spatially averaged velocity during hyperemia to that at rest,

$$\text{CFR} = \frac{\bar{u}_{\text{hyper}}}{\bar{u}_{\text{rest}}} = \frac{\frac{1}{L} \int_0^L u_{\text{hyper}}(s; \mathbf{z}) ds}{\frac{1}{L} \int_0^L u_{\text{rest}}(s; \mathbf{z}) ds}, \quad (5)$$

where L denotes the length of the vessel. Final CFR estimates are reported as the mean and 95% credible intervals of the resulting distribution, providing both a point estimate and an explicit measure of confidence.

2.2. Network architecture

The proposed framework consists of three physics-informed subnetworks that model contrast intensity, axial velocity, and effective dispersion, respectively. Each subnetwork is implemented as a fully connected multilayer perceptron with hyperbolic tangent activations. The intensity network takes spatial position, time, and a shared latent variable as input, whereas the velocity network depends on spatial position and the latent variable, and the dispersion network depends on spatial position, time, and the latent variable. Compact architectures were sufficient for stable convergence, with the intensity network comprising three hidden layers of 64 units, and the velocity and dispersion networks comprising two hidden layers of 32 units each.

To enforce physical plausibility, velocity and dispersion outputs are explicitly bounded through sigmoid-based parameterizations. Velocity predictions are constrained to physiologically realistic ranges after normalization, while dispersion coefficients are restricted to a predefined interval consistent with prior estimates of contrast transport in the coronary arteries. These bounds are applied identically to both the rest state and the hyperemic state.

Rest and hyperemic conditions are modeled using a dual-branch architecture in which each physiological state is represented by an independent set of subnetworks for intensity, velocity, and dispersion, while sharing a common latent variable $\mathbf{z}$ (Fig. 3). This shared latent representation is inferred via variational inference and captures global, unobserved factors such as image quality, contrast injection variability, and acquisition noise. By coupling the two physiological states through the same latent variable, the uncertainty is coherently propagated between rest and hyperemia, ensuring that the quality of degraded data in one condition induces correlated uncertainty in the other and producing coupled rest–hyperemia uncertainty estimates for CFR.

The latent variable is modeled using a Gaussian variational posterior with learned mean and variance, regularized to a specified prior via

a Kullback–Leibler divergence penalty. Sampling from this posterior during training enables uncertainty-aware estimation of flow-related quantities without requiring ground-truth velocity supervision. All subnetworks are trained jointly by minimizing a composite loss comprising physics-based residuals of the advection–diffusion equation, data fidelity terms on observed angiographic intensities, and the variational regularization term.

Rest and hyperemic angiograms were acquired in matched right anterior oblique–cranial (RAO–CRA) projections, and point correspondence between frames within a single video was established using point-tracking algorithms (LocoTrack; Cho et al. (2024)), allowing consistent spatial parameterization across physiological states. Within each video, intra-acquisition motion is compensated along the LAD centerline by LocoTrack-based point tracking; between the rest and hyperemic acquisitions, frames are matched by RAO–CRA projection but not by cardiac phase, so residual phase mismatch between the two states is absorbed into the modeled measurement/identifiability uncertainty captured by the shared latent $\mathbf{z}$ (see Uncertainty Quantification, below).

2.3. Training protocol

The framework is optimized *independently for every patient (or every synthetic case)*, with no parameter sharing between cases. This per-patient training design eliminates concerns about generalization between heterogeneous populations and about data leakage between training and test sets, but means that each new case requires its own short optimization on a single GPU.

For each case, all three subnetworks (intensity I , velocity u , dispersion D) and the variational posterior parameters ($\mu_z, \log \sigma_z^2$) are jointly optimized with the Adam optimizer at an initial learning rate of 10^{-3} , with a cosine-annealing learning-rate schedule over the full training horizon. Each iteration draws $N_f = 2000$ collocation points per state (sampled uniformly on the (s, t) rectangle) for the PDE residual, $N_d = 2000$ data points per state (sampled uniformly from the kymograph grid) for the data-fidelity term, and one reparameterized latent sample $\mathbf{z} = \mu_z + \sigma_z \odot \epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, for the KL term.

Intensity inputs are normalized per kymograph to the unit interval prior to training. Spatial and temporal coordinates are normalized by the arc-length of the imaged LAD segment and by the recording duration respectively, separately for the rest and hyperemic states; the bounds on u and D in the sigmoid-parameterized outputs are rescaled consistently. The PDE residual is evaluated by automatic differentiation

on the dimensionless coordinates and does not require explicit Dirichlet or Neumann boundary terms; the inlet and outlet behavior is implicit in the data-fidelity loss.

For the model-selection grid (Tables 6–22), each combination of prior standard deviation $\sigma_p \in \{0.5, 1, 2, 5, 10, 15\}$ and KL weight $\lambda_{KL} \in \{10^{-4}, 10^{-3}, 10^{-2}\}$ was trained under the same protocol; the prior mean was held fixed at $\mu_p = 45$ for all configurations. We emphasize that the per-case uncertainty is carried by the *posterior* standard deviation of the latent rather than by the prior mean; μ_p and σ_p are model-selection hyperparameters, and the latent itself remains an abstract per-case variable whose posterior spread reflects measurement/identifiability uncertainty (image quality, motion, incomplete filling). The configuration selected for the clinical evaluation reported in Section 3.3 used $\sigma_p = 2$, $\lambda_{KL} = 10^{-3}$, latent dimension $Z_{\text{dim}} = 1$, and a training horizon chosen to converge the composite loss under the cosine schedule (see Tables 6–22 of the Supplement for the full grid). We additionally *propose* a composite selection criterion balancing accuracy, coverage, sharpness, and threshold-based decision performance, stated explicitly in Section 3 but not applied to the results reported here.

To characterize sensitivity to initialization and to the stopping point, we repeated the per-patient inference on ten matched synthetic cases across five random seeds at the operating configuration. The inferred CFR was stable across seeds, with a median coefficient of variation of 0.8% (mean 1.1%, maximum 3.8%). Varying the training horizon over 7000, 10,000, and 13,000 epochs changed the inferred CFR by a median of 0.6% (7000 vs. 10,000 epochs) and 0.1% (13,000 vs. 10,000 epochs), indicating that the inference has effectively converged by the chosen horizon.

2.4. Uncertainty quantification

During inference, $N = 100$ Monte Carlo samples of $\mathbf{z}$ were drawn to approximate the posterior over flow fields. For each sample i , $\text{CFR}_i = \bar{u}_{\text{hyper}}^{(i)} / \bar{u}_{\text{rest}}^{(i)}$ was computed, and the mean and percentile-based 95% credible intervals were reported.

2.5. Statistical analysis

The correlation between predicted and invasively measured CFR was assessed using Spearman rank correlation. Agreement between methods was evaluated using Bland–Altman analysis (Bland and Altman, 1986) and the concordance correlation coefficient (CCC). Calibration of predicted uncertainty intervals was assessed by computing the fraction of ground-truth (synthetic) or invasive (clinical) CFR values contained within the model's 95% credible intervals. For synthetic datasets where true values were known, the uncertainty–error correlation was calculated to quantify whether higher predicted uncertainty corresponded to greater absolute error.

2.6. Synthetic data generation

To enable controlled evaluation of model accuracy, robustness, and uncertainty calibration under known ground truth, we generated 1000 synthetic coronary angiogram sequences using a physics-based forward model of contrast transport. The advection–diffusion equation was numerically solved using a method-of-lines framework with high-resolution finite-volume advection and finite-difference diffusion, integrated in time with an explicit Runge–Kutta ODE solver (Fig. 4). This synthetic framework allowed systematic variation of flow dynamics and image quality while retaining direct access to the true underlying parameters.

The propagation of contrast along the vessel centerline was modeled as a one-dimensional transport process governed by advection and dispersion, reflecting contrast motion with blood flow and its spreading due to velocity heterogeneity and mixing. Specifically, contrast intensity $I(s, t)$ was generated by solving Eq. (3). Physiologically plausible

velocity profiles were randomly generated for each sequence. Mean resting velocities were sampled between 10 and 20 cm/s for typical cases, while 25% of the sequences were designated as atypical stress-test cases with elevated resting velocities of 20–40 cm/s. Spatial velocity variations were introduced using smooth, localized modulations along the vessel to mimic focal acceleration or deceleration. Hyperemic flow was simulated by scaling resting velocities using a randomly sampled CFR, drawn from 1.5 to 4.0 for typical cases and 1.0 to 2.0 for low-CFR cases. The sampled CFR range spans the invasive CFR distribution of the 20-patient clinical cohort, and the two distributions are statistically indistinguishable by a two-sample Kolmogorov–Smirnov test ($D = 0.19$, $p = 0.45$; Fig. 5).

The contrast injection was modeled as a finite-duration inlet pulse with a randomized onset time and duration, while the length of the vessel (6–10 cm) and the total simulation time (3–6 s) were independently sampled for rest and hyperemia. The simulated durations (3–6 s) match the temporal window typical of clinical acquisitions. The dispersion coefficient was sampled over an order-of-magnitude physiological range to reflect inter-patient variability. All simulated intensity fields were normalized to the unit interval.

To emulate real-world angiographic acquisition, synthetic kymographs were subsequently corrupted with realistic imaging artifacts, including shot noise, spatial blur, low-frequency intensity inhomogeneity, banding artifacts, localized contrast dropout, and temporal down-sampling consistent with clinical frame rates. Overall image degradation was modulated by a scalar noise factor spanning mild to severe conditions.

In this base (*matched-model*) dataset the forward model shares the same 1D advection–diffusion form, Eq. (3), that is enforced as a soft constraint during inference, and hyperemic sequences are constructed by scaling the resting velocity field by a sampled CFR. To probe sensitivity to this construction, we additionally generated two *model-mismatch* datasets whose data-generating physics is deliberately richer than the inference PDE. In the first, contrast transport is solved in the full mass-conservation form along a vessel with physiologically tapering cross-sectional area $A(s)$ (a 20%–40% area reduction over a 6–10 cm proximal–mid LAD segment),

$$\frac{\partial I}{\partial t} + u(s) \frac{\partial I}{\partial s} = D \frac{\partial^2 I}{\partial s^2} + \frac{D}{A} \frac{dA}{ds} \frac{\partial I}{\partial s}, \quad u(s) = \frac{u_{\text{in}} A_{\text{in}}}{A(s)}, \quad (6)$$

where the velocity profile $u(s)$ follows from mass conservation and the additional geometric term $(D/A)(dA/ds) \partial_s I$ encodes the area-driven transport that a constant-area model omits. PUNCH inference is applied without modification, enforcing only the reduced constant-area residual of Eq. (3); the tapering term and the area-driven velocity gradient are therefore unknown to the inference model, creating a genuine geometric mismatch between the data-generating and the inference physics. Ground-truth CFR is read out post hoc as the ratio of spatial-mean velocities rather than imposed on the inference. The resulting matched-versus-mismatch comparison is reported in Section 3 (Table 1). We additionally generated a *tapering + side-branch* set that augments the tapered geometry with 2–4 side-branch outflow sinks at locations s_i along the segment, each draining a fraction α_i of the arriving flow so that $u(s)A(s) = u_{\text{in}}A_{\text{in}} \prod_{j: s_j \leq s} (1 - \alpha_j)$; the resulting downstream velocity reductions and contrast loss are likewise absent from the constant-area inference model. This set therefore combines two simultaneous mismatch sources (tapering geometry and branch sinks).

2.7. Clinical data acquisition

All angiographic and invasive measurements were acquired as part of routine clinical care and subsequently used for retrospective analysis from the University of Michigan Health System. Cases included those in whom CFR testing was done for routine clinical indications. Standard coronary angiography was performed in the right anterior oblique–cranial projection at the clinically common frame rate of 10

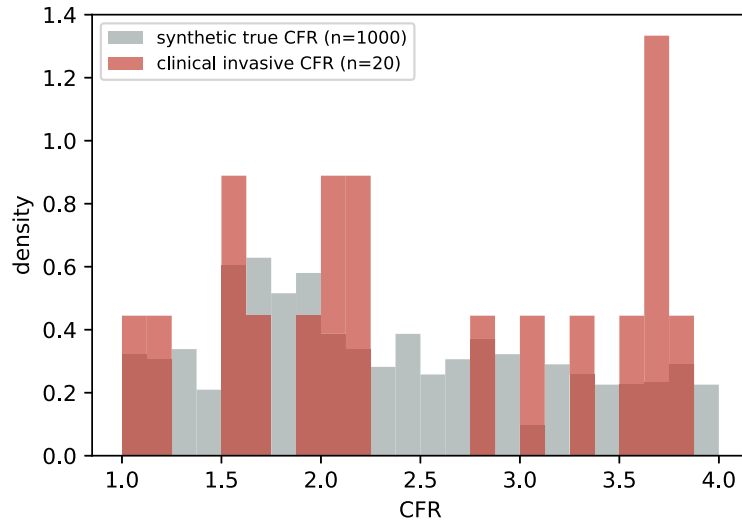


Fig. 5. Distribution of sampled synthetic CFR ($n = 1000$) and invasive clinical CFR ($n = 20$). The synthetic sampling range covers the clinical range, and the two distributions are statistically indistinguishable (two-sample Kolmogorov–Smirnov $D = 0.19$, $p = 0.45$).

or 15 frames per second. We used rest and hyperemic angiographic acquisitions from matched projections to minimize foreshortening and motion-related artifacts and to ensure consistent spatial correspondence between physiological states. Invasive CFR was measured by bolus thermodilution following the established protocol (Barbato et al., 2004; Fearon et al., 2003). Briefly, three injections of 3 mL room-temperature saline were delivered into the proximal LAD at rest, and the mean resting transit time Tm_{rest} was computed as the average of the three transit-time curves recorded by the distal thermistor of a pressure/temperature wire. Pharmacological hyperemia was then induced by intravenous adenosine at $140 \mu\text{g kg}^{-1} \text{min}^{-1}$; three additional saline boluses were delivered during the hyperemic plateau, yielding Tm_{hyp} . Invasive CFR was then computed as $CFR_{inv} = Tm_{rest}/Tm_{hyp}$, since mean transit time is inversely proportional to flow. Bolus thermodilution has been reported to yield systematically higher CFR values than continuous thermodilution (Gallinoro et al., 2023; Jansen et al., 2023; Fawaz et al., 2024); we report this as context for interpreting agreement with PUNCH in Section 3.3 rather than as an explanation for any specific bias.

2.8. Image processing and intensity extraction

The angiogram of each patient was processed in a standardized workflow:

- 1. Automated vessel segmentation.** LAD centerlines were extracted using a modified SAM2-UNet (Xiong et al., 2026) architecture with dynamic snake convolution layers (Qi et al., 2023) which was optimized using binary LAD segmentation masks. We manually verified the centerline tracking of the LAD for each patient.
- 2. Point tracking.** Point correspondence was established between frames within a single video, allowing consistent measurements of image intensity using the LocoTrack algorithm (Cho et al., 2024).
- 3. Contrast isolation.** Multi-scale Hessian filtering was applied to enhance contrast intensity and suppress background structures.
- 4. Spatial calibration.** DICOM metadata were used to convert pixel coordinates to millimeters.
- 5. Intensity field construction.** Filtered frames were stacked over time to form a kymograph of $I(s, t)$ (Fig. 2), where s denotes the arc-length along the centerline of the vessel.

Collocation points for evaluating physics residuals were uniformly sampled along the vessel and across time frames within regions of reliable tracking. The supervised data points corresponded to all spatiotemporal locations with valid Hessian-filtered intensity measurements.

For the clinical cases the kymograph uses all frames of the acquired sequence, with no manual selection of a start frame, end frame, or sub-window, so the temporal window is set by the acquisition rather than by the operator. Because CFR is determined by the contrast-front slope rather than by the number of baseline or washout frames, reasonable variation of the start frame, end frame, or sequence length that still brackets contrast passage leaves the inferred CFR essentially unchanged.

3. Results

3.1. Clinical interpretation of uncertainty

In the proposed framework, each CFR estimate is accompanied by a model-predicted credible interval that reflects data quality rather than population variability. Narrow credible intervals indicate that the observed angiographic contrast dynamics are internally consistent and strongly constrain the inferred flow parameters. Conversely, wide intervals signal ambiguity arising from factors such as motion artifacts, incomplete contrast filling, or low signal-to-noise ratio. Clinically, this distinction is useful for triage: a low CFR estimate with a narrow model-derived uncertainty interval may warrant further investigation for coronary microvascular dysfunction, whereas high-uncertainty estimates suggest that selecting alternative angiogram projections, repeat imaging, or adjunctive physiological testing may be warranted. By explicitly communicating uncertainty, the framework enables risk-aware interpretation of angiography-derived physiology and aims to reduce the likelihood of false certainty from degraded data.

3.2. Validation on synthetic kymographs

Point-estimate accuracy was optimal in an intermediate regime, with MAE values as low as 0.035 (Table 7) and RMSE below 0.10 (Table 8) for moderate prior variance and KL regularization. In this regime, Deming regression analysis (Linnet, 1993) demonstrated near-unity slopes (0.96–0.98; Table 10) and small intercepts (Table 11), indicating minimal proportional bias and good agreement with ground truth throughout the dynamic range. The mean bias remained small and consistently negative across all configurations (Table 12), reflecting a mild conservative tendency that increased with stronger regularization.

Beyond aggregate error metrics, we evaluated the diagnostic performance of both posterior means and credible interval bounds at the standard CFR classification threshold of 2.5, consistent with the clinical CMD analysis in Section 3.3. The sensitivity and specificity of the posterior mean exceeded 0.95 in most configurations (Tables 13 and 14), while the lower and upper credible bounds maintained high specificity (Tables 16 and 18) and sensitivity exceeding 0.90 for moderate regularization strengths (Tables 15 and 17). These results indicate that, on synthetic data, the posterior-mean and credible-bound estimates preserve clinically meaningful threshold-based discrimination at the CFR threshold; this concerns classification accuracy rather than interval calibration, which we examine below.

We examined the effect of latent dimensionality in the variational posterior for representative configurations with prior standard deviation = 10 and prior standard deviation = 2 at $\lambda_{KL} = 10^{-3}$ (Tables 19 and 21). Increasing the latent dimension beyond one to two dimensions resulted in progressively worse performance across multiple metrics, including reduced coverage, increased error, and decreased sensitivity of the lower credible bound. These degradations occurred without compensatory gains in uncertainty sharpness, indicating overparameterization of the latent space. In contrast, low-dimensional latent representations ($Z_{dim} = 1-2$) achieved the best balance between accuracy, calibration, and interval sharpness.

We additionally assessed robustness to reduced temporal sampling by varying the effective frame rate of the angiographic sequences for the same representative configurations (Tables 20 and 22). For the operating configuration selected for the clinical evaluation ($\sigma_p = 2$, $\lambda_{KL} = 10^{-3}$, Table 22), decreasing the effective frame rate from 15 to 7.5 frames per second produced non-negligible degradation in point-estimate accuracy: MAE rose from 0.034 to 0.086 and RMSE from 0.068 to 0.158 (an approximately 2.5-fold increase), and the sensitivity of the lower credible bound decreased from 0.95 to 0.86, while 95% coverage remained essentially unchanged and below nominal (0.672 to 0.674). Uncertainty intervals widened with reduced temporal resolution (sharpness increasing roughly in proportion to the error), indicating that the model expresses greater uncertainty as the information content of the input decreases rather than producing overconfident estimates. Clinical acquisitions at 10–15 fps fall within the tested range, but the present results should be read as a preliminary characterization at the operating point rather than as evidence of frame-rate invariance.

Collectively, these results demonstrate that the proposed framework achieves accurate CFR estimation under controlled synthetic conditions, while explicitly exposing the trade-offs between accuracy and uncertainty across all evaluated hyperparameters. We note explicitly that, on this matched-model synthetic set, the selected operating configuration achieves coverage below the nominal 95% level (Table 6), and no configuration in the hyperparameter grid reaches nominal coverage; we therefore characterize the raw posterior intervals as *under-dispersed*, and discuss the recalibration step required to correct this in the Limitations. A reliability diagram across nominal credible levels (Fig. 6) confirms this under-dispersion over the full range of nominal coverage, with an expected calibration error (ECE) of 0.17; the post-hoc split-conformal recalibration described in the Limitations restores nominal 95% coverage.

The model's uncertainty increases predictably with image degradation, which makes it useful for ranking estimates by reliability and for quality control; given the under-coverage noted above, we do not present it as a calibrated clinical interval (see Limitations). The operating point used for the clinical evaluation ($\sigma_p = 2$, $\lambda_{KL} = 10^{-3}$, latent dimension $Z_{dim} = 1$) was selected from the grid in Tables 6–22 on the basis of point-estimate accuracy (MAE) together with the raw coverage and sharpness reported there. To weigh accuracy against calibration and decision performance rather than accuracy alone, we additionally propose a composite model-selection criterion for future use, combining (i) mean absolute error, (ii) coverage at the 95% level, (iii) mean 95% interval width as a sharpness penalty, and (iv) threshold-based

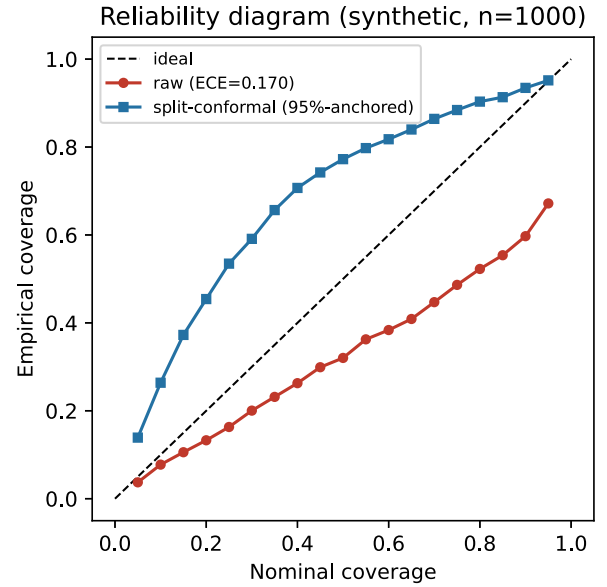


Fig. 6. Reliability diagram on the matched synthetic set ($n = 1000$, operating configuration $\sigma_p = 2$, $\lambda_{KL} = 10^{-3}$). Empirical coverage falls below the nominal line at every level (raw expected calibration error ECE = 0.17), indicating systematic under-dispersion of the raw posterior intervals. The split-conformal scaling described in the Limitations is anchored to restore nominal 95% coverage.

classification accuracy at the clinical cut-off CFR < 2.5. We state this criterion explicitly so that the operating-point trade-off is transparent; it is proposed as a model-selection criterion and is not applied to the results reported here.

To address the concern that the synthetic evaluation may be optimistic because the data-generating physics shares the advection–diffusion form enforced at inference (the “inverse-crime” risk), we additionally evaluated PUNCH on two model-mismatch synthetic sets whose forward models are structurally richer than the inference PDE: a *tapering* set (full mass-conservation transport with spatially varying cross-sectional area $A(s)$, Eq. (6)), and a *tapering + side-branch* set that additionally drains a fraction of the arriving flow into 2–4 side-branch outflow sinks along the segment (Methods, Synthetic data generation). PUNCH was applied without modification at the identical operating configuration used for the clinical evaluation ($\sigma_p = 2$, $\lambda_{KL} = 10^{-3}$, $Z_{dim} = 1$, 15 fps). The two mismatch sets are generated independently (each with its own random draws of velocities, CFR, geometry, and acquisition parameters), and each is reported on a common 256 cases so the two are directly comparable; a small number of non-converged optimizations were excluded. Table 1 reports the result. Under tapering the Deming slope is essentially unchanged ($0.977 \rightarrow 0.968$) and the correlation remains strong ($r = 0.980$), with mean absolute error rising from 0.034 to 0.088 CFR units; the tapering + side-branch set gives slope 0.978, $r = 0.996$, and MAE 0.054. Because the two mismatch sets are independent random draws rather than the same cases with branches added, the absolute error differs between them (the side-branch set happens to contain easier draws), so these numbers should not be read as “more mismatch yields lower error” for a fixed case; the controlled and consistent finding is that under *both* mismatches the Deming slope and correlation are essentially preserved (slope ≈ 0.97 – 0.98 , $r \geq 0.98$) while interval coverage falls relative to the matched model (matched 0.672; tapering 0.535; tapering + side-branch 0.574). The dominant effect of model mismatch is thus on uncertainty coverage, reinforcing rather than contradicting our finding that the raw posterior intervals

Table 1

Model-mismatch validation: CFR recovery under (a) vessel tapering and (b) tapering plus side-branch outflow. PUNCH is applied at the identical clinical operating configuration to the matched-model set (constant-area, the same advection-diffusion PDE enforced at inference), a tapering set (full mass-conservation, $A(s)$, Eq. (6)), and a tapering + side-branch set (2–4 outflow sinks; Methods). Each mismatch set is subsampled to 256 usable cases; the matched set is reported in full. Bias denotes mean (predicted – true) CFR; lower MAE/RMSE and a Deming slope closer to 1 indicate better recovery.

Condition	N	MAE	RMSE	Deming slope	Pearson r	95% coverage
Matched (constant-area)	1000	0.034	0.068	0.977	0.997	0.672
Tapering ($A(s)$)	256	0.088	0.184	0.968	0.980	0.535
Tapering + side-branch	256	0.054	0.092	0.978	0.996	0.574

are under-dispersed, while point-estimate and ranking accuracy remain robust; the agreement reported below is therefore not an artifact of the matched-model assumption.

To attribute PUNCH's performance to its individual components, we ran an ablation on the first 256 synthetic cases at the clinical operating configuration, removing one component at a time (Table 2). Removing the PDE constraint (data-fit only) changes MAE from 0.032 to 1.380 and the Deming slope to 0.049 ($r = 0.135$); because the velocity and dispersion fields enter the objective *only* through the physics residual, this constraint is essential — without it those fields receive no training signal and the inferred CFR is uninformative. A deterministic variant (no variational latent, no KL term) gives slightly *better* point accuracy (MAE 0.021 vs. 0.032) but collapses the predictive interval (coverage 0.000): the variational formulation therefore trades a small amount of point accuracy for usable, error-ranking uncertainty rather than improving the point estimate itself. Replacing the shared rest/hyperemia latent with independent latents leaves point accuracy essentially unchanged (Deming slope 0.982); its slightly higher coverage (0.746) is obtained with wider, less sharp intervals (mean width 0.095 vs. 0.070) rather than better calibration, while the shared latent provides a single, more parsimonious source of coupled rest/hyperemia uncertainty. Finally, an MC-dropout uncertainty model in place of the variational posterior (dropout rate 0.1, 100 Monte Carlo samples, identical architecture and physics loss) yields markedly smaller CFR uncertainty (mean 95% interval width 0.001 vs. 0.070 for the variational posterior; coverage 0.004). This degeneracy is structural rather than a tuning artifact: per-point dropout noise averages out over the spatial domain, whereas the shared latent perturbs the velocity field coherently, so the variational posterior yields more faithful CFR uncertainty for this spatially-aggregated quantity. We did not pursue a deep-ensemble or full Bayesian-PINN posterior, as both multiply the per-patient inference cost by the ensemble/chain size and are incompatible with the few-minute operating point.

3.3. Clinical validation

The clinical evaluation here is a feasibility-scale, single-center study (20 patients, LAD assessment in matched right anterior oblique–cranial projections); the requirements for broader validation are addressed in the Limitations.

Clinical validation was performed in a cohort of 20 patients who underwent a standard invasive CFR assessment using the bolus thermodilution technique. For each patient, three repeated measurements were acquired at rest and during pharmacologically induced hyperemia using intravenous adenosine at $140 \mu\text{g kg}^{-1} \text{min}^{-1}$. Invasive CFR was calculated as the ratio of mean hyperemic to mean resting flow values (Barbato et al., 2004; Fearon et al., 2003). The CFR estimates derived from PUNCH were obtained independently for each case and summarized by their posterior mean and standard deviation.

To quantify experimental uncertainty in the invasive CFR measurements, a nonparametric bootstrap approach was employed. Given

Table 2

Ablation and uncertainty-method comparison on the first 256 synthetic cases at the clinical operating configuration, against matched-model ground truth. Each row removes or replaces one component: –PDE (data-fit only), deterministic (no variational latent/KL), separate rest/hyperemia latents (no shared latent), and MC-dropout uncertainty (in place of the variational posterior). Lower MAE/RMSE and a Deming slope and coverage closer to the full model indicate a smaller contribution from that component; markedly different values indicate an essential one. The full-model coverage here (0.669) is computed on the 248-case common subset shared by all variants; it corresponds to the same operating point as the 0.672 reported on the model-mismatch set (Table 1) and the 0.671 in the hyperparameter grid (Table 6), the small differences reflecting the different case subsets.

Variant	MAE	RMSE	Deming slope	Pearson r	95% cov.
Full PUNCH	0.032	0.058	0.980	0.998	0.669
– PDE constraint	1.380	1.600	0.049	0.135	0.012
Deterministic (– variational)	0.021	0.039	0.995	0.999	0.000
Separate latents	0.034	0.056	0.982	0.998	0.746
MC-dropout UQ	0.081	0.133	0.941	0.992	0.004

three repeated measurements at rest and three during hyperemia, all possible replacement combinations were enumerated for each condition (i.e., $3^3 = 27$ bootstrap samples per state). For each bootstrap realization, CFR was computed as the ratio of the mean hyperemic to the mean resting values. This resulted in $27 \times 27 = 729$ bootstrap CFR estimates per patient, from which the invasive CFR mean and the standard deviation of the sample were computed. This approach captures the uncertainty arising from finite repeated measurements, while preserving the dependence structure inherent to the CFR calculation.

The PUNCH estimates demonstrated a strong association with invasive measurements, with a Spearman rank correlation of $\rho = 0.89$ ($p < 10^{-6}$). Ordinary least squares (OLS) regression yielded an R^2 of 0.79, indicating that a substantial proportion of the variance in invasive CFR was explained by the PUNCH predictions (Fig. 7b).

To assess agreement while accounting for measurement error in both modalities, Deming regression was performed. This analysis produced a slope of 0.83 and an intercept of 0.11 (PUNCH = $0.83 \times$ Invasive + 0.11), consistent with near-unity proportionality and minimal systematic deviation across the observed CFR range. For comparison, the OLS regression resulted in a slope of 0.75 and an intercept of 0.30 ($p < 10^{-6}$). The mean difference between PUNCH and invasive CFR was -0.31 , while the mean relative difference was -11.4% .

The agreement was further evaluated using the Bland–Altman analysis (Fig. 7c). The mean difference between PUNCH and invasive CFR was -0.31 , with a standard deviation of 0.48. The corresponding 95% limits of agreement ranged from -1.25 to 0.63 , and all 20 cases fell within these limits. Relative difference analysis revealed a mean bias of -11.4% (SD $\pm 20.4\%$), with 10 of 20 cases (50%) within $\pm 20\%$ agreement (Fig. 7c). Collectively, these findings indicate a modest negative bias, reflecting a tendency toward lower CFR estimates relative to invasive measurements. It should be noted that bolus thermodilution has been reported to yield systematically higher CFR values than continuous thermodilution (Gallinoro et al., 2023; Jansen et al., 2023; Fawaz et al., 2024). This methodological tendency toward higher CFR may contribute, in part, to the modest negative bias observed when comparing PUNCH estimates to CFR values measured using bolus thermodilution.

In addition to point estimates, the PUNCH framework provides posterior uncertainty predicted by the model for each CFR estimate. In Fig. 7d, PUNCH posterior means and 95% credible intervals are shown alongside invasive CFR means and bootstrap-derived standard deviations. Because the invasive bolus-thermodilution reference is itself measured with substantial intra-patient variability (three repeated injections per state), we assess agreement against that variability: the PUNCH estimate falls within the invasive 95% test–retest interval in 15 of 20 patients, and the Bland–Altman limits of agreement ($\pm 38\%$

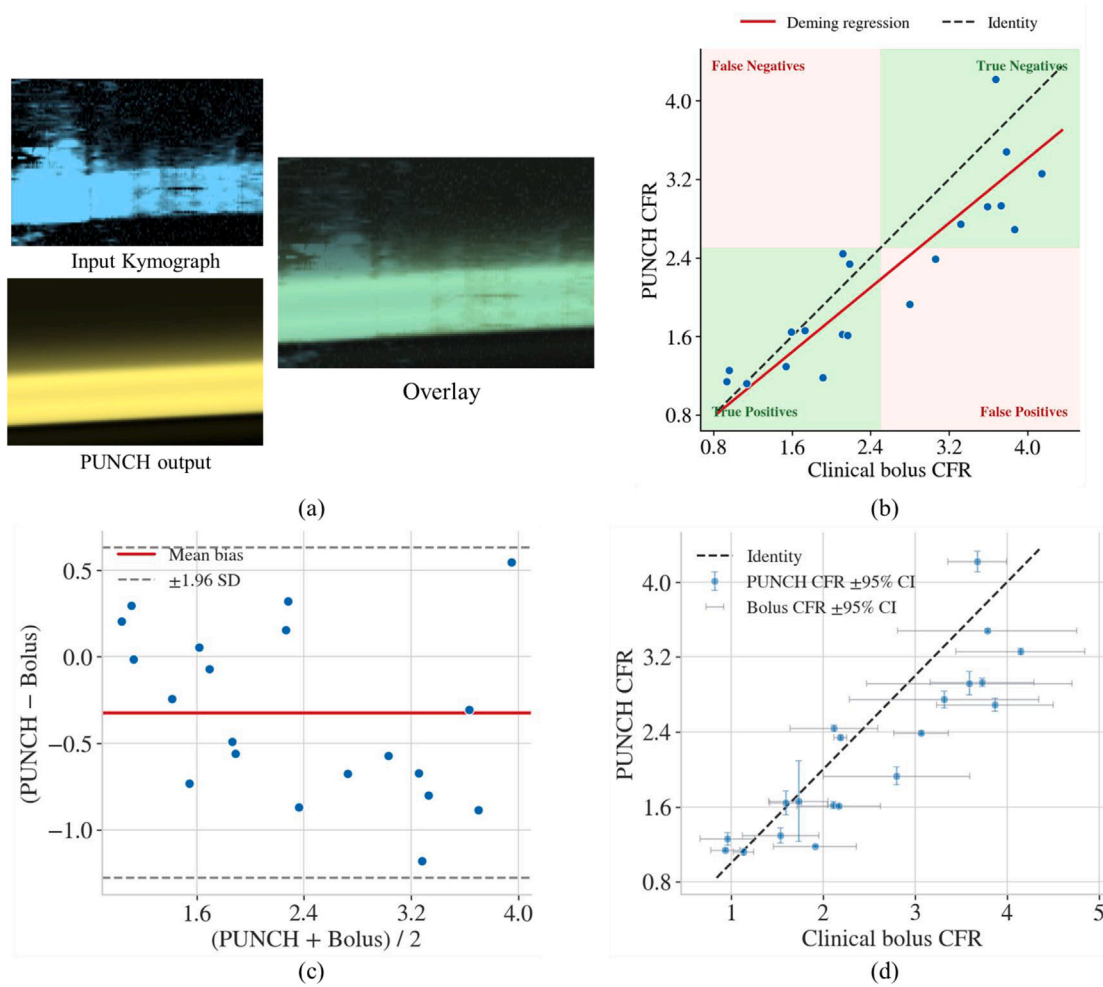


Fig. 7. PUNCH demonstrates strong agreement with invasive CFR assessment. (a) Input kymograph (blue) to PUNCH, output kymograph (yellow) and overlay of input and output kymograph. (b) Correlation plot comparing PUNCH-derived CFR posterior means against invasive bolus thermodilution measurements in 20 patients shows strong linear association (Spearman rank correlation $\rho = 0.89$, $p < 10^{-6}$; Deming regression slope = 0.83, intercept = 0.11). (c) Bland-Altman analysis reveals mean bias of -0.31 (SD ± 0.48) with 95% limits of agreement $[-1.25, 0.63]$, indicating modest systematic underestimation. All 20 cases fall within agreement bounds. (d) Individual patient predictions with 95% credible intervals (shaded regions) shown alongside invasive CFR means and bootstrap-derived standard deviations; the two interval families are of comparable magnitude (the formal calibration of the PUNCH intervals is examined in the Discussion). Error bars represent PUNCH posterior standard deviations and invasive measurement variability across three repeated acquisitions.

of mean CFR) are comparable to the documented reproducibility of bolus thermodilution itself ($\approx \pm 33\%$). The formal calibration of the PUNCH intervals, which are sharper than the reference, is examined in the Discussion.

To complement the correlation and agreement analysis with a threshold-based assessment, we evaluated PUNCH for the binary classification of coronary microvascular dysfunction at the standard clinical cut-off CFR < 2.5 (Table 3). Against invasive CFR, PUNCH achieved a sensitivity of 1.00, specificity 0.78, positive predictive value 0.85, negative predictive value 1.00, overall accuracy 0.90, and area under the ROC curve 0.97 ($n = 20$). Both misclassifications were false positives, arising where PUNCH's mild negative bias of -0.31 (noted above) pushed the estimate below the cut-off (invasive CFR 2.8 and 3.06, estimated as 1.93 and 2.39); this underscores that a recalibration step (Limitations) would be advisable before any threshold-based clinical use. We emphasize that these estimates derive from a single-center cohort of only 20 patients and are therefore statistically fragile (wide confidence intervals); they are reported as feasibility-scale evidence, not as established diagnostic performance.

Table 3

Diagnostic performance of PUNCH for CMD classification at the standard cut-off CFR < 2.5 on the 20-patient cohort, against invasive bolus-thermodilution CFR. 95% confidence intervals are from a 2000-sample case bootstrap; their width (e.g. specificity [0.50, 1.00]) reflects the small cohort, and the comparison additionally inherits the known variability of bolus thermodilution.

Metric	Value ($n = 20$)	95% CI
Confusion (TP, FP, FN, TN)	11, 2, 0, 7	–
Sensitivity	1.00	[1.00, 1.00]
Specificity	0.78	[0.50, 1.00]
Positive predictive value	0.85	[0.62, 1.00]
Negative predictive value	1.00	[1.00, 1.00]
Accuracy	0.90	[0.75, 1.00]
AUC	0.97	[0.88, 1.00]

To place these results against a simpler image-based comparator, we evaluated a naive front-tracking baseline on the same 20-patient cohort, in which CFR is estimated as the ratio of the hyperemic to the resting

Table 4

Image-based baseline comparison on the 20-patient clinical cohort. CFR from naive Hough-slope front tracking (ratio of hyperemic to resting front slope) versus PUNCH, both against invasive bolus-thermodilution CFR. The front tracker returned no usable CFR for 10 of 20 patients (infinite or non-physical slopes); its metrics are computed on the 10 patients with a finite, positive ratio, while PUNCH is reported on all 20.

Method	Usable N	Spearman ρ	Pearson r	MAE	Bias
Hough front-tracking	10/20	-0.03	-0.19	2.26	-0.16
PUNCH	20/20	0.89	0.89	0.47	-0.31

contrast-front slope obtained by a Hough transform of the kymograph ($\text{CFR}_{\text{Hough}} = \text{slope}_{\text{hyp}}/\text{slope}_{\text{rest}}$), without any physics constraint or uncertainty model. The front tracker returned a usable (finite, positive) CFR for only 10 of 20 patients; for the remaining 10 it failed outright, because the discretized Hough slope was infinite (a near-vertical, near-instantaneous front) or non-physical (negative). On the 10 patients where it was defined, the baseline showed no association with invasive CFR (Spearman $\rho = -0.03$, Pearson $r = -0.19$, MAE = 2.26), whereas PUNCH achieved $\rho = 0.89$, $r = 0.89$, MAE = 0.47 across all 20 (Table 4). This indicates that the agreement reported above is not attainable by direct slope tracking and is attributable to the physics-constrained inference rather than to the kymograph representation alone.

4. Discussion

We present a physics-informed, uncertainty-aware framework for wire-free estimation of coronary flow reserve (CFR) from paired resting and hyperemic coronary angiography, evaluated here as a proof-of-concept feasibility study for downstream coronary microvascular dysfunction (CMD) assessment. Our approach integrates (1) physics-based learning directly from unlabeled angiographic sequences with per-patient inference; (2) variational inference that produces uncertainty estimates that increase under image degradation and correlate with estimation error on synthetic data; and (3) a dual-branch architecture that jointly models resting and hyperemic states to produce coupled rest-hyperemia CFR uncertainty estimates. The ablation in Section 3 attributes these contributions cleanly: the physics constraint is what drives point-estimate accuracy (removing it is catastrophic), whereas the variational latent contributes the uncertainty estimates rather than the accuracy (a deterministic variant is marginally more accurate but provides no usable intervals). We therefore frame physics-constrained inference as the accuracy contribution and the variational latent as a separate, uncertainty contribution. The complete inference pipeline executes in approximately three minutes per patient on a single NVIDIA A40 GPU, with no requirement for prior population-level training, supporting per-patient analysis from images that are already being acquired. The contributions are methodological: per-patient, label-free, physics-constrained inference of CFR with propagated uncertainty, evaluated here as a proof-of-concept feasibility study.

Disease burden and clinical opportunity. Coronary microvascular dysfunction represents a major public health challenge affecting millions of patients worldwide. CMD is associated with a nearly 4-fold increase in all-cause mortality (OR 3.93, 95% CI 2.91–5.30) and a 5-fold increase in major adverse cardiac events (OR 5.16, 95% CI 2.81–9.47) compared to patients without microvascular disease (Gdowski et al., 2020). Patients with CMD experience substantially elevated risks of cardiovascular death, myocardial infarction, hospitalization for heart failure, and stroke (Kelshiker et al., 2022; Taqueti and Di Carli, 2018), with approximately 30%–40% of patients with non-obstructive coronary artery disease demonstrating evidence of CMD upon systematic evaluation (Kunadian et al., 2021). Among patients with angina and non-obstructive coronary arteries, the burden is particularly pronounced,

with estimates suggesting that 3–4 million patients in the United States alone have signs and symptoms of ischemia despite no evidence of obstructive atherosclerosis (Bailey Merz et al., 2017; Tjoe et al., 2021).

However, despite this substantial burden of the disease, CMD remains profoundly underdiagnosed. Approximately 4 million elective coronary angiograms are performed annually in Europe and the United States combined (Beştemir et al., 2023; Ford et al., 2018), representing one of the most common invasive cardiovascular procedures worldwide. By contrast, specialized invasive CMD assessments using bolus or continuous thermodilution are performed in only hundreds to a few thousand patients annually, concentrated in a small number of specialized centers with dedicated expertise (Barbato et al., 2004). This striking disparity — millions of angiograms versus thousands of functional microvascular assessments — leaves the vast majority of patients without a definitive functional diagnosis. Standard angiography, while providing excellent anatomical resolution of the epicardial vessels, does not provide direct information about microvascular function, perpetuating diagnostic uncertainty in a high-risk population.

potential clinical role, contingent on validation. If validated in larger, multi-center, multi-territory cohorts with formally evaluated diagnostic performance against clinically established CFR thresholds, a wire-free, angiography-derived CFR estimate could complement existing wire-based assessment, particularly in settings where specialized invasive microvascular testing is not routinely available. The framework still requires paired rest/hyperemic acquisitions with pharmacological vasodilation, so it does not displace the need for an angiographic procedure and a vasodilator protocol; what it changes is the requirement for additional intravascular instrumentation. CMD is disproportionately underdiagnosed in women, who more often present with angina without obstructive disease and less often undergo invasive physiological testing; if the validation steps described in the Limitations succeed, wire-free assessment may help reduce this gap, but the present study is not powered to demonstrate that effect. We retain “honest AI” as a design principle rather than a performance claim: the system reports an interval, but on the current synthetic and clinical evaluations the raw 95% intervals are under-dispersed, so the principle is partly aspirational pending the recalibration and external validation described below.

Comparison to existing paradigms. Wire-free, angiography-derived assessments of coronary physiology fall along several axes: the imaging input (angiography alone vs. CT/MR), whether labeled training data are required, the computational cost per patient, the physiological quantity recovered (epicardial pressure index, microvascular resistance index, or flow), and whether the method propagates uncertainty. Table 5 summarizes how PUNCH is positioned along these axes relative to representative pressure-derived (QFR/vFFR), microvascular-resistance (Angio-IMR), transit-based (TIMI/corrected TIMI frame count), classical indicator-dilution, CFD-based, and supervised data-driven approaches.

Computational fluid dynamics (CFD)-based methods require high-resolution CT imaging, three-dimensional vessel reconstruction, and extensive computation (often hours per patient), fundamentally limiting clinical scalability (Taylor et al., 2013; Morris et al., 2016). Supervised machine learning approaches require large labeled datasets that simply do not exist for challenging phenotypes such as CMD, where invasive ground truth measurements are available only in small registries (Itu et al., 2016; Li et al., 2022). In contrast, our framework learns directly from paired rest/hyperemic angiograms without labeled flow data, computes patient-specific solutions in minutes, and provides probabilistic estimates with quantified uncertainty; capabilities absent in prior approaches. The per-patient training paradigm eliminates concerns about generalization between heterogeneous populations or data leakage between training and test sets. The PUNCH 95% credible intervals were of comparable or somewhat smaller width than the intra-patient bootstrap variability of repeated invasive thermodilution

Table 5

Positioning of PUNCH among representative angiography-derived and related approaches to coronary functional assessment.

Method	Input	Labels	Runtime	Output	UQ
QFR/vFFR	Angio	No	Seconds	FFR (pressure)	No
Angio-IMR	Angio	No	Seconds	IMR (resistance)	No
TIMI/cTFC	Angio	No	Seconds	Transit/flow ratio	No
Indicator dilution (Zierler, 2000)	Angio	No	Seconds	Transit time	No
CFD-derived FFR	CT	No	Hours	FFR (pressure)	No
Supervised CFR (Itu et al., 2016)	CT/Angio	Yes	Seconds	FFR/flow	No
PUNCH (this work)	Paired Angio	No	~3 min	CFR (flow ratio)	Yes

measurements (median standard deviation approximately 12%) (Barbato et al., 2004); however, given that the raw posterior intervals are under-dispersed on synthetic data, this comparison should be interpreted as one of *magnitude* between two different uncertainty quantities rather than as evidence that PUNCH is more reproducible than the invasive reference.

Limitations. The current framework assumes quasi-one-dimensional flow along the vessel centerline; complex anatomies involving severe bifurcations, vessel overlap, or extreme tortuosity may violate this approximation and reduce inference fidelity. In the imaged proximal-mid LAD this assumption is most defensible, but side-branch outflow, vessel tapering, and branch-related contrast loss are not modeled in the inference PDE and may bias inferred velocity and CFR. We quantified both effects directly with model-mismatch synthetic generators: a tapering generator (full mass-conservation transport with spatially varying $A(s)$) and a tapering-plus-side-branch generator (2–4 outflow sinks), both evaluated against the matched-model set in Table 1. CFR recovery remained robust in point estimate and ranking under both, with the mismatch acting principally on interval coverage; the remaining unexercised mismatch source is a non-multiplicative (microvascular-resistance) construction of the hyperemic state, which we do not exercise here. The vessel centerline is extracted automatically from the SAM2-UNet segmentation, with no manual-editing stage. We manually checked that our method reliably locates the LAD centerline. However, the kymograph provides an intermediate visualization of contrast transport between vessel segmentation and velocity analysis, and would allow departure of the tracking points from the vessel lumen to be checked before CFR values are estimated. Systematic characterization of these failure modes, together with extension to multi-branch coronary geometries and explicit ECG-gated cardiac-phase alignment between the rest and hyperemic acquisitions, represents an important direction for future methodological development. We further note that the synthetic data used in this work were generated by the same family of 1D advection–diffusion equations imposed as a soft constraint at inference, and that hyperemic sequences were constructed by scaling the resting velocity field by a sampled CFR; the matched-model synthetic evaluation therefore tests identifiability and uncertainty behavior under ideal conditions and may overestimate real-world performance. To bound this optimism we performed model-mismatch synthetic evaluations in which the forward model is intentionally richer than the inference PDE: full mass-conservation transport with a tapering cross-sectional area $A(s)$ (Eq. (6)), and a further set that adds side-branch outflow sinks; the resulting changes in CFR recovery are reported in Table 1. The one mismatch source we do not exercise is a non-multiplicative construction of the hyperemic state from a microvascular-resistance model (rather than by multiplying the resting field by a sampled CFR). Beyond model generalization, larger multi-center studies will be required to establish clinically actionable diagnostic thresholds, evaluate robustness across diverse patient populations and anatomical variants, and

relate PUNCH-derived physiological estimates to downstream clinical decision-making and outcomes. The clinical evaluation here is a feasibility-scale, single-center cohort of 20 patients, restricted to the LAD in matched right anterior oblique–cranial projections; it is not sufficient to support claims about CMD diagnostic performance, generalization across vessel territories, or robustness to projection angle and image quality. The wide 95% Bland–Altman limits of agreement and the modest -0.31 (-11.4%) bias also matter for a threshold-sensitive metric like CFR: borderline cases near the standard 2.5 cut-off can shift classification, as we observe directly in the diagnostic analysis (Table 3), where both misclassifications were false positives driven by this negative bias. We report the confusion matrix, sensitivity, specificity, AUC, PPV, and NPV at the 2.5 threshold, but emphasize that with $n = 20$ these are feasibility-scale estimates and that a recalibration step would be advisable before any threshold-based clinical interpretation. Regarding uncertainty calibration, the synthetic grid search shows that the raw 95% posterior intervals are under-dispersed, with coverage substantially below nominal at the operating configuration selected for the clinical evaluation. We assessed a post-hoc *split-conformal* recalibration, fitting a single multiplicative factor on a held-out synthetic calibration split and validating on a disjoint synthetic test split: this raised synthetic test coverage from 0.68 to 0.94 (the calibrated factor is $\approx 4\times$ the nominal 1.96, quantifying the degree of under-dispersion), confirming that the uncertainty is calibratable in-domain. For the clinical cohort we characterize agreement directly. The raw PUNCH intervals are sharp (median 95% half-width ≈ 0.04 CFR units), narrower than the bolus-thermodilution reference’s own measurement variability, so the invasive value falls inside the PUNCH 95% interval in only 3 of 20 cases; that is, PUNCH is more precise than the reference rather than disagreeing with it. Assessed against the reference’s variability, the PUNCH estimate falls within the invasive 95% test–retest interval in 15 of 20 patients, and the Bland–Altman limits of agreement ($\pm 38\%$ of mean CFR) are comparable to thermodilution’s own reproducibility ($\approx \pm 33\%$). The point estimates agree well with invasive CFR (Deming slope 0.83, Spearman $\rho = 0.89$); the limitation is that the model’s intervals are under-dispersed (overconfident), not that the estimates disagree. This under-dispersion is informative: it indicates that the dominant clinical error arises from sources *outside* the model’s posterior — the variability of the invasive bolus-thermodilution reference itself, together with real anatomy and acquisition factors — which a recalibration fit on synthetic identifiability uncertainty cannot capture. We therefore interpret PUNCH uncertainty as capturing measurement/identifiability uncertainty (image quality, motion, incomplete contrast filling, propagated through the shared latent variable), useful for relative ranking and quality control; it does not capture physiological/biological variability of CFR, nor epistemic uncertainty about the 1D model form, and we do not claim calibrated clinical credible intervals. As an image-based comparator we report a naive Hough-slope front-tracking baseline on the clinical cohort (Section 3, Table 4); it returned a usable CFR for only half the patients and showed no association with invasive CFR, which we take as evidence that the agreement achieved by PUNCH

is attributable to the physics-constrained inference rather than to the kymograph representation alone. We do not include a TIMI/corrected-TIMI frame-count or a supervised deep-learning comparator. Frame counting requires a dedicated workflow and a standardized contrast-injection/acquisition protocol that were not part of this retrospective dataset, for which only the raw angiographic sequences are available; for a single imaged segment it captures the same contrast-front transit information that the Hough baseline already uses. A supervised comparator cannot be trained for this phenotype because no labeled angiographic CFR dataset of useful size exists, an absence that is itself part of the motivation for a physics-informed, label-free approach. The contributions of the PDE loss, the variational latent, the shared latent across rest and hyperemia, and the choice of variational inference over Monte Carlo dropout are quantified in the ablation and uncertainty-method comparison of Table 2. In this work, we focus on the LAD as it is the most common artery studied in coronary microvascular dysfunction testing in routine clinical practice. The method can also be extended to other coronary territories, such as the right coronary and left circumflex arteries, with due calibration.

Broader implications. Methodologically, embedding the governing physics of contrast transport into a per-patient inference problem with explicit uncertainty quantification offers an alternative to label-hungry supervised learning when ground-truth measurements for the target phenotype are scarce. Whether this approach extends usefully to other medical imaging-based inverse problems — perfusion imaging, tracer kinetics, or dynamic contrast studies across vascular beds — is an open question we do not resolve here; the present study should be read as a feasibility demonstration on a single anatomical territory, with extension to multi-branch coronary geometries and to other imaging modalities identified as separate future studies.

CRedit authorship contribution statement

Sukirt Thakur: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marcus Roper:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Yang Zhou:** Writing – review & editing, Software. **Dmitry Yu. Isaev:** Writing – review & editing, Software. **Reza Akbarian Bafghi:** Writing – review & editing, Software. **Brahmajee K. Nallamothu:** Writing – review & editing, Resources, Investigation. **C. Alberto Figueroa:** Writing – review & editing, Investigation. **Srinivas Paruchuri:** Writing – review & editing, Resources, Project administration. **Scott Burger:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Carlos Collet:** Writing – review & editing, Investigation. **Maziar Raissi:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude (Anthropic) in order to edit and improve the language and readability of the manuscript. The ideas, methodology, implementation, and analyses presented are the authors' own. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research was funded internally by AngioInsight, Inc. The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

Sukirt Thakur, Yang Zhou, Dmitry Yu. Isaev, Srinivas Paruchuri, and Scott Burger are employees of and hold equity interests in AngioInsight, Inc., a company developing artificial intelligence solutions in cardiac health. Marcus Roper and Maziar Raissi hold an equity interest in AngioInsight, Inc. as consultants. Brahmajee K. Nallamothu and C. Alberto Figueroa are employed by the University of Michigan and hold an equity interest in AngioInsight, Inc. as consultants and co-founders. The other authors have no competing interests to declare.

Acknowledgments

The authors thank the clinical and catheterization laboratory staff at the University of Michigan Health System for their assistance in acquiring the imaging and physiology data used in this retrospective analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.media.2026.104273>.

Data availability

Clinical angiographic data cannot be shared publicly due to patient privacy. The code that generates the synthetic angiogram dataset, including the forward advection–diffusion solver, the model-mismatch generators (vessel tapering and side-branch outflow), the parameter sampling, and the noise and corruption pipeline, can be shared upon reasonable request, so that the synthetic data and the model-mismatch, ablation, and uncertainty analyses can be reproduced in full. The per-patient physics-informed inference solver and its trained components were developed by, and are proprietary to, AngioInsight, Inc., which funded the work and is carrying the method toward clinical deployment; this component is therefore not released as open source. To keep the method transparent: the network architecture, the exact operating-point hyperparameters (prior mean and variance of the variational posterior, KL weight, optimizer settings, number of training epochs, and collocation- and data-point counts), the data normalization procedure, the boundary-condition handling for the PDE residual, and pseudocode for the training and per-patient inference loops are provided.

References

- Bairey Merz, C.N., Pepine, C.J., Walsh, M.N., Fleg, J.L., 2017. Ischemia and no obstructive coronary artery disease (INOCA). *Circulation* 135 (11), 1075–1092. <http://dx.doi.org/10.1161/CIRCULATIONAHA.116.024534>, URL: <https://www.ahajournals.org/doi/10.1161/circulationaha.116.024534>.
- Ball, C., Pontone, G., Rabbat, M., 2018. Fractional flow reserve derived from coronary computed tomography angiography datasets: The next frontier in noninvasive assessment of coronary artery disease. *BioMed. Res. Int.* 2018, 1–8. <http://dx.doi.org/10.1155/2018/2680430>, URL: <https://www.hindawi.com/journals/bmri/2018/2680430/>.
- Barbato, E., Aarnoudse, W., Aengevaeren, W.R., Werner, G., Klauss, V., Bojara, W., Herzfeld, I., Oldroyd, K.G., Pijls, N.H.J., De Bruyne, B., Week 25 study group, 2004. Validation of coronary flow reserve measurements by thermodilution in clinical practice. *Eur. Heart J.* 25 (3), 219–223. <http://dx.doi.org/10.1016/j.ehj.2003.11.009>.
- Beştemir, A., Apaydın, Z., Yaşar Kılınc, A., 2023. Analysis of coronary angiography and revascularization rates made over 5 years in public institutions in Türkiye. *Anatol. J. Cardiol.* 27 (9), 529–533. <http://dx.doi.org/10.14744/AnatolJCardiol.2023.3112>, URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10510416/>.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1 (8476), 307–310.
- Cho, S., Huang, J., Nam, J., An, H., Kim, S., Lee, J.-Y., 2024. Local all-pair correspondence for point tracking. In: *Computer Vision – ECCV 2024*. In: *Lecture Notes in Computer Science*, vol. 15068, Springer, pp. 306–325. http://dx.doi.org/10.1007/978-3-031-72684-2_18.

- De Bruyne, B., Baudhuin, T., Melin, J.A., Pijls, N.H., Sys, S.U., Bol, A., Paulus, W.J., Heyndrickx, G.R., Wijns, W., 1994. Coronary flow reserve calculated from pressure measurements in humans. Validation with positron emission tomography. *Circulation* 89 (3), 1013–1022. <http://dx.doi.org/10.1161/01.cir.89.3.1013>.
- Evans, L.C., 2010. Partial Differential Equations. American Mathematical Society.
- Fawaz, S., Marin, F., Khan, S.A., Simpson, R.F.G., Kotronias, R.A., Chai, J., Oxford Acute Myocardial Infarction (OxAMI) Study Investigators, Al-Janabi, F., Jagathesan, R., Konstantinou, K., MohdNazri, S.R., Clesham, G.J., Tang, K.H., Cook, C.M., Channon, K.M., Banning, A.P., Davies, J.R., Karamasis, G.V., De Maria, G.L., Keeble, T.R., 2024. Comparison of bolus versus continuous thermodilution derived indices of microvascular dysfunction in revascularized coronary syndromes. *IJC Heart Vasc.* 51, 101374. <http://dx.doi.org/10.1016/j.ijcha.2024.101374>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S235290672400040X>.
- Fearon, W.F., Balsam, L.B., Farouque, H.M.O., Robbins, R.C., Fitzgerald, P.J., Yock, P.G., Yeung, A.C., 2003. Novel index for invasively assessing the coronary microcirculation. *Circulation* 107 (25), 3129–3132. <http://dx.doi.org/10.1161/01.CIR.0000080700.98607.D1>, URL: <https://www.ahajournals.org/doi/10.1161/01.cir.0000080700.98607.d1>.
- Ford, T.J., Stanley, B., Good, R., Rocchiccioli, P., McEntegart, M., Watkins, S., Eteiba, H., Shaukat, A., Lindsay, M., Robertson, K., Hood, S., McGeoch, R., McDade, R., Yip, E., Sidik, N., McCartney, P., Corcoran, D., Collison, D., Rush, C., McConachie, A., Touyz, R.M., Oldroyd, K.G., Berry, C., 2018. Stratified medical therapy using invasive coronary function testing in Angina: The CorMicA trial. *J. Am. Coll. Cardiol.* 72, 2841–2855. <http://dx.doi.org/10.1016/j.jacc.2018.09.006>.
- Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning. In: ICML'16, JMLR.org, pp. 1050–1059.
- Gallinoro, E., Bertolone, D.T., Fernandez-Peregrina, E., Paolillo, P., Bermpis, K., Esposito, G., Gomez-Lopez, A., Candreva, A., Mileva, N., Belmonte, M., Mizukami, T., Fournier, S., Vanderheyden, M., Wyffels, E., Bartunek, J., Sonck, J., Barbato, E., Collet, C., De Bruyne, B., 2023. Reproducibility of bolus versus continuous thermodilution for assessment of coronary microvascular function in patients with ANOCA. *EuroIntervention* 19 (2), e155–e166. <http://dx.doi.org/10.4244/EIJ-D-22-00772>, URL: <https://eurointervention.pconline.com/doi/10.4244/EIJ-D-22-00772>.
- Gdowski, M.A., Murthy, V.L., Doering, M., Monroy-Gonzalez, A.G., Slart, R., Brown, D.L., 2020. Association of isolated coronary microvascular dysfunction with mortality and major adverse cardiac events: A systematic review and meta-analysis of aggregate data. *J. Am. Heart Assoc.* 9 (9), e014954. <http://dx.doi.org/10.1161/JAHA.119.014954>, URL: <https://www.ahajournals.org/doi/10.1161/JAHA.119.014954>.
- Itu, L., Rapaka, S., Passerini, T., Georgescu, B., Schwemmer, C., Schoebinger, M., Flohr, T., Sharma, P., Comaniciu, D., 2016. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *J. Appl. Physiol.* 121 (1), 42–52. <http://dx.doi.org/10.1152/jappphysiol.00752.2015>.
- Jansen, T.P.J., De Vos, A., Paradies, V., Dimitriu-Leen, A., Crooijmans, C., Elias-Smale, S., Rodwell, L., Maas, A.H.E.M., Smits, P.C., Pijls, N., Van Royen, N., Damman, P., 2023. Continuous versus bolus thermodilution-derived coronary flow reserve and microvascular resistance reserve and their association with angina and quality of life in patients with angina and nonobstructive coronaries: A head-to-head comparison. *J. Am. Heart Assoc.* 12 (16), e030480. <http://dx.doi.org/10.1161/JAHA.123.030480>, URL: <https://www.ahajournals.org/doi/10.1161/JAHA.123.030480>.
- Kelshiker, M.A., Seligman, H., Howard, J.P., Rahman, H., Foley, M., Nowbar, A.N., Rajkumar, C.A., Shun-Shin, M.J., Ahmad, Y., Sen, S., Al-Lamee, R., Petraro, R., Coronary Flow Outcomes Reviewing Committee, 2022. Coronary flow reserve and cardiovascular outcomes: A systematic review and meta-analysis. *Eur. Heart J.* 43 (16), 1582–1593. <http://dx.doi.org/10.1093/eurheartj/ehab775>.
- Kingma, D.P., Welling, M., 2014. Auto-encoding variational Bayes. <http://dx.doi.org/10.48550/arXiv.1312.6114>, URL: <http://arxiv.org/abs/1312.6114>.
- Kunadian, V., Chieffo, A., Camici, P.G., Berry, C., Escaned, J., Maas, A.H.E.M., Prescott, E., Karam, N., Appelman, Y., Fraccaro, C., Buchanan, G.L., Manzo-Silberman, S., Al-Lamee, R., Regar, E., Lansky, A., Abbott, J.D., Badimon, L., Duncker, D.J., Mehran, R., Capodanno, D., Baumbach, A., 2021. An EAPCI expert consensus document on ischaemia with non-obstructive coronary arteries in collaboration with European society of cardiology working group on coronary pathophysiology & microcirculation endorsed by coronary vasomotor disorders international study group. *EuroIntervention*. J. EuroPCR Collab. the Work. Group Interv. Cardiol. Eur. Soc. of Cardiol. 16 (13), 1049–1069. http://dx.doi.org/10.4244/EIJY20M07_01.
- Li, Y., Wu, Y., He, J., Jiang, W., Wang, J., Peng, Y., Jia, Y., Xiong, T., Jia, K., Yi, Z., Chen, M., 2022. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. *Eur. Radiol.* 32 (9), 6037–6045. <http://dx.doi.org/10.1007/s00330-022-08761-z>, URL: <https://link.springer.com/10.1007/s00330-022-08761-z>.
- Linka, K., Schäfer, A., Meng, X., Zou, Z., Karniadakis, G.E., Kuhl, E., 2022. Bayesian physics informed neural networks for real-world nonlinear dynamical systems. *Comput. Methods Appl. Mech. Engrg.* 402, 115346. <http://dx.doi.org/10.1016/j.cma.2022.115346>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0045782522004327>.
- Linnet, K., 1993. Evaluation of regression procedures for methods comparison studies. *Clin. Chem.* 39 (3), 424–432.
- Morris, P.D., Narracott, A., von Tengg-Kobligh, H., Silva Soto, D.A., Hsiao, S., Lungu, A., Evans, P., Bressloff, N.W., Lawford, P.V., Hose, D.R., Gunn, J.P., 2016. Computational fluid dynamics modelling in cardiovascular medicine. *Heart* 102 (1), 18–28. <http://dx.doi.org/10.1136/heartjnl-2015-308044>, URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4717410/>.
- Murthy, V.L., Naya, M., Foster, C.R., Hainer, J., Gaber, M., Dorbala, S., Charytan, D.M., Blankstein, R., Di Carli, M.F., 2012. Coronary vascular dysfunction and prognosis in patients with chronic kidney disease. *JACC Cardiovasc. Imaging* 5 (10), 1025–1034. <http://dx.doi.org/10.1016/j.jcmg.2012.06.007>.
- Ong, P., Camici, P.G., Beltrame, J.F., Crea, F., Shimokawa, H., Sechtem, U., Kaski, J.C., Bairey Merz, C.N., Coronary Vasomotor Disorders International Study Group (COVADIS), 2018. International standardization of diagnostic criteria for microvascular angina. *Int. J. Cardiol.* 250, 16–20. <http://dx.doi.org/10.1016/j.ijcard.2017.08.068>.
- Qi, Y., He, Y., Qi, X., Zhang, Y., Yang, G., 2023. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, IEEE, pp. 6070–6079.
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. <http://dx.doi.org/10.1016/j.jcp.2018.10.045>, URL: <https://www.sciencedirect.com/science/article/pii/S0021999118307125>.
- Rezende, D.J., Mohamed, S., Wierstra, D., 2014. Stochastic backpropagation and approximate inference in deep generative models. In: Proceedings of the 31st International Conference on Machine Learning. In: ICML'14, JMLR.org, pp. 11278–11286.
- Taqueti, V.R., Di Carli, M.F., 2018. Coronary microvascular disease pathogenic mechanisms and therapeutic options: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* 72 (21), 2625–2641. <http://dx.doi.org/10.1016/j.jacc.2018.09.042>.
- Taylor, C.A., Fonte, T.A., Min, J.K., 2013. Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: Scientific basis. *J. Am. Coll. Cardiol.* 61 (22), 2233–2241. <http://dx.doi.org/10.1016/j.jacc.2012.11.083>.
- Tjoe, B., Barsky, L., Wei, J., Samuels, B., Azarbal, B., Merz, C.N.B., Shufelt, C., 2021. Coronary microvascular dysfunction: Considerations for diagnosis and treatment. *Cleve. Clin. J. Med.* 88 (10), 561–571. <http://dx.doi.org/10.3949/ccjm.88a.20140>.
- Xiong, X., Wu, Z., Tan, S., Li, W., Tang, F., Chen, Y., Li, S., Ma, J., Li, G., 2026. SAM2-UNet: Segment anything 2 makes strong encoder for natural and medical image segmentation. *Vis. Intell.* 4 (1), 2. <http://dx.doi.org/10.1007/s44267-025-00106-w>, URL: <http://arxiv.org/abs/2408.08870>.
- Yang, Y., Perdikaris, P., 2019. Adversarial uncertainty quantification in physics-informed neural networks. *J. Comput. Phys.* 394, 136–152. <http://dx.doi.org/10.1016/j.jcp.2019.05.027>, URL: <https://www.sciencedirect.com/science/article/pii/S0021999119303584>.
- Yang, H., Zhang, J., Nallamothu, B.K., Garikipati, K., Figueroa, C.A., 2026. Assessing coronary microvascular dysfunction using angiography-based data-driven methods. *Comput. Methods Appl. Mech. Engrg.* 452, 118743. <http://dx.doi.org/10.1016/j.cma.2026.118743>, URL: <https://www.sciencedirect.com/science/article/pii/S0045782526000174>.
- Zierler, K.L., 1962. Theoretical basis of indicator-dilution methods for measuring flow and volume. *Circ. Res.* 10 (3), 393–407. <http://dx.doi.org/10.1161/01.RES.10.3.393>, URL: <https://www.ahajournals.org/doi/10.1161/01.RES.10.3.393>.
- Zierler, K., 2000. Indicator dilution methods for measuring blood flow, volume, and other properties of biological systems: A brief history and memoir. *Ann. Biomed. Eng.* 28 (8), 836–848. <http://dx.doi.org/10.1114/1.1308496>, URL: <http://link.springer.com/10.1114/1.1308496>.
- Zou, Z., Meng, X., Karniadakis, G.E., 2025. Uncertainty quantification for noisy inputs-outputs in physics-informed neural networks and neural operators. *Comput. Methods Appl. Mech. Engrg.* 433, 117479. <http://dx.doi.org/10.1016/j.cma.2024.117479>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0045782524007333>.