

Data-efficient learning of anomalous diffusion with wavelet representations: Enabling direct learning from experimental trajectories

Gongyi Wang , Yu Zhang , and Zihan Huang *

School of Physics and Electronics, Hunan University, Changsha 410082, China



(Received 10 December 2025; accepted 18 April 2026; published 5 May 2026)

Machine learning (ML) has become a versatile tool for analyzing anomalous diffusion trajectories, yet most existing pipelines are trained on large collections of simulated data. In contrast, experimental trajectories, such as those from single-particle tracking (SPT), are typically scarce and may differ substantially from the idealized models used for simulation, leading to degradation or even breakdown of performance when ML methods are applied to real data. To address this mismatch, we introduce a wavelet-based representation of anomalous diffusion that enables data-efficient learning directly from experimental recordings. This representation is constructed by applying six complementary wavelet families to each trajectory and combining the resulting wavelet modulus scalograms. We first evaluate the wavelet representation on simulated trajectories from the andi-datasets benchmark, where it clearly outperforms both feature-based and trajectory-based methods with as few as 1000 training trajectories and still retains an advantage on large training sets. We then use this representation to learn directly from experimental SPT trajectories of fluorescent beads diffusing in F-actin networks, where the wavelet representation remains superior to existing alternatives for both diffusion-exponent regression and mesh-size classification. In particular, when predicting the diffusion exponents of experimental trajectories, a model trained on 1200 experimental tracks using the wavelet representation achieves significantly lower errors than state-of-the-art deep learning models trained purely on 10^6 simulated trajectories. We associate this data efficiency with the emergence of distinct scale fingerprints disentangling underlying diffusion mechanisms in the wavelet spectra. These findings position the wavelet representation as a data-efficient basis for machine-learning analysis of anomalous diffusion, reducing reliance on simulation-based training and enabling more direct use of experimental trajectories.

DOI: [10.1103/sx22-92gh](https://doi.org/10.1103/sx22-92gh)

I. INTRODUCTION

Anomalous diffusion [1–4], a widespread phenomenon where transport deviates from classical Brownian motion, is prevalent across various complex systems such as biological tissues [5–8], crowded environments [9–15], complex fluids [16–19], and financial markets [20–22]. Effective analysis of anomalous diffusion is essential for revealing the dynamics and interactions within these systems. Beyond traditional statistical methods [23–32], machine-learning (ML) techniques [33–44] have recently gained significant attention for their powerful ability to characterize anomalous diffusion trajectories, with applications including diffusion-exponent prediction [45–52], diffusion model classification [53–60], and state transition detection (i.e., trajectory segmentation) [61–80]. These ML methods have greatly enhanced the precision of anomalous diffusion analysis and have the capacity to address challenges that conventional methods struggle to overcome.

Currently, the prevailing ML approaches for analyzing anomalous diffusion follow a *simulation-based* paradigm [81,82], where models are trained on large synthetic datasets generated from idealized diffusion models. Once trained, these models are then transferred to real experimental data for analysis. By carefully designing simulated diffusion scenarios, the simulation-based paradigm leverages large amounts of labeled synthetic data, allowing ML models to capture complex patterns and relationships that would be difficult to discern using traditional methods [83–91]. However, this paradigm relies heavily on the assumption that simulated data closely mirror real experimental conditions, which is often not the case due to factors such as limited sample sizes, differences in diffusion processes, and distribution shifts between synthetic and real data [92–94]. As a result of this simulation-to-reality gap, models trained on simulated data may not generalize well to real-world scenarios, leading to performance degradation or even failure when applied to experimental data.

Despite these inherent limitations of the simulation-based paradigm when transferring to experimental data, it has so far remained the primary route for applying ML to anomalous diffusion. This situation is largely driven by practical constraints on experimental data. In typical experiments like single-particle tracking (SPT) [55,95–101], the total number of recorded trajectories is itself limited [102]. Moreover, since obtaining high-quality labels for these trajectories is

*Contact author: huangzih@hnu.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

substantially more difficult than generating labeled synthetic data, only a small fraction of experimental trajectories can be reliably annotated. Under such data constraints, the two main representation families for anomalous diffusion trajectories, namely, feature-based [60,103–106] and trajectory-based [46,55–57] representations, encounter specific limitations in the absence of simulation-based training, albeit for opposite reasons:

(1) Feature-based representations map each trajectory to a vector of handcrafted descriptors with clear physical or statistical meaning, such as the mean-squared displacement (MSD) and p -variation statistics. These physics-informed features allow models to remain reasonably reliable with limited experimental data, but their expressive power is constrained by the predefined feature set, so performance quickly saturates at a relatively low level even as more data are added [61].

(2) In contrast, trajectory-based representations operate directly on raw trajectories or step sequences, typically using deep sequence models to learn data-driven features. These high-capacity models can achieve excellent performance when trained on large datasets, but are extremely data hungry and tend to severely underperform in the small-data regime (e.g., on the order of 10^3 experimental trajectories) [46].

Taken together, these observations highlight that overcoming the limitations of the simulation-based paradigm cannot be achieved simply by accumulating more simulated data or scaling existing ML methods. Instead, relaxing this reliance on large synthetic datasets requires a trajectory representation of anomalous diffusion that combines physics-informed multiscale features with sufficient expressive capacity, enabling data-efficient and stable learning from limited experimental data while still fully exploiting large training sets when available. Wavelet transforms, with their inherent ability to localize signal structure simultaneously in time and scale, are natural candidates for constructing such a representation.

Indeed, wavelet and spectral methods have a substantial history in the analysis of anomalous diffusion and single-particle trajectories, though they have been deployed in a fundamentally different manner from the approach proposed here. Chen *et al.* [98] applied the Haar continuous wavelet transform (CWT) with adaptive universal thresholding to SPT data, using the wavelet coefficients as a diagnostic tool to segment trajectories into distinct dynamical states such as active transport and passive diffusion. Krapf *et al.* [26] developed an exact analytical framework for the single-trajectory power spectral density of fractional Brownian motion, establishing how spectral scaling encodes the anomalous-diffusion exponent. More broadly, wavelet coefficient moments and their scaling properties have been used to characterize self-similar processes [107]. A unifying feature of these prior approaches is that wavelet or spectral coefficients are ultimately reduced to lower-dimensional descriptors or summary statistics (such as thresholding decisions, scaling exponents, or spectral slopes), which are then used for inference. The representation we introduce below departs from this paradigm: Rather than extracting predefined statistics from the wavelet coefficients, it retains the full two-dimensional (2D) scalogram as an image-like input to supervised learning models, thereby allowing the downstream model to discover task-relevant patterns directly from the time-scale plane.

In this work, we pursue this idea by constructing a wavelet-based encoding of anomalous diffusion trajectories. By mapping trajectories to wavelet modulus scalograms on the time-scale plane, the characteristic multiscale structure of random motions is rendered explicit and becomes more amenable to learning. On simulated trajectories from the andi-datasets benchmark [81], the wavelet representation consistently outperforms both feature-based and trajectory-based approaches in diffusion-exponent regression and diffusion-model classification, even when trained on as few as 1000 trajectories. On experimental SPT trajectories of fluorescent beads diffusing in F-actin networks [55], models built on the wavelet representation remain superior to alternative representations for both diffusion-exponent regression and mesh-size classification. Notably, when predicting the experimental diffusion exponents, a model trained on 1200 experimental trajectories using the wavelet representation achieves substantially lower errors than state-of-the-art (SOTA) deep learning models following the simulation-based paradigm (trained purely on 10^6 simulated trajectories). In addition, the characteristic scale fingerprints in the wavelet spectra are analyzed, which provide physical insights into the data-efficient behavior of the wavelet representation. These results demonstrate that the wavelet representation enables data-efficient learning of anomalous diffusion, thereby allowing direct learning from experimental trajectories and offering a promising direction toward developing alternatives to the simulation-based paradigm.

The rest of this paper is organized as follows. Section II presents the proposed wavelet representation in detail. Section III summarizes the simulated and experimental datasets, the associated learning tasks, and the evaluation metrics used in this work. Section IV reports the main results on simulated trajectories from the andi-datasets benchmark. Section V presents the results on experimental SPT trajectories of fluorescent beads in F-actin networks, specifically contrasting the proposed direct learning approach against the prevailing simulation-based paradigm. Section VI addresses interpretability by analyzing scale fingerprints in the wavelet spectra. Finally, we draw our conclusions in Sec. VII. Supporting information on wavelet scale ablation, experimental trajectory length distributions, detailed model-wise performance, noise optimization for simulation-based training, and wavelet family selection is provided in the Appendixes.

II. WAVELET REPRESENTATION

In this section, we describe the proposed wavelet-based representation of anomalous diffusion in detail. For brevity, we refer to it as the *wavelet representation* throughout this paper. We start from a single standardized one-dimensional (1D) trajectory $x(t)$ of length L and compute its CWT [108,109] with respect to a mother wavelet $\psi(t)$. The CWT is defined as

$$W_\psi(a, b) = \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) \frac{dt}{\sqrt{|a|}}, \quad (1)$$

where $a > 0$ denotes the scale, b is the translation (time position), and the superscript $*$ denotes complex conjugation. For a fixed wavelet family, the complex coefficients $W_\psi(a, b)$ form a 2D time-scale representation of the trajectory, and we use their modulus $|W_\psi(a, b)|$ to construct a wavelet modulus

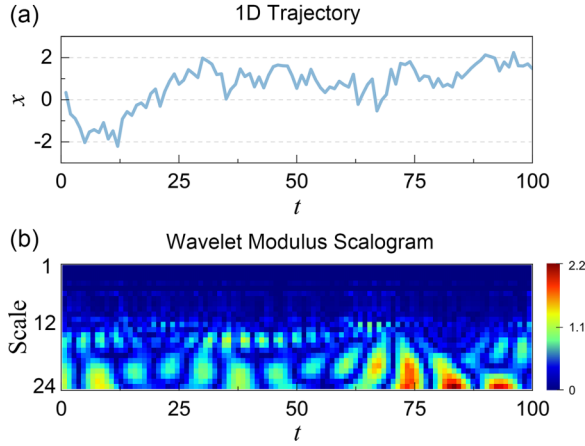


FIG. 1. Representative example of the continuous wavelet transform applied to a 1D trajectory. (a) Standardized 1D trajectory $x(t)$ of length $L = 100$. (b) Corresponding wavelet modulus scalogram $|W_\psi(a, b)|$ computed with the real Morlet wavelet over 24 discrete scales.

scalogram. In this work, all CWTs are implemented using the Python package PyWavelets [110]. A representative example is shown in Fig. 1: A standardized 1D trajectory with $L = 100$ [Fig. 1(a)] is transformed into a modulus scalogram evaluated at 24 scales [Fig. 1(b)] using the real Morlet wavelet [111]. The color in this scalogram encodes the magnitude of $|W_\psi(a, b)|$, with localized high-amplitude patches revealing the multiscale structure of the underlying random motion.

Six standard continuous mother wavelets are employed to construct the wavelet representation in this work:

- (1) Real Morlet wavelet [111]: A sinusoidal carrier modulated by a Gaussian envelope, widely used for time-frequency analysis of oscillatory signals.
- (2) Mexican hat wavelet [112]: Proportional to the second derivative of a Gaussian function, acting as a band-pass, pulse-like kernel sensitive to local peaks and curvature changes.
- (3) Complex Gaussian derivative wavelet [113]: A complex-valued wavelet defined as the P -th ($P = 1$ in this work) derivative of a complex Gaussian function, providing simultaneous access to amplitude and phase information.
- (4) Complex Morlet wavelet [114]: A Gaussian envelope multiplied by a complex sinusoidal carrier, with explicitly tunable bandwidth and center frequency.

TABLE I. Continuous wavelet families used to construct the wavelet representation. Here, *Abbrev.* lists the shorthand names of wavelet families used throughout the text, *PyWT identifier* gives the corresponding callable name in PyWavelets, and *Parameters* specifies the wavelet parameters used in our implementation. j denotes the imaginary unit.

Continuous wavelet family	Abbrev.	PyWT identifier	Analytical form of $\psi(t)$	Parameters
Real Morlet	morl	morl	$\psi(t) = e^{-t^2/2} \cos(5t)$	—
Mexican hat	mexh	mexh	$\psi(t) = \frac{2}{\sqrt{3}\pi^{1/4}} e^{-t^2/2} (1 - t^2)$	—
Complex Gaussian derivative	cgau	cgau1	$\psi(t) \propto \frac{d^P}{dt^P} (e^{-jt} e^{-t^2})$	Derivative order $P = 1$
Complex Morlet	cmor	cmor1.5-1.0	$\psi(t) = \frac{1}{\sqrt{\pi B}} e^{-t^2/B} e^{j2\pi C t}$	$B = 1.5, C = 1.0$
Shannon	shan	shan1.5-1.0	$\psi(t) = \sqrt{B} \frac{\sin(\pi B t)}{\pi B t} e^{j2\pi C t}$	$B = 1.5, C = 1.0$
Frequency B-spline	fbsp	fbsp3-0.5-0.5	$\psi(t) = \sqrt{B} \left[\frac{\sin(\pi B \frac{t}{M})}{\pi B \frac{t}{M}} \right]^M e^{j2\pi C t}$	$M = 3, B = 0.5, C = 0.5$

(5) Shannon wavelet [115]: A band-limited sinc function modulated by a complex exponential, corresponding to an idealized rectangular passband in the frequency domain.

(6) Frequency B-spline wavelet [116]: A wavelet with a B-spline-shaped passband in the frequency domain, realizing a smooth, band-limited kernel with controllable main-lobe shape and smoothness.

The detailed analytical forms of these mother wavelets, the abbreviations used throughout the text, and the corresponding PyWavelets identifiers and parameter settings are summarized in Table I.

Next, we adopt a fixed set of 24 scales for all CWT computations in this work:

$$\{0.1, 0.2, \dots, 0.9, 1, 2, \dots, 15\}, \quad (2)$$

which spans from fine to coarse resolutions on the time-scale plane. For the wavelet families considered here, the pseudofrequency associated with a given scale a is approximately inversely proportional to a , so smaller scales tend to emphasize faster local fluctuations, whereas larger scales capture slower, more global structure. For discretely sampled trajectories with unit sampling interval, such a pseudofrequency interpretation is only approximate, and scales $a < 1$ may formally correspond to pseudofrequencies beyond the Nyquist limit. However, since our goal is to construct an ML representation, we use the CWT primarily as a multiscale feature extractor rather than as a precise spectral estimator. To enrich the wavelet representation with short-time information, we explicitly include subunit scales $a \in \{0.1, 0.2, \dots, 0.9\}$ to realize very short-support wavelet filters that are sensitive to variations over windows of one to a few time steps, such as abrupt changes in displacement or local fluctuations in step-to-step variability. On top of these fine scales, we use integer scales $a = 1, 2, \dots, 15$ to capture slower variations in the overall spreading behavior and larger-scale trends along the trajectory. Preliminary experiments on simulated trajectories further indicate that extending the integer-scale range beyond $a = 15$ does not bring systematic improvement, whereas augmenting the integer scales with the subunit scales yields a consistent gain in downstream performance. We therefore adopt the combined set in Eq. (2) as our default scale grid for all subsequent experiments. Detailed ablation results that support this choice are reported in Appendix A.

After specifying the mother wavelets and the scale set in Eq. (2), we now describe how the wavelet representation is

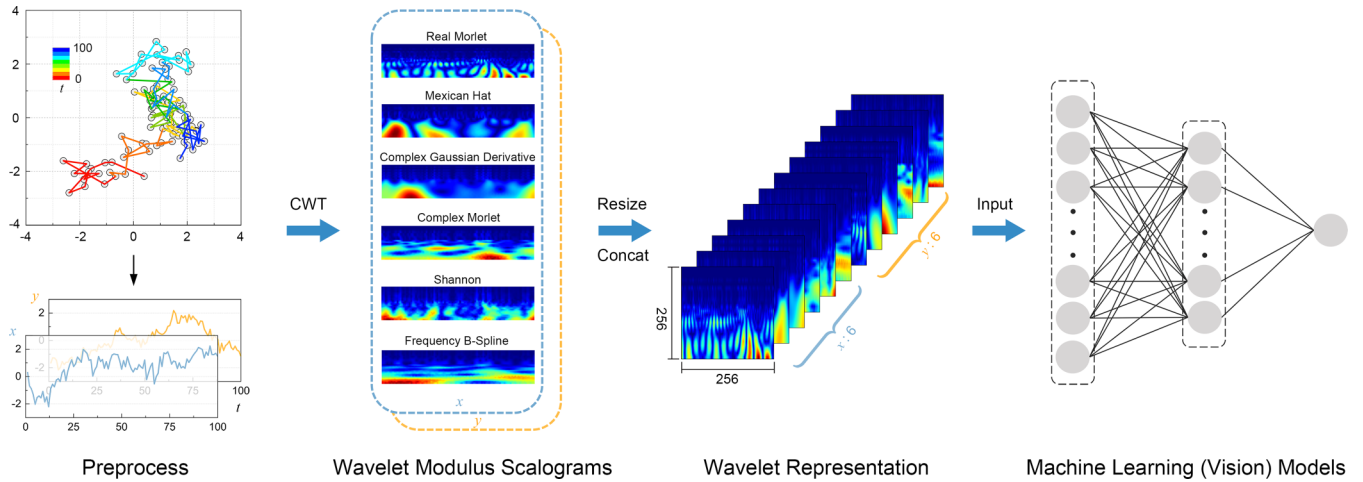


FIG. 2. Schematic construction of the wavelet representation. A multidimensional trajectory is first split into its 1D components, which are standardized. Subsequently, continuous wavelet transforms with six mother wavelets on the 24-scale grid of Eq. (2) produce wavelet modulus scalograms, which are resampled to a common 256×256 resolution and concatenated into a $6d$ -channel tensor. This wavelet representation is then used as the input to downstream supervised (vision) models.

constructed for a generic trajectory, as illustrated schematically in Fig. 2. The detailed steps are as follows:

(1) We consider a d -dimensional trajectory $\mathbf{x}(t) \in \mathbb{R}^d$ of length L (with $d = 1, 2, 3$ in this work) and decompose it into its d 1D components. Each 1D component is then standardized independently by a per-trajectory normalization to zero mean and unit variance.

(2) For each standardized 1D component, we compute the CWT with each of the six wavelet families in Table I, evaluated on the 24-scale grid in Eq. (2). This yields, for each wavelet, a wavelet modulus scalogram $|W_\psi(a, b)|$ defined on a discrete time-scale grid of size $24 \times L$. For a d -dimensional trajectory, this step produces $6d$ scalograms in total.

(3) Each scalogram is subsequently resampled onto a common 256×256 grid in the time-scale plane. The resulting $6d$ scalograms are then concatenated along the channel dimension to form a $6d$ -channel tensor, which we refer to as the wavelet representation of the trajectory.

(4) This multichannel representation serves as the input to downstream supervised models (in particular, computer vision models) for learning from the trajectories, without requiring any further task-specific feature engineering.

This procedure yields a multiscale, imagelike representation of trajectories that can be coupled with standard supervised models without additional feature engineering. We note that the computational overhead of constructing the wavelet representation is modest: With GPU-accelerated model inference, the total per-trajectory processing time of the wavelet pipeline is comparable to that of the feature-based pipeline. A detailed analysis of computational cost and scalability for all three representation families is provided in Appendix B. In the following section, we detail the datasets and learning tasks used to assess its performance.

III. DATASETS AND TASKS

In this section, we present the anomalous diffusion datasets and learning tasks used to evaluate the proposed wavelet

representation, together with the corresponding evaluation metrics. Section III A describes the simulated trajectories generated with the *andi-datasets* library [81] and the associated tasks of diffusion-exponent regression and diffusion-model classification. Section III B introduces the experimental SPT trajectories of fluorescent beads diffusing in F-actin networks with different mesh sizes [55] and the corresponding task design.

A. Simulated dataset and benchmark tasks

The simulated anomalous diffusion trajectories used in this work are generated with the open-source *andi-datasets* library, originally developed for the AnDi Challenge and now widely used as a benchmark for testing analysis methods on anomalous diffusion data. This library implements five prototypical stochastic models of anomalous transport covering a broad range of diffusion exponents α :

(1) Annealed transient time motion (ATTM) [117]: A non-ergodic process in which the particle undergoes Brownian motion with a diffusivity that is randomly renewed in time or space, leading to effective exponents in the range $0.05 \leq \alpha \leq 1$.

(2) Continuous-time random walk (CTRW) [118]: A random-walk model with broadly distributed waiting times between successive steps, giving rise to subdiffusive dynamics with $0.05 \leq \alpha \leq 1$.

(3) Fractional Brownian motion (FBM) [119]: A Gaussian process driven by fractional Gaussian noise with power-law temporal correlations, which can generate both subdiffusive and superdiffusive behavior, $0.05 \leq \alpha < 2$.

(4) Lévy walk (LW) [120]: A superdiffusive model in which step durations and jump lengths are coupled, producing non-Gaussian displacement statistics and exponents in the range $1 \leq \alpha \leq 2$.

(5) Scaled Brownian motion (SBM) [121]: A diffusion process with a deterministically time-dependent diffusion coefficient, yielding $0.05 \leq \alpha \leq 2$.

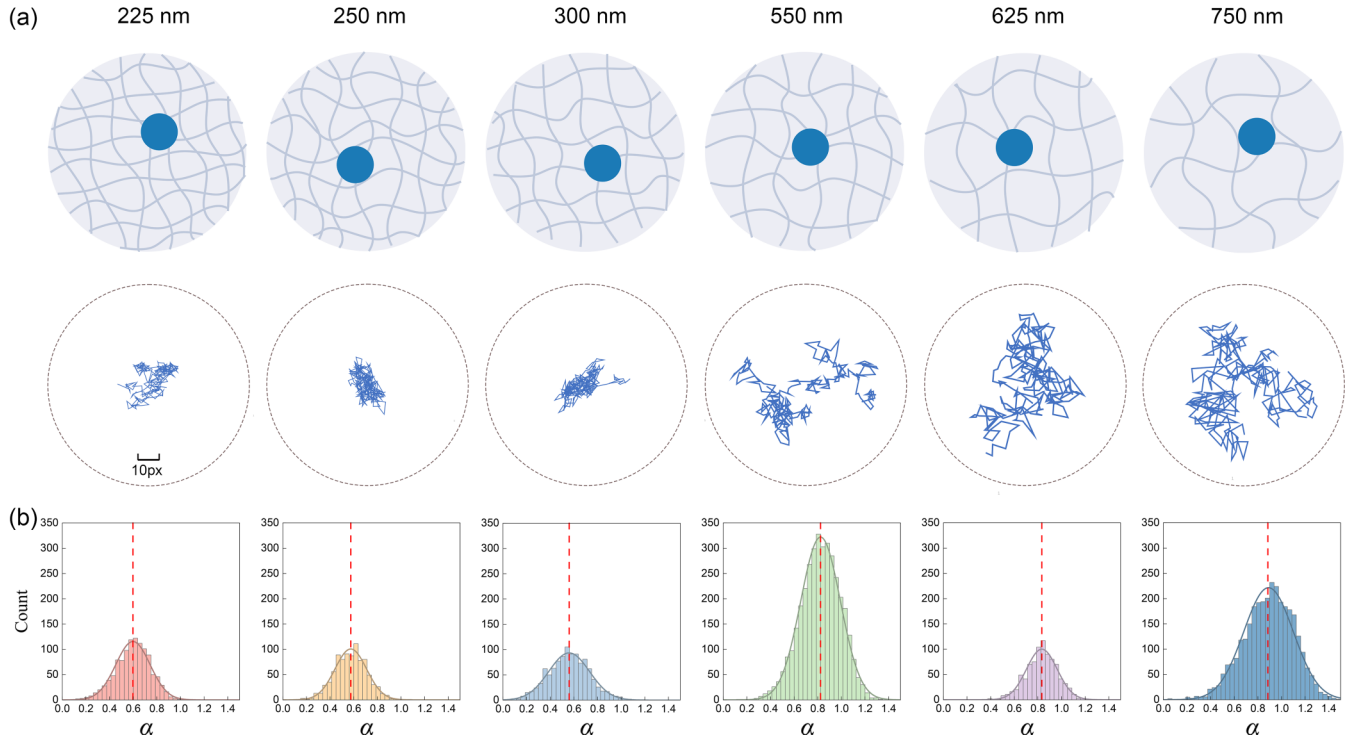


FIG. 3. (a) Schematic illustration of F-actin networks with mesh sizes $\xi \approx 225, 250, 300, 550, 625$, and 750 nm (top row), together with representative 2D bead trajectories recorded under each condition (bottom row). (b) Distributions of the diffusion exponent α estimated from TA-MSD fits for all trajectories at each mesh size; histograms show the counts of trajectories vs α , and the red dashed line in each panel marks the mean exponent for that condition.

Here, the diffusion exponent α characterizes the time scaling of MSD ($\text{MSD} \sim t^\alpha$). Simulated trajectories are generated from these five diffusion models in one, two, or three spatial dimensions with uniformly spaced time steps and optional additive Gaussian localization noise.

On this simulated dataset, we consider two supervised learning tasks. The first is the *diffusion-exponent regression* task, where the goal is to infer the diffusion exponent α of a single trajectory. Performance on this regression task is quantified by the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\alpha_{i,p} - \alpha_{i,GT}|, \quad (3)$$

where $\alpha_{i,p}$ and $\alpha_{i,GT}$ denote the predicted and ground-truth exponents of the i th trajectory, respectively, and N is the number of test trajectories. The second is the *diffusion-model classification* task, in which each trajectory must be assigned to one of the five diffusion models {ATTM, CTRW, FBM, LW, SBM}. As in the AnDi Challenge, we evaluate the classification performance using the microaveraged F1 score:

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (4)$$

Here, precision and recall are computed by pooling true positives (TPs), false positives (FPs), and false negatives (FNs) over all five classes, and are given by

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

B. Experimental SPT dataset and task design

The experimental SPT dataset in this work originates from the study by Granik *et al.* [55], which contains fluorescence trajectories of micron-sized beads embedded in reconstituted F-actin networks with controlled mesh sizes. In their experiments, fluorescent polystyrene beads with diameter $d_{\text{PS}} \approx 550$ nm are dispersed in entangled F-actin gels prepared at different actin monomer concentrations, yielding networks with characteristic mesh sizes $\xi \approx 225, 250, 300, 550, 625$, and 750 nm. The mesh size for each condition is estimated from established microrheology calibrations. By tuning the mesh size, the authors modulate the crowding and viscoelastic properties of the medium, thereby altering the diffusive behavior of beads. 2D trajectories are then acquired by video fluorescence microscopy. Illustrations of different mesh sizes and their corresponding representative trajectories in this dataset are shown in Fig. 3(a).

The SPT trajectories in this dataset exhibit substantial heterogeneity in length across the six mesh sizes. As reported in Appendix C, for the three smaller meshes (225, 250, and 300 nm), most trajectories contain more than 10^3 time steps, with a broad tail extending up to several thousand steps. For the larger meshes (550, 625, and 750 nm), the trajectories are typically shorter, with lengths concentrated in the range of a few hundred time steps. Moreover, to characterize the dynamical behavior under each condition, we estimate the diffusion exponent α for every trajectory by fitting the slope of its time-averaged mean-squared displacement (TA-MSD) in log-log coordinates. The resulting distributions of α are

TABLE II. ML models used for different data representations. Here, *Representation* lists the type of input representation, *ML model* gives the model family, *Implementation details* specifies the hyperparameters or implementation used in our experiments, and *Abbrev.* lists the shorthand names of ML models used throughout the text.

Representation	ML model	Implementation details	Abbrev.
Wavelet	EfficientNet [33]	timm/efficientnet_b5.sw_in12k_ft_in1k [122]	EffNet
	RegNet [34]	timm/regnetz_040_h.ra3_in1k [123]	RegNet
Feature	Vision transformer [35]	timm/vit_huge_patch14_224.orig_in21k [124]	ViT
	LightGBM [36]	n_estimators=100, max_depth=5	LGB
	XGBoost [37]	n_estimators=100, max_depth=5	XGB
	Random forest [38]	n_estimators=100, max_depth=None	RF
	Linear regression [40]	/	LinR
	Logistic regression [41]	C=1.0	LogR
Trajectory	Transformer [39]	Ref. [73]	Trm
	WADNet [46]	Ref. [125]	WADNet
	LSTM [42]	3 recurrent layers, hidden size 64	LSTM
	GRU [43]	3 recurrent layers, hidden size 64	GRU

shown in Fig. 3(b). For all six mesh sizes, the α values are broadly distributed around a well-defined peak. As the mesh size increases from 225 to 750 nm, the peak of the distribution shifts toward larger α , with mean exponents (highlighted by red dashed lines) increasing from clearly subdiffusive values below unity at small meshes to values close to $\alpha \approx 1$ at large meshes.

To adapt these experimental trajectories to supervised learning, we construct fixed-length segments and define two learning tasks. For each mesh size ξ , the pool of available long trajectories is first randomly split into disjoint training, validation, and test sets. After this trajectory-level separation, long tracks are partitioned into nonoverlapping 2D segments of a target length $L \in \{50, 100, 150, 200, 250\}$. All segments are then standardized (zero mean, unit variance per segment) and randomly shuffled within each subset. Each segment inherits the mesh-size label of its parent trajectory as well as the diffusion exponent α estimated from its TA-MSD. Based on these labels, two supervised tasks are considered, in direct analogy with the simulated case in Sec. III A:

(1) The first is *diffusion-exponent regression*, where the target is the TA-MSD-based exponent α of each segment. Model performance is quantified by the MAE, as defined in Eq. (3).

(2) The second is *mesh-size classification*, in which each segment is assigned to one of the mesh-size conditions $\xi \in \{225, 250, 300, 550, 625, 750\}$ nm. Performance is evaluated using the micro-averaged F1 score, following the definition in Eq. (4).

IV. RESULTS ON SIMULATED TRAJECTORIES

In this section, we systematically compare the proposed wavelet representation with feature-based and trajectory-based representations on the simulated benchmark described in Sec. III A. The detailed choice of features used in the feature-based representation can be found in Table 3 of Ref. [83], whereas the trajectory-based representation is obtained by applying per-trajectory standardization to the raw trajectories. For each representation family, we couple the input encoding with strong supervised learners (feature vectors

with decision-tree or linear models, raw trajectories with deep sequence networks, and wavelet representations with modern vision architectures). The specific ML models used for each representation, together with the abbreviations adopted in this work, are summarized in Table II. It is worth mentioning that the three vision models listed in Table II are initialized with ImageNet-pretrained weights. An ablation study comparing pretrained and trained-from-scratch versions of these architectures (Appendix D) shows that pretraining provides a moderate boost in the small-data regime ($N_{\text{train}} \leq 10^4$) but becomes negligible for $N_{\text{train}} \geq 10^5$, where both settings converge to the same performance. This convergence indicates that the high asymptotic accuracy is an intrinsic property of the wavelet representation, while pretraining serves as a practical accelerator that improves sample efficiency when training data are limited.

First, we evaluate these learners on diffusion-exponent regression (MAE) and diffusion-model classification (micro-F1) for 2D simulated trajectories of fixed length $L = 100$, across a range of training set sizes N_{train} . Here, Figs. 4(a) and 4(b) present a comprehensive performance comparison of ML models listed in Table II across a wide spectrum of training set sizes, ranging from 10^3 to 10^6 trajectories. Each heatmap reports test set scores (on 2×10^5 trajectories) as a function of N_{train} . Across this entire range and for both tasks, models built on the wavelet representation consistently achieve lower MAE and higher F1 than those using feature-based or trajectory-based representations. This advantage is particularly pronounced in the small-data regime, but it remains visible even when the training set contains as many as 10^6 simulated trajectories.

In addition, the comparison between feature-based and trajectory-based representations highlights a distinct trade-off regarding data scalability. With limited training data, feature-based models generally outperform trajectory-based deep sequence models, reflecting the benefit of explicit physics-informed descriptors in the low-sample regime. However, their performance quickly saturates as the dataset size increases. As the training set grows ($N_{\text{train}} \geq 10^5$), the trajectory-based models gradually narrow the gap and eventually outperform the feature-based ones, highlighting their

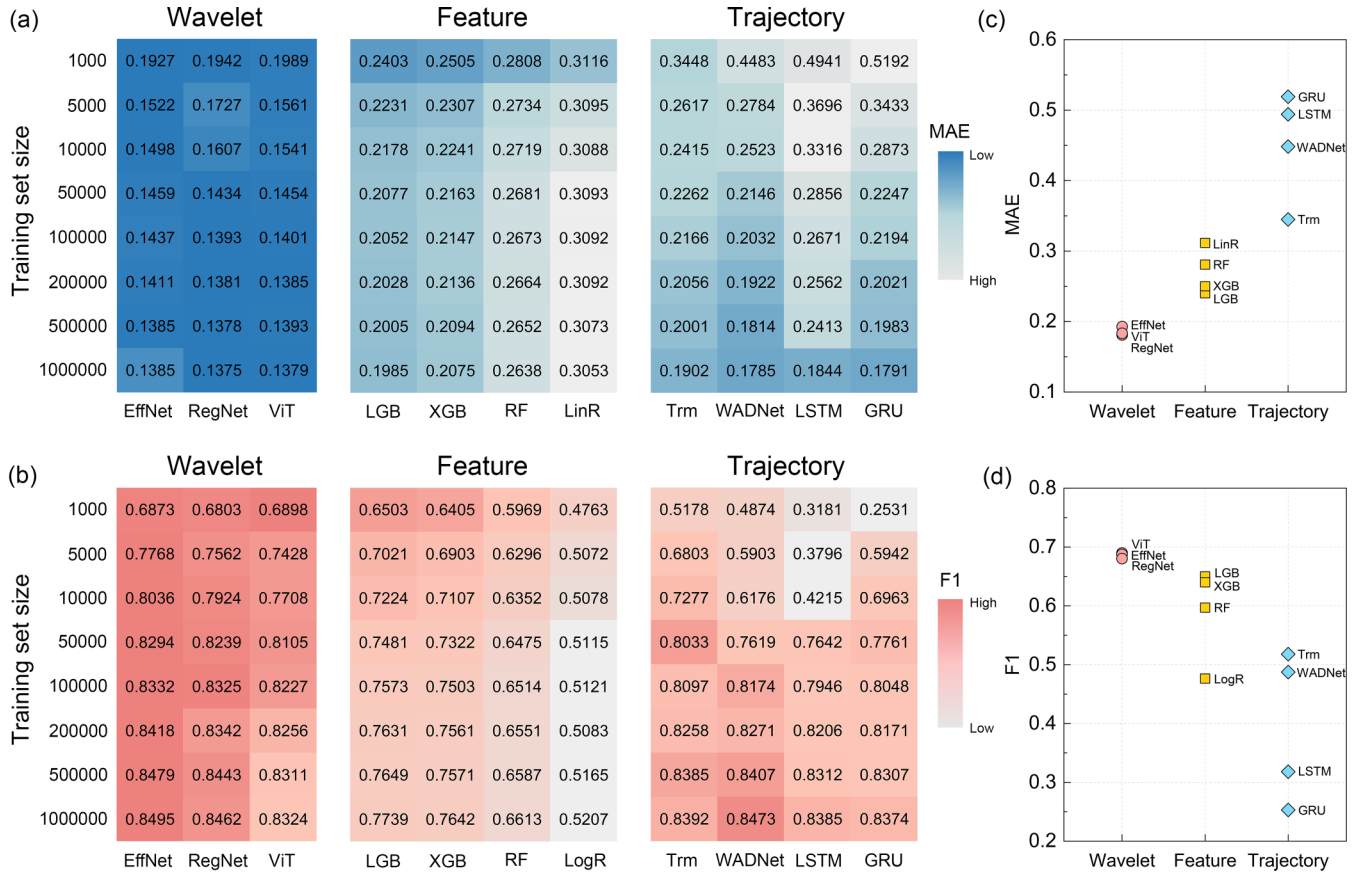


FIG. 4. Performance comparison of wavelet-based, feature-based, and trajectory-based representations on 2D simulated trajectories of length $L = 100$. (a), (b) Heatmaps displaying the performance on the test set across varying training set sizes N_{train} . (a) MAE for diffusion-exponent regression (blue scale; darker indicates lower error). (b) Micro-averaged F1 scores for diffusion-model classification (red scale; darker indicates higher F1). (c), (d) Detailed performance comparison at the small-data regime $N_{\text{train}} = 1000$, highlighting a clear advantage of wavelet-based models over both feature-based and trajectory-based alternatives.

higher asymptotic capacity when sufficient data are available. Nevertheless, they remain clearly inferior to the wavelet-based approach.

Moreover, to further illustrate data efficiency, Figs. 4(c) and 4(d) depict the performance distribution of individual models at a fixed training size of $N_{\text{train}} = 1000$. A clear hierarchy emerges that underscores the data-efficient nature of the wavelet representation: wavelet-based models cluster in the high-performance region (low MAE, high F1), followed by feature-based models, while trajectory-based models struggle to generalize from such limited data.

Based on these comparative results, we establish a standardized setup for the subsequent experiments, which focus specifically on the challenging small-data regime ($N_{\text{train}} = 1000$). We select the top-performing model from each representation family to serve as a representative baseline for further comparison:

- (1) EfficientNet (EffNet) is selected to represent the *wavelet* approach.
- (2) LightGBM (LGB) is chosen as the representative for *feature*-based learning.
- (3) Transformer (Trm) is selected to represent *trajectory*-based sequence modeling.

Unless otherwise specified, these three models will serve as the benchmarks for their respective representation families in the remainder of this work.

Having established the baseline performance, we now analyze the three representation families in more detail in the small-data regime ($N_{\text{train}} = 1000$). We examine four aspects: stability against sample variation, dependence on trajectory length, generalization across spatial dimensionality, and robustness to localization noise, with the results summarized in Fig. 5. We begin by assessing the stability of each representation under random resampling of the training set. To this end, we construct 50 independent training subsets of size $N_{\text{train}} = 1000$ and evaluate the models on a fixed test set. The resulting boxplots in Figs. 5(a1) and 5(a2) reveal a clear contrast: While the wavelet-based and feature-based models exhibit tight, compact distributions, the trajectory-based model suffers from significant variance and numerous outliers. This result underscores that deep sequence models become highly sensitive to sample selection when constrained by limited data, whereas the wavelet representation provides a robust embedding that is largely invariant to data resampling.

We then characterize the dependence on trajectory length L and spatial dimensionality. As shown in Figs. 5(b1) and

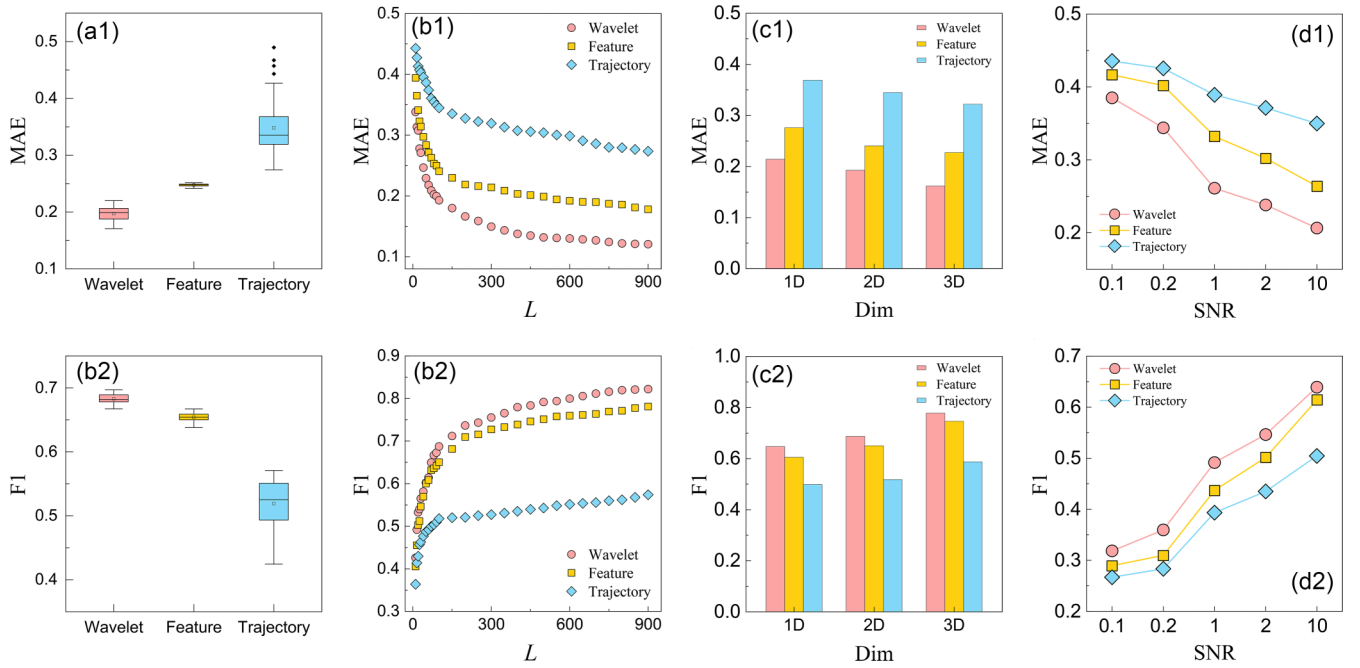


FIG. 5. Detailed performance analysis on simulated trajectories in the small-data regime ($N_{\text{train}} = 1000$). (a1), (a2) Boxplots of test MAE and F1 scores over 50 independent training subsets. Wavelet-based and feature-based models exhibit high stability (tight distributions), contrasting with the high variance of trajectory-based models. (b1), (b2) Dependence on trajectory length L ; the wavelet representation outperforms the baselines across all lengths. (c1), (c2) MAE and F1 scores for 1D, 2D, and 3D trajectories, confirming consistent performance rankings across spatial dimensions. (d1), (d2) Performance vs signal-to-noise ratio (SNR); the wavelet representation demonstrates higher noise resilience compared to feature-based and trajectory-based alternatives.

5(b2), although performance naturally improves with increasing L across all methods due to the accumulation of temporal information, the wavelet representation maintains a decisive lead throughout the entire range. This advantage extends to the spatial domain as well; Figs. 5(c1) and 5(c2) confirm that the performance hierarchy is preserved across 1D, 2D, and 3D systems, indicating that the efficacy of the wavelet approach is intrinsic to the representation rather than an artifact of a specific dimensional setting. Finally, we investigate the robustness of the representation to localization noise. We add zero-mean Gaussian noise with standard deviation σ_n to the simulated trajectories, corresponding to signal-to-noise ratios $\text{SNR} = 1/\sigma_n$, and evaluate performance across a range of noise levels. As shown in Figs. 5(d1) and 5(d2), while performance inevitably degrades for all models as the SNR decreases, the wavelet-based model remains markedly more resilient, maintaining the lowest MAE and highest F1 even in the noisiest conditions, whereas trajectory-based models deteriorate sharply. This suggests that the wavelet transform effectively filters high-frequency fluctuations while preserving the essential multiscale features required for diffusion analysis.

This comprehensive superiority is further corroborated by a detailed comparison across individual diffusion models (ATTM, CTRW, FBM, LW, and SBM), presented in Appendix E. Irrespective of the underlying physical mechanism, the wavelet representation consistently achieves the lowest regression errors and highest classification scores, confirming that its advantage is universal across the different types of anomalous diffusion considered. Collectively, these findings on simulated trajectories demonstrate that the wavelet

representation is more data efficient and robust than existing alternatives, providing a strong rationale for its application to experimental trajectories where data scarcity and noise are unavoidable.

V. RESULTS ON EXPERIMENTAL SPT TRAJECTORIES

Building on the analysis of simulated benchmarks, we now address the critical challenge of learning directly from real experimental trajectories. In this section, we compare the three representation families (wavelet-based, feature-based, and trajectory-based representations) on SPT recordings of fluorescent beads diffusing in F-actin networks, employing EffNet, LGB, and Trm as their representative learners. Distinct from the simulation-based paradigm that leverages large synthetic datasets, our approach here relies exclusively on experimental trajectories for both training and evaluation. To this end, we deliberately restrict the number of labeled tracks to mimic the data-scarce and noise-dominated conditions characteristic of real SPT experiments.

To rigorously evaluate the representations in the data-scarce regime, we curate small experimental subsets that closely mirror the simulated conditions in Sec. IV. For each of the six mesh-size conditions ($\xi \in \{225, 250, 300, 550, 625, 750\}$ nm), we randomly sample 200 trajectories to form a balanced training set, yielding a total of $N_{\text{train}} = 1200$ experimental tracks. This scale aligns with the small-data regime ($N_{\text{train}} = 1000$) investigated in the simulated benchmarks. For performance evaluation, we construct a separate, larger test set by sampling 800 independent trajectories per condition

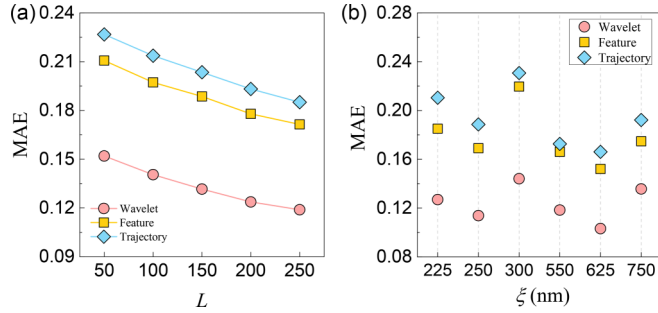


FIG. 6. Results of diffusion-exponent regression on experimental SPT trajectories in F-actin networks with $N_{\text{train}} = 1200$. (a) Test MAE as a function of segment length L for wavelet-based, feature-based, and trajectory-based representations; the wavelet representation consistently yields the lowest error across all lengths. (b) Test MAE resolved by mesh size ξ at $L = 200$, showing that the wavelet representation maintains the smallest error for all network conditions.

(totaling $N_{\text{test}} = 4800$). To assess the robustness of each representation across different trajectory durations, we repeat this sampling procedure for five distinct segment lengths $L \in \{50, 100, 150, 200, 250\}$. The ensuing analysis is organized into two parts. In Sec. V A, we benchmark the three representation families directly on these limited experimental datasets. Subsequently, in Sec. V B, we challenge the prevailing simulation-based paradigm by contrasting the wavelet-based model trained only on the small experimental set against SOTA deep networks trained on massive simulated trajectories.

A. Benchmarking representations via direct learning from limited experimental trajectories

In this subsection, we first benchmark the three representation families on the diffusion-exponent regression and mesh-size classification tasks defined in Sec. III B, using the small experimental training and test sets described above ($N_{\text{train}} = 1200$, $N_{\text{test}} = 4800$).

For the diffusion-exponent regression, Fig. 6(a) reports the MAE as a function of segment length L . While prediction error naturally decreases with L for all methods, the wavelet representation establishes a distinct advantage, achieving the lowest MAE across the entire range. Notably, the feature-based representation, although strong on simulated benchmarks, exhibits a pronounced gap in this experimental setting. This suggests that with only 1200 training samples, handcrafted statistical descriptors suffer from fluctuations inherent to experimental noise, whereas the wavelet transform extracts more robust signatures. The trajectory-based model yields the largest errors, reinforcing the difficulty of training deep sequence models on raw coordinates without large datasets. The robustness of the wavelet representation is further highlighted in Fig. 6(b), which decomposes the MAE by mesh size ξ at $L = 200$. Regression difficulty varies across mesh sizes, and all models show elevated errors at specific conditions (e.g., $\xi = 300$ and 750 nm), yet the wavelet representation consistently attains the lowest MAE. This indicates that the wavelet representation provides a robust basis for

diffusion analysis that is largely insensitive to the specific viscoelastic properties of the environment.

In the mesh-size classification task, discerning the exact environmental condition proves challenging in the small-data regime due to the small variations between mesh sizes (as small as 25 nm) and the presence of experimental noise. Consequently, absolute F1 scores remain moderate for all representations (Fig. 7). Nevertheless, the wavelet representation establishes a clear lead over the alternatives. Figure 7(a) shows that while the wavelet representation's F1 score scales naturally with L , reaching ≈ 0.6 at $L = 250$, the feature-based and trajectory-based representations plateau at lower levels (F1 ≈ 0.40 – 0.45). The confusion matrices for $L = 200$, presented in Figs. 7(b)–7(d), offer physical insight into this disparity.

The experimental system involves beads of diameter $d_{\text{PS}} \approx 550$ nm diffusing in meshes of varying sizes. For the smaller meshes ($\xi \in \{225, 250, 300\}$ nm), where $\xi < d_{\text{PS}}$, the beads are strongly confined. In this regime, all three representations capture the signatures of confinement relatively well, as evidenced by the recognizable diagonal structures in the top-left 3×3 blocks of the confusion matrices. However, the wavelet representation [Fig. 7(b)] yields the sharpest separation with the least leakage between adjacent classes, indicating that it successfully resolves the subtle variations in confinement degree that feature-based and trajectory-based statistics miss. For the larger meshes ($\xi \in \{550, 625, 750\}$ nm), the mesh size becomes comparable to or larger than the bead diameter. In this regime, the bead motion is less strongly constrained and relatively closer to free diffusion, making the trajectories more statistically similar and difficult to differentiate. Reflecting this physical ambiguity, both wavelet-based and feature-based models [Figs. 7(b) and 7(c)] exhibit “blocky” confusion regions among these large-mesh classes, indicating that distinguishing between these subtle variations in mesh size is challenging for any method. Crucially, however, the trajectory-based model [Fig. 7(d)] suffers from a more severe failure mode: It displays significant off-diagonal dispersion that extends beyond the large-mesh block, frequently misclassifying large-mesh trajectories (less confined) as belonging to small-mesh conditions (confined). This contrast highlights that while the large-mesh regime is difficult for all, the wavelet representation remains the most robust, preserving the distinction between confined and less-confined states better than the alternatives.

B. Wavelet representations versus the simulation-based paradigm

As highlighted in Sec. I, the general strategy for the machine-learning analysis of anomalous diffusion relies on the simulation-based paradigm. In this paradigm, ML models are trained on large synthetic datasets and then transferred to experiments. Despite exploiting effectively unlimited labeled trajectories, their performance is ultimately constrained by the simulation-to-reality gap, because idealized diffusion models cannot fully reproduce the complex noise patterns, viscoelastic memory, and transient heterogeneities present in real recordings. Consequently, simply increasing the volume of synthetic training data yields diminishing returns once performance is limited by this distributional mismatch rather than

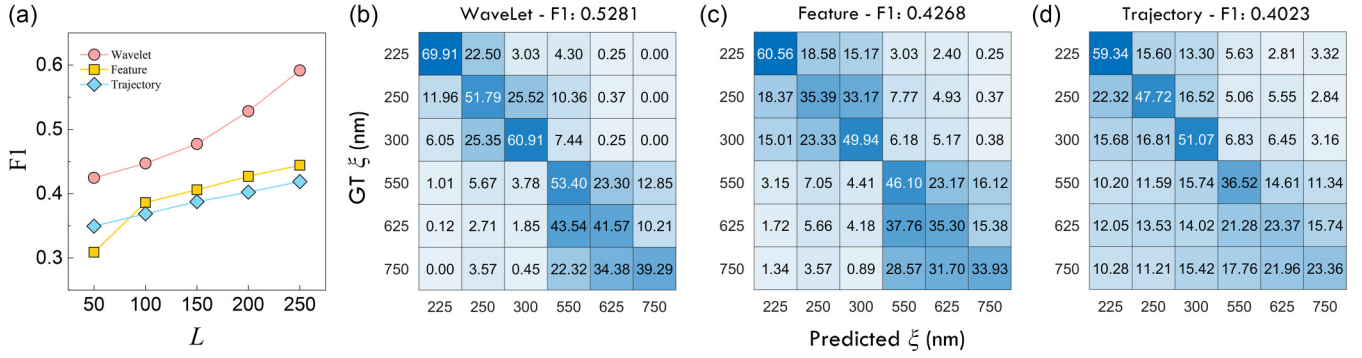


FIG. 7. Mesh-size classification on experimental SPT trajectories in F-actin networks with $N_{\text{train}} = 1200$. (a) Micro-averaged F1 score as a function of segment length L for wavelet-based, feature-based, and trajectory-based representations; the wavelet representation achieves the highest F1 for all L . (b), (d) Confusion matrices for $L = 200$ for the wavelet-based, feature-based, and trajectory-based representations, respectively [rows: ground-truth (GT) mesh size $\xi \in \{225, 250, 300, 550, 625, 750\}$ nm; columns: predicted ξ ; entries in percent].

sample size. In this subsection, we challenge this paradigm by conducting a direct head-to-head comparison. We contrast the performance of the WADNet, a SOTA deep model from the AnDi Challenge, trained on up to 10^6 simulated trajectories, against our wavelet-representation-based model (EffNet), trained directly on a small set of 1200 experimental segments.

To push this simulation-based baseline to its performance limit, we construct two optimized synthetic training sets for the WADNet, each expanded to 10^6 trajectories to represent the saturation level of the paradigm:

(1) Sim-FBM: Consisting exclusively of FBM trajectories. Since the original experimental study [55] identified FBM as the primary diffusion mechanism of the beads, this configuration provides the simulation model with strong physical priors, effectively creating an “idealized” training scenario informed by prior literature.

(2) Sim-All: A balanced mixture of the five standard anomalous diffusion models (ATTM, CTRW, FBM, LW, SBM) to promote generalization and reduce potential model bias.

Crucially, to reduce the domain shift between simulation and reality, we add zero-mean Gaussian noise into these simulated tracks. The noise intensity σ_n is selected by grid search to best match the localization error of the experimental setup (see Appendix F). These massive, noise-matched simulated trajectories are designed to approximate an upper bound on the performance achievable within the simulation-based paradigm.

Before proceeding to the formal comparison, we first assess whether performance has saturated with respect to training data by examining how it scales for the two competing approaches at $L = 200$, as illustrated in Fig. 8(a). For the simulation-based models, increasing the synthetic training set size from 10^4 to 10^6 trajectories yields only marginal improvements on the experimental test set, leading to a clear performance plateau. This observation indicates that the performance of WADNet is limited by a persistent mismatch between simulated and real diffusion dynamics, rather than by a lack of training samples. In contrast, the wavelet representation shows no sign of saturation. As the experimental training set size increases from $N_{\text{train}} = 1200$ to $N_{\text{train}} = 7200$, the prediction error consistently decreases, indicating that the model

has not yet reached its performance ceiling. Consequently, the comparison presented below pits the approximately saturated performance of the simulation-based paradigm against the early-stage, data-constrained performance of the wavelet representation, with the specific comparison regimes highlighted by dashed boxes in Fig. 8(a).

With the saturation behavior established, we now turn to the direct performance comparison. Figure 8(b) presents the diffusion-exponent regression results for the WADNet trained on the large $N_{\text{train}} = 10^6$ simulated datasets versus the experimental models trained on the restricted $N_{\text{train}} = 1200$ subset. The wavelet representation consistently attains the lowest MAE, significantly outperforming both the Sim-FBM and Sim-All baselines across all trajectory lengths. Despite the three orders of magnitude disparity in training data volume, the wavelet representation extracts more relevant physical information from a few thousand real segments than the deep learning baselines learn from millions of imperfectly matched simulated trajectories. This superiority is robustly preserved across diverse physical environments, as shown in the mesh-size decomposition in Fig. 8(c) (for $L = 200$). Regardless of the specific viscoelastic condition, the wavelet representation maintains the lowest error, whereas Sim-FBM and Sim-All consistently yield larger errors by a visible margin. It is worth noting that in the experimental context, the ground-truth value of α is operationally defined by the TA-MSD fit rather than a theoretical generation parameter. Therefore, the superior performance of the wavelet representation reflects, at least in part, its ability to reproduce this established experimental standard, effectively bypassing the simulation-to-reality gap that limits simulation-trained models.

However, attributing this superiority solely to the elimination of the simulation-to-reality gap would be an oversimplification. The performance of feature-based and trajectory-based representations trained on the same 1200 experimental segments provides a critical test of this hypothesis. Theoretically, direct experimental learning benefits from inherent label consistency: The training targets (TA-MSD estimates) share the same distribution, noise characteristics, and estimator biases as the test set, whereas simulation-based training relies on theoretical ground-truth labels that may not perfectly map to empirical estimators. Yet, despite this intrinsic advantage

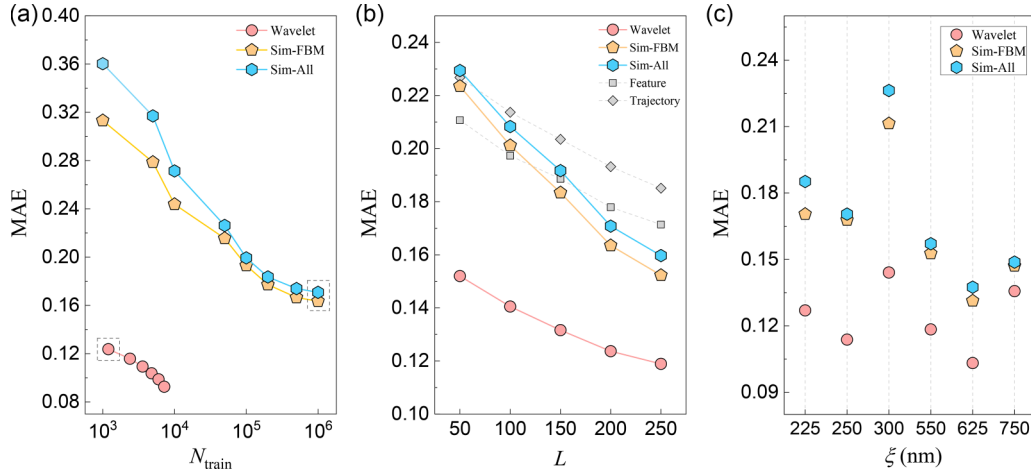


FIG. 8. Wavelet representations vs the simulation-based paradigm on the experimental SPT test set ($N_{\text{test}} = 4800$). (a) MAE for the diffusion-exponent regression at fixed segment length $L = 200$ as a function of training set size N_{train} . Dashed boxes highlight the training regimes used for the comparisons in panels (b) and (c). (b) MAE as a function of segment length L for the wavelet representation, Sim-FBM, Sim-All, feature-based, and trajectory-based models. Across all lengths, the wavelet representation attains the lowest MAE. (c) MAE resolved by mesh size ξ for $L = 200$, where the wavelet model achieves the smallest error for all network conditions.

of in-distribution training, Fig. 8(b) shows that the feature- and trajectory-based models yield error rates on par with the simulation-based WADNet, failing to replicate the substantial gain achieved by the wavelet model. This finding is pivotal: It strongly suggests that the performance leap cannot be attributed solely to in-distribution training on experimental data. Instead, it confirms that the breakthrough relies on the specific synergy between the on-distribution nature of direct experimental learning and the high data efficiency of the wavelet representation. Collectively, these results demonstrate that the wavelet representation constitutes a viable alternative to the simulation-based paradigm, enabling high-precision analysis by learning directly from limited experimental recordings without relying on large simulated datasets.

VI. INTERPRETABILITY: THE PHYSICAL ORIGIN OF DATA EFFICIENCY

The superior data efficiency and generalization capability of the wavelet representation, as demonstrated in the preceding sections, warrant a physical explanation. Why does transforming a trajectory into a time-scale scalogram allow a machine-learning model to learn more effectively from limited data than utilizing the raw trajectory or hand-crafted features? Our central thesis is that the CWT explicitly disentangles the multiscale physical mechanisms driving anomalous diffusion, such as nonstationarity, temporal correlations, and transient heterogeneities. These mechanisms are often hidden in the time domain or averaged out by global statistical descriptors, whereas the wavelet scalogram renders them explicit as distinct geometric patterns in the time-scale plane.

To understand the physical content of the scalograms, we first interpret the continuous wavelet transform in the language of stochastic processes. For a trajectory $x(t)$, the coefficient $W_\psi(a, b)$ at scale a and time b is the inner product between $x(t)$ and a time-localized analyzing wavelet centered at b with an effective support of order a . Its modulus

$|W_\psi(a, b)|$ therefore measures the magnitude of fluctuations of $x(t)$ on timescales of order a in a neighbourhood of b . In this sense, the scalogram $|W_\psi(a, b)|$ can be regarded as a *local multiscale structure function*: At each point (a, b) , it probes the amplitude of fluctuations on a characteristic timescale a .

As a representative example, standard results from wavelet analysis for self-similar processes show that the moments of coefficients scale as $\langle |W_\psi(a, b)|^q \rangle_b \propto a^{\zeta(q)}$ [107]. Specifically, the scaling of the modulus closely tracks the diffusion exponent α [e.g., $\zeta(2) = \alpha + 1$ for FBM]. Thus, the vertical profile of the scalogram encodes information equivalent to the MSD, but resolved locally. Crucially, unlike the global MSD that averages over time, the time localization of the CWT preserves nonstationary features, such as diffusivity changes or trapping events. This capability creates distinct 2D “scale fingerprints” that reveal the underlying physical mechanisms, which we examine in detail below for both simulated diffusion models and experimental SPT trajectories.

A. Scale fingerprints of anomalous diffusion models

To illustrate how these scale fingerprints manifest in practice, we examine the wavelet spectra of the five prototypical models implemented in the andi-datasets library. As our machine-learning pipeline utilizes a stack of six complementary wavelet families, a specific criterion is required to select the most appropriate channel for visual interpretation. Guided by the ablation study in Appendix G, we adopt a best-in-class strategy for this purpose: For each diffusion model, we select the mother wavelet that attains the highest diagonal classification accuracy for that class in the single-wavelet setting. Consequently, the representative scalograms are generated using the morl for ATTM, cgau for CTRW and LW, and mexh for FBM and SBM, which are presented in Fig. 9 (with trajectory length $L = 100$). Although the raw trajectories often appear qualitatively similar, their wavelet spectra display highly structured, model-specific fingerprints driven by

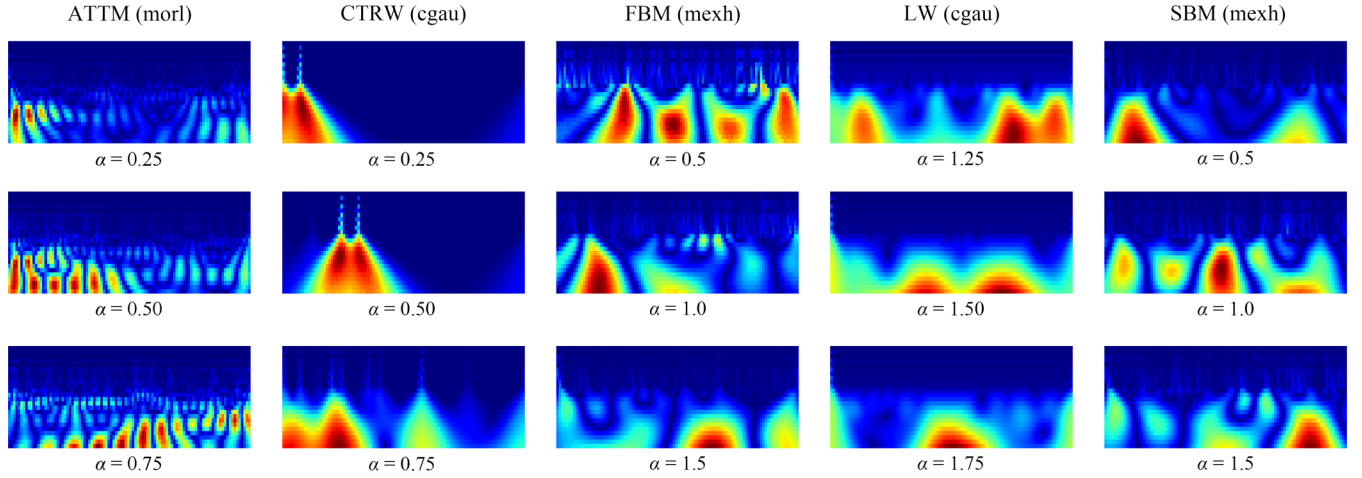


FIG. 9. Representative wavelet scalograms (trajectory length $L = 100$) for five anomalous diffusion models (columns) under varying anomalous exponents α (rows). To visualize the most discriminative features, each model is paired with its optimal mother wavelet (noted in parentheses), selected based on the highest single-wavelet classification accuracy (see Appendix G).

distinct physical mechanisms underlying each type of anomalous diffusion:

(1) ATTM (from diffusivity switching to patchy patterns): In the ATTM model, the particle undergoes Brownian motion with a diffusion coefficient $D(t)$ that resets stochastically. Within a constant- D interval, the process exhibits stationary increments; however, sharp transitions in D cause simultaneous magnitude jumps across all scales. This mechanism manifests in the scalogram as a “patchy” texture (Fig. 9, first column), where rectangular blocks of uniform intensity are separated by vertical edges. For classification, these edges serve as a distinct signature of medium heterogeneity. Regarding parameter estimation, the contrast between these blocks can be qualitatively associated with the heterogeneity of the diffusion process in our simulations; as α increases (from top to bottom), the joint statistics of diffusivities and residence times change so that the texture tends to become more uniform and less segmented. This gradual homogenization allows the vision models to regress α based on the decaying prominence of the patches.

(2) CTRW (from trapping events to vertical gaps): The subdiffusive CTRW is governed by heavy-tailed waiting times $\Psi(t) \sim t^{-\sigma}$, where diffusion exponent $\alpha = \sigma - 1$. During a trapping event, the particle is effectively immobilized, which suppresses wavelet coefficients across all scales. In the scalogram (Fig. 9, second column), these events appear as deep vertical “gaps” or cones of silence. This sparse pattern is the primary cue for diffusion-model classification. Crucially for exponent regression, the statistics of the width and density of these gaps depend on the anomalous exponent α through the waiting-time distribution. As shown in Fig. 9, a lower α (top row, $\alpha = 0.25$) corresponds to a heavier tail in the waiting time distribution, resulting in wider and more prominent gaps compared to the higher α case (bottom row, $\alpha = 0.75$), where the texture becomes denser with more closely spaced activity cones.

(3) FBM (from self-similarity to spectral slopes): FBM is a Gaussian process characterized by long-range correlations defined by the Hurst exponent H . Unlike ATTM or CTRW, FBM scalograms (Fig. 9, third column) are statistically homo-

geneous along the time axis but exhibit a strong dependence on the scale axis. The classification relies on this temporal uniformity. For regression, the model leverages the “spectral slope” along the vertical axis. A low α (subdiffusive, top row) implies negatively correlated increments, concentrating energy at fine scales (high frequencies) and creating a rough texture. Conversely, a high α (superdiffusive, bottom row) implies positive correlations, shifting energy to coarse scales (low frequencies) and resulting in a smoother texture with high-intensity regions at the bottom of the scalogram.

(4) LW (from ballistic flights to coarse-scale ridges): Lévy walks consist of ballistic flights coupled with power-law distributed durations. A constant-velocity flight represents a linear trend, which manifests as an intense, horizontal ridge concentrated at coarse scales (Fig. 9, fourth column). While the presence of these ridges distinguishes LW from the localized features of CTRW, their persistence encodes the diffusion exponent. As α increases toward the ballistic limit (comparing top row $\alpha = 1.25$ to bottom row $\alpha = 1.75$), the flight durations become longer. Consequently, the coarse-scale ridges in the scalogram become more continuous and elongated in time, providing a robust visual feature for estimating α .

(5) SBM (from aging to temporal gradients): For SBM, the diffusivity scales deterministically as $D(t) \sim t^{\alpha-1}$. This aging process breaks time-translation invariance, creating a global gradient in scalogram intensity along the time axis (Fig. 9, fifth column), which serves as the unique identifier for SBM. The direction and steepness of this gradient allow for precise regression of α . For subdiffusive cases ($\alpha < 1$, top row), the diffusivity decays over time, causing the scalogram to fade from bright to dark. In contrast, for superdiffusive cases ($\alpha > 1$, bottom row), the diffusivity increases, resulting in a gradient that brightens from early to late times. This inversion of the temporal gradient provides a clear, monotonic cue for the vision models to determine the anomalous exponent.

Collectively, these analyses make precise what we mean by “scale fingerprints”: Each physical mechanism gives rise to a characteristic arrangement of intensity in the time-scale plane, such as patchy domains from diffusivity switching, sparse gaps from heavy-tailed trapping, vertically

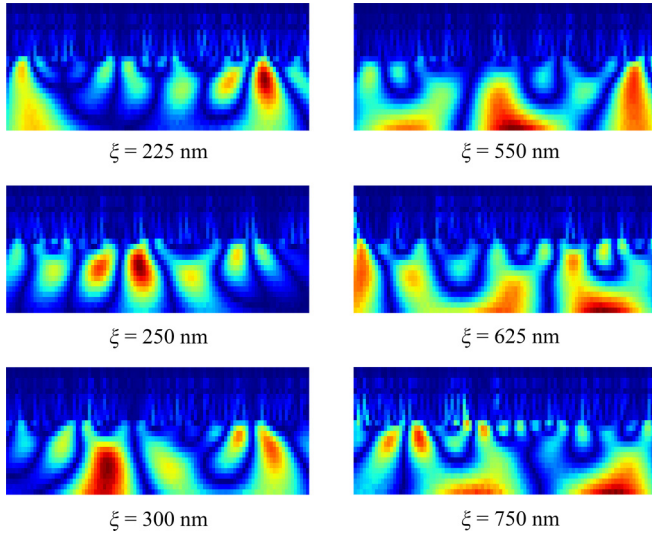


FIG. 10. Evolution of wavelet scalograms for experimental SPT trajectory segments ($L = 100$) in F-actin networks with increasing mesh sizes ξ .

homogeneous spectra with scale-dependent slopes from long-range correlations, coarse-scale ridges from ballistic flights, and temporal gradients from aging. In contrast to raw trajectories, where these mechanisms are often obscured by stochastic fluctuations, the wavelet representation disentangles them into a small set of robust multiscale patterns that are naturally amenable to pattern recognition by vision architectures. These fingerprints simultaneously encode model identity through their qualitative structure and the anomalous exponent through systematic deformations along the columns of Fig. 9. Therefore, the wavelet representation reduces the effective complexity of the learning problem and enables vision models to achieve high accuracy in both diffusion-model classification and diffusion-exponent regression even in the small-data regime. In the next subsection, we show that analogous scale fingerprints also emerge in the experimental SPT trajectories in F-actin networks and that they carry clear signatures of confinement and mesh-size dependence.

B. Experimental validation: fingerprints of confinement in F-actin networks

To validate whether the concept of scale fingerprints extends beyond theoretical prototypes to complex experimental systems, we analyze the SPT trajectories of fluorescent beads diffusing in F-actin networks. In this system, tracer diffusion is modulated by the mesh size ξ of the network, which imposes steric constraints and yields a tunable degree of confinement. As in the previous subsection, we visualize single-channel wavelet spectra in order to highlight qualitative patterns. Given that previous studies [55] have characterized these SPT trajectories as FBM, we employ the mexh to generate the experimental scalograms, as this kernel demonstrates superior efficacy for FBM in our simulated benchmarks (see Appendix G).

Figure 10 presents the wavelet spectra of experimental trajectory segments ($L = 100$) with varying mesh sizes

ranging from $\xi = 225$ nm to $\xi = 750$ nm. A visual inspection reveals that the mesh size acts as a control parameter that systematically alters the texture of the scalograms, mirroring the “rough-to-smooth” transition described in the theoretical FBM analysis (refer to Fig. 9, third column). For dense networks with small mesh sizes $\xi \in \{225, 250, 300\}$ nm where $\xi < d_{ps}$, the scalograms exhibit a “rough” and fragmented texture, particularly at fine scales. Physically, this texture reflects the strong steric confinement imposed by the dense network. Although the underlying mechanism is caging, the resulting rapid directional reversals phenomenologically resemble the antipersistent correlations of subdiffusive FBM. In the wavelet domain, these rapid reversals manifest as high-frequency interruptions that preclude the formation of continuous structures, resulting in a rough, vertically striated pattern. Note that, similar to the simulated FBM, these scalograms are approximately homogeneous along the time axis and do not exhibit pronounced long-time trends, distinguishing them from the transient patchiness of ATTM or the temporal gradients of SBM.

As the mesh size increases to $\xi \in \{550, 625, 750\}$ nm where $\xi \geq d_{ps}$, a morphological transition occurs. The distinct vertical striations gradually merge into smoother, continuous horizontal ridges that span extended temporal durations. At $\xi = 750$ nm, the texture at the coarse scales (bottom of the scalogram) becomes significantly brighter and more coherent. This evolution is consistent with the relaxation of confinement constraints. As the pores enlarge, the bead motion approaches free diffusion, effectively reducing the apparent antipersistence observed at finer scales. This shift effectively redistributes spectral energy from fine to coarse scales, visually reproducing the signature of higher- α regimes.

These findings elucidate the physical origin of the wavelet representation’s data efficiency. By converting subtle temporal correlations into salient, high-contrast visual features, the wavelet representation effectively lowers the learning barrier. The model does not need to implicitly derive the complex statistical laws of viscoelasticity from massive datasets of raw coordinates; instead, it can directly leverage these explicit textural motifs that robustly encode the system’s state. This direct mapping from physical constraints (mesh size) to textural features (roughness versus smoothness) significantly reduces the effective complexity of the learning problem, enabling the vision backbone to generalize accurately even when trained on limited examples.

VII. CONCLUSION

In this work, we have introduced a wavelet-based representation of anomalous diffusion trajectories and demonstrated that, when combined with standard supervised models, it enables data-efficient learning directly from experimental recordings. By mapping trajectories to multichannel wavelet modulus scalograms computed from six complementary continuous wavelet families, we obtain an imagelike, multiscale encoding in which the temporal structure of random motion is rendered explicit in the time-scale plane. On simulated trajectories from the andi-datasets benchmark, models utilizing the wavelet representation maintain superior accuracy in diffusion-exponent regression and model classification with

as few as 1000 training trajectories, while remaining robust to localization noise and variations in trajectory length. Most notably, when applied to experimental SPT data of fluorescent beads in F-actin networks, our approach achieves a decisive advantage. A vision model trained on merely 1200 experimental segments using the wavelet representation yields significantly lower regression errors than SOTA deep learning models trained on 10^6 simulated trajectories. These results challenge the prevailing reliance on massive synthetic datasets as the primary route to high-performance ML for anomalous diffusion, suggesting instead that a physics-informed representation coupled with small-scale, in-distribution experimental data offers a more effective strategy. We note that these conclusions have been validated specifically for diffusion-exponent regression and diffusion-model/mesh-size classification on the datasets considered here. Whether the observed data efficiency extends to substantially different experimental systems (e.g., intracellular transport [126,127], crowded colloidal suspensions [128,129]) or to other analysis tasks remains an open question that warrants future investigation.

This data efficiency of the wavelet representation originates from the ability of the continuous wavelet transform to explicitly disentangle the multiscale mechanisms governing random motion. As our interpretability analysis reveals, complex dynamic behaviors, such as diffusivity switching in ATTM, heavy-tailed trapping in CTRW, or antipersistent correlations in FBM, are rendered into distinct, stationary geometrical patterns or “scale fingerprints” in the time-scale plane. By converting temporal statistical properties into robust visual textures, the wavelet representation reduces the effective complexity of the learning problem, allowing standard vision architectures to identify physical states without implicitly reconstructing statistical laws from raw coordinates. The supporting codes in this work are accessible at our GitHub repository [130].

Looking forward, the wavelet representation opens several promising avenues for future research. One immediate avenue is the application of this framework to trajectory segmentation (or change-point detection) in heterogeneous environments. Given the wavelet transform’s inherent sensitivity to local transient features (manifested as vertical edges or texture shifts in the scalograms), adapting object detection or segmentation networks to the wavelet domain could potentially enable precise detection of state transitions even within extremely short trajectories, a regime where current statistical sliding-window methods [61–70] often struggle. We emphasize, however, that this extension remains prospective; it has not yet been demonstrated and will require dedicated benchmarks on both simulated and experimental segmentation tasks in future work.

A second, perhaps more fundamental direction is to use this representation for the unsupervised discovery of diffusion behaviors. Current ML approaches are largely supervised, limiting analysis to a predefined set of theoretical models (e.g., the five standard models used in this work). However, experimental systems often exhibit complex dynamics that do not fit neatly into these categories. Since the wavelet representation provides a rich and distinct fingerprint for different dynamical universality classes, its combination with unsu-

pervised clustering or manifold learning techniques could, in principle, allow researchers to construct a purely data-driven taxonomy of random motion. While yet to be empirically verified, such an approach holds the potential to enable the identification and characterization of previously unknown diffusion modes directly from experimental data, paving the way for a deeper understanding of transport phenomena in complex biological and soft matter systems.

ACKNOWLEDGMENTS

We thank Wenjie Cai for helpful discussions. This work was supported by the National Natural Science Foundation of China (Grant No. 12104147) and the industry-university cooperation project with DataComp AI (Project No. H202591470640).

DATA AVAILABILITY

The data that support the findings of this article are openly available [130].

APPENDIX A: ABLATION STUDY ON THE WAVELET SCALE SELECTION

In the main text, we fix the set of CWT scales to $\{0.1, 0.2, \dots, 0.9, 1, 2, \dots, 15\}$, which combines a block of subunit scales with a finite range of integer scales. In this Appendix, we provide an ablation study that motivates this choice and, in particular, the upper cutoff at $a = 15$ and the inclusion of the subunit scales $a < 1$.

In this ablation study, 2D simulated trajectories of fixed length $L = 100$ and $L = 500$ generated with the andi-datasets library are considered. As in the main text, we evaluate the wavelet representation on two benchmark tasks: diffusion-exponent regression (MAE) and diffusion-model classification (micro-F1), using a training set of 1000 trajectories together with a test set of size 2×10^5 . We vary the maximum integer scale $a_{\max}$ over the values $a_{\max} \in \{7, 11, 15, 19, 23, 27, 31\}$ and compare three scale configurations:

- (1) Subunit-only scale, $\{0.1, 0.2, \dots, 0.9\}$;
- (2) Integer-only scale, $\{1, 2, \dots, a_{\max}\}$;
- (3) Combined scale, $\{0.1, 0.2, \dots, 0.9, 1, 2, \dots, a_{\max}\}$.

Results of the ablation study are presented in Fig. 11. Here, panels (a) and (b) report the MAE for exponent regression at $L = 100$ and $L = 500$, respectively, as a function of $a_{\max}$; panels (c) and (d) show the corresponding micro-F1 scores for diffusion-model classification. Obviously, for both trajectory lengths and both tasks, the performance curves for both the integer-only and the combined scale configurations exhibit a clear optimum around $a_{\max} = 15$. Increasing $a_{\max}$ beyond this point does not provide systematic improvement and in some cases degrades the performance of wavelet representation, which indicates that very coarse scales contribute little additional useful information while potentially introducing redundancy or edge effects.

In addition, the comparison between integer-only and combined scales shows that augmenting the integer scales with the subunit scales $a \in \{0.1, 0.2, \dots, 0.9\}$ consistently improves performance: Across all tested $a_{\max}$ and both tasks, the

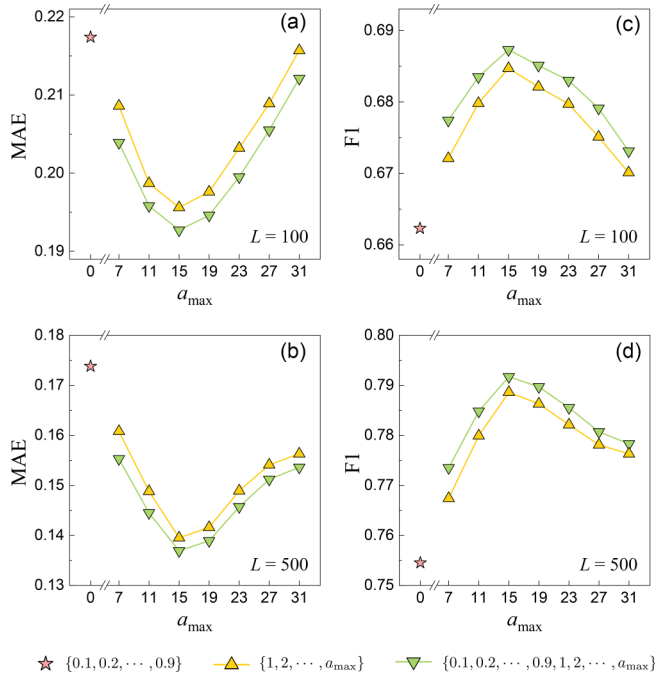


FIG. 11. Ablation study of the CWT scale selection on 2D simulated trajectories. Panels (a) and (b) show the MAE of diffusion-exponent regression for trajectories of length $L = 100$ and $L = 500$, respectively, as a function of the maximum integer scale $a_{\max}$; panels (c) and (d) show the corresponding micro-F1 scores for diffusion-model classification. Both regression and classification performance are optimized around $a_{\max} = 15$, and the combined scale consistently outperforms the integer-only one.

combined-scale representation attains lower MAE and higher F1 than using only integer scales. Moreover, as highlighted by the stars in Fig. 11, the subunit-only scale already achieves nontrivial performance levels in both MAE and F1, which demonstrates that the wavelet coefficients at $a < 1$ indeed encode meaningful short-time structure.

Based on this ablation, we adopt the combined scale grid $\{0.1, 0.2, \dots, 0.9, 1, 2, \dots, 15\}$ as the default in all experiments reported in the main text. This choice balances coverage of short- and long-time behavior, avoids unnecessary proliferation of very coarse scales, and is empirically near optimal for both exponent regression and model classification.

APPENDIX B: COMPUTATIONAL COST ANALYSIS OF THE WAVELET REPRESENTATION

To evaluate the practical overhead of the wavelet representation, we benchmark the end-to-end computational time per trajectory for the three representation families on 2D simulated trajectories across a range of lengths $L \in \{50, 100, \dots, 500\}$. For each method, we separately measure the preprocessing time (transforming a raw trajectory into the model-ready input) and the model inference time. Preprocessing is performed on the CPU (Intel Core i7-10700K); neural network inference (EffNet and Trm) is measured on an NVIDIA TITAN RTX with batch size 32, while LightGBM inference runs natively on CPU. The total time t_c denotes the end-to-end computational time from raw trajectory input to

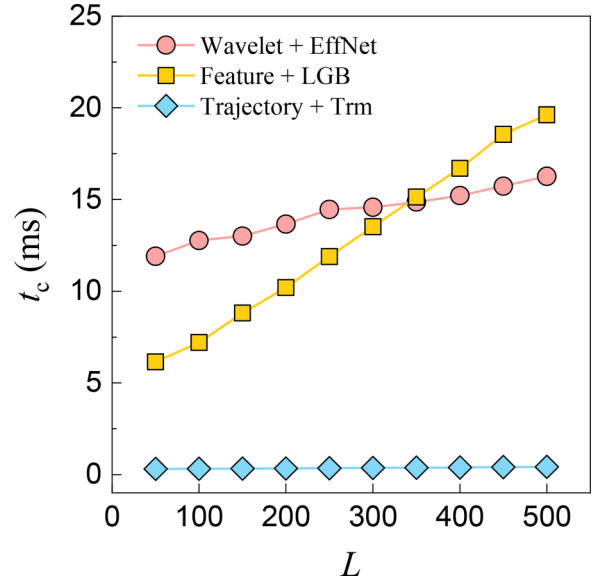


FIG. 12. End-to-end computational time per trajectory (t_c) as a function of trajectory length L for the three representation families. For neural-network-based models (EffNet and Trm), t_c includes CPU preprocessing and GPU-accelerated inference; for the tree-based model (LGB), both preprocessing and inference are performed on CPU. Hardware details are specified in Table III.

final prediction output, measured under each method's standard deployment setting (GPU for neural networks, CPU for tree models). All reported timings are averaged over 3000 independent trajectories.

Table III summarizes the results for $L = 100, 300$, and 500 . The wavelet preprocessing (~ 9 – 14 ms) is dominated by the six CWT computations and the subsequent resampling to 256×256 resolution; this cost grows sublinearly with L because the number of wavelet scales is fixed. In contrast, the feature-based preprocessing (~ 7 – 19 ms) grows more steeply, as several handcrafted descriptors (e.g., p -variation and autocovariance statistics) involve computations that scale superlinearly with trajectory length. The trajectory-based preprocessing is negligible (< 0.3 ms), as it consists only of per-trajectory standardization. Regarding model inference, the EffNet backbone incurs a notable cost on CPU (~ 62 ms), but this is reduced to ~ 2.3 ms with standard GPU acceleration, which is the natural deployment setting for vision models of this scale. We therefore define the total time t_c using GPU inference for neural networks and CPU inference for tree models, reflecting the standard deployment of each method. Figure 12 displays the total time t_c as a function of L for all three methods. The wavelet and feature-based pipelines cross near $L \approx 350$: Below this length the feature-based approach is faster, while above it the wavelet representation becomes more efficient. The trajectory-based pipeline remains the fastest throughout, though this speed advantage must be weighed against its substantially lower data efficiency and accuracy as demonstrated in the main text. For reference, processing the entire experimental test set of $N_{\text{test}} = 4800$ trajectories with the wavelet pipeline takes approximately 1 min on this hardware configuration, confirming the wavelet representation as a practical choice for limited experimental recordings.

TABLE III. Computational time per trajectory for different representations across varying trajectory lengths (L). All preprocessing and CPU inference are performed on an Intel Core i7-10700K; GPU inference is measured on an NVIDIA TITAN RTX with batch size 32.

Method	L	Preprocessing (CPU) (ms)	Inference (CPU) (ms)	Inference (GPU) (ms)	Total time, t_c (ms)
Wavelet + EffNet	100	9.44	61.37	2.34	12.77
	300	11.51	61.86	2.31	14.58
	500	13.65	62.38	2.32	16.27
Feature + LGB	100	6.67	0.46	N/A	7.20
	300	12.88	0.44	N/A	13.52
	500	18.97	0.43	N/A	19.62
Trajectory + Trm	100	0.12	1.49	0.19	0.32
	300	0.15	3.35	0.20	0.37
	500	0.21	6.79	0.21	0.43

APPENDIX C: LENGTH DISTRIBUTIONS OF EXPERIMENTAL SPT TRAJECTORIES

To document the heterogeneity in trajectory length in the experimental SPT dataset, Fig. 13 summarizes the length distributions of trajectories recorded in F-actin networks with mesh sizes $\xi = 225, 250, 300, 550, 625$, and 750 nm. Panels (a)–(c) correspond to the three smaller meshes (225, 250, and 300 nm), where most trajectories contain more than 10^3 time steps and some extend up to several thousand steps. Panels (d)–(f) show the distributions for the larger meshes (550, 625, and 750 nm), where the trajectories are typically shorter,

with lengths concentrated in the range of a few hundred time steps. These unbalanced length distributions motivate the fixed-length segmentation procedure described in Sec. III B when constructing datasets for supervised learning.

APPENDIX D: EFFECT OF PRETRAINED WEIGHTS ON WAVELET-REPRESENTATION-BASED VISION MODELS

To disentangle the contribution of the wavelet representation from that of pretrained vision backbones, we compare

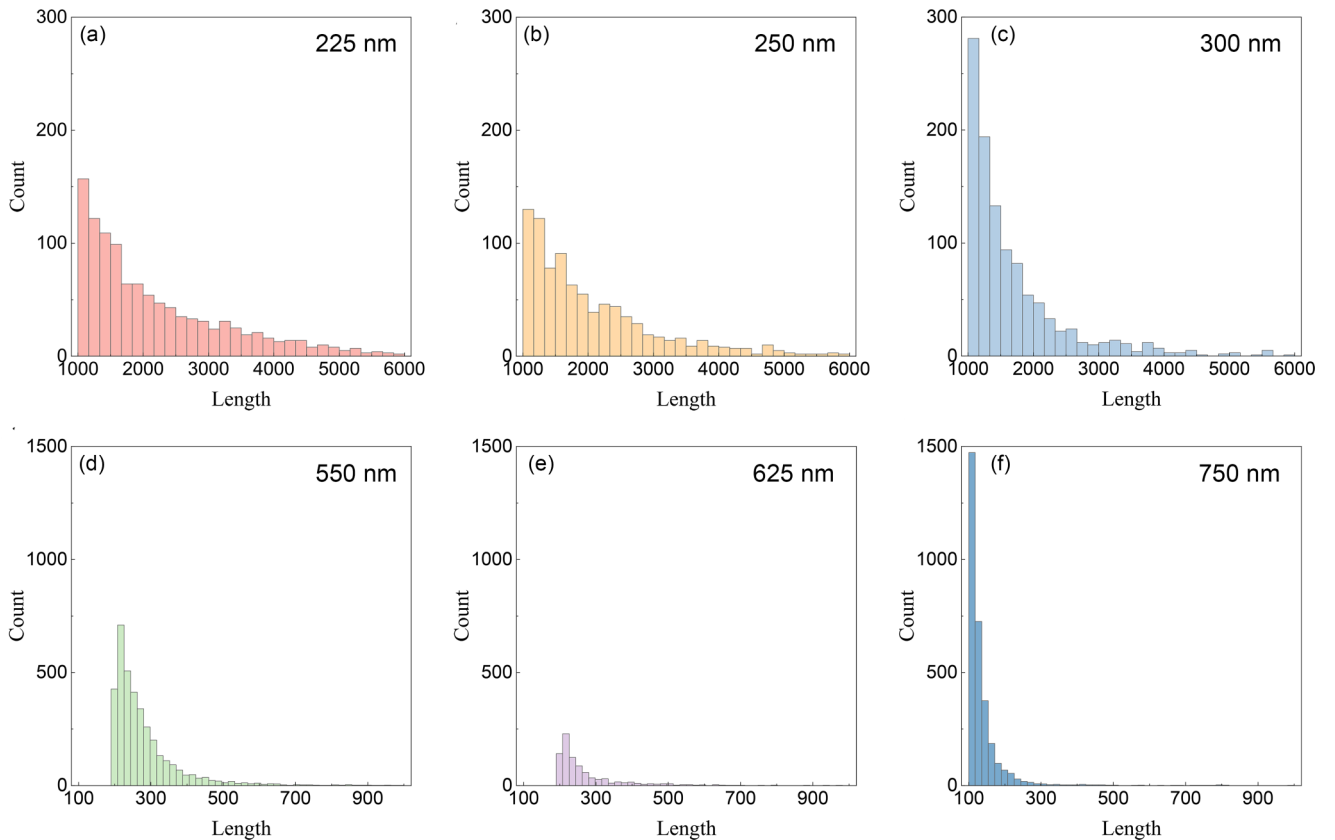


FIG. 13. Length distributions of experimental SPT trajectories in F-actin networks with mesh sizes $\xi = 225$ [panel (a)], 250 [panel (b)], 300 [panel (c)], 550 [panel (d)], 625 [panel (e)], and 750 nm [panel (f)]. For each condition, the histogram shows the number of recorded trajectories as a function of their length (number of time steps).

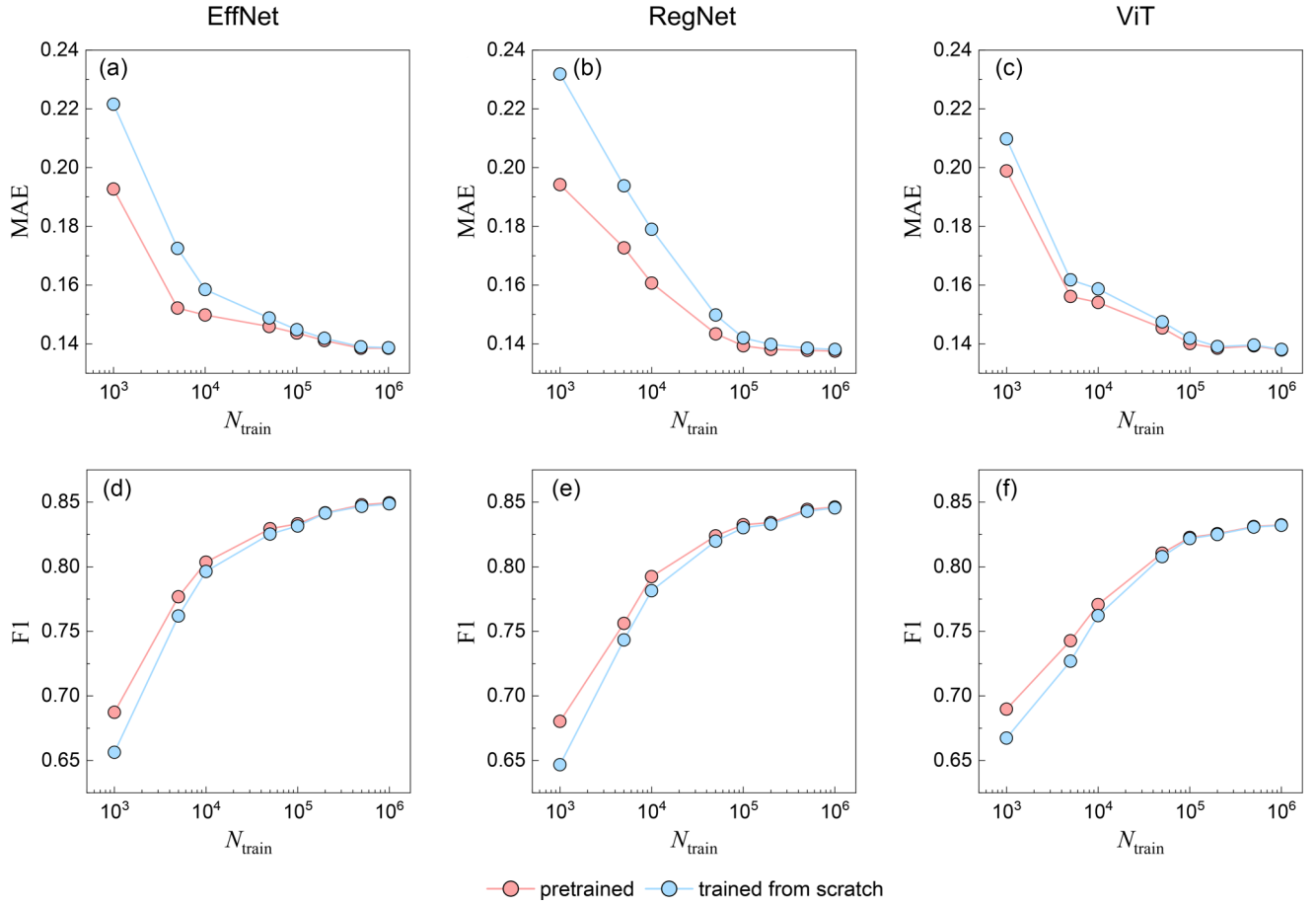


FIG. 14. Effect of ImageNet-pretrained weights on wavelet-representation-based vision models. Top row (a)–(c): MAE for diffusion-exponent regression; bottom row (d)–(f): micro-averaged F1 for diffusion-model classification, both as a function of training set size N_{train} on 2D simulated trajectories of length $L = 100$. Columns correspond to EffNet [panels (a) and (d)], RegNet [panels (b) and (e)], and ViT [panels (c) and (f)].

the performance of three vision architectures (EffNet, RegNet, and ViT) initialized with ImageNet-pretrained weights versus random initialization (trained from scratch), across a range of training set sizes N_{train} on 2D simulated trajectories of length $L = 100$. Figure 14 presents the results for both diffusion-exponent regression (MAE, top row) and diffusion-model classification (F1, bottom row). A consistent trend emerges across all three architectures and both tasks. In the large-data regime ($N_{\text{train}} \geq 10^5$), models trained from scratch converge to essentially the same performance as their pretrained counterparts, with negligible differences in both MAE and F1. This convergence demonstrates that the high asymptotic accuracy is an intrinsic property of the wavelet representation: Given sufficient training data, the vision models learn equally effective features from the scalograms regardless of their initialization, confirming that the multiscale structure exposed by the wavelet transform is the primary driver of performance.

In the small-data regime ($N_{\text{train}} \leq 10^4$), pretrained models consistently yield lower MAE and higher F1 than their from-scratch counterparts. This advantage reflects the benefit of general-purpose visual features (e.g., edge and texture detectors) learned from natural images, which provide a favorable initialization for extracting patterns from wavelet scalograms when task-specific training data are scarce. Notably, this

transfer-learning benefit is consistent across all three architectures, suggesting that it arises from a generic compatibility between scalogram textures and low-level features learned on natural images rather than from any architecture-specific property. Since pretrained weights are readily available and straightforward to deploy, this moderate small-data boost can be obtained at no additional cost in practice. Taken together, these results confirm that pretraining serves as a practical accelerator that enhances sample efficiency when training data are limited, while the wavelet representation itself determines the ultimate performance ceiling.

APPENDIX E: PERFORMANCE ANALYSIS ON SIMULATED TRAJECTORIES ACROSS DIFFUSION MODELS

To assess the universality of the wavelet representation, we further dissect the performance of the three representation families across five specific diffusion models in the and-datasets library: ATTM, CTRW, FBM, LW, and SBM. The results for 2D simulated trajectories of length $L = 100$ in the small-data regime ($N_{\text{train}} = 1000$) are summarized in Fig. 15. For the diffusion-exponent regression task [Fig. 15(a)], a consistent hierarchy is observed irrespective of the underlying

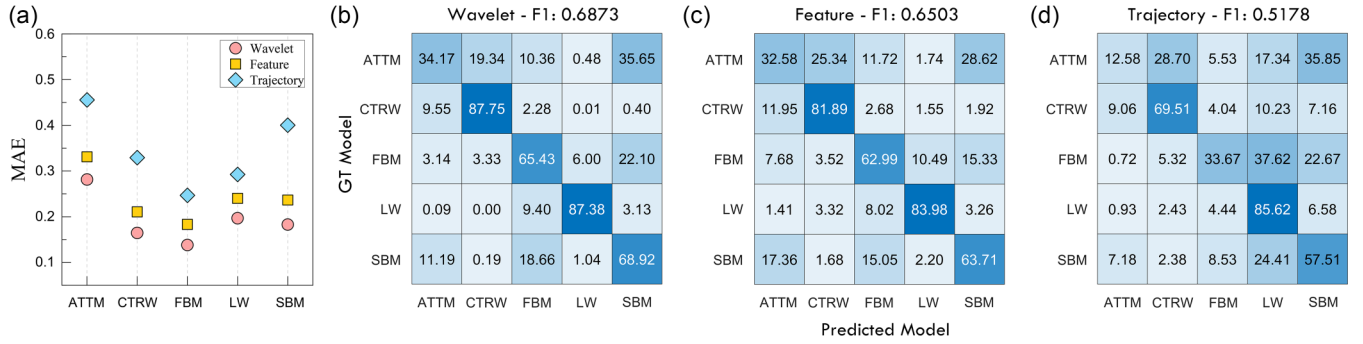


FIG. 15. Model-wise performance comparison on 2D simulated trajectories of length $L = 100$ in the small-data regime ($N_{\text{train}} = 1000$). (a) MAE for diffusion-exponent regression across five specific models (ATTM, CTRW, FBM, LW, SBM). The wavelet representation consistently achieves the lowest error compared to feature-based and trajectory-based ones. (b), (d) Confusion matrices for diffusion-model classification using wavelet-based, feature-based, and trajectory-based representations, respectively. Rows represent GT labels, and columns represent predictions; values indicate the percentage of correct classifications (diagonal) and misclassifications (off-diagonal). The wavelet representation shows the strongest diagonal dominance and highest overall F1 score.

physical mechanism: The wavelet representation achieves the lowest regression error for every model. Notably, the wavelet advantage is preserved across diverse dynamical regimes, ranging from nonergodic processes (e.g., CTRW, ATTM) to ergodic Gaussian processes with long-range correlations (e.g., FBM), demonstrating that its efficacy is not tied to a specific type of anomalous diffusion.

For the diffusion-model classification task, the results for the three representations are shown as confusion matrices in Figs. 15(b)–15(d). The wavelet representation [Fig. 15(b)] exhibits the strongest diagonal dominance, yielding the highest overall micro-F1 score (0.6873). Misclassifications are largely confined to physically ambiguous pairs, such as the confusion between ATTM and SBM models, which is a known challenge for short trajectories due to their similar statistical signatures. The feature-based representation [Fig. 15(c)] shows a qualitatively similar confusion pattern but with systematically lower diagonal accuracy (F1: 0.6503). In sharp contrast, the trajectory-based representation [Fig. 15(d)] suffers from significant off-diagonal dispersion (F1: 0.5178). It struggles particularly with differentiating FBM from LW and SBM, as evidenced by the high misclassification rates in the corresponding rows.

Taken together, these model-specific analyses confirm that the superior performance of the wavelet representation is robust and universal. It effectively captures the distinct multiscale features of each diffusion mechanism, thereby outperforming feature-based and trajectory-based baselines in both regression and classification tasks.

APPENDIX F: OPTIMIZATION OF LOCALIZATION NOISE INTENSITY FOR SIMULATION-BASED MODELS

To reduce the distributional mismatch between the simulated training data and the experimental recordings, we

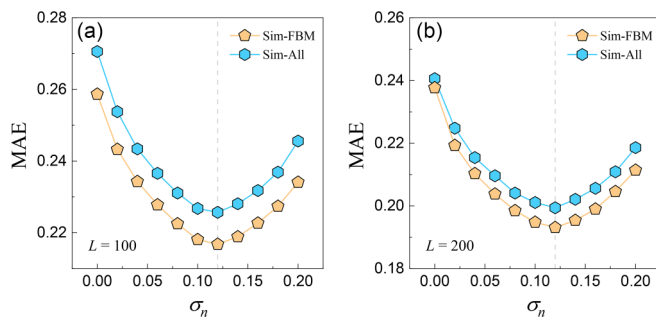


FIG. 16. MAE on the experimental test set with segments of length $L = 100$ [panel (a)] and $L = 200$ [panel (b)], as a function of the noise standard deviation σ_n . In both cases, the MAE displays a U-shaped dependence on σ_n with a consistent minimum at $\sigma_n = 0.12$ (vertical dashed line).

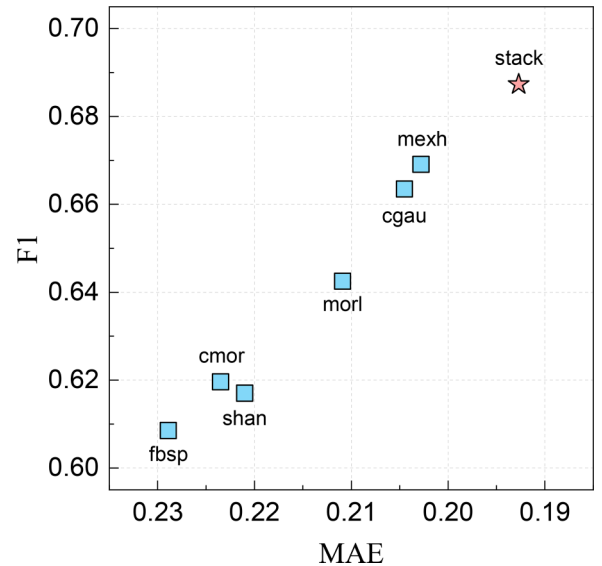


FIG. 17. Performance comparison between the full multichannel wavelet representation (“stack”) and single-wavelet representations derived from individual wavelet families on the test set. The x axis represents the MAE for diffusion-exponent regression, and the y axis represents the micro-averaged F1 score for diffusion-model classification. The stacked representation achieves the best overall performance, demonstrating the benefit of ensemble learning from complementary wavelet bases.

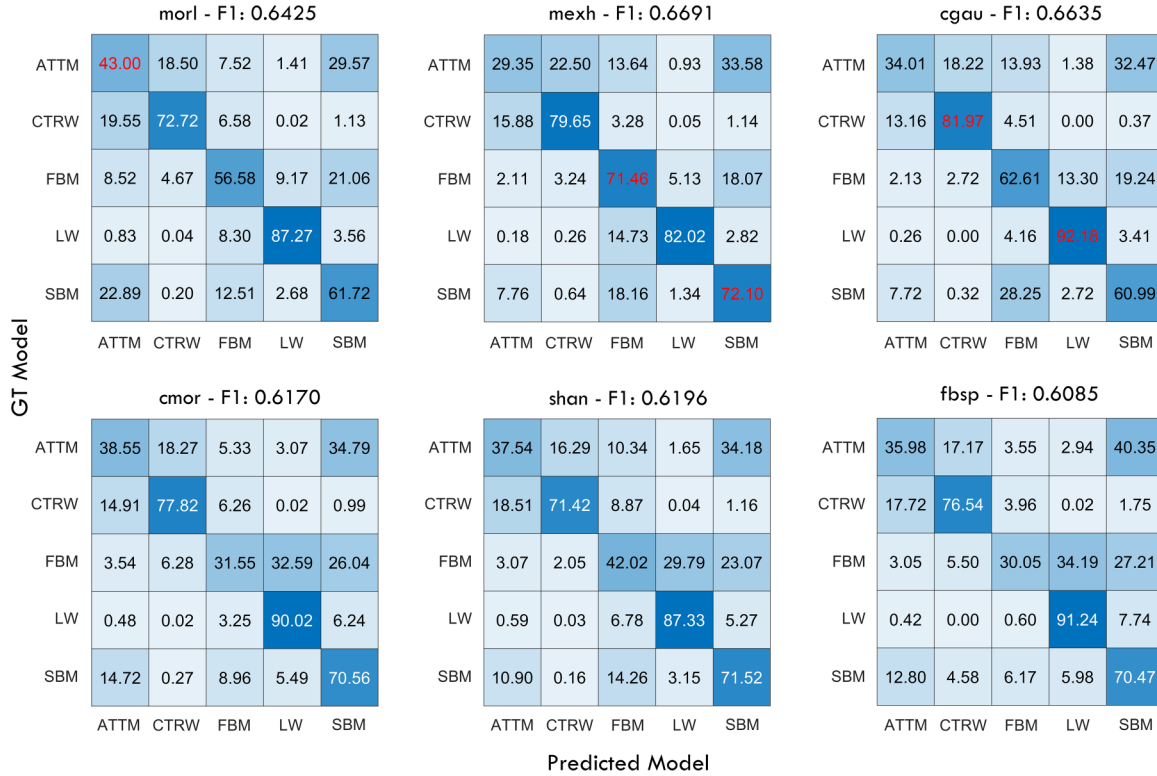


FIG. 18. Confusion matrices for diffusion-model classification using single-wavelet representations. Each panel corresponds to a model trained exclusively on scalograms from one wavelet family. The overall micro-F1 score is indicated in the title of each matrix. Red numbers highlight the highest diagonal accuracy achieved for each specific diffusion model across the different wavelet families. These maxima guide the selection of representative scalograms for visualization: morl is selected for ATTM, cgau for CTRW and LW, and mexh for FBM and SBM.

explicitly add Gaussian localization noise into the simulation-based training pipeline. The noise intensity, parametrized by the standard deviation σ_n , is determined via a systematic grid search designed to minimize the MAE of the simulation-based models on the experimental test set. We construct a series of synthetic training sets for both the Sim-FBM and Sim-All configurations, each containing 10^5 trajectories. For each configuration, we generate distinct versions by adding zero-mean Gaussian noise with standard deviations $\sigma_n \in \{0.00, 0.02, \dots, 0.20\}$. We then train the WADNet model on each noise-augmented dataset and evaluate its diffusion-exponent regression performance (MAE) on the full experimental test set. This procedure is repeated for trajectory lengths $L = 100$ and $L = 200$ to ensure the robustness of the optimal parameter choice across different timescales.

Figure 16 summarizes the results of this optimization. For both $L = 100$ [Fig. 16(a)] and $L = 200$ [Fig. 16(b)], the MAE exhibits a clear U-shaped dependence on σ_n . In the absence of noise ($\sigma_n = 0$), the models suffer from large errors due to the lack of resilience against experimental imperfections. As noise is introduced, performance improves significantly, reaching a distinct minimum before degrading again as the noise overwhelms the signal. Crucially, for both the Sim-FBM and Sim-All models, and across both trajectory lengths, the optimal performance is consistently achieved at $\sigma_n = 0.12$ (highlighted by dashed lines). Consequently, we fix this optimal noise level ($\sigma_n = 0.12$) for all large-scale

($N_{\text{train}} = 10^6$) simulation-based training described in the main text.

APPENDIX G: ABLATION STUDY ON WAVELET FAMILIES AND SELECTION OF REPRESENTATIVE SCALOGRAMS

The wavelet representation proposed in this work is constructed by stacking the modulus scalograms from six distinct continuous wavelet families: real Morlet (morl), Mexican hat (mexh), complex Gaussian derivative (cgau), complex Morlet (cmor), Shannon (shan), and frequency B-spline (fbsp). The downstream vision models are always trained on this six-wavelet representation in order to exploit the complementary information provided by the different wavelets. For the interpretability analysis in Sec. VI, however, we need to display individual scalograms when discussing scale fingerprints. This necessitates a criterion for selecting the most representative wavelet family for each diffusion model. In this Appendix, we analyze the contribution of individual wavelet families to justify the multichannel design and to guide the selection of these representative visualizations.

To assess the necessity of combining multiple wavelets, we first quantify the contribution of each wavelet family in isolation. Using 2D simulated trajectories with $N_{\text{train}} = 1000$ and $L = 100$, we train separate EffNet models on single-wavelet representations, where the scalogram of only one mother wavelet is provided as input. Figure 17 compares the

test performance of these single-wavelet models with that of the full six-wavelet “stack” model. The stacked representation (marked by the star) consistently outperforms any single wavelet family in both diffusion-exponent regression (lowest MAE) and diffusion-model classification (highest F1), thereby validating the multichannel stacking strategy employed in the main text.

For the purpose of visual interpretation, we then identify, for each diffusion model, the wavelet family that performs best when used alone. To this end, we analyze the class-wise performance using the confusion matrices of the single-wavelet classifiers, as shown in Fig. 18. We adopt a best-in-class selection strategy: For each diffusion model, we select the mother wavelet that achieves the highest diagonal accuracy. The corresponding entries in Fig. 18 are highlighted in red. This strategy yields the following choices:

(1) ATTM: morl attains the highest diagonal accuracy 43.00% for ATTM.

(2) CTRW and LW: cgau achieves the best performance for both CTRW and LW, with diagonal accuracies of 81.97% and 92.18%, respectively.

(3) FBM and SBM: mexh gives the highest diagonal accuracy for FBM and SBM, with values of 71.46% and 72.10%, respectively.

These wavelet families are therefore used to generate the representative scalograms in Fig. 9 in Sec. VI: morl for ATTM, cgau for CTRW and LW, and mexh for FBM and SBM. We stress that this choice is pragmatic and guided by the single-wavelet classification performance. The full six-wavelet representation remains superior for all quantitative tasks, but the selected wavelets provide clear and model-specific scale fingerprints that are well suited for qualitative interpretation.

-
- [1] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai, Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, *Phys. Chem. Chem. Phys.* **16**, 24128 (2014).
 - [2] J. Klafter and I. M. Sokolov, Anomalous diffusion spreads its wings, *Phys. World* **18**, 29 (2005).
 - [3] C. Manzo, G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, M. Lewenstein, and R. Metzler, Preface: Characterisation of physical processes from anomalous diffusion data, *J. Phys. A: Math. Theor.* **56**, 010401 (2023).
 - [4] B. Wang, J. Kuo, S. C. Bae, and S. Granick, When Brownian diffusion is not Gaussian, *Nat. Mater.* **11**, 481 (2012).
 - [5] R. Illukkumbura, T. Bland, and N. W. Goehring, Patterning and polarization of cells by intracellular flows, *Curr. Opin. Cell Biol.* **62**, 123 (2020).
 - [6] M. Woring, I. Izeddin, C. Favard, and H. Berry, Anomalous subdiffusion in living cells: Bridging the gap between experiments and realistic models through collaborative challenges, *Front. Phys.* **8**, 134 (2020).
 - [7] J. Miné-Hattab, V. Recamier, I. Izeddin, R. Rothstein, and X. Darzacq, Multi-scale tracking reveals scale-dependent chromatin dynamics after DNA damage, *Mol. Biol. Cell* **28**, 3323 (2017).
 - [8] S. Sadegh, J. L. Higgins, P. C. Mannion, M. M. Tamkun, and D. Krapf, Plasma membrane is compartmentalized by a self-similar cortical actin meshwork, *Phys. Rev. X* **7**, 011031 (2017).
 - [9] T. Skóra, F. Vaghefikia, J. Fitter, and S. Kondrat, Macromolecular crowding: How shape and interactions affect diffusion, *J. Phys. Chem. B* **124**, 7537 (2020).
 - [10] B. M. Regner, D. Vučinić, C. Domnisoru, T. M. Bartol, M. W. Hetzer, D. M. Tartakovsky, and T. J. Sejnowski, Anomalous diffusion of single particles in cytoplasm, *Biophys. J.* **104**, 1652 (2013).
 - [11] P. G. Meyer and R. Metzler, Time scales in the dynamics of political opinions and the voter model, *New J. Phys.* **26**, 023040 (2024).
 - [12] A. Dechant, F. Kindermann, A. Widera, and E. Lutz, Continuous-time random walk for a particle in a periodic potential, *Phys. Rev. Lett.* **123**, 070602 (2019).
 - [13] T. Orré, A. Joly, Z. Karatas, B. Kastberger, C. Cabriel, R. T. Böttcher, S. Lévêque-Fort, J.-B. Sibarita, R. Fässler, B. Wehrle-Haller, *et al.*, Molecular motion and tridimensional nanoscale localization of kindlin control integrin activation in focal adhesions, *Nat. Commun.* **12**, 3104 (2021).
 - [14] X. Dai, X. Zhang, L. Gao, Z. Xu, and L.-T. Yan, Topology mediates transport of nanoparticles in macromolecular networks, *Nat. Commun.* **13**, 4094 (2022).
 - [15] L. Dabelow, S. Bo, and R. Eichhorn, Irreversibility in active matter systems: Fluctuation theorem and mutual information, *Phys. Rev. X* **9**, 021009 (2019).
 - [16] T. G. Mason and D. A. Weitz, Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids, *Phys. Rev. Lett.* **74**, 1250 (1995).
 - [17] F. W. Tavares, P. M. Ndiaye, E. K. Lenzi, L. R. Evangelista, H. V. Ribeiro, and R. S. Zola, Anomalous diffusion and sorption-desorption process in complex fluid systems, *Commun. Nonlinear Sci.* **90**, 105411 (2020).
 - [18] Y. Nakayama, Y. Matsuoka, and T. Kajiwara, Simulating Brownian motion in thermally fluctuating viscoelastic fluids by using the smoothed profile method, *J. Comput. Phys.* **509**, 113035 (2024).
 - [19] S. Granick, Motions and relaxations of confined liquids, *Science* **253**, 1374 (1991).
 - [20] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix, and H. E. Stanley, Economic fluctuations and anomalous diffusion, *Phys. Rev. E* **62**, R3023 (2000).
 - [21] J. Masoliver, M. Montero, and G. H. Weiss, Continuous-time random-walk model for financial distributions, *Phys. Rev. E* **67**, 021112 (2003).
 - [22] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, and D. Sornette, Multifractal analysis of financial markets: A review, *Rep. Prog. Phys.* **82**, 125901 (2019).
 - [23] V. Sposini, D. Krapf, E. Marinari, R. Sunyer, F. Ritort, F. Taheri, C. Selhuber-Unkel, R. Benelli, M. Weiss, R. Metzler, and G. Oshanin, Towards a robust criterion of anomalous diffusion, *Commun. Phys.* **5**, 305 (2022).
 - [24] R. Metzler, Brownian motion and beyond: First-passage, power spectrum, non-Gaussianity, and anomalous diffusion, *J. Stat. Mech.* (2019) 114003.

- [25] O. Vilk, E. Aghion, R. Nathan, S. Toledo, R. Metzler, and M. Assaf, Classification of anomalous diffusion in animal movement data using power spectral analysis, *J. Phys. A: Math. Theor.* **55**, 334004 (2022).
- [26] D. Krapf, N. Lukat, E. Marinari, R. Metzler, G. Oshanin, C. Selhuber-Unkel, A. Squarcini, L. Stadler, M. Weiss, and X. Xu, Spectral content of a single non-Brownian trajectory, *Phys. Rev. X* **9**, 011019 (2019).
- [27] S. Burov, J.-H. Jeon, R. Metzler, and E. Barkai, Single particle tracking in systems showing anomalous diffusion: The role of weak ergodicity breaking, *Phys. Chem. Chem. Phys.* **13**, 1800 (2011).
- [28] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy, and R. Metzler, Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: Maximum-likelihood model selection applied to stochastic-diffusivity data, *Phys. Chem. Chem. Phys.* **20**, 29018 (2018).
- [29] S. Park, S. Thapa, Y. Kim, M. A. Lomholt, and J.-H. Jeon, Bayesian inference of Lévy walks via hidden Markov models, *J. Phys. A: Math. Theor.* **54**, 484001 (2021).
- [30] S. Thapa, S. Park, Y. Kim, J.-H. Jeon, R. Metzler, and M. A. Lomholt, Bayesian inference of scaled versus fractional Brownian motion, *J. Phys. A: Math. Theor.* **55**, 194003 (2022).
- [31] J. Krog and M. A. Lomholt, Bayesian inference with information content model check for Langevin equations, *Phys. Rev. E* **96**, 062106 (2017).
- [32] J. Krog, L. H. Jacobsen, F. W. Lund, D. Wüstner, and M. A. Lomholt, Bayesian model selection with fractional Brownian motion, *J. Stat. Mech.* (2018) 093501.
- [33] M. Tan and Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97 (PMLR, Long Beach, CA, 2019), pp. 6105–6114.
- [34] P. Dollár, M. Singh, and R. Girshick, Fast and accurate model scaling, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2021), pp. 924–932.
- [35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16×16 words: Transformers for image recognition at scale, in *International Conference on Learning Representations (ICLR)* (OpenReview.net, Amherst, MA, 2021).
- [36] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, in *31st Conference on Neural Information Processing Systems (NIPS)*, California, 2017).
- [37] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, 2016).
- [38] L. Breiman, Random forests, *Mach. Learn.* **45**, 5 (2001).
- [39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems* (Curran Associates Inc., Red Hook, 2017), pp. 5998–6008.
- [40] F. Galton, Regression towards mediocrity in hereditary stature, *J. Anthropol. Inst. Great Brit. Ireland* **15**, 246 (1886).
- [41] D. R. Cox, The regression analysis of binary sequences, *J. R. Stat. Soc.: Series B Stat. Method.* **20**, 215 (1958).
- [42] A. Graves, Long short-term memory, in *Supervised Sequence Labelling with Recurrent Neural Networks*, Studies in Computational Intelligence, Vol. 385 (Springer, Berlin, Heidelberg, 2012), pp. 37–45.
- [43] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, Doha, Qatar, 2014), pp. 1724–1734.
- [44] N. Korabel and T. A. Waigh, Deep learning for heterogeneous anomalous dynamics in cellular and molecular biology, *Cell Rep. Phys. Sci.* **6**, 102891 (2025).
- [45] G. Muñoz-Gil, M. A. Garcia-March, C. Manzo, J. D. Martín-Guerrero, and M. Lewenstein, Single trajectory characterization via machine learning, *New J. Phys.* **22**, 013010 (2020).
- [46] D. Li, Q. Yao, and Z. Huang, WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet), *J. Phys. A: Math. Theor.* **54**, 404003 (2021).
- [47] H. Verdier, M. Duval, F. Laurent, A. Cassé, C. L. Vestergaard, and J.-B. Masson, Learning physical properties of anomalous random walks using graph neural networks, *J. Phys. A: Math. Theor.* **54**, 234001 (2021).
- [48] H. Verdier, F. Laurent, A. Cassé, C. L. Vestergaard, and J.-B. Masson, Variational inference of fractional Brownian motion with linear computational complexity, *Phys. Rev. E* **106**, 055311 (2022).
- [49] H. Seckler and R. Metzler, Bayesian deep learning for error estimation in the analysis of anomalous diffusion, *Nat. Commun.* **13**, 6717 (2022).
- [50] Y. Ahsini, B. Reverte, and J. A. Conejero, AI-driven consensus: Modeling multi-agent networks with long-range interactions through path-Laplacian matrices, *Appl. Sci.* **15**, 5064 (2025).
- [51] R. Haidari and A. N. Kapanidis, Pointwise prediction of protein diffusive properties using machine learning, *J. Phys. Photonics* **7**, 035025 (2025).
- [52] J. Park, N. Sokolovska, C. Cabriel, I. Izeddin, and J. Miné-Hattab, Bottom-up iterative anomalous diffusion detector (BI-ADD), *J. Phys. Photonics* **7**, 045027 (2025).
- [53] P. Kowalek, H. Loch-Olszewska, Ł. Łaszczuk, J. Opała, and J. Szwabiński, Boosting the performance of anomalous diffusion classifiers with the proper choice of features, *J. Phys. A: Math. Theor.* **55**, 244005 (2022).
- [54] H. Loch-Olszewska and J. Szwabiński, Impact of feature choice on machine learning classification of fractional anomalous diffusion, *Entropy* **22**, 1436 (2020).
- [55] N. Granik, L. E. Weiss, E. Nehme, M. Levin, M. Chein, E. Perlson, Y. Roichman, and Y. Shechtman, Single-particle diffusion characterization by deep learning, *Biophys. J.* **117**, 185 (2019).
- [56] E. A. AL-hada, X. Tang, and W. Deng, Classification of stochastic processes by convolutional neural networks, *J. Phys. A: Math. Theor.* **55**, 274006 (2022).
- [57] N. Firbas, Óscar Garibó-i Orts, M. Ángel Garcia-March, and J. A. Conejero, Characterization of anomalous diffusion

- through convolutional transformers, *J. Phys. A: Math. Theor.* **56**, 014001 (2023).
- [58] DeepSPT, GitHub (2020), https://github.com/AnDiChallenge/AnDi2020_TeamD_DeepSPT.
- [59] NOA, GitHub (2020), https://github.com/AnDiChallenge/AnDi2020_TeamH_NOA.
- [60] H. D. Pinholt, S. S.-R. Bohr, J. F. Iversen, W. Boomsma, and N. S. Hatzakis, Single-particle diffusional fingerprinting: A machine-learning framework for quantitative analysis of heterogeneous diffusion, *Proc. Natl. Acad. Sci. USA* **118**, e2104624118 (2021).
- [61] C. Manzo, Extreme learning machine for the characterization of anomalous diffusion from single trajectories (AnDi-ELM), *J. Phys. A: Math. Theor.* **54**, 334002 (2021).
- [62] S. Bo, F. Schmidt, R. Eichhorn, and G. Volpe, Measurement of anomalous diffusion using recurrent neural networks, *Phys. Rev. E* **100**, 010102 (2019).
- [63] A. Argun, G. Volpe, and S. Bo, Classification, inference and segmentation of anomalous diffusion with recurrent neural networks, *J. Phys. A: Math. Theor.* **54**, 294003 (2021).
- [64] P. Dosset, P. Rassam, L. Fernandez, C. Espenel, E. Rubinstein, E. Margeat, and P.-E. Milhiet, Automatic detection of diffusion modes within biological membranes using back-propagation neural network, *BMC Bioinf.* **17**, 197 (2016).
- [65] Y. Zhang, F. Ge, X. Lin, J. Xue, Y. Song, H. Xie, and Y. He, Extract latent features of single-particle trajectories with historical experience learning, *Biophys. J.* **122**, 4451 (2023).
- [66] T. Wagner, A. Kroll, C. R. Haramagatti, H.-G. Lipinski, and M. Wiemann, Classification and segmentation of nanoparticle diffusion trajectories in cellular micro environments, *PLoS One* **12**, e0170165 (2017).
- [67] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, and M. Bathe, Inferring transient particle transport dynamics in live cells, *Nat. Methods* **12**, 838 (2015).
- [68] Y. Matsuda, I. Hanasaki, R. Iwao, H. Yamaguchi, and T. Niimi, Estimation of diffusive states from single-particle trajectory in heterogeneous medium using machine-learning methods, *Phys. Chem. Chem. Phys.* **20**, 24099 (2018).
- [69] A. Gentili and G. Volpe, Characterization of anomalous diffusion classical statistics powered by deep learning (CONDER), *J. Phys. A: Math. Theor.* **54**, 314003 (2021).
- [70] M. Arts, I. Smal, M. W. Paul, C. Wyman, and E. Meijering, Particle mobility analysis using deep learning and the moment scaling spectrum, *Sci. Rep.* **9**, 17160 (2019).
- [71] J. Pineda, B. Midtvedt, H. Bachimanchi, S. Noé, D. Midtvedt, G. Volpe, and C. Manzo, Geometric deep learning reveals the spatiotemporal features of microscopic motion, *Nat. Mach. Intell.* **5**, 71 (2023).
- [72] X. Qu, Y. Hu, W. Cai, Y. Xu, H. Ke, G. Zhu, and Z. Huang, Semantic segmentation of anomalous diffusion using deep convolutional networks, *Phys. Rev. Res.* **6**, 013054 (2024).
- [73] B. Requena, S. Masó, J. Bertran, M. Lewenstein, C. Manzo, and G. Muñoz-Gil, Inferring pointwise diffusion properties of single trajectories with deep learning, *Biophys. J.* **122**, 4360 (2023).
- [74] Q. Martinez, C. Chen, J. Xia, and H. Bahai, Sequence-to-sequence change-point detection in single-particle trajectories via recurrent neural network for measuring self-diffusion, *Transp. Porous Media* **147**, 679 (2023).
- [75] H. Seckler and R. Metzler, Change-point detection in anomalous-diffusion trajectories utilising machine-learning-based uncertainty estimates, *J. Phys. Photonics* **6**, 045025 (2024).
- [76] Z. Chen, L. Geffroy, and J. S. Biteen, NOBIAS: Analyzing anomalous diffusion in single-molecule tracks with non-parametric Bayesian inference, *Front. Bioinform.* **1**, 742073 (2021).
- [77] H. Kabbech and I. Smal, Identification of diffusive states in tracking applications using unsupervised deep learning methods, in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (IEEE, Piscataway, NJ, 2022), pp. 1–4.
- [78] H. Kabbech and I. Smal, TrackSegNet: A tool for trajectory segmentation into diffusive states using supervised deep learning, *J. Open Source Softw.* **9**, 6157 (2024).
- [79] J. Bae and H. Jeong, Segment-based task-specific deep learning framework for detecting state transitions in anomalous diffusion trajectories, *J. Phys. Photonics* **7**, 025027 (2025).
- [80] Y. Ahsini, M. Escoto, and J. A. Conejero, AnomalousNet: A hybrid approach with attention U-Nets and change point detection for accurate characterization of anomalous diffusion in video data, *J. Phys. Photonics* **7**, 045015 (2025).
- [81] G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, T. Bland, S. Bo, J. A. Conejero, N. Firas, *et al.*, Objective comparison of methods to decode anomalous diffusion, *Nat. Commun.* **12**, 6253 (2021).
- [82] G. Muñoz-Gil, H. Bachimanchi, J. Pineda, B. Midtvedt, G. Fernández-Fernández, B. Requena, Y. Ahsini, S. Asghar, J. Bae, F. J. Barrantes, *et al.*, Quantitative evaluation of methods to analyze motion changes in single-particle experiments, *Nat. Commun.* **16**, 6749 (2025).
- [83] W. Cai, Y. Hu, X. Qu, H. Zhao, G. Wang, J. Li, and Z. Huang, Machine learning analysis of anomalous diffusion, *Eur. Phys. J. Plus* **140**, 183 (2025).
- [84] H. Seckler, J. Szwabiński, and R. Metzler, Machine-learning solutions for the analysis of single-particle diffusion trajectories, *J. Phys. Chem. Lett.* **14**, 7910 (2023).
- [85] G. Fernández-Fernández, C. Manzo, M. Lewenstein, A. Dauphin, and G. Muñoz-Gil, Learning minimal representations of stochastic processes with variational autoencoders, *Phys. Rev. E* **110**, L012102 (2024).
- [86] S. Asghar, Q.-X. Pei, G. Volpe, and R. Ni, Efficient rare event sampling with unsupervised normalizing flows, *Nat. Mach. Intell.* **6**, 1370 (2024).
- [87] S. W. Bender, M. W. Dreisler, M. Zhang, J. Kæstel-Hansen, and N. S. Hatzakis, SEMORE: SEgmentation and MORphological fingerprinting by machine learning automates super-resolution data analysis, *Nat. Commun.* **15**, 1763 (2024).
- [88] L. A. Saavedra, A. Mosqueira, and F. J. Barrantes, A supervised graph-based deep learning algorithm to detect and quantify clustered particles, *Nanoscale* **16**, 15308 (2024).
- [89] W. Cai, G. Wang, Y. Zhang, X. Qu, and Z. Huang, Reinforcement learning for active matter, *Biophysics Rev.* **6**, 031302 (2025).
- [90] J. Janczura, P. Kowalek, H. Loch-Olszewska, J. Szwabiński, and A. Weron, Classification of particle trajectories in living cells: Machine learning versus statistical testing hypothesis

- for fractional anomalous diffusion, *Phys. Rev. E* **102**, 032402 (2020).
- [91] A. Lanza, X. Qu, and S. Bo, Recurrent neural network analysis of single trajectories switching between anomalous diffusion states, *J. Phys. Photonics* **7**, 045001 (2025).
- [92] R. J. Ober, S. Ram, and E. S. Ward, Localization accuracy in single-molecule microscopy, *Biophys. J.* **86**, 1185 (2004).
- [93] F. Simon, L. E. Weiss, and S. van Teeffelen, A guide to single-particle tracking, *Nat. Rev. Methods Primers* **4**, 66 (2024).
- [94] X. Feng, H. Sha, Y. Zhang, Y. Su, S. Liu, Y. Jiang, S. Hou, S. Han, and X. Ji, Reliable deep learning in anomalous diffusion against out-of-distribution dynamics, *Nat. Comput. Sci.* **4**, 761 (2024).
- [95] H. Sha, Y. Wu, Y. Zhang, R. Liu, X. Feng, H. Li, Z. Wang, X. Zhang, and S. Hou, Single molecule spectrum dynamics imaging with 3D target-locking tracking, *Nat. Commun.* **16**, 8686 (2025).
- [96] B. Wang, S. M. Anthony, S. C. Bae, and S. Granick, Anomalous yet Brownian, *Proc. Natl. Acad. Sci. USA* **106**, 15160 (2009).
- [97] B. Wang, J. Guan, S. M. Anthony, S. C. Bae, K. S. Schweizer, and S. Granick, Confining potential when a biopolymer filament reptates, *Phys. Rev. Lett.* **104**, 118301 (2010).
- [98] K. Chen, B. Wang, J. Guan, and S. Granick, Diagnosing heterogeneous dynamics in single-molecule/particle trajectories with multiscale wavelets, *ACS Nano* **7**, 8634 (2013).
- [99] J. Malinowski, M. Kostrzewa, M. Balcerek, W. Tomczuk, and J. Szewabiński, CINNAMON: A hybrid approach to change point detection and parameter estimation in single-particle tracking data, *J. Phys. Photonics* **7**, 035008 (2025).
- [100] F. Laurent, H. Verdier, M. Duval, A. Serov, C. L. Vestergaard, and J.-B. Masson, TRamWay: Mapping physical properties of individual biomolecule random motion in large-scale single-particle tracking experiments, *Bioinformatics* **38**, 3149 (2022).
- [101] H. Verdier, F. Laurent, A. Cassé, C. L. Vestergaard, C. G. Specht, and J.-B. Masson, Simulation-based inference for non-parametric statistical comparison of biomolecule dynamics, *PLoS Comput. Biol.* **19**, e1010088 (2023).
- [102] C. Zhang, R. Liu, Z. Ding, P. Lu, W. Tian, Y. Zhao, J. He, and S. Hou, Review of machine learning for single-particle tracking: Methods, challenges, and biophysical insights, *Chem. Biomed. Imaging* (2025).
- [103] A. Chhabra and R. V. Jensen, Direct determination of the $f(\alpha)$ singularity spectrum, *Phys. Rev. Lett.* **62**, 1327 (1989).
- [104] A. Weron, K. Burnecki, E. J. Akin, L. Solé, M. Balcerek, M. M. Tamkun, and D. Krapf, Ergodicity breaking on the neuronal surface emerges from random switching between diffusive states, *Sci. Rep.* **7**, 5404 (2017).
- [105] M. Mangalam, R. Metzler, and D. G. Kelty-Stephen, Ergodic characterization of nonergodic anomalous diffusion processes, *Phys. Rev. Res.* **5**, 023144 (2023).
- [106] H. Seckler, R. Metzler, D. G. Kelty-Stephen, and M. Mangalam, Multifractal spectral features enhance classification of anomalous diffusion, *Phys. Rev. E* **109**, 044133 (2024).
- [107] J.-F. Muzy, E. Bacry, and A. Arneodo, Wavelets and multifractal formalism for singular signals: Application to turbulence data, *Phys. Rev. Lett.* **67**, 3515 (1991).
- [108] A. Grossmann and J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM J. Math. Anal.* **15**, 723 (1984).
- [109] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Theory* **36**, 961 (2002).
- [110] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, PyWavelets: A Python package for wavelet analysis, *J. Open Source Software* **4**, 1237 (2019).
- [111] C. Torrence and G. P. Compo, A practical guide to wavelet analysis, *Bull. Am. Meteorol. Soc.* **79**, 61 (1998).
- [112] D. Marr and E. Hildreth, Theory of edge detection, *Proc. R. Soc. B* **207**, 187 (1980).
- [113] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance* (CRC Press, Boca Raton, FL, 2017).
- [114] A. Teolis and J. J. Benedetto, *Computational Signal Processing with Wavelets* (Springer, Boston, MA, 1998), Vol. 182.
- [115] C. Cattani, Shannon wavelets theory, *Math. Probl. Eng.* **2008**, 164808 (2008).
- [116] M. Unser, A. Aldroubi, and M. Eden, On the asymptotic convergence of B-spline wavelets to Gabor functions, *IEEE Trans. Inf. Theory* **38**, 864 (2002).
- [117] P. Massignan, C. Manzo, J. A. Torreno-Pina, M. F. García-Parajo, M. Lewenstein, and G. J. Lapeyre, Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium, *Phys. Rev. Lett.* **112**, 150603 (2014).
- [118] H. Scher and E. W. Montroll, Anomalous transit-time dispersion in amorphous solids, *Phys. Rev. B* **12**, 2455 (1975).
- [119] B. B. Mandelbrot and J. W. V. Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10**, 422 (1968).
- [120] J. Klafter and G. Zumofen, Lévy statistics in a Hamiltonian system, *Phys. Rev. E* **49**, 4873 (1994).
- [121] S. C. Lim and S. V. Muniandy, Self-similar Gaussian processes for modeling anomalous diffusion, *Phys. Rev. E* **66**, 021114 (2002).
- [122] Hugging Face, https://huggingface.co/timm/efficientnet_b5.sw_in12k_ft_in1k.
- [123] Hugging Face, https://huggingface.co/timm/regnetz_040_h.ra3_in1k.
- [124] Hugging Face, https://huggingface.co/timm/vit_huge_patch14_224.orig_in21k.
- [125] WADNet, GitHub, <https://github.com/huangzih/WADNet>.
- [126] M. Liu, Q. Li, L. Liang, J. Li, K. Wang, J. Li, M. Lv, N. Chen, H. Song, J. Lee, *et al.*, Real-time visualization of clustering and intracellular transport of gold nanoparticles by correlative imaging, *Nat. Commun.* **8**, 15646 (2017).
- [127] A. Sabri, X. Xu, D. Krapf, and M. Weiss, Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells, *Phys. Rev. Lett.* **125**, 058101 (2020).
- [128] J. Szymanski and M. Weiss, Elucidating the origin of anomalous diffusion in crowded fluids, *Phys. Rev. Lett.* **103**, 038102 (2009).
- [129] T. Sentjabrskaja, E. Zaccarelli, C. De Michele, F. Sciortino, P. Tartaglia, T. Voigtman, S. U. Egelhaaf, and M. Laurati, Anomalous dynamics of intruders in a crowded environment of mobile obstacles, *Nat. Commun.* **7**, 11133 (2016).
- [130] Z. Huang, Wavelet representation [Data set], Zenodo (2026), <https://doi.org/10.5281/zenodo.19648037>.