

# Reliable deep learning in anomalous diffusion against out-of-distribution dynamics

Received: 7 April 2024

Accepted: 9 September 2024

Published online: 11 October 2024



Xiaochen Feng<sup>1,5</sup>, Hao Sha<sup>1,5</sup>, Yongbing Zhang<sup>1,5</sup>✉, Yaoquan Su<sup>2</sup>, Shuai Liu<sup>2</sup>, Yuan Jiang<sup>1</sup>, Shangguo Hou<sup>3</sup>, Sanyang Han<sup>2</sup> & Xiangyang Ji<sup>4</sup>✉

Anomalous diffusion plays a crucial rule in understanding molecular-level dynamics by offering valuable insights into molecular interactions, mobility states and the physical properties of systems across both biological and materials sciences. Deep-learning techniques have recently outperformed conventional statistical methods in anomalous diffusion recognition. However, deep-learning networks are typically trained by data with limited distribution, which inevitably fail to recognize unknown diffusion models and misinterpret dynamics when confronted with out-of-distribution (OOD) scenarios. In this work, we present a general framework for evaluating deep-learning-based OOD dynamics-detection methods. We further develop a baseline approach that achieves robust OOD dynamics detection as well as accurate recognition of in-distribution anomalous diffusion. We demonstrate that this method enables a reliable characterization of complex behaviors across a wide range of experimentally diverse systems, including nicotinic acetylcholine receptors in membranes, fluorescent beads in dextran solutions and silver nanoparticles undergoing active endocytosis.

Diffusion is an important phenomenon across numerous scientific domains, including physics<sup>1</sup>, chemistry<sup>2</sup>, biology<sup>3</sup>, geophysics<sup>4</sup>, materials science<sup>5</sup> and finance<sup>6</sup>. To comprehend normal diffusion in the molecular nature, Brownian motion—a classical model of random walks—is commonly regarded as a cornerstone concept in diffusion-related research. However, due to factors such as inhomogeneous environments<sup>7</sup>, transient binding<sup>8</sup> and molecular crowding<sup>9</sup>, diffusion in real-world settings sometimes exhibits nonlinear, unstable and memory-dependent properties, deviating from the classical assumptions of normal diffusion. This behavior is recognized as anomalous diffusion<sup>10</sup>. The molecular dynamics of anomalous diffusion represents a deviation from Brownian motion, whose mean-squared displacement (MSD) no longer grows linearly but with power law to time  $t$

( $\text{MSD} \propto t^a, a \neq 1$ )<sup>11</sup>. To describe anomalous diffusion, many types of models were established such as fractional Brownian motion (FBM)<sup>12</sup>, continuous-time random walks (CTRW)<sup>13</sup>, Lévy walks (LW)<sup>14</sup>, scaled Brownian motion (SBM)<sup>15</sup> and annealed transient time motion (ATTM)<sup>16</sup>. With well-defined mechanisms, these models equip researchers with sophisticated mathematical tools to elucidate complex dynamics in diverse systems.

To prevent the established models from being incorrectly associated, it is critical to develop methods capable of accurately categorizing observed trajectories into their corresponding models<sup>17</sup>. Prevalent approaches rely on statistical observables to assess diffusion patterns. For example, time-averaged MSD<sup>18</sup> and co-difference<sup>19</sup> are used to distinguish non-ergodic from ergodic diffusion; and the  $p$ -variation test<sup>20</sup>

<sup>1</sup>School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. <sup>2</sup>Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. <sup>3</sup>Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China. <sup>4</sup>Department of Automation, Tsinghua University, Beijing, China. <sup>5</sup>These authors contributed equally: Xiaochen Feng, Hao Sha, Yongbing Zhang. ✉e-mail: [ybzhang08@hit.edu.cn](mailto:ybzhang08@hit.edu.cn); [xyji@tsinghua.edu.cn](mailto:xyji@tsinghua.edu.cn)

and velocity autocorrelation<sup>21</sup> are applied to distinguish between FBM and CTRW. Although these methods are deployed easily, it is difficult to achieve satisfactory performance when higher levels of recognition granularity and scope are required. With the rapid advancement of artificial intelligence, deep learning has changed the landscape of this task. The Anomalous Diffusion (AnDi) Challenge<sup>22</sup> held in 2020 sets a benchmark for the recognition of anomalous diffusion, in which most deep-learning approaches showed remarkable success. The impressive performance of deep learning has encouraged its widespread adoption in recent diffusion-related research.

However, the reliability of deep learning in anomalous diffusion recognition remains a concern. As the distribution of real-world data is uncertain, observed trajectories may either be in-distribution (ID) samples covered by the training distribution or out-of-distribution (OOD) samples that deviate from it. The OOD trajectories will be incorrectly identified as one of ID models by the network. Consequently, applying deep learning to the analysis of real-world dynamics could result in unreliable recognition. To enhance the reliability of predictions, Bayesian neural networks (BNN) have been proposed for estimating the uncertainty of anomalous diffusion recognition<sup>23</sup>, thereby mitigating the likelihood of model misuse. However, their effectiveness is demonstrated primarily in confident ID recognition and does not adequately consider OOD dynamics. For detecting OOD dynamics, the unsupervised autoencoder has emerged as a promising method, leveraging the reconstruction loss to identify anomalies<sup>24</sup>. Nevertheless, integrating this method with ID anomalous diffusion recognition remains challenging. To the best of our knowledge, there are few studies considering OOD samples in anomalous diffusion recognition, although OOD dynamics is ubiquitous in the real-world scenarios. Current deep-learning-based methods therefore still fall short of the reliability required for scientific research and are far from achieving robust anomalous diffusion recognition expected to identify both in-distribution and OOD dynamics.

As deep learning consistently encounters the potential threat from OOD samples, OOD detection is a critical aspect for ensuring the safety of AI systems. There has been a rich line of research on OOD detection in recent years, which can be roughly divided into two categories: (1) data-driven methods that utilize outlier exposure to let the network learn OOD features in a supervised way<sup>25</sup>; and (2) training-driven methods that are based on the source data without outliers, using source data to train the network for effective ID representations to distinguish OOD samples<sup>26</sup>. These methods have been widely highlighted in the fields of computer vision, speech recognition and natural language processing, but they experience serious performance degradation when applied to fine-grained recognition tasks<sup>27</sup>. As the task of anomalous diffusion recognition lacks natural outliers (we can hardly judge whether the collected trajectories are OOD) and is inherently fine-grained (the associated random walk patterns are complex), it is challenging for both data- and training-driven methods to learn effective OOD features and ID representations for OOD dynamics detection. Moreover, as researchers generally expect that methods for distinguishing OOD dynamics will not worsen the performance of anomalous diffusion recognition, it is crucial to establish comprehensive evaluations for OOD dynamics detection and set up a baseline for solutions.

In this work, we developed a general evaluation framework and a powerful baseline method for OOD dynamics detection, identifying opportunities to achieve reliable deep learning in anomalous diffusion. Our innovative framework and method is adaptable for dynamics in real-world scenarios, broadening its applicability in practical research contexts. The proposed method integrates three key perspectives: data augmentation, training regularization and confidence computation, which can improve the performance of deep learning in both accurate anomalous diffusion recognition and effective OOD dynamics detection. For experimental observations with unknown distributions,

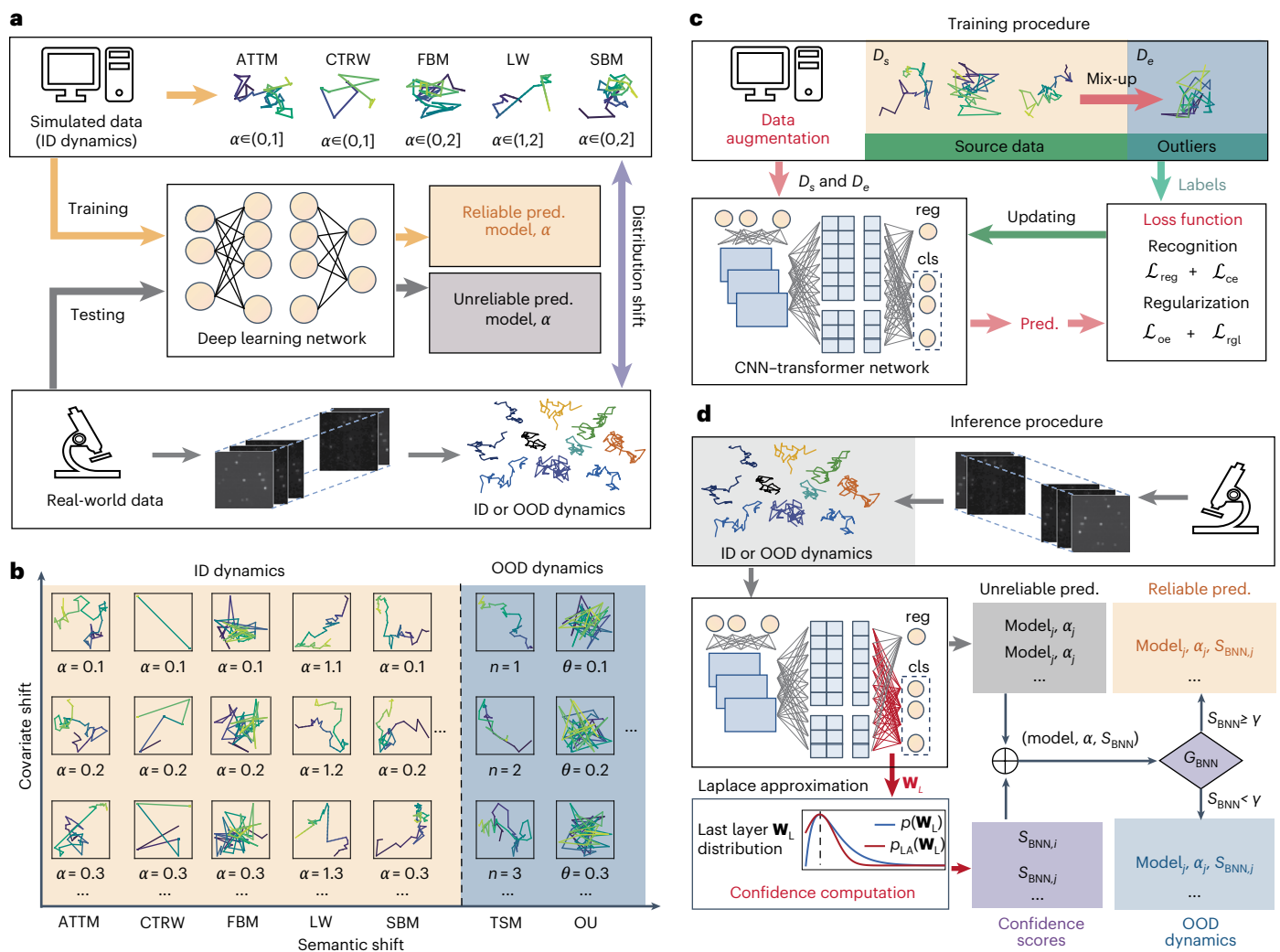
our work provides a guidance to produce reliable deep-learning predictions and avoid misinterpretation. We demonstrate that deep learning integrated with function of OOD detection is a promising tool for the analysis of real dynamic data, providing reliable and interesting insight into the feature of unknown dynamics and complex systems.

## Results

The deployment process of deep-learning-based methods for anomalous diffusion recognition is depicted in Fig. 1a. These methods operate under the closed-world assumption, where the test data are drawn from the same distribution as the source data for training, known as ID. Deep learning can therefore only perform well in identifying anomalous diffusion models with distributions that align with the source data, while incorrectly predicting other types of dynamics (Supplementary Fig. 1). Given that the distribution of observed data is uncertain, deep-learning predictions will be unreliable in real-world scenarios. To elucidate this problem, we divide the differences among trajectories from various distributions into semantic (for example, trajectories drawn from different dynamics) and covariate shifts (for example, trajectories originating from the same dynamics but presenting in different forms, such as FBMs with different anomalous exponents  $\alpha$ ). In Fig. 1b–d, we propose our OOD dynamics-detection method that is designed to improve the reliability of deep learning for anomalous diffusion, enabling the network to accurately identify ID diffusion models and simultaneously detect the OOD sample that undergoes a semantic shift. In the data preparation, we employ the concept of Mixup<sup>28</sup>—a technique that creates new examples by blending pairs of data and labels—to generate simulated outliers with semantic shift, thereby broadening the training distribution. For the training procedure, we develop a multi-component loss function that uses generated outliers to enhance the network's boundary perception of the source data domain  $\mathcal{D}_s$ , thereby improving regularization. Furthermore, inspired by the success of BNNs in estimating the uncertainty of anomalous diffusion recognition<sup>23</sup>, we applied the last-layer Laplace approximation (LLA)<sup>29</sup>—a Bayesian method suitable for already-trained networks—to compute confidence scores. Unlike traditional Bayesian frameworks that rely on maximum a posteriori (MAP) estimates, the LLA captures more information about the uncertainty surrounding the MAP estimate and thus reduces the overconfidence in OOD detection. To conduct a comprehensive evaluation, we use the *andi*-datasets package provided by AnDi Challenge to generate simulated datasets with ground-truth values of anomalous exponents and diffusion models (including ATTM, CTRW, FBM, LW and SBM). The backbone of the network uses a CNN–transformer-based architecture (Supplementary Fig. 2)<sup>30</sup>. For the task of OOD dynamics detection, additional types of motions are introduced, including the directed Brownian motion (DBM)<sup>31</sup>, the two-state model (TSM)<sup>32</sup>, the combined anomalous diffusion motion (CBM)<sup>33</sup>, the Sinai diffusion model (SINAI)<sup>34</sup> and the Ornstein–Uhlenbeck (OU) process<sup>35</sup>. It should be noted that the TSM in our study combines Brownian and obstructed diffusion mechanisms, whereas the CBM combines distinct anomalous diffusion mechanisms.

## Benchmark of OOD dynamics detection

We use the same models in the AnDi Challenge as ID samples to benchmark the performance in anomalous diffusion recognition. Two evaluation strategies are employed regarding OOD dynamics definition in both the training and testing procedures. We designate all models from the AnDi Challenge (A.C.F.L.S, including ATTM, CTRW, FBM, LW and SBM) as ID and simulate other common dynamics (for example, DBM, TSM, CBM, SINAI and OU) as OOD to perform OOD dynamics detection. We also adopt cross-validation. We randomly select a model in the A.C.F.L.S as OOD, and the remaining models as ID. This strategy allows for a more comprehensive exploration of both ID and OOD domains. Therefore, ten types of simulated datasets (ID–OOD) are



**Fig. 1 | OOD dynamics detection for anomalous diffusion recognition.**

**a**, Schematic representation of deep learning for anomalous diffusion recognition. The network is trained on simulated trajectories generated from theoretically established diffusion models (for example, ATT, CTRW, FBM, LW, SBM) with specific anomalous exponent  $\alpha$ , and is tested on real-world observed trajectories with unknown dynamics. The potential distribution shift between the dynamics of simulated and real-world trajectories can lead to unreliable deep-learning predictions. The color code of the trajectories represents time.

**b**, Illustration of ID and OOD dynamics. Reliable deep learning is expected to effectively recognize ID dynamics as well as detect OOD dynamics. Parameter  $n$  in TSM represents the number of state jumps, whereas  $\theta$  in OU represents the decay

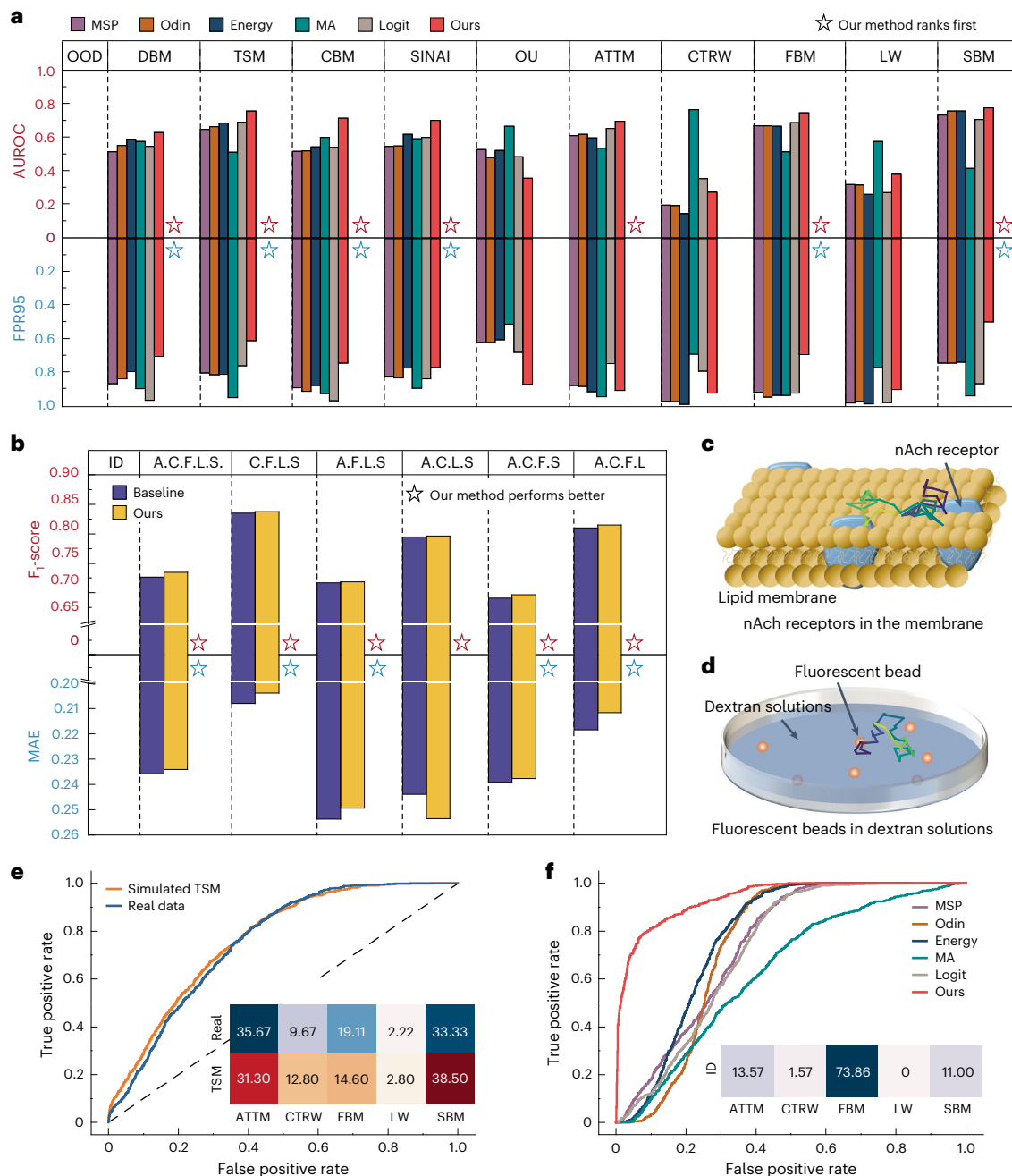
rate. **c**, Training procedure for our OOD dynamics detection. Data augmentation broadens the training distribution by introducing outliers from the source data using a specialized mix-up strategy. During parameter updating, the loss function for outlier exposure  $\mathcal{L}_{oe}$  and L2-norm loss function  $\mathcal{L}_{rgl}$  regularize the network to learn OOD features and achieve MAP parameter estimation.

**d**, The inference procedure for our OOD dynamics detection. The Laplace approximation on the last layer for classification turns the MAP-trained classifier (the weights of the classifier's last layer are denoted  $\mathbf{W}_L$ ) into a BNN in a post-hoc manner. The MAP predictions with BNN-based confidence scores  $S_{BNN}$  below the threshold  $\gamma$  will be detected as OOD dynamics by the detector  $G_{BNN}$ .

generated for recognizing anomalous diffusion with OOD dynamics. All trajectories in the benchmark are simulated with: dimension  $m = 2$ ; length  $w = 50$ ; and signal-to-noise ratio (SNR)  $\in \{1, 2, 10\}$ . We benchmark our method against other typical OOD-detection algorithms in anomalous diffusion recognition (Supplementary Note 1), where the performance of OOD detection is measured with the following metrics: the area under the receiver operating characteristic curve (AUROC), and the false positive rate of OOD examples when the true positive rate of ID examples is 95% (FPR95). Among ten scenarios for OOD dynamics detection (Fig. 2a), our method achieves the best metrics (the highest AUROC and the lowest FPR95) in six of them, and the top two rankings in AUROC in two of the remaining four scenarios. Considering the ID recognition, our method only slightly decreases the prediction of the anomalous exponent in A.C.F.L.S while improving the performance in all other cases (Fig. 2b and Supplementary Fig. 3). Benchmark results

demonstrate the superiority of our method for enabling deep learning with better performance in both OOD dynamics detection and ID anomalous diffusion recognition (Supplementary Note 2).

To further validate the efficacy of our method on real-world data (Fig. 2c,d), we apply it to experimental single-particle tracking (SPT) data of nicotinic acetylcholine (nACh) receptors, which are reported to be formulated as a TSM<sup>33</sup>. Our method with the network trained on A.C.F.L.S detects the SPT data as OOD with metrics that closely align with the simulated TSM trajectories, highlighting the robustness of our method for real-world OOD trajectories (Fig. 2e). Furthermore, the previously considered sub-diffusive ( $\alpha < 1$ ) FBM behavior of the fluorescent bead in dextran solutions is reassessed. Based on the network trained on A.C.F.L.S (FBM is in the source domain), our method identifies this diffusion as FBM with a mean anomalous exponent of 0.78 at a probability of 73.86%, validating that the experimental data



**Fig. 2 | Results of OOD dynamics detection. a**, Benchmark for OOD dynamics detection. All OOD detection methods are evaluated in the same (ID–OOD) datasets with metrics of AUROC and FPR95. The datasets include (A.C.F.L.S–DBM), (A.C.F.L.S–TSM), (A.C.F.L.S–CBM), (A.C.F.L.S–SINAI), (A.C.F.L.S–OU), (C.F.L.S–ATTM), (A.F.L.S–CTRW), (A.C.L.S–FBM), (A.C.F.S–LW) and (A.C.F.L–SBM), where A.C.F.L.S represents the ID dataset with dynamics  $\in \{\text{ATTM, CTRW, FBM, LW, SBM}\}$ . The other methods used in the benchmark (MSP, Odin, Energy, MA, Logit) are explained in Supplementary Note 1. **b**, The ability of our method to recognize ID anomalous diffusion compared with a baseline employing a normal training objective on the basis of the source data and cross-entropy. The  $F_1$ -score of anomalous diffusion classification and MAE of anomalous exponent

regression are used as metrics for ID recognition. **c,d**, Schematic representation of experimental SPT data. The SPT data for the nACh receptors (**c**) are acquired from the public dataset, whereas the SPT data of fluorescent beads in dextran solutions (**d**) are acquired from our home-built SPT microscopy (Supplementary Fig. 4). **e**, The ability of our method to detect OOD dynamics across data pairs, including (A.C.F.L.S–SPT data of nACh receptors) and (A.C.F.L.S–TSM). The confusion matrix shows the classifier's output for the SPT data and TSMs. **f**, Comparative benchmark for detecting SPT data of fluorescent beads in dextran solutions, where the network is trained on A.C.L.S. The confusion matrix shows the classifier's output for the free diffusion in dextran solutions, where the classifier is trained on A.C.F.L.S.

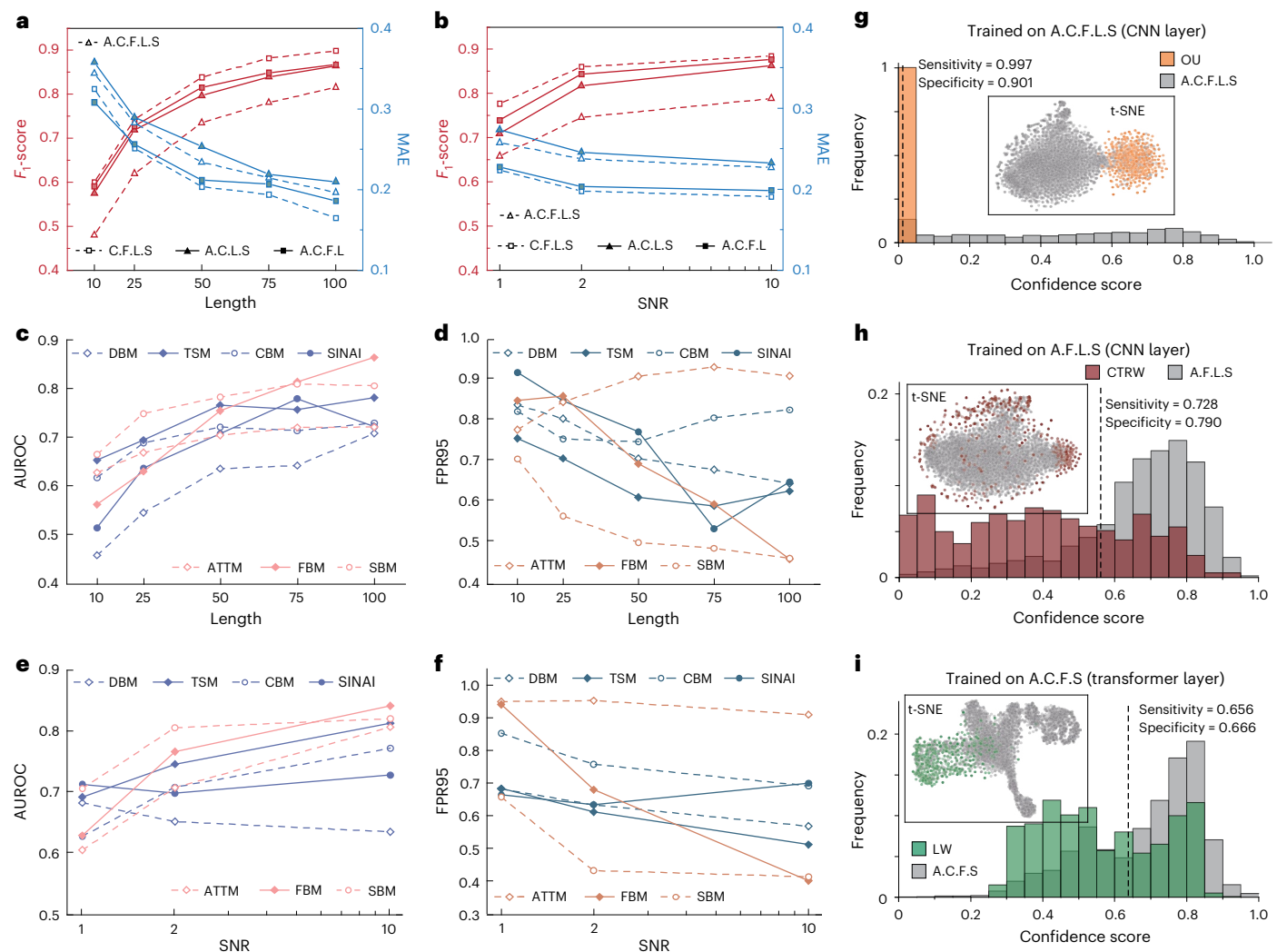
conform to the description in the literature (FBM with an averaged  $\alpha = 0.82$ )<sup>36</sup>. We then benchmark OOD dynamics detection on the dataset (A.C.L.S–diffusion in dextran solutions) and our method ranks first with AUROC/FPR95 of 0.94/0.31 (Fig. 2f). These results demonstrate the effectiveness of our method in identifying realistic diffusion types not included in the source domain. Moreover, deep learning with our

method provides further evidence supporting the interpretation of specific motor behaviors.

### Enhancement for OOD dynamics detection

The quality of the datasets is vital for the ultimate performance of deep learning. The reliability of deep learning in anomalous diffusion





**Fig. 3 | Evaluation for OOD dynamics detection with enhancement.** **a**,  $F_1$ -score for anomalous diffusion classification and MAE for anomalous exponent regression as a function of trajectory length, where the trajectory SNR  $\in \{1, 2, 10\}$ . **b**,  $F_1$ -score for ID diffusion model classification and MAE for ID anomalous exponent regression as a function of trajectory SNR, where the trajectory length is 50 points. **c**, AUROC (**c**) and FPR95 (**d**) for OOD dynamics detection as a function of trajectory length, where the trajectory SNR  $\in \{1, 2, 10\}$ . **e**, AUROC (**e**) and FPR95 (**f**) for OOD dynamics detection as a function of the trajectory SNR,

where trajectory length is 50 points. **g**–**i**, 1D linear discriminant analysis projections and t-SNE visualizations of OOD OUs (**g**), CTRWs (**h**), LWs (**i**) and their corresponding ID datasets A.C.F.L.S (**g**), A.F.L.S (**h**) and A.C.F.S (**i**). The dashed lines represent thresholds that maximize the sum of sensitivity (the proportion of correctly identified positives) and specificity (the proportion of correctly identified negatives). The confidence score is defined as the normalized distance between the feature vector of the input and the center of the hypersphere.

is therefore closely related to the observation system, for which the spatiotemporal resolution and photon count play a crucial role. A higher spatiotemporal resolution means more points observed per unit time, whereas the photon count determines the efficiency and accuracy of the trajectory extraction. As the impacts can be approximately mapped to the length and signal-to-noise ratio (SNR) of the simulated trajectories, we further evaluate the relationship between our method and the observation system based on the simulation. For ID recognition, our method shows improved accuracy with increasing length and SNR, which is consistent with conventional expectations (Fig. 3a,b). We conduct further experiments on the OOD dynamics that is effectively detected in the benchmark. As the length increases, the AUROC performance improves in all cases and FPR95 performance only declines when detecting OOD CBMs and ATTMs (Fig. 3c,d). With higher SNRs, only OOD DBMs and SINAI experience slight declines in AUROC and FPR95, respectively (Fig. 3e,f). Thus, at least one metric (AUROC or FPR95) improves in all scenarios. Furthermore, along with high-quality data, the metrics achieve relatively high

values (averaged AUROC/FPR95 = 0.85/0.54 for 100-point length and SNR = 10), confirming the importance of implementing reliable deep learning in trustworthy dynamics analysis with the enhancement of observation systems.

Detection for certain OOD dynamics (OU, CTRW, LW) remains ineffective (AUROC < 0.5) even with improved data quality (Supplementary Fig. 5), where inference based on the output layer hardly distinguishes them. However, as shown in Fig. 3g–i, these OOD dynamics may exhibit distinct distribution in the feature space of specific hidden layers. We thus modify our confidence computation part by enclosing the feature vectors of ID dynamics within a hypersphere, and design the OOD detector  $G_{\text{dis}}$  on the basis of the distance between the input's feature vector and the hypersphere's center (see the 'Confidence computation' section in the Methods). Using  $G_{\text{dis}}$  on the basis of the last CNN layer, detection for OOD OUs and CTRWs achieves AUROC/FPR95 of 0.96/0.07 and 0.81/0.72, respectively (Supplementary Fig. 6). For OOD Lévy walks, AUROC/FPR95 = 0.69/0.84, where  $G_{\text{dis}}$  is based on the last transformer layer. As various feature layers possess distinct

perception capabilities for OOD dynamics, the results demonstrate that our method is flexible and not inherently ineffective against specific OOD dynamics.

### OOD detection allows for reliable dynamic interpretation

Employing deep learning without OOD detection for diffusion analysis may engender misinterpretations of the observed phenomena. To evaluate the reliability of our method for deep-learning-based dynamic interpretation, we conduct experiments showing what will happen when feeding OOD into the network, which is trained for the task of anomalous diffusion model classification and exponent regression. In-distribution models are typically expected to exhibit higher scores than OOD models, thus reducing predictions that are misrepresentative. It is noteworthy that the network inevitably outputs incorrect models for OOD inputs and we can assess the risk of erroneous interpretations by measuring the deviation of the predicted anomalous exponent from the ground-truth value. In the subsequent analysis, we categorize OOD models on the basis of their detection efficacy, which can be divided into OOD models with effective and ineffective detection.

**OOD models with effective detection.** When taking ATTM, FBM or SBM as the OOD model, the network using our method effectively identifies such models as OOD, as evidenced by their lower mean confidence scores compared with their corresponding ID models (C.F.L.S, A.C.L.S, A.C.F.L) (Fig. 4c–e). For OOD ATTMs, the mean absolute error (MAE) for  $\alpha$  prediction on ID models is 0.19, whereas it is 0.49 for OOD ATTMs (Fig. 4b). There is a large number of subdiffusive OOD ATTMs that are incorrectly predicted as superdiffusive ( $\alpha > 1$ ) dynamics (Fig. 4a). For OOD FBMs, as SBM and ATTM are compatible with non-ergodic behavior<sup>22</sup>, a risky scenario can be observed in which most ergodic OOD FBMs are predicted as SBM and ATTM when  $\alpha$  approaches 1 (Fig. 4d). For OOD SBMs, a substantial number of superdiffusive OOD SBMs are predicted as subdiffusive, and the MAE for  $\alpha$  prediction is 0.28 lower than that of ID models (Fig. 4a,b). Considering that ATTMs, FBMs and SBMs with  $\alpha = 1$  converge to standard Brownian motion, their low confidence scores effectively prevent normal diffusion from being misidentified as specific anomalous diffusion (Supplementary Note 4). Overall, these findings underscore the severe risks of misinterpretations if such OOD dynamics is not distinguished, highlighting the importance of our effective detection.

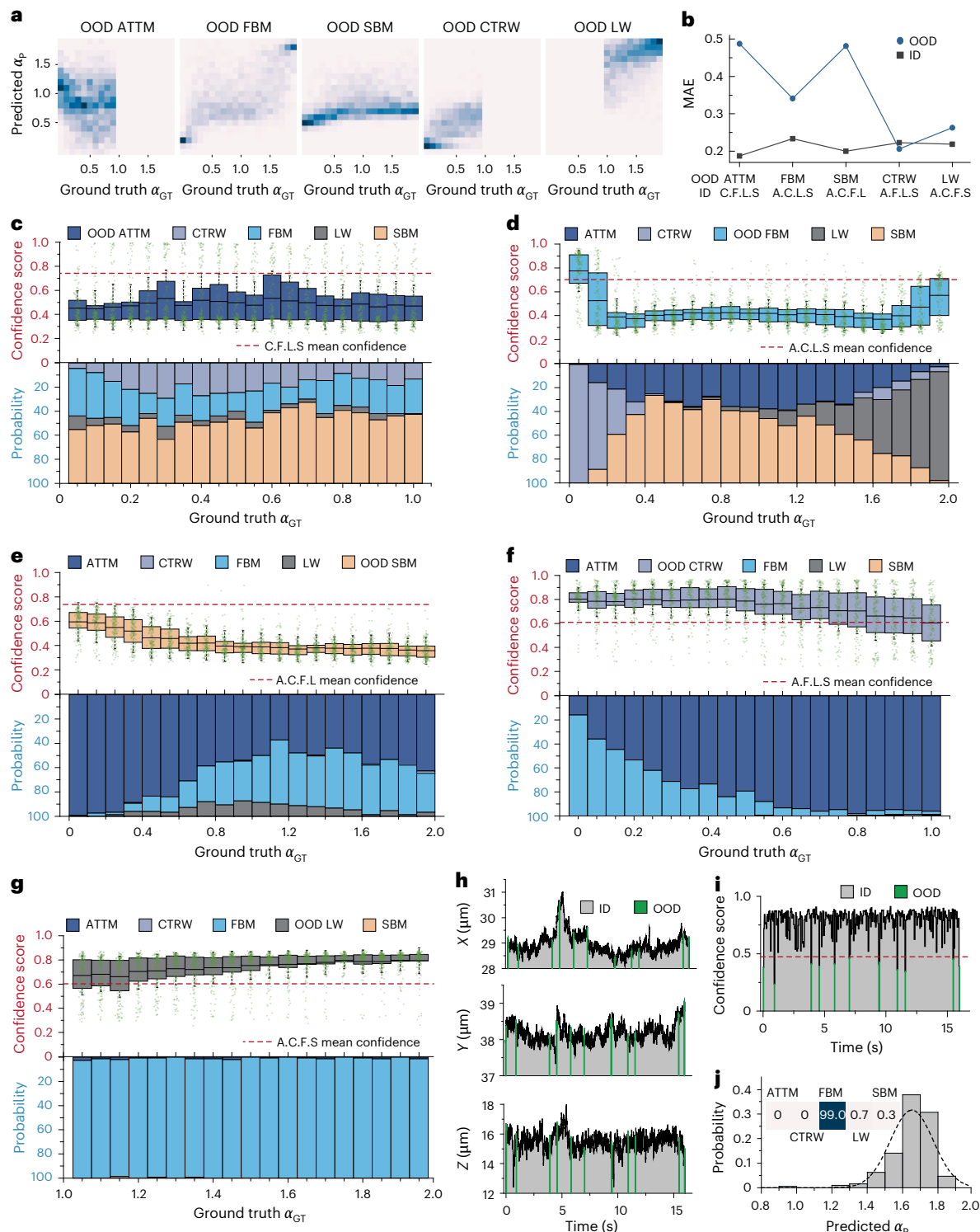
**OOD models with ineffective detection.** When the OOD model is CTRW or LW, the mean confidence score is close to that of their corresponding ID models (A.F.L.S or A.C.F.S), indicating limited detection capability with our method (Fig. 4f,g). However, the majority of subdiffusive OOD CTRWs and superdiffusive OOD LWs are accurately identified as subdiffusion and superdiffusion, respectively (Fig. 4a). As depicted in Fig. 4b, the prediction of  $\alpha$  for OOD CTRWs and LWs shows improved performance compared with other OOD models with effective detection. Notably, OOD CTRWs achieve an even lower MAE than the ID models (A.F.L.S). Furthermore, subdiffusive OOD CTRWs tend to be identified as subdiffusive FBMs when  $\alpha \rightarrow 0$ , and as ATTMs when  $\alpha \rightarrow 1$  (Fig. 4f). These identifications are understandable considering the fact that CTRW without jumps corresponds to an FBM with  $\alpha = 0$ , and both CTRW and ATTM exhibit hidden waiting times<sup>23</sup>. Regarding OOD LWs, most are predicted to be superdiffusive FBMs (Fig. 4g), which inherently resembles LWs with non-local, scale-free diffusion properties, leading to hyperballistic behavior over long timescales<sup>37</sup>. Hence, notwithstanding the impediments of our method in detecting such OOD models, their high confidence scores signify relatively low-risk misinterpretations, given that the predictions adequately capture the dynamic characteristics to a certain extent.

**OOD dynamics detection in experimental observation.** To substantiate our simulation-based findings, we extend our OOD dynamics detection to an experimental scenario involving the active endocytosis of silver nanoparticles. Past studies have indicated that the majority of endocytosed metal nanoparticles exhibit a pearl-on-a-string motion, characterized by a combination of random and linear walk patterns<sup>38</sup>. This motion model serves to represent the transportation facilitated by endocytic vesicles and motor proteins. To analyze the endocytosed silver nanoparticle trajectory, we divide the observed continuous, long trajectory into 320 non-overlapping trajectories, each containing 50 data points. Deep learning in conjunction with our OOD dynamics detection is then used for further recognition and the network is trained on A.C.F.L.S without the pearl-on-a-string motion. Results show that 305 trajectories are identified as superdiffusive FBM (Fig. 4j), whereas 12 are detected as OOD (Fig. 4h,i). As superdiffusive FBM is proven to represent a strong trend in persistent long-range correlations of particle movements<sup>11,39</sup>, it aligns well with the linear walk pattern in pearl-on-a-string motion. Furthermore, the superdiffusive FBM is widely used to model the movement of cellular cargoes actively transported by motor proteins<sup>40</sup>, indicating the fact that motor proteins promote the transport of the silver nanoparticle mediated by endocytic vesicles. Therefore, despite the potential distribution shift from the source domain, the deep-learning predictions can remain low-risk and trustworthy for representing experimental trajectories identified as ID by our method. Meanwhile, during the detected periods of OOD dynamics, there are abrupt drops in confidence compared with the adjacent moments (Fig. 4i). This indicates that the dynamics of the endocytosed silver nanoparticle may undergo unknown alterations, potentially related to free or confined diffusion discussed in previous research<sup>38,41</sup>. Given the prominence of abnormal behaviors in research, our method can serve as a valuable warning system for such behaviors, preventing them from being falsely identified or ignored by deep learning.

## Discussion

Deep-learning-based methods for anomalous diffusion have demonstrated advantages in accuracy; however, such advantages are confined to ID trajectories. Our work extends OOD detection to deep learning for anomalous diffusion and provides a general framework for evaluation. Benchmark results demonstrate the effectiveness of our method in both OOD dynamics detection and ID recognition. Enhanced spatiotemporal resolution and increased photon counts in observation systems are expected to further boost performance. However, it can be observed that certain OOD dynamics are even more difficult to detect (Fig. 3d,e), indicating potential overfitting to features optimized for ID recognition. This phenomenon underscores the need for regularization methods designed specifically for OOD dynamics detection. Furthermore, deep learning with our method successfully detects high-risk OOD dynamics, preventing inappropriate applications of anomalous diffusion. The detected samples can serve as valuable empirical evidence for mathematicians and physicists in formulating novel dynamics. For identified ID dynamics, predictions will exhibit characteristics similar to their actual properties. Considering the fact that real-world data rarely follow established models exactly, this capability of deep learning with OOD dynamics detection carries important implications for practical dynamics analysis. Furthermore, as the undetected OOD dynamics exhibits similarities with the predicted model, we can potentially use the failures of deep learning as an effective tool to identify where human modeling should delve deeper to uncover connections among qualitatively distinct diffusion models.

As the prevalence of large models with extensive training samples grows, the importance of OOD detection in the routine AI applications seems to wane. However, our work highlights its critical value in anomalous diffusion recognition. On one hand, achieving a comprehensive



**Fig. 4 | Analysis for OOD dynamics detection. a**, Probability distribution of the predicted ( $\alpha_p$ ) versus ground-truth anomalous diffusion exponent ( $\alpha_{GT}$ ) for OOD ATT, FBM, SBM, CTRW and LW detection. **b**, Mean absolute error of anomalous exponent regression across different OOD models (ATT, FBM, SBM, CTRW, LW) and their corresponding ID models (C.F.L.S, A.C.L.S, A.C.F.L, A.F.L.S, A.C.F.S). One can infer the prediction risk of different OOD models, where the predictions for OOD CTRWs and SBMs exhibit the lowest and highest risks, respectively. **c–g**, Confidence scores and model predictions for OOD ATTs (**c**), FBMs (**d**), SBMs (**e**), CTRWs (**f**) and Lévy walks (**g**) across different ground-truth anomalous exponents  $\alpha_{GT}$ . For the box plots showing confidence scores,  $n = 150$  independent simulations were conducted for each box, with corresponding data points

overlaid. Middle line, median; box edges, 25th and 75th percentiles; whiskers, range within mean  $\pm$  s.d. The dashed lines represent the mean confidence of their corresponding ID models. **h, i**, Performance of our method in analyzing experimental SPT data of the endocytosed silver nanoparticles. The SPT data (**h**) are 3D and the network is trained on ID models of A.C.F.L.S. The trajectory segments corresponding to the green areas are detected as OOD. The dashed line in the confidence performance (**i**) represents the confidence threshold  $\gamma$  for OOD detection (details for the confidence threshold selection are shown in Supplementary Note 3). **j**, Distribution of predicted anomalous diffusion exponents  $\alpha_p$  and classified models for trajectories identified as ID by our method.

enumeration of the motion patterns is an elusive goal. Hence, we need continually strive to derive new patterns of motion from empirical observations. On the other hand, the consequences of misidentifying dynamics are severe, potentially leading researchers to mischaracterize the underlying truths of their observations. Although the potential value of OOD dynamics detection is substantial, our benchmark reveals that no single OOD detection methodology reigns supreme in all (ID–OOD) cases, indicating substantial opportunities for enhancement. The benchmark currently focuses on fixed-length trajectories. However, addressing trajectories of varying lengths will introduce additional challenges. As shorter ID trajectories are prone to incorrect OOD dynamics detection (Supplementary Note 5), there is a great demand for reliable detection of single observable across time<sup>42</sup>. Furthermore, although ergodicity breaking in certain anomalous diffusion processes poses serious limitations for classical statistics-based approaches<sup>22</sup>, deep learning with OOD dynamics detection has shown robust performance (Supplementary Note 6). We therefore believe that OOD dynamics detection will witness a resurgence in the future. With more high-performance methods for OOD dynamics detection, deep-learning-driven identification of anomalous diffusion is poised to be a reliable research tool to precisely identify and explore the underlying connections among complex processes.

## Methods

### Theoretical models

All simulated dynamics used in our work are generated by the following models.

**ATM.** Annealed transient time motion features a kind of combined normal diffusion whose diffusion coefficient  $D$  varies in time. Assuming that the  $n$ th normal diffusion state is  $D_n$  and the duration is  $\tau_n$ , the simulation samples  $D_n$  and  $\tau_n$  from distribution  $P(D) \sim D^{\sigma-1}$  (with  $\sigma > 0$ ) and  $P(\tau) \sim \delta(\tau - D^{-\gamma})$  (with  $\sigma < \gamma < \sigma + 1$ ), respectively. Such generated trajectories of ATM have been verified belong to anomalous subdiffusion with  $\alpha = \sigma/\gamma$ .

**CTRW.** Continuous-time random walks is one of the most popular mathematical models in dynamic analysis, of which both the waiting time  $\tau$  and spatial displacement  $\Delta x$  of each step are stochastic and independent. In our case, we consider  $\tau$  follows a power-law distribution  $\psi(\tau) \sim \tau^{-1-\sigma}$  and  $\Delta x$  is sampled from a Gaussian distribution  $N(0, \sqrt{D})$ . This simulation of CTRW features anomalous subdiffusion with  $\alpha = \sigma$ , ( $0 < \sigma < 1$ ).

**FBM.** As a classic random process, FBM emerges as a powerful tool for modeling a host of natural time series, which is the only self-similar Gaussian process with stationary increments. Unlike Brownian motion, the increments of FBM are not pure Gaussian noise, but fractional Gaussian noise (fGn), which is given by

$$\langle \xi_{\text{fGn}}(t) \xi_{\text{fGn}}(t + \tau) \rangle = \alpha(\alpha - 1) K_\alpha \tau^{\alpha-2}, \quad (1)$$

where  $\alpha$  is the anomalous exponent related to the Hurst exponent  $H$ , commonly used in the FBM literature via  $\alpha = 2H$ , ( $0 < H < 1$ ), and  $K_\alpha$  is the generalized diffusion constant.

**LW.** Lévy walks is a special CTRW that features anomalous superdiffusion by introducing a functional dependence between waiting times  $\tau$  and spatial displacements  $\Delta x$ . For the simulation,  $\tau$  follows the same distribution as CTRW, with  $\psi(\tau) \sim \tau^{-1-\sigma}$ ; furthermore, the coupling between  $\Delta x$  and  $\tau$  is  $\psi(\Delta x, \tau) = \frac{1}{2} \delta(\Delta x - v\tau) \psi(\tau)$ , where  $v$  is the velocity. As a result, LW corresponds to anomalous superdiffusion with

$$\alpha = \begin{cases} 2 & \text{if } 0 < \sigma < 1 \\ 3 - \sigma & \text{if } 1 < \sigma < 2 \end{cases} \quad (2)$$

**SBM.** Scaled Brownian motion is a popular model based on the time-dependent diffusivity  $K(t)$ , which can be described by the Langevin equation with the white, zero mean Gaussian noise  $\xi(t)$  as follows

$$\frac{dx(t)}{dt} = \sqrt{2K(t)} \xi(t), \quad (3)$$

where  $K(t)$  has a power-law dependence  $K(t) = \alpha K_\alpha t^{\alpha-1}$ . The MSD can be calculated by a set of tracers and this ensemble-averaged MSD (EA-MSD) is expected to converge to the true value. For SBM, the corresponding EA-MSD is  $\langle x^2(t) \rangle \simeq K_\alpha t^\alpha$  ( $0 < \alpha \leq 2$ ).

**DBM.** Directed Brownian motion is a dynamic model used to describe active transport. It can be understood as standard Brownian motion (with diffusion coefficient  $D$ ) with linear drift (with constant velocity  $v$ ), whose EA-MSD follows  $\langle x^2(t) \rangle \simeq 2Dt + (vt)^2$ .

**TSM.** The TSM combines two types of diffusion: free diffusion state  $S_1$  (standard Brownian motion  $W_t$ ) and obstructed diffusion state  $S_2$  (in which Brownian motion occurs within a confined region  $[-L, L]$  and can become immobilized at the boundaries), based on the switching diffusion concept, which is established for the translational diffusion of the nicotinic acetylcholine receptor at the plasma membrane<sup>32</sup>. For the simulation, the number of state jumps is sampled from a uniform distribution  $U(0, 5)$ , and the TSM is given by

$$x(t) = \begin{cases} x(t-1) + \Delta W_t & \text{if } S_1 \text{ or } (S_2 \text{ and } |x(t-1) + \Delta W_t| < L) \\ x(t-1) & \text{if } S_2 \text{ and } |x(t-1) + \Delta W_t| \geq L. \end{cases} \quad (4)$$

**CBM.** As several research studies have found that the individual real-world trajectory sometimes shifts to another diffusion model when the external environment changes, CBM is presented to feature such trajectories combined with two diffusion models (within ATM, CTRW, FBM, LW and SBM). In our case, the change point between diffusion regimes is sampled from the uniform distribution  $U(5, n-5)$ , where  $n$  is the length of the trajectory.

**SINAI.** SINAI is a prominent representative of logarithmic time-evolution random walk, which is often associated with systems with irregular or discontinuous potential energy landscapes. In terms of simulation, it can be generated in the same way as CTRW by using a different waiting time probability density function in the form  $\psi(\tau) \sim 1/(\tau \log^{1+\gamma} \tau)$ , leading to the EA-MSD  $\langle x^2(t) \rangle \simeq \log^\gamma t$ . For  $\gamma = 4$ , simulated trajectories share the same behavior and the scale of MSD as those of SINAI<sup>11</sup>.

**The OU process.** The OU process is a Gaussian, continuous-time stochastic process with exponentially decaying autocorrelations, and is widely used to model the behavior of systems subject to random fluctuations. The differential equation followed by the OU is given by

$$dx(t) = -\theta x(t) dt + \sigma dW_t, (\theta, \sigma > 0), \quad (5)$$

where  $W_t$  is the standard Brownian motion,  $\sigma^2$  is the variation of  $W_t$ , and  $\theta$  is the decay rate. In our simulation,  $\sigma = 1$  and  $\theta \sim U(0, 2)$ .

### Deep learning for anomalous diffusion recognition and OOD detection

**Network architecture.** Short trajectories are most common in anomalous diffusion recognition because of the limited photostability and brightness of a typical optical microscopic system. As convolutional transformers have been proven to be superior in short trajectory recognition<sup>30</sup>, we choose this architecture to complete the diffusion recognition task. The input is the trajectory  $\mathbf{X} \in \mathbb{R}^{w \times m} = [\vec{x}_1; \vec{x}_2; \dots; \vec{x}_w]$ ,



where  $w$  and  $m$  are the length and dimension of the trajectory, respectively. The analysis of anomalous diffusion always includes the model classification and exponent regression, so that multitask learning is suitable for deployment. Besides, the performance of anomalous exponent regression for OOD samples can serve as an effective metric to evaluate the risk of OOD dynamics. The anomalous diffusion model and exponent are therefore predicted by the network simultaneously in our work. The network consists of three main modules: convolutional, self-attention and linear modules. The convolutional module follows the ResNet<sup>43</sup> architecture and uses 1D convolution kernels to extract features. The extracted features undergo further processing via a self-attention module with six transformer encoders, enabling the network to identify long-range correlations within the data. Finally, there are two fully connected layers in the linear module for anomalous diffusion model classification and exponent regression, respectively (see Supplementary Note 7 for details).

**Data augmentation.** To make the network learn OOD features, we propose a general data augmentation method for trajectory data based on the theory of Mixup<sup>28</sup>. This module expands training distribution at the data level and will enhance the network's OOD dynamics detection capabilities (Supplementary Note 2). Specifically, assuming that the source data  $D_s = \{(X_1, 1), (X_2, 2) \dots (X_N, N)\}$  has a total of  $N$  categories of anomalous diffusion trajectories, two types of the generated outliers can be expressed as

$$\tilde{X}_1(t) = \lambda X_i(t) + (1 - \lambda) X_j(t), \quad (6)$$

$$\tilde{X}_2(t) = \begin{cases} X_i(t) & \text{if } 0 < t < \lambda w \\ X_j(t) & \text{if } \lambda w \leq t \leq w, \end{cases} \quad (7)$$

where  $i \neq j$  and  $\lambda \sim \text{beta}(\beta, \beta)$ . Thus, the expanded source data with outliers is  $D_e = \{(X_1, 1), (X_2, 2) \dots (X_N, N), (\tilde{X}_1, N+1), (\tilde{X}_2, N+1)\}$  and the  $N$ -classification task becomes an  $N+1$  classification task. For both training and testing, a Z-score normalization is applied to preprocess each input trajectory to alleviate the effect of the diffusion coefficient on anomalous diffusion recognition. In the actual deployment of our work, the training dataset  $D_{\text{train}}$  has the same distribution as  $D_e$ , with 30,000 samples for each category ( $\beta = 25$ ).

**Training regularization.** The training dataset  $D_{\text{train}}$  has  $N+1$  categories after data augmentation. Assuming that  $\mathcal{D}_s$  is the distribution of source data, there are  $N$  categories and a total of  $K$  trajectories in  $\mathcal{D}_s$ . The rest category represents the generated outliers. We develop a training objective that ensures the classification performance in  $\mathcal{D}_s$  and uses outliers to enhance the boundary sense of  $\mathcal{D}_s$ . The combination of this module and our data augmentation allows the network to improve OOD detection capabilities while maintaining or even enhancing ID recognition performance (Supplementary Note 2). The training objective  $\mathcal{L}$  is a sum of three terms: regression ( $\mathcal{L}_{\text{reg}}$ ), classification ( $\mathcal{L}_{\text{cls}}$ ) and L2-norm ( $\mathcal{L}_{\text{rfl}}$ ) losses,

$$\mathcal{L} = \rho_1 \mathcal{L}_{\text{reg}} + \rho_2 \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{rfl}}. \quad (8)$$

The regression loss  $\mathcal{L}_{\text{reg}}$  computes the simple mean-squared error (MSE) between the predicted and true anomalous exponents. As there are no ground-truth anomalous exponents for the generated outliers, we ignore the contribution of outliers to the MSE during the training process by mixing outliers into  $D_{\text{train}}$  and computing  $\mathcal{L}_{\text{reg}}$  as follows

$$\mathcal{L}_{\text{reg}} = \sum_{i=1}^K \left( \alpha_{\text{source}}^{\text{GT},i} - \alpha_{\text{source}}^{\text{pred},i} \right)^2, \quad (9)$$

where  $\alpha_{\text{source}}^{\text{GT}}$  and  $\alpha_{\text{source}}^{\text{pred}}$  are the ground-truth and predicted anomalous exponent of  $D_{\text{train}}$  in  $\mathcal{D}_s$ , respectively. The classification loss  $\mathcal{L}_{\text{cls}}$  consists of cross-entropy loss  $\mathcal{L}_{\text{ce}}$  for each of expanded training data  $D_e$  and outlier exposure loss  $\mathcal{L}_{\text{oe}}$  for regularization,

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{ce}} + \rho_3 \mathcal{L}_{\text{oe}}, \quad (10)$$

where  $\mathcal{L}_{\text{ce}}$  is a function of the true classification labels and the predicted probability map. With data augmentation, the task changes from  $N$  classification into  $N+1$  classification and  $\mathcal{L}_{\text{ce}}$  is given by

$$\mathcal{L}_{\text{ce}} = - \sum_{i=1}^n p(x_i) \log(q(x_i)), \quad (11)$$

where  $p(x_i), q(x_i) \in \mathbb{R}^{N+1}$  represents the true and predicted distribution of each input  $x_i$  from  $D_{\text{train}}$ , respectively. The purpose of  $\mathcal{L}_{\text{oe}}$  is to allow the outlier to be isolated from  $\mathcal{D}_s$  by the network as much as possible, even if it is misclassified into  $\mathcal{D}_s$ , enabling the network to detect OOD samples in a more fine-grained manner:

$$\mathcal{L}_{\text{oe}} = \lambda \sum_{j=1}^m p(\tilde{x}_j) \log(q(\tilde{x}_j)), \quad (12)$$

where  $p(\tilde{x}_j), q(\tilde{x}_j) \in \mathbb{R}^{N+1}$  represent the true and predicted distribution of each outlier  $\tilde{x}_j$ , respectively;  $\lambda$  is the  $(N+1)$ th element of the softmax vector, corresponding to the probability that  $\tilde{x}_j$  is classified as outlier.

The L2-norm loss is formulated as  $\mathcal{L}_{\text{rfl}} = \|\mathbf{W}\|_2$ , where  $\mathbf{W}$  is the weight matrix of all layers in the network. In this work, the hyperparameters  $\rho_1, \rho_2, \rho_3$  are set to 0.3, 0.7, 0.5, respectively. As accurate recognition of  $D_{\text{train}}$  in the source domain  $\mathcal{D}_s$  is the original task, the validation dataset  $D_{\text{valid}}$  is sampled from  $\mathcal{D}_s$ , with 3,000 trajectories available for each category. For each training epoch, the network parameters with the highest classification performance on  $D_{\text{valid}}$  are saved.

**Confidence computation.** We introduce Bayesian neural network to alleviate overconfidence for OOD dynamics; however, the implementation of BNN is always expensive to train and hard to scale to modern networks<sup>44</sup>. For simple and cost-efficient, yet competitive BNN, Kristiadi and colleagues proposed LLLA<sup>29</sup>, which has proven to capture network uncertainty effectively and is readily applicable to already-trained networks. In LLLA, one applies Laplace approximation post-hoc only on the last layer and the other layers retain the estimate of MAP, protecting the predictive performance of pre-trained networks while equipping networks with robust uncertainty estimates. Thus, the network trained with our data augmentation and regularization can be further deployed with BNN by simply using LLLA, which effectively improves OOD dynamics detection (Supplementary Note 2). Specifically, we infer the weights of the last classification layer  $\mathbf{W}_L \in \mathbb{R}^{k \times d}$  obey Gaussian distribution

$$p(\mathbf{W}_L | D_s) = \mathcal{N}(\mathbf{W}_L | \mathbf{W}_{L, \text{MAP}}, \mathbf{H}^{-1}), \quad (13)$$

where  $k = N+1$  and  $d$  are the number of categories and parameters of the last classification layer, respectively;  $\mathbf{H} \in \mathbb{R}^{dk \times dk}$  is the Hessian of the negative log-posterior  $-\log(D_s | \mathbf{W}_L)$  with regard to  $\mathbf{W}_L$  at  $\mathbf{W}_{L, \text{MAP}}$ . This Hessian could be easily obtained via automatic differentiation and be formulated as  $\mathbf{H}^{-1} \approx \mathbf{V} \otimes \mathbf{U}$  with Kronecker-factored Laplace approximation<sup>45</sup>. Assuming the input vector of the last layer is  $\phi \in \mathbb{R}^d$  and  $\phi^T$  denotes its transpose, the distribution of the network output for classification  $\tilde{f}_{\text{cls}} = \mathbf{W}_L \phi$  can be expressed as

$$p(\tilde{f}_{\text{cls}} | D_s) = \mathcal{N}(\tilde{f}_{\text{cls}} | \mathbf{W}_{L, \text{MAP}} \phi, (\phi^T \mathbf{V} \phi) \mathbf{U}). \quad (14)$$

Therefore, the probability that the network predicts the trajectory  $\vec{x}$  belonging to the  $i$ th category is

$$p(y = i | \vec{x}, D_s) = \int \text{softmax}(\vec{f}, i) \mathcal{N}(\vec{f} | \mathbf{W}_{L, \text{MAP}} \phi, (\phi^T \mathbf{V} \phi) \mathbf{U}) d\vec{f} \quad (15)$$

where  $p(y = N + 1)$  is the probability that  $\vec{x}$  is recognized as the outlier. The confidence score with BNNs  $S_{\text{BNN}}$  is defined as  $S_{\text{BNN}} = \max(p(y = 1), p(y = 2), \dots, p(y = N))$  and the BNNs-based detector  $G_{\text{BNN}}$  is given as

$$G_{\text{BNN}} = \begin{cases} 0 & \text{if } S_{\text{BNN}}(\vec{x}) < \gamma \\ 1 & \text{if } S_{\text{BNN}}(\vec{x}) \geq \gamma, \end{cases} \quad (16)$$

where  $\gamma$  is the confidence threshold and  $G_{\text{BNN}}$  assigns label 0 for OOD dynamics and label 1 for ID dynamics. The full computation of  $S_{\text{BNN}}$  is shown in Supplementary Algorithm 1. Moreover, a distance-based detector  $G_{\text{dis}}$  is proven to be effective in detecting OOD dynamics that  $G_{\text{BNN}}$  fails to identify, which can be expressed as

$$G_{\text{dis}}(\vec{x}, \vec{\alpha}, R^2) = \begin{cases} 0 & \text{if } \|F(\phi_n(\vec{x})) - \vec{\alpha}\|_2 > R^2 \\ 1 & \text{if } \|F(\phi_n(\vec{x})) - \vec{\alpha}\|_2 \leq R^2 \end{cases} \quad (17)$$

where  $\phi_n(\cdot)$  is the  $n$ th layer output;  $\vec{\alpha}$  and  $R^2$  are parameters of the hypersphere trained from  $D_{\text{dis}} = \{\phi_n(\vec{x}_i), \vec{x}_i \in D_s\}$  using SVDD<sup>46</sup> with slack variable  $\xi$  and penalty constant  $C$ ;  $F(\cdot)$  is a nonlinear mapping relationship given by the following optimization problem

$$\begin{aligned} \min & \left( R^2 + C \sum_{i=1}^n \xi_i \right) \\ \text{s.t. } & \|F(\phi_n(\vec{x}_i)) - \vec{\alpha}\|_2 \leq R^2 + \xi_i \text{ and } \xi_i \geq 0 \forall_i. \end{aligned} \quad (18)$$

where s.t. indicates subject to.

### Assessment metrics

Given the expected improvements in both ID dynamics recognition and OOD dynamics detection, we utilize several metrics to evaluate and compare the performance of our method with others. These metrics include the MAE,  $F_1$ -score, AUROC and FPR95.

**MAE.** The MAE is a frequently employed metric for evaluating regression tasks; it measures the average extent of deviation between the predictions of the network and the desired outputs. We use the MAE to quantify the performance of the network in predicting anomalous exponents:

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |\alpha_{\text{GT}, i} - \alpha_{\text{pred}, i}|, \quad (19)$$

where  $M$  is the number of trajectories in the test dataset, and  $\alpha_{\text{GT}, i}$  and  $\alpha_{\text{pred}, i}$  represent the ground-truth and the predicted values of the anomalous exponent of the  $i$ th trajectory, respectively.

**$F_1$ -score.** The  $F_1$ -score comprehensively considers both the precision and recall performance of the classification network and is equally applicable for multiclassification tasks. To mitigate potential bias caused by uneven proportions of dynamics categories in the dataset, we employ the micro-averaged  $F_1$ -score in our study:

$$F_1 = \frac{2\text{TP}_{\text{source}}}{2\text{TP}_{\text{source}} + \text{FP}_{\text{source}} + \text{FN}_{\text{source}}}, \quad (20)$$

where  $\text{TP}_{\text{source}}$ ,  $\text{FP}_{\text{source}}$  and  $\text{FN}_{\text{source}}$  represent true positives, false positives and false negatives, respectively, calculated over the test dataset in  $D_s$ . The closer the  $F_1$ -score value is to 1, the better the network's classification performance is when faced with ID dynamics.

**AUROC.** AUROC is the area under the ROC curve, which is a comprehensive measure of the network's ability to detect OOD samples. The ROC curve represents the false positive rate  $\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$  on the x-axis and the true positive rate  $\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$  on the y-axis, where OOD samples are considered positive and ID samples are negative; TP, FP and FN represent true positives, false positives and false negatives calculated over the whole dataset, respectively. The AUROC value ranges from 0 to 1, where 0.5 represents random OOD detection ability and 1 represents perfect OOD detection ability.

**FPR95.** FPR95 is the value of FPR (represents the probability that the ID sample is incorrectly identified as OOD by the network) when the TPR (represents the probability that the OOD sample is correctly detected by the network) reaches 95%. A lower FPR95 value indicates the network's better performance in detecting OOD dynamics.

### Experimental data acquisition

In our work, we deal with three types of actual data: the dynamics of nicotinic acetylcholine receptors, fluorescent beads in dextran solutions and silver nanoparticles undergoing endocytosis. The trajectories of nicotinic acetylcholine receptors come from the dataset published by Maizón et al. available at [https://github.com/hectorbm/DL\\_anomalous\\_diffusion](https://github.com/hectorbm/DL_anomalous_diffusion) (ref. 32). The procedure developed by Golan and Sherman<sup>47</sup> was applied to filter out immobile trajectories.

The trajectories of fluorescent beads (BangsLab, FSDG002, FSSY002) in dextran solution and silver nanoparticles (Xfnano, 103716) undergoing endocytosis were obtained using our home-built SPT system<sup>48,49</sup>. In the SPT system, we implement an active feedback tracking approach for real-time particle tracking (further details are elaborated in Supplementary Note 8). Briefly, we use the electro-optic deflector and tunable acoustic gradient lens to swiftly scan a  $1 \times 1 \times 2 \mu\text{m}$  three-dimensional region surrounding the target particle. Position of the particle is then determined based on intensity measurements, subsequently enabling the piezoelectric device to actively track the trajectory of particle in real time.

The dextran solution is obtained by dissolving dextran (500 kDa, Sigma) in millipore water at concentrations of about  $430 \text{ mg ml}^{-1}$  (30% w/w). The tracer beads were added from a predissolved solution, resulting in a typical concentration of about 2 pM. For SPT of the silver nanoparticle undergoing endocytosis, Hela cells were cultured overnight in a confocal dish with Dulbecco's Modified Eagle Medium containing 10% fetal bovine serum and 1% double penicillin/streptomycin. Before tracking, cells were set on the ice for 5 min to inhibit the endocytosis. Silver NPs were diluted in DPBS to  $10 \mu\text{g ml}^{-1}$  and then incubated in the confocal dish on ice for 10 min and then washed twice with DPBS. During tracking, cells were kept in the 1 ml living cell buffer.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data that support the findings of this work are available on Code Ocean at <https://doi.org/10.24433/CO.6518632.v5> (ref. 50). Source data are provided with this paper.

### Code availability

Our reproduction code and the relevant documentation are available on Code Ocean at <https://doi.org/10.24433/CO.6518632.v5> (ref. 50).

### References

1. Kindermann, F. et al. Nonergodic diffusion of single atoms in a periodic potential. *Nat. Phys.* **13**, 137–141 (2017).

2. Zhang, Y. & Hess, H. Chemically-powered swimming and diffusion in the microscopic world. *Nat. Rev. Chem.* **5**, 500–510 (2021).
3. Bronstein, I. et al. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.* **103**, 018102 (2009).
4. Rempel, A., Waddington, E., Wettlaufer, J. & Worster, M. Possible displacement of the climate signal in ancient ice by premelting and anomalous diffusion. *Nature* **411**, 568–571 (2001).
5. Poletayev, A. D., Dawson, J. A., Islam, M. S. & Lindenberg, A. M. Defect-driven anomalous transport in fast-ion conducting solid electrolytes. *Nat. Mater.* **21**, 1066–1073 (2022).
6. Scalas, E. The application of continuous-time random walks in finance and economics. *Phys. A* **362**, 225–239 (2006).
7. Chechkin, A. V., Seno, F., Metzler, R. & Sokolov, I. M. Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities. *Phys. Rev. X* **7**, 021002 (2017).
8. Krapf, D. Mechanisms underlying anomalous diffusion in the plasma membrane. *Curr. Top. Membr.* **75**, 167–207 (2015).
9. Weiss, M., Elsner, M., Kartberg, F. & Nilsson, T. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys. J.* **87**, 3518–3524 (2004).
10. Sokolov, I. M. & Klafter, J. From diffusion to anomalous diffusion: a century after Einstein's Brownian motion. *Chaos* **15**, 26103 (2005).
11. Metzler, R., Jeon, J.-H., Cherstvy, A. G. & Barkai, E. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **16**, 24128–24164 (2014).
12. Mandelbrot, B. B. & van Ness, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437 (1968).
13. Scher, H. & Montroll, E. W. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B* **12**, 2455 (1975).
14. Klafter, J. & Zumofen, G. Lévy statistics in a Hamiltonian system. *Phys. Rev. E* **49**, 4873 (1994).
15. Lim, S. C. & Muniandy, S. V. Self-similar Gaussian processes for modeling anomalous diffusion. *Phys. Rev. E* **66**, 021114 (2002).
16. Massignan, P. et al. Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium. *Phys. Rev. Lett.* **112**, 150603 (2014).
17. Sokolov, I. M. Models of anomalous diffusion in crowded environments. *Soft Matter* **8**, 9043–9052 (2012).
18. Mangalam, M., Metzler, R. & Kelty-Stephen, D. G. Ergodic characterization of nonergodic anomalous diffusion processes. *Phys. Rev. Res.* **5**, 023144 (2023).
19. Ślęzak, J., Metzler, R., & Magdziarz, M. Codifference can detect ergodicity breaking and non-Gaussianity. *New J. Phys.* **21**, 053008 (2019).
20. Magdziarz, M., Weron, A., Burnecki, K. & Klafter, J. Fractional Brownian motion versus the continuous-time random walk: a simple test for subdiffusive dynamics. *Phys. Rev. Lett.* **103**, 180602 (2009).
21. Li, J. Role of ergodicity, aging and Gaussianity in resolving the origins of biomolecule subdiffusion. *Phys. Chem. Chem. Phys.* **24**, 16050–16057 (2022).
22. Muñoz-Gil, G. et al. Objective comparison of methods to decode anomalous diffusion. *Nat. Commun.* **12**, 6253 (2021).
23. Seckler, H. & Metzler, R. Bayesian deep learning for error estimation in the analysis of anomalous diffusion. *Nat. Commun.* **13**, 6717 (2022).
24. Muñoz-Gil, G., i Corominas, G. G. & Lewenstein, M. Unsupervised learning of anomalous diffusion data: an anomaly detection approach. *J. Phys. A* **54**, 504001 (2021).
25. Hendrycks, D., Mazeika, M. & Dietterich, T. Deep anomaly detection with outlier exposure. In *Proc. International Conference on Learning Representations (ICLR)*, (2019).
26. Hendrycks, D. & Gimpel, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *Proc. International Conference on Learning Representations (ICLR)*, (2016).
27. Yang, J. et al. OpenOOD: benchmarking generalized out-of-distribution detection. *Adv. Neural Inf. Process. Syst.* **35**, 32598–32611 (2022).
28. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. Mixup: beyond empirical risk minimization. In *Proc. International Conference on Learning Representations (ICLR)*, (2018).
29. Kristiadi, A., Hein, M. & Hennig, P. Being Bayesian, even just a bit, fixes overconfidence in ReLU networks. In *Proc. 37th International Conference on Machine Learning* 5436–5446 (PMLR, 2020).
30. Firbas, N., Garibo-i-Orts, Ò., Garcia-March, M. Á. & Conejero, J. A. Characterization of anomalous diffusion through convolutional transformers. *J. Phys. A* **56**, 014001 (2023).
31. Elston, T. C. A macroscopic description of biomolecular transport. *J. Math. Biol.* **41**, 189–206 (2000).
32. Maizón, H. B. & Barrantes, F. J. A deep learning-based approach to model anomalous diffusion of membrane proteins: the case of the nicotinic acetylcholine receptor. *Brief. Bioinform.* **23**, bbab435 (2022).
33. Qu, X. et al. Semantic segmentation of anomalous diffusion using deep convolutional networks. *Phys. Rev. Res.* **6**, 013054 (2024).
34. Sinai, Y. G. The limiting behavior of a one-dimensional random walk in a random medium. *Theory Probab. Appl.* **27**, 256–268 (1983).
35. Uhlenbeck, G. E. & Ornstein, L. S. On the theory of the Brownian motion. *Phys. Rev.* **36**, 823–841 (1930).
36. Ernst, D., Hellmann, M., Köhler, J. & Weiss, M. Fractional Brownian motion in crowded fluids. *Soft Matter* **8**, 4886–4889 (2012).
37. Zaburdaev, V., Denisov, S. & Klafter, J. Lévy walks. *Rev. Mod. Phys.* **87**, 483 (2015).
38. Guarneri, D. et al. Transport across the cell-membrane dictates nanoparticle fate and toxicity: a new paradigm in nanotoxicology. *Nanoscale* **6**, 10264–10273 (2014).
39. Krapf, D. et al. Spectral content of a single non-Brownian trajectory. *Phys. Rev. X* **9**, 011019 (2019).
40. Balcerek, M., Burnecki, K., Thapa, S., Wytomańska, A. & Chechkin, A. Fractional Brownian motion with random Hurst exponent: accelerating diffusion and persistence transitions. *Chaos* **32**, 093114 (2022).
41. Jin, H., Heller, D. A. & Strano, M. S. Single-particle tracking of endocytosis and exocytosis of single-walled carbon nanotubes in NIH-3T3 cells. *Nano Lett.* **8**, 1577–1585 (2008).
42. Miller, J. How many participants? How many trials? Maximizing the power of reaction time studies. *Behav. Res. Methods* **56**, 2398–2421 (2024).
43. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
44. Daxberger, E. et al. Laplace redux—effortless Bayesian deep learning. In *35th Conference on Neural Information Processing Systems* 20089–20103 (NeurIPS, 2021).
45. Ritter, H., Botev, A. & Barber, D. A scalable Laplace approximation for neural networks. In *Proc. International Conference on Learning Representations (ICLR)*, (2018).
46. Ruff, L. et al. Deep one-class classification. In *Proc. International Conference on Learning Representations* Vol. 80, 4393–4402 (ICLR, 2018).
47. Golan, Y. & Sherman, E. Resolving mixed mechanisms of protein subdiffusion at the T cell plasma membrane. *Nat. Commun.* **8**, 15851 (2017).
48. Sha, H., Li, H., Zhang, Y. & Hou, S. Deep learning-enhanced single-molecule spectrum imaging. *APL Photon.* **8**, 096102 (2023).
49. Hou, S., Exell, J. & Welshe, K. Real-time 3D single molecule tracking. *Nat. Commun.* **11**, 3607 (2020).

50. Xiao, F. Reliable deep learning in anomalous diffusion against out-of-distribution dynamics. *Code Ocean* <https://doi.org/10.24433/CO.6518632.v5> (2024).

## Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (grant nos. 62031023 and 62331011), and the Shenzhen Science and Technology Project (grant no. GXWD20220818170353009).

## Author contributions

X.F. conducted the experiments, analyzed the results and wrote the paper, with feedback from all authors. H.S. and S. Hou led the construction of optical system. Y.Z. and X.J. supervised the project. X.F., H.S. and Z.Y. proposed the initial idea. Y.S. took part in the experiments. Y.J. took part in the code design. S.L. and S. Han contributed to the results analysis.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43588-024-00703-7>.

**Correspondence and requests for materials** should be addressed to Yongbing Zhang or Xiangyang Ji.

**Peer review information** *Nature Computational Science* thanks Damian G. Kelty-Stephen, Diego Krapf and Gorka Muñoz-Gil for their contribution to the peer review of this work. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024



Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated  |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Simulated data were generated using the python package: andi-datasets 2.0.3. Experimental data were previously published ( <a href="https://github.com/hectorbm/DL_anomalous_diffusion">https://github.com/hectorbm/DL_anomalous_diffusion</a> ) and collected by our home-built tracking system ( <a href="https://doi.org/10.24433/CO.6518632.v1">https://doi.org/10.24433/CO.6518632.v1</a> ).
Data analysis	python=3.10.9, pytorch=2.0.1, and pyod=1.0.9 were used for model training. To analyze results, we used numpy=1.5.0, pandas=1.5.3, and backpack=0.1. For generating trajectories, we used andi-datasets 2.0.3. For plotting, we used matplotlib=0.1.6 and seaborn=0.12.2. scikit-learn was used to calculate AUROC and F1-score. All codes for simulations and analysis are available on Code Ocean <a href="https://doi.org/10.24433/CO.6518632.v1">https://doi.org/10.24433/CO.6518632.v1</a>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The trajectories of fluorescent beads in sucrose and Ag nanoparticles with active endocytosis were obtained by our experiment collection. All simulated and experimental datasets are available on Code Ocean <https://doi.org/10.24433/CO.6518632.v1>

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="The study do not involve human participants."/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="-"/>
Population characteristics	<input type="text" value="-"/>
Recruitment	<input type="text" value="-"/>
Ethics oversight	<input type="text" value="-"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="For dynamics of nicotinic acetylcholine receptors, sample size was determined by previous published datasets (https://github.com/hectorbm/DL_anomalous_diffusion). For dynamics of fluorescent beads in dextran solutions and Ag nanoparticles undergoing active endocytosis, 700 and 320 trajectories of length 50 are observed by our home-built tracking system (https://doi.org/10.24433/CO.6518632.v1)."/>
Data exclusions	<input type="text" value="No data was excluded."/>
Replication	<input type="text" value="All data and code are open source to help with the reproducibility of computational analysis."/>
Randomization	<input type="text" value="For experimental data, our work is a descriptive observational study and does not involve multiple experimental groups, hence randomization was not applicable."/>
Blinding	<input type="text" value="The trajectory analyzed in this study was used to illustrate the reliability of deep learning and not to test a specific hypothesis. Therefore, blinding is not relevant for this work."/>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

## Methods

- n/a Involved in the study
- ☒ ☐ Antibodies
- ☐ ☒ Eukaryotic cell lines
- ☒ ☐ Palaeontology and archaeology
- ☒ ☐ Animals and other organisms
- ☒ ☐ Clinical data
- ☒ ☐ Dual use research of concern
- ☒ ☐ Plants

- n/a Involved in the study
- ☒ ☐ ChIP-seq
- ☒ ☐ Flow cytometry
- ☒ ☐ MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Hela (from ATCC).

Authentication

The cell line is not authenticated.

Mycoplasma contamination

The cell line is not tested for mycoplasma contamination.

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified cell line was used.

## Plants

Seed stocks

-

Novel plant genotypes

-

Authentication

-