Journal of Physics: Photonics

OPEN ACCESS

RECEIVED
7 December 2024

REVISED
6 March 2025

ACCEPTED FOR PUBLICATION
10 March 2025

PUBLISHED
20 March 2025

PAPER

Temporal convolutional networks work as general feature extractors for single-particle diffusion analysis

Lucas A Saavedra and Francisco J Barrantes*

Laboratory of Molecular Neurobiology, Biomedical Research Institute (BIOMED) UCA–CONICET, Av. Alicia Moreau de Justo 1600, C1107AFF Buenos Aires, Argentina

* Author to whom any correspondence should be addressed.

E-mail: francisco_barrantes@uca.edu.ar

Keywords: convolutional network, single-molecule, deep learning, 2d trajectory analysis, change-point detection, acetylcholine receptor, pointwise analysis

Supplementary material for this article is available online

Abstract

The application of deep learning (DL) approaches to the study of particle diffusion is becoming increasingly widespread, achieving highly accurate predictive performance that is often challenging to attain with traditional analytical tools. DL has made successful incursions in the field of single-particle tracking of biological molecules in 2- and 3-dimensional spaces. In this work, we introduce a WaveNet-based deep neural network (WadNet)- and convolution-based DL network coined Wad temporal convolutional networks (WadTCN) and its transformer-based variant to decode single-value or pointwise prediction of dynamic properties of bidimensional trajectories. We trained WadTCN to be tested in all 1st and 2nd Andi Challenge tasks with bidimensional trajectories, examined how WadTCN performed when learning is transferred from one network to another to accelerate training, and evaluated its capacity to segment trajectories with a data-driven approach, at variance with other works that rely on mean changes or parameter thresholding to determine change points. WadTCN was further compared to a pure temporal convolutional network and RANDI, a top-performing recurrent neural network in the 1st Andi Challenge. We also assessed the ability of WadTCN to classify and segment single-particle trajectories of a membrane-associated protein, the nicotinic acetylcholine receptor, in the plane of a live cell plasma membrane. WadTCN outperformed similar analytical tools in predicting anomalous diffusion exponents and demonstrated notable robustness in more challenging tasks like the segmentation of trajectories into two states and prediction of the diffusion coefficient of diffusing molecules. Furthermore, most layers of WadTCN perform as general feature extractors, suggesting its applicability in other tasks requiring training acceleration.

1. Introduction

In 1905, Pearson introduced the concept of 'random walk', a mathematical model to describe the random sequence of steps of an observable (herein, particle) along a trajectory (1905). This concept is widely employed in fields as diverse as biology and economy (Muñoz-Gil *et al* 2021b). It is usual to describe the collective motion of particles with the mean-square displacement (MSD) function, which a statistical mechanics concept that measures the deviation of the position of a particle with respect to a reference position or, more intuitively, how much space the particles explore at specific time points. The MSD function follows the relationship MSD $\propto t^{\alpha}$, α being the anomalous exponent. When the MSD follows a linear relationship ($\alpha = 1$), the trajectory is characterized by the Brownian model consisting of independent displacements following a gaussian distribution (Krapf 2015). When $\alpha \neq 1$, the diffusion is anomalous. Given $\alpha > 1$ or $\alpha < 1$ the particle behaves super- or sub-diffusively, respectively.

Several models have been proposed to describe anomalous diffusion, among which are the continuous-time random walk (CTRW; subdiffusive motion characterized by random waiting times between

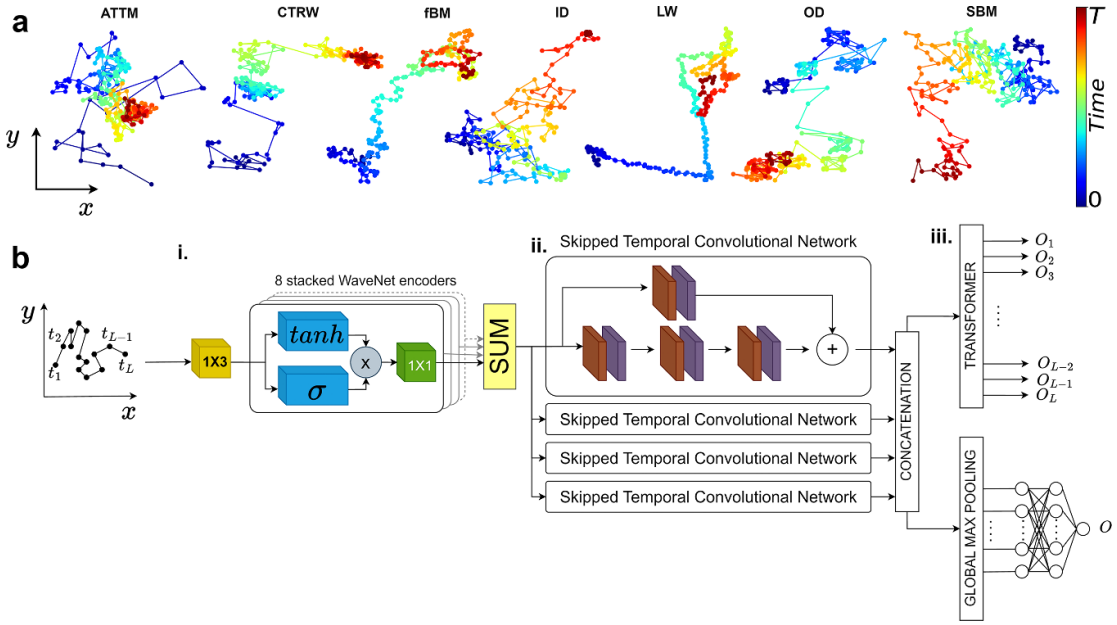

Figure 1. Schematic diagram illustrating the concatenated steps of the proposed architecture and examples of simulated trajectories. (a) Examples of simulated 2D trajectories following ATTM, CTRW, fBM, ID, LW, OD, and SBM models. Simulation details can be found in supplementary material. (b) Detailed structure of WadTCN using as input a trajectory of length L and duration $t_L - t_1$. In the case of global max pooling subsequent to feature concatenation, only one output O is predicted; when the transformer is used, several L outputs are predicted. O_i can be a number or category for regression or classification tasks, respectively.

steps) (Scher and Montroll 1975, Krapf 2015), Lévy Walk (LW; the superdiffusive case of CTRW) (Shlesinger *et al* 1982, Klafter and Zumofen 1994), annealed transient time motion (ATTM; Brownian diffusion with random diffusion coefficients along time) (Massignan *et al* 2014, Manzo *et al* 2015), scaled brownian motion (SBM; motion with increasing or decreasing diffusion coefficient for $\alpha > 1$ or $\alpha < 1$, respectively) (Lim and Muniandy 2002), and fractional Brownian motion (fBM; correlated displacements following $\alpha < 1$ and $\alpha > 1$ for anti- or correlated steps, respectively) (Mandelbrot and Van Ness 1968). Examples of simulated trajectories following the models discussed in this work can be found in figure 1(a); the detailed mathematical description of the models can be found in supplementary material. Although MSD is a valuable tool to quantitatively estimate α and derive the diffusion coefficient D, and qualitatively (e.g. model classification) analyze trajectories (Qian *et al* 1991), its use is problematic for two reasons: (i) more often than not localization noise introduced by experiments can provoke inaccurate estimation of α and (ii) the number and length of the trajectories are difficult to obtain in sufficiently large quantities to be able to accurately characterize the system under study (Muñoz-Gil *et al* 2021b). Hidden Markov Models were also proposed to tackle trajectory characterization (Slator *et al* 2015, Slator and Burroughs 2018) as well as Bayesian methods (Krog *et al* 2018) and feature-based analyses (Aghion *et al* 2021). Deep learning (DL) has provided an alternative and more reliable approach than such methods in terms of accuracy (Muñoz-Gil *et al* 2021b).

A concatenated DL-based analytical approach was recently introduced to analyze the diffusion of proteins in two-dimensions (Buena-Maizón and Barrantes 2022). The approach was applied to a real case scenario, i.e. the analysis of single-particle 2D-diffusion of a neurotransmitter receptor protein in the plasma membrane, using experimental data obtained with single-molecule localization microscopy and single-particle tracking (SPT) techniques. Three physical models were considered: fBM, CTRW, and a two-state model, i.e. trajectories undergoing free Brownian motion hampered by transient sojourns of obstructed diffusion (OD model) (Mosqueira *et al* 2018, 2020, Grebenkov 2019, Buena-Maizón and Barrantes 2022). The analytical method, which employed temporal convolutional networks (TCNs) (Bai *et al* 2018, Granik *et al* 2019), showed an outstanding prediction performance on short trajectories.

The 1st Andi Challenge, aimed at obtaining an objective comparison of different machine learning (ML) and DL approaches (Muñoz-Gil *et al* 2021b), included other physical models of diffusion, such as LW, ATTM, SBM, but not the two-state model included in the study of Buena-Maizón and Barrantes (2022). The methods covered in the challenge were tested for inference of α (Task 1), model classification (Task 2), and single-point trajectory segmentation (Task 3). One of the Andi challenge results showed that long-short term memory (LSTM)-based networks (Hochreiter and Schmidhuber 1997) outperform TCNs, as exemplified by the LSTM network with a WaveNet encoder (van den Oord *et al* 2016, Li *et al* 2021). However, the

combination of the encoder with TCNs was not explored, thus possibly excluding a similar but more efficient classification and regression performance than LSTM alone (Bai *et al* 2018, Yu *et al* 2019). Subsequent developments have resulted in new DL techniques to tackle anomalous diffusion with modern architectures. Conejero and coworkers took advantage of pretrained computer vision models to differentiate among various theoretical models and predict the anomalous exponent of short trajectories upon converting them into images (Garibo-i-Orts *et al* 2023). These authors also have used transformers, a modern architecture for dealing with sequences (Vaswani *et al* 2017), for the same purpose (Firbas *et al* 2023). However, methods presented in the 1st Andi Challenge and those published thereafter are limited to predicting a single value with compatibility for homogeneous but not heterogeneous trajectories.

There was recently a 2nd Andi Challenge to compare methods capable of predicting dynamic properties at every point of trajectories (Muñoz-Gil *et al* 2023). Manzo and coworkers presented STEP, a method based on transformers to analyze the pointwise dynamics of a trajectory and inferring, for example, how the diffusion coefficient changes with time (Requena *et al* 2023). (Qu *et al* 2024) presented a network which combines U-Net and WaveNet encoders repeatedly, which can be inefficient due to the sequential nature of WaveNet encoders. U-Nets were also proposed as processing channels in Deep-SPT (Kæstel-Hansen *et al* 2024) and Bayesian Deep Learning was proposed to leverage change-point prediction (Seckler and Metzler 2024). Although these works showed an outstanding change-point detection (CPD) performance, their reported methodologies have an important constraint: CPD; the task of detecting at which point of the trajectories diffusional properties change) in these works relies on a sole dynamic property (e.g. the diffusion coefficient) and overlook changes in other properties (e.g. diffusion state). Furthermore, CPD is carried out using parameter-based thresholding methods or offline CPD algorithms (see Truong *et al* 2020).

The present work introduces three novel contributions. Firstly, we present a TCN with a WaveNet encoder as an input pre-processor called WadTCN to tackle single-value prediction (both classification and regression). This step is tested with the Tasks 1 and 2 proposed in the 1st Andi Challenge and is further compared with RANDI, a consistently top-3 ranked recurrent neural network (RNN) launched in the first Andi Challenge (Argun *et al* 2021, Muñoz-Gil *et al* 2021b), and with the TCN-only approach implemented in our laboratory (Buena-Maizón and Barrantes 2022). Secondly, we show how WadTCN can easily be extended to pointwise inference and work as a CPD algorithm with a transformer (Vaswani *et al* 2017), using as input the trajectory instead of the diffusion coefficient and anomalous exponent inferred with other neural networks; this aspect is tested with task 3 of the 1st Andi Challenge and the trajectories-based tasks of the 2nd Andi Challenge. Thirdly, we show that the first layers of WadTCN can enhance training times and be leveraged to act as general feature extractors that can be extensively reutilized for other purposes.

In order to analyze the translational dynamics of a typical integral membrane protein, the nicotinic acetylcholine receptor (nAChR), in the plane of the plasma membrane of live CHO-K1/A5 cells, we extended the number of physical models to seven: those included in the Andi Challenge (ATTM, SBM, LW, fBM, and CTRW), the aforementioned two-state model (Obstructed-Brownian, OD) (Mosqueira *et al* 2018, 2020, Grebenkov 2019, Buena-Maizón and Barrantes 2022), and a second two-state model (trapped diffusion, TD) which considers a particle oscillating between trapped/immobilized diffusion and Brownian diffusion (Eggeling 2015). For this analysis, we used biological scaled simulations of the cases considered in the Andi Challenge and transfer learning to significantly accelerate the training process. We furthermore used the simulations provided by the 2nd Andi Challenge to interrogate the diffusion coefficient in biological trajectories that follow the fBM-type of diffusion and segment long trajectories into the four states interrogated in the second competition: immobilized (ID), confined (CD), free (FR), and directed (DI). To the best of our knowledge, this is the first reported architecture to undertake regression and classification tasks with more than five theoretical models in parallel and analyze supervised multiple-CPDs in single-particle trajectories.

2. Methods

2.1. The concatenated DL algorithm for single-value prediction

The concatenated DL architecture for the 1st Andi Challenge presented here is length-specific (i.e. all architectures detailed in this section are trained with a specific trajectory length). First, the trajectory is classified into one of 7 models: CTRW, fBM, LW, SBM, ATTM, OD, and TD. For the purpose of comparison with other architectures and to search for hyperparameters, OD and TD were excluded and only used to analyze actual biological trajectories obtained from superresolution microscopy data of plasma membrane receptors. Subsequently, trajectories falling into the fBM and SBM modalities are sub-classified into superdiffusive, subdiffusive, or 'likely-Brownian' (i.e. nearly Brownian diffusion) as implemented in Buena-Maizón and Barrantes (2022) for fBM. Finally, the Hurst exponent ($H = \alpha/2$), which also indicates how fast a molecule diffuses (Mandelbrot and Van Ness 1968), is calculated for all trajectories using a

model-specific predictor (i.e., a predictor that works exclusively for a given model). Furthermore, two additional architectures are included: one to predict the diffusion coefficient of those trajectories classified as fBM, and a second to perform pointwise prediction on TD trajectories to determine whether the particle is trapped or freely diffusing. The new concatenated architecture is composed of 3 classifiers, 10 regression length-specific models, and 1 segmenter. The segmenter is the only architecture using a transformer layer inside this workflow and not applying change point detection through prominence, as described below. The algorithm workflow is shown in supplementary figure 1.

2.2. The concatenated DL algorithm for pointwise prediction

The DL algorithm for pointwise inference prediction of the behavior of trajectories, as introduced in the 2nd Andi Challenge, is composed of 4 predictive networks: (i) anomalous exponent, (ii) diffusion coefficient, (iii) state, and (iv) change-points. The latter predicts the probability that a point within a trajectory is a change point. If only one change point is predicted, we take the point with maximum probability of being a change point. If multiple change points are detected, we convert this set of predictions into change point positions detecting probability peaks with a minimum height of 0.9 and a prominence 0.2 using the Python package SciPy (Virtanen *et al* 2020). The prominence is a widely used concept in geography and applicable to biology (Griffié *et al* 2015): it is the peak height relative to neighboring peaks. Those peaks not filtered by the prominence-based detector are considered to be the change-point positions. Once change points are detected, segments are extracted and assigned the average anomalous exponent, diffusion coefficient, and state predicted by networks (i), (ii) and (iii), respectively, within these segments. The state predictor classifies each point into 4 states: immobile, confined, free, and directed, as detailed and proposed in the 2nd Andi Challenge (Muñoz-Gil *et al* 2023). Supplementary figure 2 shows the methodology for pointwise predictions.

2.3. Implementation of the proposed architectures

WadTCN consists of three parts (figure 1(b)): (i) a WaveNet encoder stack, (ii) a skipped TCN (sTCN), i.e. the combination of TCNs with Skipped Connections, and (iii) a multilayer perceptron (MLP) or a simplified version of the Transformer proposed in Requena *et al* (2023) which returns a single-value or pointwise (a value for each trajectory point) prediction, respectively. In the case of classification, the output is the number of categories to be classified with the SoftMax activation function. In the case of regression, the output comes from a neuron which can use any activation function. In this work, most of the layers of the trained networks use the rectified linear unit as the activation function (ReLU) (Glorot *et al* 2011). Those networks trained for single-value regression tasks use the scaled exponential linear unit activation function (Klambauer *et al* 2017) in the second to last layer.

WadTCN takes as input the stacked normalized displacements Δx and Δy of one 2D trajectory along the 'x' and 'y' axes except for models designed to calculate the anomalous exponent, in which case the trajectories take normalized trajectories as input. In the case of biological trajectories, the time difference between subsequent steps of the track, Δt, is added because the two displacement arrays have not been normalized and, furthermore, it increases the classification accuracy. Next, a maximum of 64 (1×3) convolutional filters are applied to extract the features processed by a stack of WaveNet encoders like the one used in Li *et al* (2021). In the present work, we only use 8 stacked encoders. The idea behind stacking is that the results of one encoder are passed on to the next encoder and are subsequently summed. The summed results are passed through several TCNs with skip connections that prevent vanishing gradients (He *et al* 2016) as previously shown to work well for anomalous diffusion analysis (Gajowczyk and Szwabiński 2021). Finally, all the features produced by the sTCN are concatenated, evaluated by global max pooling (Lin *et al* 2014), and passed through a shallow MLP. In the case of single-value regression tasks, we added two simple LSTM layers after the pooling layer since we noticed that it improved the performance of anomalous exponent prediction on short trajectories. A graphical description of WadTCN is shown in figure 1(b). All networks were developed with a TensorFlow Python package (Abadi *et al* 2016) and are available in the GitHub repository of this paper. Hyperparameter search results are shown in supplementary figures 3–6.

2.4. Statistical analysis

Multi-label classification performance and anomalous exponent prediction were measured with micro F1-Score and mean absolute error (MAE), both implemented by the Scikit-Learn software package. The non-parametric Kruskal-Wallis test was used to compare two or more distributions. If $p < 0.05$, the null hypothesis is rejected.

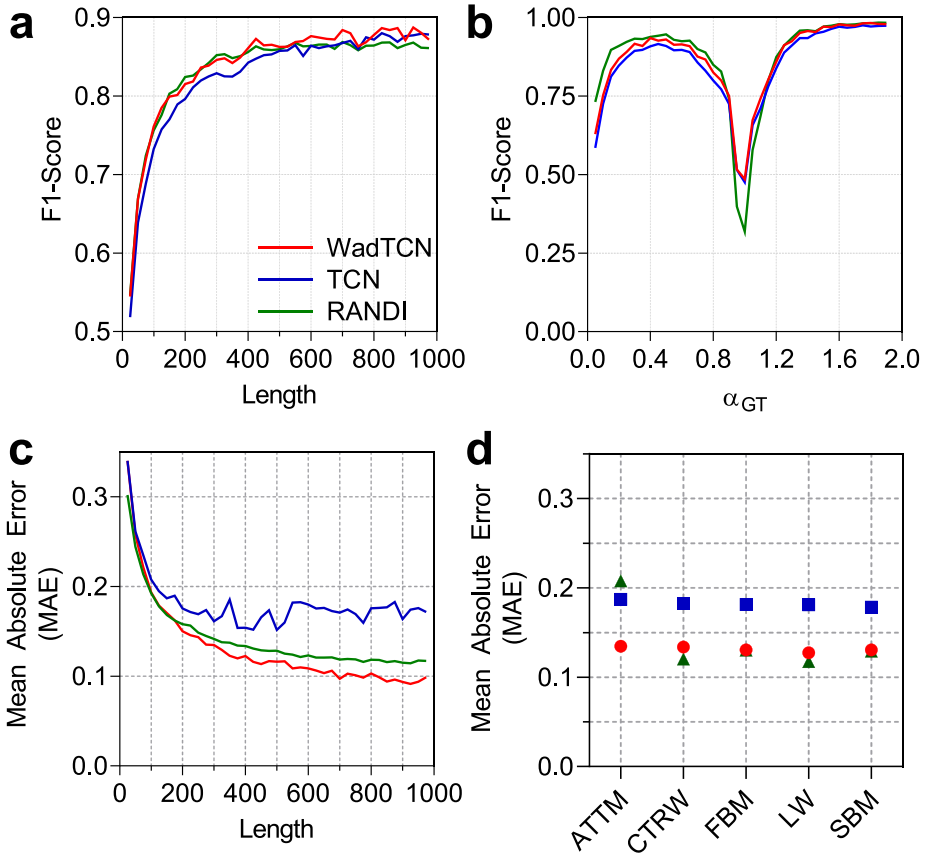

Figure 2. Predictive performance of different single-value architectures. (a) Classification performance (F1-score) of the compared architectures at different trajectory lengths using the pure TCN (Buena-Maizón and Barrantes 2022) (*blue solid line*), RANDI (Argun *et al* 2021) (*green solid line*) and WadTCN (*red solid line*). (b) Classification performance (F1-score) as a function of different anomalous exponents using the pure TCN (Buena-Maizón and Barrantes 2022) (*blue solid line*), RANDI (Argun *et al* 2021) (*green solid line*) and WadTCN (*red solid line*). 12 500 trajectories for each length between 25 and 975 steps with step size of 25 were used for (a) and (b), respectively. (c) The mean absolute error (MAE) incurred in the prediction of the anomalous exponent across different trajectory lengths using the pure TCN (Buena-Maizón and Barrantes 2022) (*blue solid line*), RANDI (Argun *et al* 2021) (*green solid line*) and WadTCN (*red solid line*). (d) The MAE of the anomalous exponent prediction across different theoretical models using the pure TCN (Buena-Maizón and Barrantes 2022) (*blue square*), RANDI (Argun *et al* 2021) (*green triangle*) and WadTCN (*red dot*).

3. Results

3.1. Performance of theoretical model classification between LSTM cells and TCNs architectures

One of the most important conclusions from the 1st Andi Challenge is that RNN architectures perform better than other architectures (Manzo *et al* 2023). The main difference between LSTM and TCN is their recurrency: it has been shown that TCNs have a longer memory than recurrent architectures (Bai *et al* 2018). Furthermore, how neural networks are trained has a direct impact on their performance. Li *et al* (2021) combined a WaveNet encoder with LSTM cells. However, the combination of a WaveNet encoder with convolutional layers and LSTM cells was not explored. Figure 2 shows a quantitative comparison between three different architectures to tackle regression, classification and segmentation of three different predictive models: a simple TCN (Buena-Maizón and Barrantes 2022), which is based on the principles described in Granik *et al* (2019), RANDI (Argun *et al* 2021), and the one proposed in this work, WadTCN. Regarding model classification, TCN performs worse than the other two architectures for short trajectories (i.e., trajectories of $\leqslant$100 steps). As shown in figure 2(a), as the length of the trajectory increases above $\sim$600, TCN outperforms RANDI. WadTCN performs similar to RANDI along different initial lengths, and it slightly exceeds the LSTM network as length increases above $\sim$500.

Different physical models can account for subdiffusive, superdiffusive, or Brownian behavior of single molecule diffusion. An example of this is fBM (Mandelbrot and Van Ness 1968). The 1st Andi Challenge examined the performance of different methodologies for classifying single-molecule trajectories under different anomalous exponents (Muñoz-Gil *et al* 2021b). Figure 2(b) shows how the classification performance varies under different anomalous behaviors. There is a noticeable drop at $\alpha = 1$ for the three architectures. Nevertheless, WadTCN and the simple TCN classify better than the LSTM-based approach

when trajectories are nearly Brownian. Supplementary figure 7 shows confusion matrices depicting how these architectures classify according to different ground-truth models. We selected lengths of 25 (shortest length of all tested trajectories) and 500 (middle length of all tested trajectories) steps to test performance.

The second task in the Andi Challenge is the prediction of the anomalous exponent in single-molecule trajectories (Muñoz-Gil *et al* 2021b). Figure 2(c) shows the mean absolute error of the predictions attained by the LSTM and WadTCN architectures, respectively. For lengths longer than 175 steps, WadTCN achieved better predictions than the LSTM architecture. For shorter trajectories, the three architectures performed similarly, RANDI being slightly superior to the other two. The main reason for the superior results obtained with WadTCN for longer lengths than 175 steps is that it performs better than RANDI for the ATTM model (figure 2(d)). ATTM was the most challenging model to infer its anomalous exponent in the Andi Challenge Competition (Muñoz-Gil *et al* 2021b). It is important to recall that WadTCN uses two LSTM layers for regression tasks. Supplementary figure 8 shows the probability distribution of the predicted vs. ground truth anomalous diffusion exponent for the three architectures. It is noteworthy that all three architectures encounter difficulties in predicting the exponent as the trajectories approach Brownian motion behavior. One reason for this is that as trajectories approach $\alpha = 1$, it becomes more difficult to distinguish their respective model from others since similarities appear in their long time statistics despite differences in their microscopic generative dynamics (Muñoz-Gil *et al* 2021b). Notwithstanding this, we observe that both TCN and WadTCN provide a more accurate inference of the exponent under this behavior, although they may both underestimate it under certain conditions. Supplementary figure 9 show the bias plot, a kernel density estimation of the differences between predicted and ground truth exponents, which indicates that none of the methodologies applied underestimates or overestimates the value of the exponent by a significant margin. Apparently, the combination of LSTM and TCNs accounts for this superior performance.

3.2. Performance of pointwise prediction

Tasks 1 and 2 of the 1st Andi Challenge both test predictive models with homogenous trajectories (i.e., trajectories following the same theoretical model through their lifetime). However, particle dynamics are not necessarily constant in time; in fact, SBM is a model where trajectories linearly decrease ($\alpha < 1$) or increase ($\alpha > 1$) their diffusion coefficient with time. Time-dependent trajectories are considered dynamically heterogeneous. The proposed architecture, designed to return a single value or category, may in fact provide an estimate of the average behavior of trajectories, akin to MSD-based analyzers (Michalet 2010). To extend the possibility of detecting changes in dynamic parameters as a function of time, we introduce a variant whereby the MLP layer is replaced by a Transformer at the end of the architecture. This transformer can easily capture long-time correlations and efficiently relate one point of the trajectory with others (Vaswani *et al* 2017). Such WadTCN variant is trained to accomplish four tasks separately: model classification, anomalous exponent, diffusion coefficient prediction, and change point (the point with highest probability) detection. The latter task is challenging when dealt with via supervised methodologies as it is considered a class imbalance problem (Aminikhanghahi and Cook 2017).

Figure 3(a) depicts how well WadTCN and RANDI work in the third task of the 1st Andi Challenge, which tests whether a method detects change points in trajectories that randomly change their theoretical model through time. WadTCN appears to outperform RANDI in detecting the change point. The anomalous exponent and theoretical model predictions of the corresponding segments are also slightly better for WadTCN. While RANDI is designed to detect only one change point, WadTCN can detect multiple changes.

Thus far, CPD is limited to only one change point. The 2nd Andi Challenge tests whether this can be extended to detect multiple change points. Since probability predictions with a minimum height of 0.9 and a maximum prominence of 0.2 classified as change points maximized the detection performance (see supplementary figure 10), these threshold values were subsequently used as change points. The idea behind implementing prominence-based peak detection is to discard isolated (potentially spurious) low probability peaks and higher surrounding peaks. The Jaccard score and RMSE of the change point detection were 0.75 ± 0.32 and 0.10 ± 0.46, respectively, indicating precise CPD. There is a noticeable improvement in the diffusion coefficient prediction as the segment length increases (figure 3(b), left). The anomalous exponent prediction error increased as it reached a segment length of 200. In the 2nd Andi Challenge, simulated trajectories had D values ranging from 10^{-12} to 10^{-6} pixel^2/frame (pixel size: 100 nm; frame time: 100 ms). This wide range is problematic given the fixed localization noise of 0.12 pixels (12 nm). For $D < 10^{-3}$ pixel^2/frame, trajectories become indistinguishable from noise, complicating analyses and causing spurious changes in dynamics, particularly in the anomalous exponent. The issue is more pronounced in segments of length 200, where trajectories remain within constant dynamic parameters, while shorter segments capture parameter changes across a broader range within a single trajectory. While the network may recognize very slow trajectories for diffusion coefficient prediction, accurately estimating the anomalous exponent in such cases remains highly challenging. Most of the trajectories in this range had extremely high noise

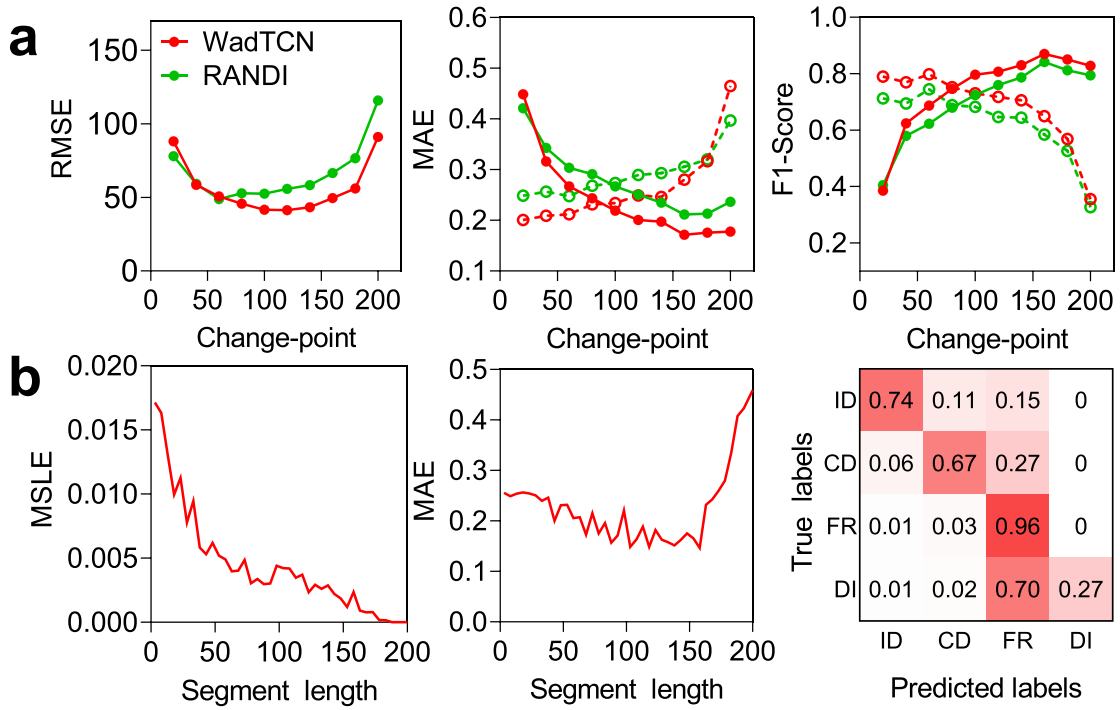

Figure 3. Predictive performance of heterogeneous trajectories. (a) Change point detection error of the segmentation task of the 1st Andi Challenge where a single change point appears. The root mean squared error (RMSE) of the predicted change point position (left), the mean absolute error (MAE) of the predicted anomalous exponent of the first (solid lines) and second (dashed lines) segments (middle), and the F1-Score of the classification of the segments theoretical model as a function of the real change point position. (b) Change point detection error of the 2nd Andi Challenge where multi change points appear. The mean squared logarithmic error (MSLE) of the predicted diffusion coefficient, the MAE of the predicted anomalous exponent, and the F1-Score of the predicted models of the detected trajectory segments are shown.

(SNR $= \frac{\sigma_{\text{displacements}}}{\sigma_{\text{noise}}} < 1$; figure 3(b) middle), thus excluding the possibility of analyzing them. As we can see in the confusion matrix displayed in figure 3(b) right, the network is capable of analyzing trajectory states except for directed diffusion owing to the imbalance of this state in the training dataset.

3.3. Transfer learning accelerates the WadTCN training process

Although WadTCN shows better performance for classification and prediction of the anomalous exponent of a given trajectory than other architectures, it is relatively slow to train. There are three reasons for this: (i) its complexity (it involves ∼800 000 weights); (ii) it uses a different training dataset on every epoch to prevent overfitting (the more complex it is, the more prone to overfit); and (iii) it is length-specific (it must be trained on every length studied). It would be convenient to reduce the complexity of WadTCN such that a comparable performance can be achieved even with a large number of weights. A reduction in complexity will also reduce proneness to overfit, enabling one to use a single training dataset for all epochs. Being able to apply a subset of the weights obtained in a WadTCN model for a given trajectory length to any trajectory length, the issues currently hindering training performance can be tackled simultaneously with transfer learning. Transfer learning (specifically, network-based deep transfer learning Tan *et al* (2018)) involves training a *base* network on a specific dataset and tasking and reutilizing the features obtained on a *target* network for a similar dataset and task (Yosinski *et al* 2014). To accomplish this, once the *base* network is trained, the first n layers from the *base* network are transferred to the first n layers of a *target* network and fixed (in DL glossary this is also known as *frozen,* i.e. those layers do not change their weights during training). However, the question arises as to the value of n, i.e. which part of the *base* network is useful for similar tasks and datasets? The response is the part corresponding to the set of layers known as *general* feature extractors (i.e. the first layers of the network, applicable to a wide range of tasks and datasets). Conversely, layers that only work for a given task and dataset (the last layers of the network) are known as *specific* feature extractors. For instance, in image-related tasks, the first layers of neural networks act as standard image filters (Lee *et al* 2009, Le *et al* 2011, Krizhevsky *et al* 2012). If we can detect where these general feature extractors are in WadTCN, we can reuse them for classification and regression independently of the length under consideration, thus accelerating the training process for all the architectures proposed in the present work.

Transfer learning was already applied in Muñoz-Gil *et al* (2020) to train a Random Forest architecture (Breiman 2001) in a given dataset and then applied without retraining to a different dataset. The same

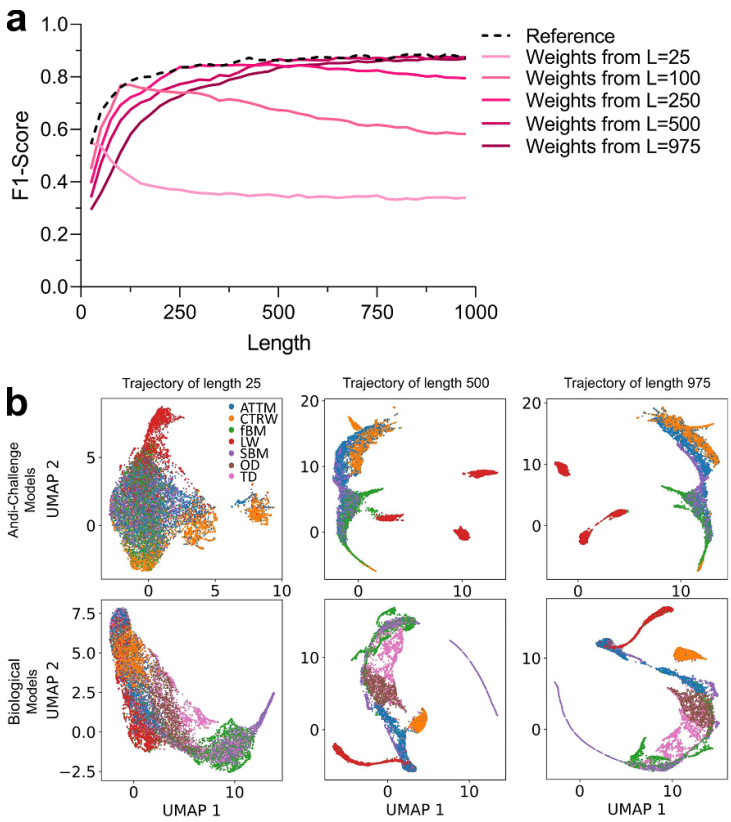

Figure 4. Transfer learning results using WadTCN. (a) Classification of Andi models using Total Transfer and weights obtained by training on different specific lengths (the reference curve is the performance obtained if one trains a WadTCN for every length of interest, the same as the red curve in figure 2(a)). (b) Reduced points obtained with UMAP, associated with simulated trajectories of length 25, 500, and 975 from different theoretical models obtained with WadTCN trained on trajectories of length 25. Here we used the UMAP tool to reduce dimensionality to a 2D representation with dimensions '*UMAP 1*' and '*UMAP 2*' in the '*x*' and '*y*' axis, respectively. The trajectories encoded by WadTCN appear as curves formed by reduced points; each reduced point represents the encoding of a specific length, and its color identifies a given theoretical model. When the curved-shaped representation is distorted by accumulation of additional points, it implies that the encodings of trajectories corresponding to the *same* physical model of diffusion aggregate into clusters. Trajectories described by the LW model appear to be the easiest model to distinguish, as their reduced points are in clusters well separated from the others.

applies to CNNs for predicting superdiffusive behavior in fBM trajectories (Kowalek *et al* 2019). Here, we analyzed how WadTCN performs if total layer transfer is done from classification models trained under different lengths, as shown in figure 4(a). Weights from models trained on short trajectories ($L = 25$ and $L = 100$) do not work on long trajectories: the performance deteriorates as the trajectory length increases. However, weights reutilized from models trained on long trajectories ($L = 250$, 500, and 975) generalize well for middle-sized and long trajectories. Despite bad performance on short trajectories, when WadTCN uses weights from very long trajectories the whole network acts as a general feature extractor.

For short trajectories, a set of layers acts as a specific feature extractor. A possible candidate for a specific feature extractor is the MLP layer, the only part of WadTCN that does not use convolutional operations (which are successful invariant feature extractors, LeCun *et al* (2015)). For the sake of simplicity, the result -now termed encoding- is a vector of length n obtained from sTCN after the input has been encoded by the WaveNet encoder. A way to visualize this encoding is the dimensionality reduction technique called uniform manifold approximation and projection (UMAP) that allows one to visualize high-dimension datasets on, for example, 2D projections, without losing their global structure (McInnes *et al* 2018). This technique was previously used to visualize the trajectory encodings in the penultimate layer of WaveNet-based deep neural network (WadNet) (Li *et al* 2021) and the latent space of a Graph Neural Network architecture to decode anomalous diffusion using 'unseen' models during the training process (Verdier *et al* 2021). In the present work, we use UMAP to visualize how trajectories that comply with different theoretical models are structured in the encoding in a bidimensional projection (details of the parameters used for UMAP are described under supplementary material). Figure 4(b) shows UMAP reduced points from encodings obtained from a WadTCN trained on trajectories of length 25 and applied to trajectories of different lengths. It is noticeable that as length increases, the more distinct and tidier is the structure of the trajectory encoding in the 2D projection. Moreover, we can visualize 'curves' formed by encodings corresponding to the same

theoretical model, which form clusters exclusively of such models. The same applies when networks are trained to classify among 7 diffusion models (including OD and TD theoretical models), as shown in figure 4(b). Due to the ordered structure displayed by the encoder after applying dimensionality reduction to 2D, it is very likely that the MLP layer acts as the specific feature extractor and the rest of the network is transferable and acts as a general feature extractor.

From here on, a trajectory length of 25 was adopted for all the architectures explored in this work, and all weights (except those from MLP layer) were frozen for models of other lengths except for binary segmentation of TD trajectories, owing to the low number of models needed to train. In this way, complexity is reduced to 250 000 weights and the training per epoch is ∼15 times faster.

3.4. Dynamics of the acetylcholine receptor at the plasma membrane

With the advent of super-resolution microscopy techniques and its combination with increasingly powerful computational analytical procedures, the translational dynamics of membrane-embedded proteins can be interrogated with increasing spatiotemporal resolution. Previous works from our laboratory have resorted to stochastic optical reconstruction microscopy (STORM) (Bates *et al* 2013) and SPT algorithms (Dedecker *et al* 2012) to describe the in-plane diffusion of the prototypic neurotransmitter receptor nAChR in living cells under different experimental conditions (Mosqueira *et al* 2018, 2020, Buena-Maizón and Barrantes 2022) using recurrent and turning angle analysis to study molecular trajectories (Weigel *et al* 2011, Burov *et al* 2013, Sadegh *et al* 2017). One of the most important conclusions of this series of studies is that the receptor covers a wide range of diffusional regimes, from subdiffusive to (seldom) superdiffusive, which can be accounted for the most part in terms of a two-state physical model (Grebenkov 2019). The dual behavior was clearly observed when analyzing the individual receptor trajectories switching between free Brownian motion and periods of confinement and confirmed by the combined turning angle and recurrence analyses (Mosqueira *et al* 2018, 2020). As the anti-correlated steps (i.e., particles tending to move backwards) were observed in the confinement sojourns, the model that best suited this behavior is the two-state OD model. A subsequent in silico analysis of the trajectories based on a deep-learning approach confirmed the goodness of the fit to a two-state OD model (Buena-Maizón and Barrantes 2022). The subdifussion of membrane-bound proteins in the plasma membrane may be the consequence of (i) confinement or (ii) immobilizing interactions.

The TD model describes a particle moving on the surface, becoming trapped (immobilized), and after a certain period of time resuming diffusion and moving freely until eventually it again becomes trapped and so on (Eggeling 2015, Ebrahimi *et al* 2023).

In this work, we used the dataset employed in Buena-Maizón and Barrantes (2022) to extend the analysis of the nAChR motion of the aforementioned work, and include additional theoretical models and only trajectories of equal length or longer than 25 steps. First, we added LW, SBM, ATTM, and TD as part of the classification analysis, classifying receptor movement among 7 categories. Trajectories classified as TD were further segmented into Brownian and trapped states. Finally, the TCN analysis was replaced by WadTCN trained with transfer learning. Details of simulation and training procedures are explained in supplementary material. Confusion matrices, classification, and regression errors are shown in supplementary figures 11–13.

Before analyzing trajectories with the neural networks, the criteria of Golan and Sherman were applied to filter out immobile trajectories based on the ratios between the radius of gyration and the mean steps size (Golan and Sherman 2017), as detailed in supplementary material. Between ∼60% and ∼76% single-particle trajectories were classified as immobile for BTX- and mAb-labeled samples, respectively.

Figure 5(a) shows trajectories classified into the seven theoretical models. The two-state OD model was found to be the most plausible model to account for the behavior of the mobile trajectories under all experimental conditions tested accounting for 30.69%–44.83%, and 34.31% and 42.36% of the trajectories, in BTX and mAb samples, respectively. It was found that the percentage of trajectories classified into ATTM was slightly lower than those classified into OD. As observed in supplementary figure 7, a confusion arises between ATTM and nearly all other theoretical models, as also noted in other works (Muñoz-Gil *et al* 2021b). We speculate that most of the trajectories classified as ATTM are, in fact, OD. One possible reason for this misclassification is that OD trajectories may be confused with ATTM trajectories moving with a diffusion coefficient below the experimental time resolution of STORM (10 ms/frame). Another factor is that trajectory length impacts on the classification performance of WadTCN; increasing the trajectory length therefore helps to better distinguish ATTM from the other models. Hence, the application of WadTCN to longer trajectories with more spatial resolution can enhance WadTCN classification. The classification results together with the classification confidence intervals for both labels and different experimental conditions are shown in figure 5(b) and supplementary tables 1 and 2.

Those trajectories classified as fBM and SBM are subclassified into subdiffusive ($0.02 < H < 0.47$), Brownian ($0.47 < H < 0.52$), or superdiffusive ($0.52 < H < 0.97$) trajectories with their corresponding neural network. In the case of fBM trajectories, subdiffusive motion predominates in BTX-labeled samples.

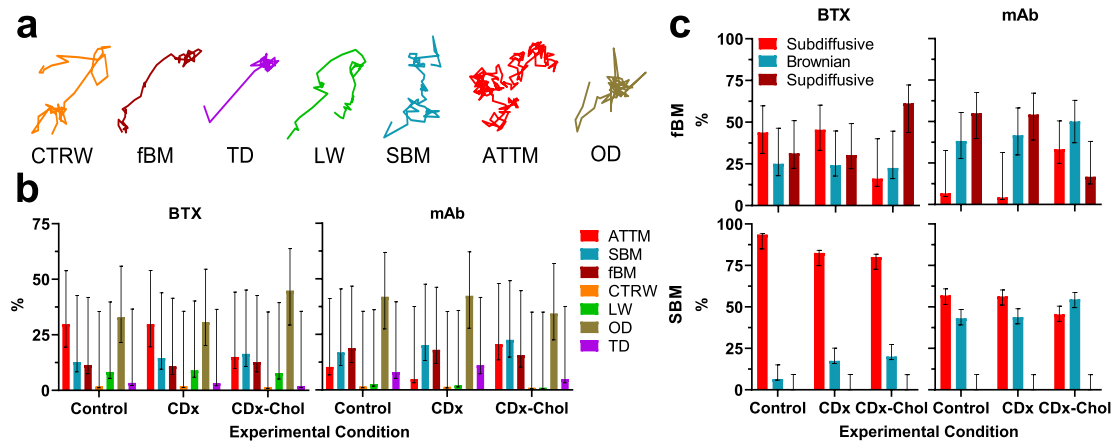

Figure 5. Single-particle trajectories classification results. (a) Examples of trajectories classified into the seven different theoretical models to classify single-particle trajectories. (b) Classification of nAChR trajectories from BTX- and mAb-labeled samples under different cholesterol depletion (CDx) or enrichment (CDx-Chol) experimental conditions using the proposed architecture. (c) Sub-model classification of nAChR trajectories from bungarotoxin (BTX) and monoclonal antibody (mAb) labelled samples using the proposed architecture. The black lines indicate the range of values below the 95% confidence interval of the classification error of all trained models of the classification network.

However, upon cholesterol enrichment of the cells (the CDx-Chol experimental condition), a significant percentage of trajectories exhibit superdiffusive behavior. In contrast, superdiffusive behavior is a predominant feature in all mAb-labeled samples, except for the CDx-Chol samples, where a high percentage of trajectories display Brownian-like behavior. In the case of SBM, the percentage of subdiffusive trajectories ranges from 80.00% to 93.54%. Conversely, for mAb, the percentage of subdiffusive trajectories falls within a range of 45.45%–56.86%. The sub-model classification results with the classification confidence intervals for both labels and different experimental conditions are shown in figure 5(c) and supplementary tables 3 and 4.

Next the Hurst exponent H was calculated for each trajectory, excluding those that were classified in the OD and TD models. This value was obtained in previous studies through fitting on the MSD curve to obtain the exponent α (Mosqueira *et al* 2018, 2020). We found a statistical difference in the Hurst exponent between control and both cholesterol-depleted (CDx) ($p < 0.05$) and cholesterol-enrichment (CDx-Chol) ($p < 0.01$) samples, with a decrease or increase in H, respectively. In the case of monoclonal antibody-labelled samples (mAb), we only found statistical differences between control and CDx, such that H decreased ($p < 0.01$). Simulations for the 2nd Andi-Challenge allow one to create a dataset of fBM trajectories with their corresponding diffusion coefficient D and use it to train a neural network to predict D from a trajectory. We calculated the diffusion coefficient of the trajectories classified as fBM and found no statistical differences between the different experimental conditions in BTX. In contrast, mAb samples show a statical difference between Control and CDx ($p < 0.05$).

TD was found to be underrepresented: 1.87% to 3.34% and 4.90% and 11.11% of the trajectories in BTX and mAb, respectively. We use WadTCN as a trajectory segmenter for trajectories classified as TD. There were no statistical differences between the residence times in the immobilized state in both BTX- and mAb-labeled single-particle trajectories.

The quantitative results with the confidence intervals for both labels and different experimental conditions are shown in figure 6 and supplementary tables 5–7. Supplementary figures 14 and 15 show the cumulative distribution functions of predicted anomalous exponents (calculated as $\alpha = H * 2$) for those trajectories that were not classified as OD and TD and supplementary figure 16 depicts how well the TD segmenter classified between the two trapped and free states.

4. Discussion

4.1. WadTCN performs better than similar analytical tools for regression tasks

Table 1 shows a comparison between the models discussed in this paper, including WadNet, the architecture on which WadTCN was based. The 1st Andi Challenge is, to date, the most objective comparison of methods for inferring and characterizing anomalous diffusion. Networks in the Andi Challenge based on LSTM cells showed an outstanding performance in both classification and regression tasks in comparison to other approaches. In this work, we show that the use of TCNs with the addition of LSTM improves the prediction of the anomalous exponent, especially in the case of the ATTM model. Furthermore, this analytical

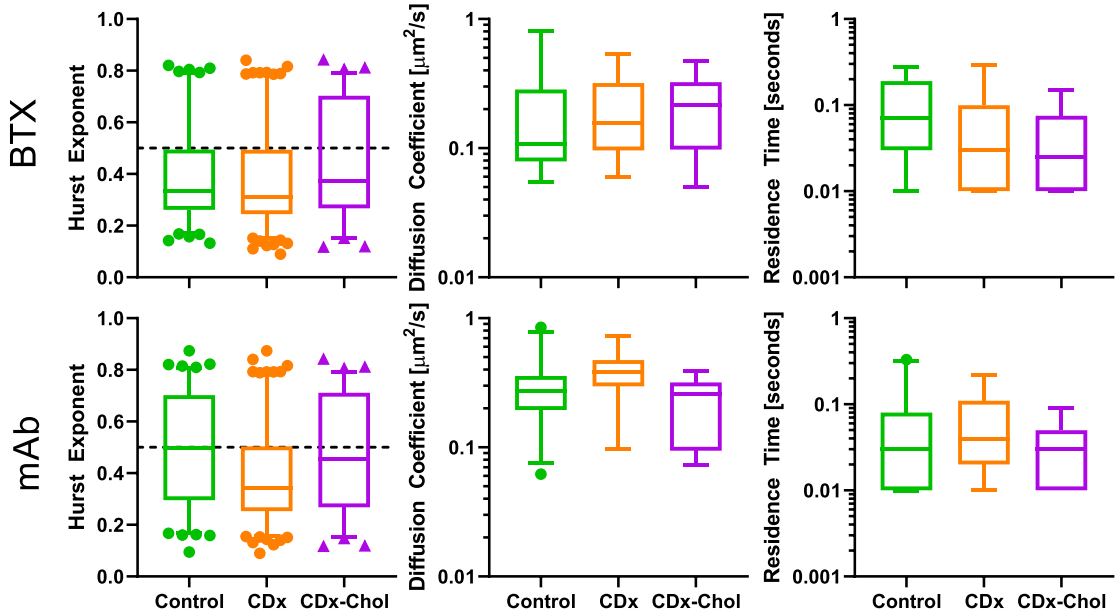

Figure 6. Trajectory parameters. Hurst exponent, diffusion coefficients and residence time in BTX- and mAb-labeled samples under different experimental conditions. The whiskers represent the interquartile range; the center line is the median; the extremes indicate the 2.5 and 97.5 percentile; the dots are outliers.

Table 1. Comparison between methodologies across the Andi tasks analyzed in this paper.

	DL Architecture			
Task	TCN	RANDI	WadNet[a]	WadTCN
1st Andi Challenge—Trajectory classification (F1-Score)	0.826	0.834	**0.909**	0.839
1st Andi Challenge—α inference (MAE)	0.182	0.142	0.132	**0.131**
1st Andi Challenge—Segment classification (F1-Score)	—	0.662	—	**0.713**
1nd Andi Challenge—Change-point inference (RMSE)	—	55	—	**41**
2nd Andi Challenge—Diffusion Coefficient inference (MSLE)	—	—	—	**0.004**
2nd Andi Challenge—α inference (MAE)	—	—	—	**0.22**
2nd Andi Challenge—Model classification (F1-Score)	—	—	—	**0.66**

[a] Values from Li *et al* (2021). Values in bold are the best score of each row.

combination surpasses the performance of RANDI, the best recurrent network in the Andi Challenge. The classification performance of WadTCN (with or without LSTMs) did not differ significantly from RANDI. However, with the addition of theoretical models such as OD and TD, the robustness of the architecture for handling more challenging classification tasks is apparent. Table 1 also reveals that, although WadNet performs better in classification, WadTCN is slightly better for inferring the anomalous exponent. In addition, the WaveNet Encoder implemented in WadNet involves more levels of processing than our encoder, leading to inefficient prediction.

4.2. The WadTCN transformer-based variant detects multiple change-point accurately

The 1st Andi Challenge evaluated models using homogeneous trajectories, but particle dynamics often vary over time, leading to heterogeneous trajectories. The architecture introduced in the present work can be modified to pointwise infer dynamical parameters capturing long-time correlations through a Transformer layer and TCNs showing segmentation capabilities (Lea *et al* 2016). This architecture handles multiple tasks, including state classification, anomalous exponent and diffusion coefficient prediction, and CPD. WadTCN showed superior performance in comparison to RANDI in detecting change points, as evidenced by results from the 1st Andi Challenge. WadTCN also predicts anomalous exponents and diffusion coefficients more accurately. Whereas RANDI detects only one change point, WadTCN can detect multiple points, a feature that was further evaluated in the 2nd Andi Challenge. With optimized height and prominence values, the WadTCN model achieved high precision in CPD (Jaccard score of 0.75 ± 0.32 and RMSE of 0.10 ± 0.46). Notably, performance in detecting diffusion coefficients improves with increasing segment length, although extreme noise (SNR < 1) in longer segments poses challenges. The network effectively classifies most trajectory states, though its accuracy in detecting directed diffusion is hindered by imbalanced training data.

Our initial contribution to the Andi Challenge competition consisted of a different methodology from the one presented here. To detect change points, we used the mean of the diffusion parameter and the anomalous exponent. Although this methodology ranks among the top 3 in the ensemble approach and among the top 10 in the single-trajectory approach, we believe it can be improved with the proposals outlined in the current paper. Furthermore, it can be combined with methodologies that obtain trajectories directly from video, as we also suggested during the competition using U-Nets (Midtvedt *et al* 2021). Details are available in our code repository, which we have included in the 'Code Availability' section in the branch `2nd_andi_challenge_original_code`.

4.3. Transfer learning accelerates diffusion-related neural networks training

Transfer learning is a technique in ML, as its name indicates, for transferring knowledge from one network to another (Yosinski *et al* 2014). Among the various possible applications (e.g. improving prediction accuracy Xiao *et al* (2019)), here we employ transfer learning to accelerate the network training process for the analysis of macromolecular trajectories in live cells. We have selected the encoding for transfer learning. Encoding is the representation obtained from the network before 'entering' the MLP layer. As we have shown, the low-dimensional UMAP representation of the encoding exhibits an organized structure and apparently non-linear separable features. This suggests that up to the MLP layer the network likely serves as a general feature extractor and can be reusable for various tasks, including classification and regression, representing a step forward in dispelling the *'black box'* preconception associated with neural networks when dealing with anomalous diffusion: via UMAP dimensionality reduction, one can interpret a global structure behind the features extracted by WadTCN (Seckler *et al* 2023).

4.4. Denoising trajectories would improve diffusion analysis

What would happen if the trajectories had no noise? Supplementary figure 16 displays confusion matrices for the classification of trajectories in the absence of noise, showing how the architecture is better able to distinguish between models than when noise is present. It is therefore convenient to remove as much noise as possible from the trajectories rather than develop more architectures without addressing the noise problem *ab initio*. The noise problem can be addressed using Kalman Filters (Kalman 1960, Wu *et al* 2010, Yüce *et al* 2011) or autoencoders (Muñoz-Gil *et al* 2021a). The latter have been successfully employed to denoise audio signals (Défossez *et al* 2020). However, caution must be applied to ensure that these denoisers do not act as 'smoothers', given their diminished capacity to detect abrupt changes.

4.5. The current TCN approach appears to confirm the two-state diffusional model previously applied to the nAChR

Previous work from our laboratory has found that nAChR lateral motion in the membrane is primarily subdiffusive, and a two-state OD model was proposed to account most satisfactorily for the motional behavior of the receptor protein (Mosqueira *et al* 2018, 2020). Here, we have extended the analysis to other physical idealizations such as ATTM, LW, SBM, and TD. LW applies exclusively to subdiffusive behavior. ATTM is characterized by diffusion coefficients that vary over time, as is SBM, with diffusivity increasing or decreasing over time. Comparison of the different models to explain the behavior of nAChR 2D diffusion shows that none can satisfactorily explain receptor translational motion better than the previously postulated two-state OD model (Mosqueira *et al* 2018, 2020). The alternative two-state model, TD, was rarely found in the validated trajectories, indicating that trapping events occurs with low frequency (Eggeling 2015). ATTM, characterized by being locally Brownian but globally anomalous, switching the diffusion coefficient in time (Massignan *et al* 2014) is the second-best model to account for diffusional dynamics. Owing to the confusion between obstructions and too slow- and Brownian-diffusion and the impact of trajectory length on WadTCN's classification performance, it is suggested that many trajectories classified as ATTM are in fact OD. This would indicate that classification can be improved using longer trajectories with higher spatial resolution.

5. Conclusion

WadTCN, a concatenated convolutional neural network combined with WaveNet encodings, is introduced here to predict dynamic properties of particle trajectories in a 2-dimensional space. The network is tested on a biological membrane-bound macromolecule, the nAChR, to analyze its diffusional behavior using trajectory data obtained through the superresolution microscopy STORM technique. WadTCN outperformed other top-performing methodologies from the first Andi-Challenge competition, particularly in predicting the anomalous exponent. We have showcased the efficacy of transfer learning in accelerating the training process while maintaining classification and regression performance. Furthermore, WadTCN

has been successfully applied to segmentation tasks and the prediction of the diffusion coefficient in short fBM trajectories (25 steps). Thus, WadTCN stands out as a versatile and robust architecture capable of addressing a range of tasks in the biophysical analysis of diffusing particles, including the many families of membrane-bound proteins and cell-surface receptors.

Data availability statement

All trained networks of this paper can be found at https://zenodo.org/record/14253907.

The data and code that support the findings of this study are openly available at the following URL/DOI: https://github.com/lucasSaavedra123/wadtcn.

Acknowledgments

Thanks are due to Héctor Buena-Maizón and Alejo Mosqueira for valuable suggestions. This work is part of an ongoing research project supported by multiple grants from the National Science and Technology Research Council of Argentina (CONICET), and PICT BID-CONICET 2015-2654 from the Ministry of Science, Technology and Innovative Production of Argentina (Mincyt) to F J B. Simulations were carried out using a Titan V GPU kindly donated by NVIDIA Corp. to F J B.

Software availability

All Python scripts and related codes used in this paper can be found at: https://github.com/lucasSaavedra123/wadtcn.

ORCID iDs

Lucas A Saavedra https://orcid.org/0000-0002-3566-1154
Francisco J Barrantes https://orcid.org/0000-0002-4745-681X

References

Abadi M *et al* 2016 TensorFlow: a system for large-scale machine learning *Proc. 12th USENIX Conf. on Operating Systems Design and Implementation* (*Savannah, GA, USA*)
Aghion E, Meyer P G, Adlakha V, Kantz H and Bassler K E 2021 Moses, Noah and Joseph effects in Lévy walks *New J. Phys.* **23** 023002
Aminikhanghahi S and Cook D J 2017 A survey of methods for time series change point detection *Knowl. Inf. Syst.* **51** 339–67
Argun A, Volpe G and Bo S 2021 Classification, inference and segmentation of anomalous diffusion with recurrent neural networks *J. Phys. A: Math. Theor.* A **54** 294003
Bai S, Kolter J Z and Koltun V 2018 An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (arXiv:1803.01271)
Bates M, Jones S A and Zhuang X 2013 Stochastic optical reconstruction microscopy (STORM): a method for superresolution fluorescence imaging *Cold Spring Harb. Protoc.* **2013** 498–520
Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
Buena-Maizón H and Barrantes F J 2022 A deep learning-based approach to model anomalous diffusion of membrane proteins: the case of the nicotinic acetylcholine receptor *Brief. Bioinform.* **23** bbab435
Burov S, Tabei S M, Huynh T, Murrell M P, Philipson L H, Rice S A, Gardel M L, Scherer N F and Dinner A R 2013 Distribution of directional change as a signature of complex dynamics *Proc. Natl Acad. Sci. USA* **110** 19689–94
Dedecker P, Duwé S, Neely R K and Zhang J 2012 Localizer: fast, accurate, open-source, and modular software package for superresolution microscopy *J. Biomed. Opt.* **17** 126008
Défossez A, Synnaeve G and Adi Y 2020 Real time speech enhancement in the waveform domain (arXiv:2006.12847)
Ebrahimi V, Stephan T, Kim J, Carravilla P, Eggeling C, Jakobs S and Han K Y 2023 Deep learning enables fast, gentle STED microscopy *Commun. Biol.* **6** 674
Eggeling C 2015 Super-resolution optical microscopy of lipid plasma membrane dynamics *Essays Biochem.* **57** 69–80
Firbas N, Garibo-i-Orts Ò, Garcia-March M A and Conejero J A 2023 Characterization of anomalous diffusion through convolutional transformers *J. Phys. A: Math. Theor.* A **56** 014001
Gajowczyk M and Szwabiński J 2021 Detection of anomalous diffusion with deep residual networks *Entropy* **23** 649
Garibo-i-Orts Ò, Firbas N, Sebastiá L and Conejero J A 2023 Gramian angular fields for leveraging pretrained computer vision models with anomalous diffusion trajectories *Phys. Rev.* E **107** 034138
Glorot X, Bordes A and Bengio Y 2011 Deep sparse rectifier neural networks *Proc. 14th Int. Conf. on Artificial Intelligence and Statistics* (Proceedings of Machine Learning Research) (https://proceedings.mlr.press/v15/glorot11a.html)
Golan Y and Sherman E 2017 Resolving mixed mechanisms of protein subdiffusion at the T cell plasma membrane *Nat. Commun.* **8** 15851
Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 Single-particle diffusion characterization by deep learning *Biophys. J.* **117** 185–92
Grebenkov D S 2019 Time-averaged mean square displacement for switching diffusion *Phys. Rev.* E **99** 032133
Griffié J, Boelen L, Burn G, Cope A P and Owen D M 2015 Topographic prominence as a method for cluster identification in single-molecule localisation data *J. Biophotonics* **8** 925–34

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*
Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
Kæstel-Hansen J, Kirchhausen T and Hatzakis N S 2024 Deep-SPT, a deep learning toolbox for single particle tracking in 3D, reveals how biological motion encodes function *Biophys. J.* **123** 43a–4a
Kalman R E 1960 A new approach to linear filtering and prediction problems *J. Basic Eng.* **82** 35–45
Klafter J and Zumofen G 1994 Levy statistics in a Hamiltonian system *Phys. Rev.* E **49** 4873–7
Klambauer G, Unterthiner T, Mayr A and Hochreiter S 2017 Self-normalizing neural networks *Proc. 31st Int. Conf. on Neural Information Processing Systems* (*Long Beach, CA, USA*)
Kowalek P, Loch-Olszewska H and Szwabiński J 2019 Classification of diffusion modes in single-particle tracking data: feature-based versus deep-learning approach *Phys. Rev.* E **100** 032410
Krapf D 2015 Mechanisms underlying anomalous diffusion in the plasma membrane *Curr. Top. Membr.* **75** 167–207
Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* (https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
Krog J, Jacobsen L H, Lund F W, Wüstner D and Lomholt M A 2018 Bayesian model selection with fractional Brownian motion *J. Stat. Mech. Theor. Exp.* **2018** 093501
Le Q, Karpenko A, Ngiam J and Ng A 2011 ICA with reconstruction cost for efficient overcomplete feature learning *Advances in Neural Information Processing Systems* (https://proceedings.neurips.cc/paper_files/paper/2011/file/233509073ed3432027d48b1a83f5fbd2-Paper.pdf)
Lea C, Vidal R, Reiter A and Hager G D 2016 Temporal convolutional networks: a unified approach to action segmentation *Computer Vision—ECCV 2016 Workshops* (*Cham*) ed G Hua and H Jégou
LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
Lee H, Grosse R, Ranganath R and Ng A Y 2009 Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations *Proc. 26th Annual Int. Conf. on Machine Learning* (*Montreal, QC, Canada*) (https://doi.org/10.1145/1553374.1553453)
Li D, Yao Q and Huang Z 2021 WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet) *J. Phys. A: Math. Theor.* **54** 404003
Lim S C and Muniandy S V 2002 Self-similar Gaussian processes for modeling anomalous diffusion *Phys. Rev.* E **66** 021114
Lin M, Chen Q and Yan S 2014 Network in network *Int. Conf. on Learning Representations* (*Banff, Alberta, Canada*)
Mandelbrot B B and Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37
Manzo C, Muñoz-Gil G, Volpe G, Angel Garcia-March M, Lewenstein M and Metzler R 2023 Preface: characterisation of physical processes from anomalous diffusion data *J. Phys. A: Math. Theor.* A **56** 010401
Manzo C, Torreno-Pina J A, Massignan P, Lapeyre G J, Lewenstein M and Garcia Parajo M F 2015 Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity *Phys. Rev.* X **5** 011021
Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J 2014 Nonergodic subdiffusion from brownian motion in an inhomogeneous medium *Phys. Rev. Lett.* **112** 150603
McInnes L, Healy J, Saul N and Großberger L 2018 UMAP: uniform manifold approximation and projection *J. Open Source Softw.* **3** 861
Michalet X 2010 Mean square displacement analysis of single-particle trajectories with localization error: brownian motion in an isotropic medium *Phys. Rev.* E **82** 041914
Midtvedt B, Helgadottir S, Argun A, Pineda J, Midtvedt D and Volpe G 2021 Quantitative digital microscopy with deep learning *Appl. Phys. Rev.* **8** 011310
Mosqueira A, Camino P A and Barrantes F J 2018 Cholesterol modulates acetylcholine receptor diffusion by tuning confinement sojourns and nanocluster stability *Sci. Rep.* **8** 11974
Mosqueira A, Camino P A and Barrantes F J 2020 Antibody-induced crosslinking and cholesterol-sensitive, anomalous diffusion of nicotinic acetylcholine receptors *J. Neurochem.* **152** 663–74
Muñoz-Gil G *et al* 2021b Objective comparison of methods to decode anomalous diffusion *Nat. Commun.* **12** 6253
Muñoz-Gil G, Bachimanchi H, Pineda J, Midtvedt B, Lewenstein M, Metzler R, Krapf D, Volpe G and Manzo C 2023 Quantitative evaluation of methods to analyze motion changes in single-particle experiments (arXiv:2311.18100)
Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 Single trajectory characterization via machine learning *New J. Phys.* **22** 013010
Muñoz-Gil G, i Corominas G G and Lewenstein M 2021a Unsupervised learning of anomalous diffusion data: an anomaly detection approach *J. Phys. A: Math. Theor.* A **54** 504001
Pearson K 1905 The problem of the random walk *Nature* **72** 294
Qian H, Sheetz M P and Elson E L 1991 Single particle tracking. Analysis of diffusion and flow in two-dimensional systems *Biophys. J.* **60** 910–21
Qu X, Hu Y, Cai W, Xu Y, Ke H, Zhu G and Huang Z 2024 Semantic segmentation of anomalous diffusion using deep convolutional networks *Phys. Rev. Res.* **6** 013054
Requena B, Masó-Orriols S, Bertran J, Lewenstein M, Manzo C and Muñoz-Gil G 2023 Inferring pointwise diffusion properties of single trajectories with deep learning *Biophys. J.* **122** 4360–9
Sadegh S, Higgins J L, Mannion P C, Tamkun M M and Krapf D 2017 Plasma membrane is compartmentalized by a self-similar cortical actin meshwork *Phys. Rev.* X **7** 011031
Scher H and Montroll E W 1975 Anomalous transit-time dispersion in amorphous solids *Phys. Rev.* B **12** 2455–77
Seckler H and Metzler R 2024 Change-point detection in anomalous-diffusion trajectories utilising machine-learning-based uncertainty estimates *J. Phys. Photonics* **6** 045025
Seckler H, Szwabiński J and Metzler R 2023 Machine-learning solutions for the analysis of single-particle diffusion trajectories *J. Phys. Chem. Lett.* **14** 7910–23
Shlesinger M F, Klafter J and Wong Y M 1982 Random walks with infinite spatial and temporal moments *J. Stat. Mech.* 499–512
Slator P J and Burroughs N J 2018 A hidden Markov model for detecting confinement in single-particle tracking trajectories *Biophys. J.* **115** 1741–54
Slator P J, Cairo C W and Burroughs N J 2015 Detection of diffusion heterogeneity in single particle tracking trajectories using a hidden Markov model with measurement noise propagation *PLoS One* **10** e0140759

Tan C, Sun F, Kong T, Zhang W, Yang C and Liu C 2018 A survey on deep transfer learning *Artificial Neural Networks and Machine Learning—ICANN 2018* (*Cham*)

Truong C, Oudre L and Vayatis N 2020 Selective review of offline change point detection methods *Signal Process.* **167** 107299

van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A and Kavukcuoglu K 2016 WaveNet: a generative model for raw audio (arXiv:1609.03499)

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł U and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* (https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)

Verdier H, Duval M, Laurent F, Cassé A, Vestergaard C L and Masson J-B 2021 Learning physical properties of anomalous random walks using graph neural networks *J. Phys. A: Math. Theor.* A **54** 234001

Virtanen P *et al* (SciPy 1.0 Contributors) 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72

Weigel A V, Simon B, Tamkun M M and Krapf D 2011 Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking *Proc. Natl Acad. Sci. USA* **108** 6438–43

Wu P H, Agarwal A, Hess H, Khargonekar P P and Tseng Y 2010 Analysis of video-based microscopic particle trajectories using Kalman filtering *Biophys. J.* **98** 2822–30

Xiao M, Shen X and Pan W 2019 Application of deep convolutional neural networks in classification of protein subcellular localization with microscopy images *Genetic Epidemiol.* **43** 330–41

Yosinski J, Clune J, Bengio Y and Lipson H 2014 How transferable are features in deep neural networks? *Proc. 27th Int. Conf. on Neural Information Processing Systems* (*Montreal, Canada*) vol 2

Yu Y, Si X, Hu C and Zhang J 2019 A review of recurrent neural networks: LSTM cells and network architectures *Neural Comput.* **31** 1235–70

Yüce M Y, Erdoğan A, Jonáš A and Kiraz A 2011 Single molecule tracking with Kalman filtering *OSA Technical Digest Frontiers in Optics 2011/Laser Science XXVII* (*San Jose, California*)