

PAPER • OPEN ACCESS

# Boosting the performance of anomalous diffusion classifiers with the proper choice of features






To cite this article: Patrycja Kowalek *et al* 2022 *J. Phys. A: Math. Theor.* **55** 244005

View the [article online](#) for updates and enhancements.

## You may also like

- [Relation of planar Hall and planar Nernst effects in thin film permalloy](#)  
D Wesenberg, A Hojem, R K Bennet et al.
- [Characterization of anomalous diffusion classical statistics powered by deep learning \(CONDOR\)](#)  
Alessia Gentili and Giorgio Volpe
- [Driving the growth of the earliest supermassive black holes with major mergers of host galaxies](#)  
Takamitsu L Tanaka

# Boosting the performance of anomalous diffusion classifiers with the proper choice of features

Patrycja Kowalek<sup>\*</sup> , Hanna Loch-Olszewska ,  
Łukasz Łaszczyk , Jarosław Opała  and  
Janusz Szwabiński 

Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław  
University of Science and Technology, 50-370 Wrocław, Poland

E-mail: [patrycja.kowalek@pwr.edu.com](mailto:patrycja.kowalek@pwr.edu.com)

Received 30 December 2021, revised 6 April 2022

Accepted for publication 5 May 2022

Published 26 May 2022



CrossMark

## Abstract

Understanding and identifying different types of single molecules' diffusion that occur in a broad range of systems (including living matter) is extremely important, as it can provide information on the physical and chemical characteristics of particles' surroundings. In recent years, an ever-growing number of methods have been proposed to overcome some of the limitations of the mean-squared displacements approach to tracer diffusion. In March 2020, the anomalous diffusion (AnDi) challenge was launched by a community of international scientists to provide a framework for an objective comparison of the available methods for AnDi. In this paper, we introduce a feature-based machine learning method developed in response to task 2 of the challenge, i.e. the classification of different types of diffusion. We discuss two sets of attributes that may be used for the classification of single-particle tracking data. The first one was proposed as our contribution to the AnDi challenge. The latter is the result of our attempt to improve the performance of the classifier after the deadline of the competition. Extreme gradient boosting was used as the classification model. Although the deep-learning approach constitutes the state-of-the-art technology for data classification in many domains, we deliberately decided to pick this traditional machine learning algorithm due to its superior interpretability. After the extension of the feature set our classifier achieved the accuracy of 0.83, which is comparable with the top methods based on neural networks.

<sup>\*</sup> Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: single particle tracking, anomalous diffusion, time series classification, machine learning, feature-based approach

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Single-particle tracking (SPT) is a popular method for observing the molecular dynamics in a wide range of materials, including the living cells [1, 2]. Typically, an SPT experiment results in trajectories of tracers, i.e. time series of their consecutive positions, the analysis of which allows one to extract properties of the molecules and their environment (e.g. density of obstacles, velocity).

The key first step in the analysis of SPT data is to connect the observed motion of particles with the available models of diffusion [3], since it already sheds light on the mechanical properties of the particles' surrounding [4]. The most common approach to characterize the diffusion of particles is the mean-squared displacement (MSD) [3, 5, 6], which measures their deviations with respect to a reference position in time. A MSD growing linearly in time corresponds to Brownian motion (a.k.a. normal diffusion), which describes the movement of a microscopical particle as a consequence of thermal forces. Any deviations from that linear behavior are called anomalous diffusion (AnDi) [7], which can occur in physical, biological, or geological systems. Those deviations are usually represented by a power-law scaling of MSD (i.e.  $\text{MSD}(t) \propto t^\alpha$ ) [3, 5, 6], with  $\alpha$  being the AnDi exponent. A sublinear MSD ( $\alpha < 1$ ) corresponds to subdiffusion. It indicates trapped particles [8, 9], particles which hit upon obstacles [10, 11] or particles moderated by viscoelastic properties of the environment [12]. Subdiffusion can take place in crowded cellular fluids and membranes [13], and can be caused by crowding and complex interactions with the cytoskeleton and macromolecular complexes: in cytoplasm [14–16], the nucleus [17] and the plasma membrane [18–20]. A superlinear MSD ( $\alpha > 1$ ) is referred to as superdiffusion. In this case, the movement of particles is usually faster than in normal diffusion and frequently directed to a specific location. Examples of superdiffusion include Lévy flights [21, 22], active cytoplasmic flows and transport mediated by molecular motors [23, 24], particle motion in random potentials or in turbulent flows [25, 26], foraging by animals and humans [27], and many others [28].

Despite its simplicity, the MSD-based analysis of SPT trajectories is truly challenging. The experimental trajectories are very often too short to extract meaningful information from the MSD. Moreover, the finite precision of tracers' localization adds a term to the MSD, which can limit the interpretation of data [29–31]. Consequently, several other approaches going beyond MSD have been introduced. For instance, the radius of gyration [32], the velocity autocorrelation function [33, 34], the time-dependent directional persistence of trajectories [35], the distribution of directional changes [36], the mean maximum excursion method [37], the fractionally integrated moving average framework [38] may efficiently replace the MSD estimator for classification purposes. The full distribution of displacements may be fitted to a mixed model in order to extract differences in diffusive behavior between subsets of particle ensembles [39]. Heterogeneity within single trajectories can be checked with hidden Markov models [40, 41].

In the last few years, machine learning (ML) has become an interesting alternative for the analysis of AnDi. This approach is very appealing because in contrast to the analytical methods it does not require explicit rules for data processing. Instead, a ML algorithm can learn those rules directly from a series of data. ML is already known to excel in different

domains including computer vision [42], speech recognition [43] and natural language processing [44]. First attempts to apply them to SPT data turned out to be very promising, even though the characterization of diffusion remains very challenging. Bayesian approach [45–47], random forests [48–52], gradient boosting [49–52], neural networks [53], and deep neural networks [49, 54–57] have been used to either classify the trajectories or to extract quantitative information about them.

The increasing number of methods, both analytical and ML-based, and the lack of an objective comparison between them have resulted in the launch of the AnDi challenge by a team of international scientists [58]. The goal of the challenge was at least two-fold: to provide a framework to benchmark the existing and new methods for the analysis of AnDi and to spur the invention of new high-performance methods. The challenge itself was divided into three main tasks: (1) inference of the AnDi exponent  $\alpha$  from the trajectories provided by the organizers, (2) classification of trajectories into five different models of diffusion, and (3) detection of points within single trajectories, at which the anomalous exponent  $\alpha$  and/or the diffusion coefficient  $D$  change. From the results of the challenge, it follows that the ML methods based on neural networks outperform the more traditional statistics-based approaches [59]. However, it should be emphasized here that the latter usually offer a deeper insight into the mechanisms of classification.

The main obstacle limiting the deployment of ML to the trajectory analysis is the availability of decent training data. The experimental trajectories are not provable (otherwise we would not need any new method of analysis). As a consequence, synthetic trajectories generated by computer simulations of various diffusion models are typically used as training data. The organizers of the AnDi challenge have identified five models of particular importance for the interpretation of experimental results: continuous time random walk (CTRW) [60], fractional Brownian motion (FBM) [61], Lévy walk (LW) [26, 62, 63], annealed transient time motion (ATTM) [64], and scaled Brownian motion (SBM) [65]. To ensure that all participants of the challenge use the same data for inventing/improving their methods, a Python package with the implementation of the models has been provided by the organizers.

This paper describes the details of the method we proposed in response to the AnDi challenge. We focus on the classification of 2D trajectories, i.e. a part of the task 2 within the challenge. In our recent researches [49, 51, 52, 57], we already investigated the applicability of both feature-based models and deep learning techniques to AnDi. The first approach requires a set of human-engineered features to characterize each trajectory. Vectors of these features are then used as input for a classification model. Deep learning methods, in contrast, extract significant features from raw data on their own, without any effort from a human expert. The deep learning approach yielded slightly better performance than the feature-based ones, in line with the general findings of the AnDi challenge, but we decided to further elaborate on the latter due to its superior interpretability. Our method turned out to be inferior to the winning teams, but still offered a reasonable performance [59]. Moreover, further investigation of the method after the challenge allowed us to significantly improve its accuracy by adding some new features and to achieve performance similar to the best methods (0.83 accuracy in task 2 versus 0.88 accuracy of the winners).

The paper is divided as follows. Diffusion models used for training and the classification method are introduced in section 2. Section 3 briefly summarizes the dataset used for training. Results of our analysis are presented in section 4. Finally, some concluding remarks are given.

## 2. Models and methods

### 2.1. Anomalous exponent $\alpha$

The most popular method of deducing the particles' type of motion from their trajectories is based on the MSD [3, 5, 6, 66], defined as

$$\text{MSD}(t) = E \left( \|X_{t+t_0} - X_{t_0}\|^2 \right), \quad (1)$$

where  $(X_t)_{t>0}$  is a particle's trajectory,  $\|\cdot\|$  stands for the Euclidean distance and  $E$  is the ensemble average. Since in many experiments only a limited number of trajectories is observed, the time-averaged MSD (TAMSD) calculated from a single trajectory is usually used as the estimator of MSD,

$$\widehat{\text{MSD}}(n\Delta t) = \frac{1}{N-n+1} \sum_{i=0}^{N-n} \|X_{i+n} - X_i\|^2. \quad (2)$$

The trajectory consists of  $N$  consecutive two dimensional positions  $X_i = (x_i, y_i)$  ( $i = 0, \dots, N$ ) recorded with a constant time interval  $\Delta t$ . The quantity  $n$  is the time lag between the initial and the final positions of the particle. If the underlying process is ergodic and has stationary increments, TAMSD converges to the theoretical MSD [67].

AnDi refers to a broad class of processes that deviate from the standard Brownian motion. Those deviations display an asymptotic power-law dependence of the MSD on the time lag,

$$\widehat{\text{MSD}}(n\Delta t) \sim K_\alpha t^\alpha, \quad (3)$$

where  $K_\alpha$  is the generalized diffusion coefficient and  $\alpha$  is the anomalous exponent. As already mentioned in the introduction, the value of the latter may be used to discriminate the trajectories into normal diffusion ( $\alpha = 1$ ), subdiffusion ( $\alpha < 1$ ), and superdiffusion ( $\alpha > 1$ ).

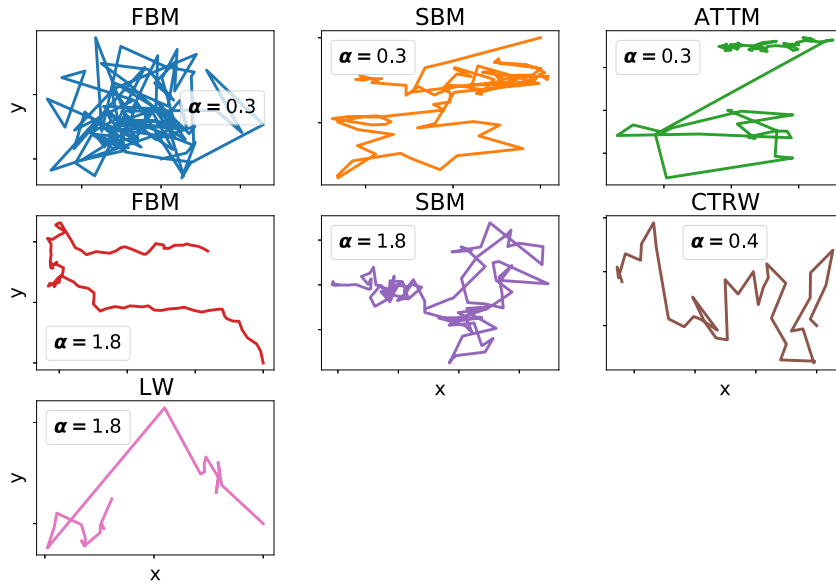
### 2.2. Stochastic models of diffusion

Particles' movements that exhibit deviations from the linear behavior of MSD may be described by multiple models, depending on some specific properties of the corresponding trajectories. Five of such processes have been identified within the AnDi challenge as crucial for the interpretation of the experimental data [58, 59]: CTRW, FBM, LW, ATTM and SBM (see figure 1 for example trajectories). In this section, the main characteristics of those models are briefly summarized.

**2.2.1. Continuous time random walk.** CTRW defines a large family of random walks with arbitrary displacement density for which the time spent between subsequent steps (i.e. the waiting time) is a stochastic variable [60]. Here, we will consider a particular instance of CTRW with the waiting times sampled from a power-law distribution  $\psi(t) \sim t^{-\sigma}$ , and the displacements drawn from Gaussian distribution with mean zero and variance  $D$ ,  $N(0, \sqrt{D})$ . In this case, we have  $\alpha = \sigma - 1$ .

**2.2.2. Fractional Brownian motion.** FBM [59] is the solution of the stochastic differential equation

$$dX_t^i = \sigma dB_t^{H,i}, \quad i = 1, 2. \quad (4)$$



**Figure 1.** Examples of trajectories generated with different anomalous exponents by the five models of diffusion: CTRW, FBM, LW, ATTM, and SBM.

Here,  $\sigma > 0$  is the scale coefficient, which relates to the diffusion coefficient  $D$  via  $\sigma = \sqrt{2D}$ .  $H \in (0, 1)$  is the Hurst parameter and  $B_t^H$  is a continuous-time, zero-mean Gaussian process starting at zero, with the following covariance function

$$E(B_t^H B_s^H) = \frac{1}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}). \quad (5)$$

The value of  $H$  determines the type of diffusion in the process. For  $H < \frac{1}{2}$ , FBM produces subdiffusion (regime with a negatively correlated noise). For  $H > \frac{1}{2}$ , FBM generates superdiffusive motion (positively correlated noise). It reduces to the normal diffusion at  $H = \frac{1}{2}$ .

**2.2.3. Lévy walk.** LW [26, 62, 63] is another model of a random walk, but in contrast to CTRW, the waiting times and step lengths are coupled. In this work, we assume that waiting times are power-law distributed (as in case of CTRW,  $\psi(t) \sim t^{-\sigma}$ ) and that the displacements and waiting times are correlated in such a way that the probability to move a certain step  $\Delta x$  at time  $t$  and stop at the new position to wait for a new random event is  $\Psi(\delta x, t) = \frac{1}{2} \psi(t) \delta(|\delta x| - vt)$ , where  $v$  is the velocity. Then, one can show that the anomalous exponent  $\alpha = 2$  for  $0 < \sigma < 1$  and  $\alpha = 3 - \sigma$  for  $1 < \sigma < 2$ . We may also observe then that the distribution of displacements is not Gaussian.

**2.2.4. Annealed transient time motion.** ATTM describes the motion of a Brownian particle with a time-varying diffusion coefficient [64]. The tracer performs Brownian motion with a random diffusion coefficient  $D_1$  for a random time  $t_1$ , then with  $D_2$  for  $t_2$  and so on. The diffusion coefficients are drawn from a power-law distribution  $P(D) \sim D^{\sigma-1}$ . If the random times are sampled from a distribution with the expected value  $E[t|D] = D^{-\gamma}$  ( $\sigma < \gamma < \sigma + 1$ ), then we have  $\alpha = \sigma/\gamma$ . According to reference [59] we will assume that  $P(t|D) \propto \delta(t - D^{-\gamma})$ .

**2.2.5. Scaled Brownian motion.** The SBM is a process described by the Langevin equation with a time-dependent diffusivity  $K(t)$ ,

$$\frac{dX(t)}{dt} = \sqrt{2K(t)}\xi(t), \quad (6)$$

where  $\xi(t)$  is white Gaussian noise. If the diffusivity has the form  $K(t) = \alpha K_\alpha t^{\alpha-1}$ , the MSD follows the power-law  $\text{MSD}(t) \propto K_\alpha t^\alpha$ .

### 2.3. Classification model

In machine learning, all classification algorithms may be divided into two classes. The traditional machine learning is a set of statistical learning methods, which do not operate on raw data. Instead, each data sample is characterized by a vector of human-engineered features or attributes. Those vectors are then used as input for the classifier. The second class consists of deep learning methods that identify and extract features on their own. The representation of data is constructed automatically and there is no need for its complex preprocessing as in the case of the feature-based methods.

Deep learning is nowadays the state-of-the-art technology for data classification in many domains and overshadows a little bit the qualities of the classical machine learning. However, the choice of a suitable classification method is usually more subtle than simply looking at its performance. A lot of deep learning methods do indeed an excellent job with respect to the predictions but are extremely complex to interpret. To give an example, the ResNet18 architecture we considered in one of our previous attempts to trajectory classification [57] originally had 11,220,420 parameters. We were able to reduce their number down to 399,556 with a positive impact on accuracy. Although it was an impressive achievement, interpretation of all of those remaining parameters is almost impossible. Thus, if one wants to comprehend the decisions made by an ML algorithm, it could be tempted to go for feature-based methods and gain on interpretability at the expense of accuracy.

Being aware of this trade-off, we deliberately decided to follow the interpretability path in our contribution to the AnDi challenge and chose the extreme gradient boosting (XGB) algorithm with decision trees as base learners [68] as the classification method. We already used similar approach [49, 51, 52], however for different sets of data.

XGB is an ensemble method, which combines multiple decision trees to obtain better predictive performance. A single decision tree [69] is fairly simple to build. The original dataset is split into smaller subsets based on the values of a given feature. The process is recursively repeated until the resulting subsets are homogeneous (all samples from the same class) or further splitting does not improve the classification performance. A splitting feature for each step is chosen according to similarity score measure [70].

Single decision trees are famous for their ease of interpretation. However, they are sensitive to even small variations of data and prone to overfitting. That is why one rather uses forests (ensembles) of trees as reliable classifiers. In the case of the XGB algorithm, the trees are built in a stage-wise fashion. We start with a single tree trained on random subsets of data and features and then at every step, a new tree is added to the ensemble. This new tree learns from the mistakes committed by the previous trees.

### 2.4. Diffusion characteristics for AnDi challenge

The main challenge of feature-based classification methods is the proper choice of features that can differentiate between samples belonging to different classes. The impact of their choice on the final performance of the classifier was already discussed in reference [52]. In this work,

**Table 1.** The features used to characterize the SPT trajectories. The original set of features for the AnDi challenge (left column, see section 2.4 for definitions) have been extended afterwards (right columns, see section 2.5 for further details) to improve the performance of the classifier.

Original features	Additional features
Anomalous exponent	D’Agostino–Pearson test statistic
Diffusion coefficient	Kolmogorov–Smirnov statistic against $\chi^2$ distribution
Asymmetry	Noah exponent
Efficiency	Moses exponent
Empirical velocity autocorrelation function	Joseph exponent
Fractal dimension	Detrending moving average
Maximal excursion	Average moving window characteristics
Mean maximal excursion	Maximum standard deviation
Mean gaussianity	
MSD	
Kurtosis	
Statistics based on $p$ -variation	
Straightness	
Trappedness	

we continue the search for a robust set of features for the classification of different diffusion modes.

In this section, we briefly introduce features that have been used for the AnDi challenge. Since we were not satisfied with the performance of the resulting classifier in task 2 [59], after the end of the challenge we further searched for features that would improve the classification results. Those additional features will be described in section 2.5. Both the original feature set for the challenge and its extension are listed in table 1.

**2.4.1. Anomalous exponent.** In our set of features, we included 4 estimates for the anomalous exponent  $\alpha$  (see equation (3)), calculated using the following methods:

- the standard estimation, based on fitting the empirical TAMSD from equation (2) to equation (3),
- Three estimation methods proposed for the trajectories with noise [71], under the assumption that the noise is normally distributed with zero mean:
  - \* using estimator

$$\hat{\alpha} = \frac{n_{\max} \sum_{n=1}^{n_{\max}} \ln(n) \ln(\widehat{\text{MSD}}(n\Delta t)) - \sum_{n=1}^{n_{\max}} \ln(n) \left( \sum_{n=1}^{n_{\max}} \ln(\widehat{\text{MSD}}(n\Delta t)) \right)}{n_{\max} \sum_{n=1}^{n_{\max}} \ln^2(n) - \left( \sum_{n=1}^{n_{\max}} \ln(n) \right)^2}, \quad (7)$$

with  $n_{\max}$  equal to 0.1 times the trajectory length, rounded to nearest lower integer (but not less than 4),

- \* simultaneous fitting of parameters  $\hat{D}$ ,  $\hat{\alpha}$  and  $\hat{\sigma}$  in the equation

$$\widehat{\text{MSD}}(n\Delta t) = 2d\hat{D}(n\Delta t)^{\hat{\alpha}} + \hat{\sigma}^2, \quad (8)$$



where  $d$  denotes dimension,  $D$  is the diffusion coefficient and  $\sigma^2$ —the variance of noise,

\* simultaneous fitting of parameters  $\hat{D}$  and  $\hat{\alpha}$  in the equation

$$\widehat{\text{MSD}}(n\Delta t) = 2d\hat{D}(\Delta t)^{\hat{\alpha}}(n^{\hat{\alpha}} - 1). \quad (9)$$

**2.4.2. Diffusion coefficient.** We used the diffusion coefficient extracted from the fit of the empirical TA-MSD to equation (3).

**2.4.3. Asymmetry.** The asymmetry of a trajectory can detect directed motion. According to Saxton [32], it can be derived from the gyration tensor, which describes the second moments of positions of a particle. For a 2D random walk of  $N$  steps, the tensor is given by

$$\mathbf{T} = \begin{pmatrix} \frac{1}{N} \sum_{j=1}^N (x_j - \langle x \rangle)^2 & \frac{1}{N} \sum_{j=1}^N (x_j - \langle x \rangle)(y_j - \langle y \rangle) \\ \frac{1}{N} \sum_{j=1}^N (x_j - \langle x \rangle)(y_j - \langle y \rangle) & \frac{1}{N} \sum_{j=1}^N (y_j - \langle y \rangle)^2 \end{pmatrix}, \quad (10)$$

where  $\langle x \rangle = (1/N) \sum_{j=1}^N x_j$  is the average of  $x$  coordinates over all steps in the random walk. We define the asymmetry as [72]

$$A = -\log \left( 1 - \frac{(\lambda_1 - \lambda_2)^2}{2(\lambda_1 + \lambda_2)} \right), \quad (11)$$

where  $\lambda_1$  and  $\lambda_2$  are the principle radii of gyration, i.e. the eigenvalues of the tensor  $\mathbf{T}$ .

**2.4.4. Efficiency.** Efficiency  $E$  measures the linearity of a trajectory. It relates the net squared displacement of a particle to the sum of squared step lengths,

$$E = \frac{|X_{N-1} - X_0|^2}{(N-1) \sum_{i=1}^{N-1} |X_i - X_{i-1}|^2}. \quad (12)$$

It may help to detect directed motion (superdiffusion).

**2.4.5. Empirical velocity autocorrelation function.** Empirical velocity autocorrelation function [73] for lag 1 and point  $n$  is defined as:

$$\chi_n = \frac{1}{N-1} \sum_{i=0}^{N-2} (X_{i+1+n} - X_{i+n})(X_{i+1} - X_i). \quad (13)$$

It can be used to distinguish subdiffusion processes. In our model, we used  $\chi_n$  for points  $n = 1$  and  $n = 2$ .

**2.4.6. Fractal dimension.** Fractal dimension is a measure of the space-filling capacity of a pattern (a trajectory in our case). According to Katz and George [74], the fractal dimension of a planar curve may be calculated as

$$D_f = \frac{\ln N}{\ln(NdL^{-1})}, \quad (14)$$

where  $L$  is the total length of the trajectory,  $N$  is the number of steps, and  $d$  is the largest distance between any two positions. It takes values around 1 for straight trajectories (i.e. directed motion), around 2 for random ones (normal diffusion), and around 3 for constrained trajectories (subdiffusion).

**2.4.7. Maximal excursion.** Maximal excursion of the particle is given by the formula:

$$\text{ME} = \frac{\max(X_{i+1} - X_i)}{X_{N-1} - X_0}. \quad (15)$$

It should detect relatively long jumps (in comparison to the overall displacement).

**2.4.8. Mean maximal excursion.** According to reference [37], the mean maximal excursion is usually a better observable than MSD to determine the anomalous exponent  $\alpha$ . Given the largest distance traveled by a particle,

$$D_N = \max |X_i - X_0|, \quad (16)$$

the mean maximal excursion is defined as its standardized value, i.e.:

$$T_n = \frac{\max(|X_i - X_0|)}{\sqrt{\hat{\sigma}_N^2(t_N - t_0)}}. \quad (17)$$

Here,  $\hat{\sigma}_N$  is a consistent estimator of the standard deviation of  $D_N$ ,

$$\hat{\sigma}_N^2 = \frac{1}{2N\delta t} \sum_{j=1}^N \|X_j - X_{j-1}\|_2^2. \quad (18)$$

**2.4.9. Mean gaussianity.** Gaussianity  $g(n)$  [75] checks the Gaussian statistics of increments of a trajectory and is defined as

$$g(n) = \frac{2\langle r_n^4 \rangle}{3\langle r_n^2 \rangle^2}, \quad (19)$$

where  $\langle r_n^k \rangle$  denotes the  $k$ th moment of the trajectory at time lag  $n$ :

$$\langle r_n^k \rangle = \frac{1}{N-n} \sum_{i=1}^{N-n} |X_{i+n} - X_i|^k. \quad (20)$$

Gaussianity for normal diffusion is equal to 0. The same result should be obtained for FBM, since its increments follow Gaussian distribution. Other types of motion should show deviations from zero.

Instead of looking at gaussianities at single time lags, we will include the mean over all lags as one of the features:

$$\langle g \rangle = \frac{1}{N} \sum_{i=1}^N g(n). \quad (21)$$

**2.4.10. Mean-squared displacement ratio.** MSD ratio gives information about the shape of the corresponding MSD curve. We will define it as

$$\text{MSDR}(n_1, n_2) = \frac{\langle r_{n_1}^2 \rangle}{\langle r_{n_2}^2 \rangle} - \frac{n_1}{n_2}, \quad (22)$$

where  $n_1 < n_2$ .  $\text{MSDR} = 0$  is zero for normal diffusion ( $\alpha = 1$ ). We should get  $\text{MSDR} \leq 0$  for sub- and  $\text{MSDR} \geq 0$  for superdiffusion. We simply took  $n_2 = n_1 + \Delta t$  and calculate an averaged ratio for every trajectory.

**2.4.11. Kurtosis.** Kurtosis gives insight into the asymmetry and peakedness of the distribution of points within a trajectory [72]. To calculate it, the position vectors  $X_i$  are projected onto the dominant eigenvector  $\vec{r}$  of the gyration tensor (10),

$$x_i^p = X_i \cdot \vec{r}. \quad (23)$$

Kurtosis is then defined as the fourth moment of  $x_i^p$ ,

$$K = \frac{1}{N} \sum_{i=1}^N \frac{(x_i^p - \bar{x}^p)^4}{\sigma_{x^p}^4}, \quad (24)$$

with  $\bar{x}^p$  being the mean projected position and  $\sigma_{x^p}$ —the standard deviation of  $x^p$ .

**2.4.12. Statistics based on  $p$ -variation.** The empirical  $p$ -variation is given by the formula [76, 77]:

$$V_m^{(p)} = \sum_{i=0}^{\frac{N}{m}-1} |X_{(i+1)m} - X_{im}|^p. \quad (25)$$

These statistics can be used to detect the fractional Lévy stable motion (including FBM). We defined 6 features based on  $V_m^{(p)}$ :

- power  $\gamma^p$  fitted to  $p$ -variation for lags 1 to 5 [76],
- statistics  $P$  used in reference [52], based on the monotonicity changes of  $V_m^{(p)}$  as a function of  $m$ :

$$P = \begin{cases} 0 & \text{if } V_m^{(p)} \text{ does not change the monotonicity,} \\ 1 & \text{if } V_m^{(p)} \text{ is convex for the highest } p \text{ for which it is not monotonous,} \\ -1 & \text{if } V_m^{(p)} \text{ is concave for the highest } p \text{ for which it is not monotonous.} \end{cases} \quad (26)$$

**2.4.13. Straightness.** Straightness  $S$  measures the average direction change between subsequent steps. It relates the net displacement of a particle to the sum of all step lengths,

$$S = \frac{|X_{N-1} - X_0|}{\sum_{i=1}^{N-1} |X_i - X_{i-1}|}. \quad (27)$$

**2.4.14. Trappedness.** With trappedness we will refer to the probability that a diffusing particle is trapped in a bounded region with radius  $r_0$ . A comparison of analytical and Monte Carlo results for confined diffusion allowed Saxton [32] to estimate this probability with

$$P(D, t, r_0) = 1 - \exp \left( 0.2048 - 0.25117 \left( \frac{Dt}{r_0^2} \right) \right). \quad (28)$$

Here,  $r_0$  is approximated by half of the maximum distance between any two positions along a given trajectory.  $D$  in equation (28) is estimated by fitting the first two points of the MSD curve (i.e. it is the so called short-time diffusion coefficient).

## 2.5. Additional diffusion characteristics

In order to improve the performance of the original classifier, we searched for further feature candidates after the AnDi challenge. The additional features extending the basic set are described below.

**2.5.1. D'Agostino-Pearson test statistic.** D'Agostino-Pearson  $\kappa^2$  test statistic [78, 79] is a goodness-of-fit measure that aims to establish whether or not a given sample comes from a normally distributed sample. It is defined as

$$\kappa^2 = Z_1(g_1) + Z_2(K), \quad (29)$$

where  $K$  is the sample kurtosis given by equation (24) and  $g_1 = m_3/m_2^{3/2}$  is the sample skewness with  $m_j$  being the  $j$ th sample central moment. The transformations  $Z_1$  and  $Z_2$  should bring the distributions of the skewness and kurtosis as close to the standard normal as possible. Their definitions may be found elsewhere [78, 79]. This feature should help to distinguish ATTM and SBM from other motions.

**2.5.2. Kolmogorov-Smirnov statistic against  $\chi^2$  distribution.** The Kolmogorov-Smirnov statistic quantifies the distance between the empirical distribution function of the sample  $F_n(X)$  and the cumulative distribution function  $G_n(X)$  of a reference distribution,

$$D_n = \sup_X |F_n(X) - G_n(X)|. \quad (30)$$

Here,  $n$  is the number of observations (i.e. the length of a trajectory). The value of this statistic for the empirical distribution of squared increments of a trajectory against the sampled  $\chi^2$  distribution has been taken as the next feature. The rationale of such choice is that for a Gaussian trajectory the theoretical distribution of squared increments is the mentioned  $\chi^2$  distribution.

**2.5.3. Noah, Moses, and Joseph exponents.** For processes with stationary increments, there are in principle two mechanisms that violate the Gaussian central limit theorem and produce anomalous scaling of MSD: long-time increment correlations or a flat-tailed increment distribution (in the latter case the second moment is divergent) [80]. These mechanisms are referred to as the Joseph and Noah effects, respectively. FBM is the prototypical process that exhibits the Joseph effect. LW on the other hand is an example of a process with the Noah effect. An anomalous scaling can also be induced by a non-stationary increment distribution [80]. In this case, we deal with the Moses effect. It should help to handle ATTM and SBM trajectories.

All three effects may be quantified by exponents, which will be used as features in our extended set of attributes. Given a stochastic process  $X_t$  and the corresponding increment process  $\delta_t(\tau) = X_{t+\tau} - X_t$ , the Joseph, Moses and Noah exponents are defined as follows.

- (a) Joseph exponent  $J$  is estimated from the ensemble average of the rescaled range statistic:

$$E \left[ \frac{\max_{1 \leq s \leq t} [X_s - \frac{s}{t} X_t] - \min_{1 \leq s \leq t} [X_s - \frac{s}{t} X_t]}{\sigma_t} \right] \sim t^J,$$

$$E \left[ \frac{R_t}{S_t} \right] \sim t^J,$$

where  $R_t$  is calculated as  $R_t = \max_{1 \leq s \leq t} [X_s - \frac{s}{t} X_t] - \min_{1 \leq s \leq t} [X_s - \frac{s}{t} X_t]$  and  $S_t$  is standard deviation of process  $X_t$ .

- (b) Moses exponent  $M$  is determined from the scaling of the ensemble probability distribution of the sum of the absolute value of increments, which can be estimated by the scaling of the median of the probability distribution of  $Y_t = \sum_{s=0}^{t-1} |\delta_s|$ :

$$E[|\delta_t - E(\delta_t)|] \sim t^{M-\frac{1}{2}}.$$

- (c) Noah exponent  $L$  is extracted from the scaling of the ensemble probability distribution of the sum of increment squares, which can be estimated by the scaling of the median of the probability distribution of  $Z_t = \sum_{s=0}^{t-1} \delta_s^2$ :

$$m[Z_t] \sim t^{2L+2M-1}.$$

**2.5.4. Detrending moving average.** The detrending moving average (DMA) statistic [81, 82] is given by

$$\text{DMA}(\tau) = \frac{1}{N - \tau + 1} \sum_{i=\tau}^N (X_i - \bar{X}_i^\tau)^2, \quad (31)$$

for  $\tau = 1, 2, \dots$ , where  $\bar{X}_i^\tau$  is a moving average of  $\tau$  observations, i.e.  $\bar{X}_i^\tau = \frac{1}{\tau+1} \sum_{j=0}^{\tau} X_{i-j}$ .

As mentioned in reference [82], the DMA-based statistical tests can help in the detection of the SBM. In our model, we used two values of DMA for each trajectory as input features, namely DMA(1) and DMA(2).

**2.5.5. Average moving window characteristics.** As the moving average methods have already been successfully applied to many problems (see for example reference [83] for change point detection), we decided to accommodate them in our feature set as well. We added eight features based on the formula

$$\text{MW} = \frac{1}{2(N+1)} \sum_{i=0}^N \left| \text{sgn} \left( \bar{X}_{i+1}^{(m)} - \bar{X}_i^{(m)} \right) - \text{sgn} \left( \bar{X}_{i+1}^{(m)} - \bar{X}_i^{(m)} \right) \right|, \quad (32)$$

where  $\bar{X}^{(m)}$  denotes a statistic of the process calculated within the window of length  $m$  and  $\text{sgn}$  is the signum function. In particular, we used the mean and the standard deviation for  $\bar{X}$  and calculated moving window (MW) with windows of lengths  $m = 10$  and  $m = 20$  separately for  $x$  and  $y$  coordinates.

**Table 2.** Parameters of the diffusion models used to generate the synthetic trajectories.

Parameter	Meaning	Parameter ranges
$\alpha$	Diffusion exponent	$\langle 0.05, 2 \rangle$
$n$	Length of a trajectory	$\langle 10, 1000 \rangle$
$\sigma_{\text{noise}}$	Standard deviation (level) of noise	$\{0.1, 0.5, 1\}$

**2.5.6. Maximum standard deviation.** The idea of the MW helped us to introduce another two features based on the standard deviation  $\sigma_m$  of the process calculated within a window of length  $m$ . They are given by

$$\text{MXM} = \frac{\max(\sigma_m(t))}{\min(\sigma_m(t))} \quad (33)$$

and

$$\text{MXC} = \frac{\max |\sigma_m(t+1) - \sigma_m(t)|}{\sigma}, \quad (34)$$

where  $\sigma$  denotes the overall standard deviation of the sample. We used the window of length  $m = 3$  and calculated the features for both coordinates separately. They should improve the detection of ATTM type of movements.

### 3. Data

XGB, i.e. the algorithm we decided to use, is an example of a supervised model. In other words, it requires a training dataset consisting of trajectories and the corresponding labels indicating their diffusion type (i.e. the diffusion model).

We used a synthetic set of trajectories produced with the `andi-dataset` package [84] to train the classifier. Once the trajectories were generated, a vector of features was calculated for each of them in the preprocessing phase. The resulting set consisting of vectors of attributes and their labels was used as input for the XGB model.

We generated  $3 \times 10^5$  trajectories for each diffusion model. Those trajectories were evenly distributed over 20 ranges of lengths, from very short ones ( $n \in \langle 10, 50 \rangle$ ) to really long ones ( $n \in \langle 951, 1000 \rangle$ ). The whole dataset was divided into training, test, and validation subsets in the proportion of 0.7:0.15:0.15, respectively. The classifier is initially fit on the training data. The validation set provides an unbiased evaluation of a model while tuning its hyperparameters. Test data is used to provide an unbiased assessment of the final model. Stratified sampling was used to ensure a proper balance of data in the subsets. It should be noted that in this article, the dataset is significantly larger than the one used for the AnDi challenge [58] (only  $7 \times 10^4$  trajectories).

The parameters used to generate the trajectories are shown in table 2. Their ranges have been constrained by the organizers of the challenge. The range of diffusion exponents prevents stationary paths. All trajectories were standardized so that the distribution of displacements over the unit time is characterized by the standard deviation equal to 1.

To better simulate experimental endeavors, each trajectory was contaminated with a finite localization error. For that purpose, a random number from a normal distribution  $N(0, \sigma_{\text{noise}})$  was added to each trajectory coordinate.

**Table 3.** Hyperparameters of the final XGB models used for classification.

Hyperparameter	Parameter value
min_child_weight	7
max_depth	10
gamma	4
eval_metric	mae
eta	0.3
base_score	1

## 4. Results

We used the implementation of the XGB from the `xgboost` module [70]. All computations were carried out on a cluster of 24 CPUs with a total memory of 25 GB. A randomized search strategy over the hyperparameter space of the classifier (the `RandomizedSearchCV` function in `scikit-learn` module [85]) has been deployed in order to optimize the model. The final values of the hyperparameters are summarized in table 3. The parameter `min_child_weight` indicates the minimum of instance weight (corresponding to the number of instances) needed in a child node of the single decision tree. If the tree partition step results in a leaf node with the sum of instance weight less than its value, the building process of the tree will give up further partitioning. `Max_depth` determines the maximum depth of a tree. The value of `gamma` indicates the minimum loss reduction to make a further split on a leaf node of the tree. The evaluation metric for the validation step is specified by `eval_metric`. Mean absolute error (mae) is used in our case. `Eta` is the learning rate used to prevent overfitting of the classifier. And the `base_score` is the global bias, i.e. the initial prediction score of all instances.

The values of the hyperparameters were optimized for the original set of features. However, the same values of hyperparameters were also used with the extended set for comparison purposes.

### 4.1. Performance of the classifiers

The performance of the classifiers in task 2 of the AnDi challenge was evaluated with the  $F_1$  score,

$$F_1 = \frac{TP}{TP + 0.5(FP + FN)}, \quad (35)$$

where TP, FP, and FN are the true positive, false positive, and false negative rates, respectively. Since we are dealing here with a multilabel classification problem, a micro-averaged  $F_1$  score was calculated to assess the aggregated contributions of all classes (i.e. the sums of all TP, FP and FN were first determined and then inserted into equation (35)). For balanced sets (same number of samples for each class), like those of the AnDi challenge, the micro-averaged  $F_1$  score coincides with the accuracy of the classifier.

Results for the performance of the classifiers are shown in table 4. As we can see, the model with the original set of features achieves an accuracy of 0.73. After adding the features described in section 2.5 we observe a significant improvement of its performance. The accuracy of the extended version (0.83) is comparable with the best methods of the AnDi challenge [59].

**Table 4.** Micro-averaged  $F_1$  score (i.e. accuracy) of both the original and the extended classifiers (see sections 2.4 and 2.5 for more details).

Base XGB model	Extended XGB model
0.73	0.83

Base XGB model						Extended XGB model						
	ATTM	37% 3363	22% 1965	3% 268	1% 62	37% 3348	ATTM	61% 5508	13% 1131	5% 418	1% 47	21% 1905
	CTRW	11% 993	85% 7702	1% 120	0% 3	2% 198	CTRW	5% 452	94% 8461	1% 79	0% 2	0% 30
	FBM	1% 104	1% 113	72% 6425	7% 621	19% 1722	FBM	3% 278	1% 93	82% 7399	2% 157	12% 1044
	LW	0% 38	0% 5	5% 472	92% 8276	2% 220	LW	0% 24	0% 2	1% 123	98% 8795	1% 63
	SBM	11% 996	1% 114	6% 517	1% 87	81% 7269	SBM	10% 893	0% 17	7% 596	0% 39	83% 7445
		ATTM	CTRW	FBM	LW	SBM		ATTM	CTRW	FBM	LW	SBM

**Figure 2.** Normalized confusion matrices of the models. Rows correspond to the true labels and columns to the predicted ones.

Inspection of the confusion matrices of the classifiers may give us additional insight into partial contributions of the overall performance. To recall, an element  $c_{ij}$  of the confusion matrix indicates how many observations known to belong to class  $i$  (true label) are predicted to be in class  $j$  (predicted labels) [68]. Normalized confusion matrices for both models are shown in figure 2. The best performance of the base model is observed for LWs. Among all LW samples in the test data, 92% of them were correctly recognized. Most of the misclassified LW trajectories were assigned to FBM motion (5%) and SBM (2%). The accuracy of the classifier for CTRW (85%) and SBM (81%) is smaller, but still quite good. The incorrectly classified CTRW trajectories have been usually confused with ATTM. In the case of SBM, many trajectories have been assigned to ATTM (11%) and FBM (6%) classes.

We have not expected such a moderate performance of the original classifier for FBM (72%). In our earlier attempts [49, 51, 52], we usually got much higher performance for this type of motion. However, it should be stressed that previously we trained our classifiers on different set of trajectories than used for the AnDi challenge. Moreover, we usually used longer trajectories. Thus, it seems that the features in the original set are lacking the power to differentiate FBM successfully from other types of motion considered in this paper, in particular from SBM (19% of misclassifications) and LW (7%).

Finally, it should be emphasized that the performance of the base classifier for ATTM is unsatisfactory (37%). Most of the ATTM trajectories are confused with SBM (37%) and CTRW (22%). It is clear that the base model cannot separate the different origins of trajectories from models with time-varying diffusion coefficient  $D$ .

Adding the features from section 2.5 to the input vectors has significantly improved the overall accuracy of the classifier. As can be seen in the right plot of figure 2, all of the partial contributions to the accuracy improved as well. Although the performance for ATTM remains weak (61% only), it is much higher than in the previous case. As for the other classes, the modified classifier achieves very good accuracies for LW (98%) and CTRW (94%) and good



Extended XGB model - F1 scores						
[10,50]	0.39	0.8	0.58	0.87	0.58	0.64
[51,100]	0.51	0.87	0.74	0.94	0.64	0.74
[101,150]	0.57	0.87	0.76	0.97	0.66	0.77
[151,200]	0.6	0.89	0.82	0.98	0.73	0.8
[201,250]	0.68	0.92	0.83	0.98	0.75	0.83
[251,300]	0.67	0.91	0.82	0.98	0.75	0.83
[301,350]	0.66	0.92	0.84	0.98	0.74	0.83
[351,400]	0.69	0.92	0.86	0.99	0.76	0.85
[401,450]	0.72	0.91	0.86	0.99	0.8	0.86
[451,500]	0.72	0.91	0.88	0.99	0.8	0.86
[501,550]	0.72	0.9	0.91	1	0.82	0.87
[551,600]	0.75	0.92	0.89	0.99	0.82	0.87
[601,650]	0.75	0.92	0.9	1	0.83	0.88
[651,700]	0.78	0.92	0.91	1	0.85	0.89
[701,750]	0.77	0.94	0.89	0.99	0.82	0.88
[751,800]	0.77	0.94	0.91	1	0.83	0.89
[801,850]	0.78	0.93	0.92	1	0.85	0.9
[851,900]	0.76	0.92	0.92	1	0.83	0.88
[901,950]	0.76	0.92	0.93	1	0.83	0.89
[951,1000]	0.76	0.93	0.91	1	0.84	0.89
	ATTM	CTRW	FBM	LW	SBM	avg/total

**Figure 3.**  $F_1$  scores as functions of different lengths of the trajectories. Columns represent  $F_1$  score for each type of the motion, rows correspond to ranges of the lengths.

ones for SBM (83%) and FBM (82%). The biggest challenge is still to distinguish ATTM and FBM from SBM trajectories.

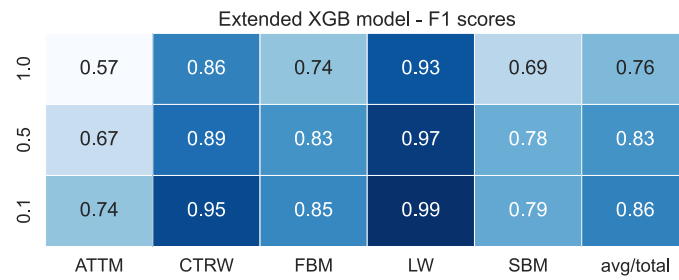
#### 4.2. Impact of trajectory lengths

In figure 3, the  $F_1$  scores for different ranges of the trajectory lengths in the case of the extended model are shown. Not surprisingly, the classifier struggles a lot with very short trajectories. In this regime, many of the paths are indistinguishable from normal diffusion. Moreover, in the case of ATTM and SBM models, the information contained in the short trajectories may be not enough for detecting the time-varying diffusion coefficient  $D$ . Interestingly, the accuracy of the FBM classification increases with the sample length to achieve around 90% for trajectories longer than 500 steps.

It is worth to mention that the difficulties with very short trajectories are not specific to the XGB in particular or to feature-based methods in general. All methods presented in the challenge were afflicted by similar problems in this regime [59].

#### 4.3. Impact of the level of noise

In figure 4, the  $F_1$  scores for different levels of noise (i.e. different values of its standard deviation) are presented in case of our extended model. While the classifier performs quite well for the CTRW and LW even in the case of high noise levels, its accuracy significantly decreases



**Figure 4.**  $F_1$  scores for different levels of trajectories' noise. Columns represent  $F_1$  scores for each type of the motion, rows correspond to standard deviations of noise.

**Table 5.** Permutation feature importance values for the XGB classifier.

Model characteristic	Mean feature importance	Std. deviation
mw_y_mean10	0.1089	0.0015
mw_x_mean10	0.1013	0.0008
dagostino_y	0.0779	0.0016
dagostino_x	0.0727	0.0010
M	0.0645	0.0003
mw_y_mean20	0.0516	0.0009
mw_x_mean20	0.0497	0.0007
max_std_y	0.0422	0.0013
max_std_x	0.0402	0.0010
alpha	0.0295	0.0008
p_var_1	0.0283	0.0005
fractal_dimension	0.0272	0.0007
mean_gaussianity	0.0240	0.0004
vac_lag_1	0.0211	0.0013
ksstat_chi2	0.0194	0.0011
max_std_change_x	0.0178	0.0012
max_std_change_y	0.0111	0.0007
max_ts	0.0078	0.0007
L	0.0074	0.0005
D	0.0068	0.0007

for the remaining diffusion models. It seems that the classifier is not able to capture enough characteristics of ATTM, SBM and FBM types of movements from noise trajectories in order to successfully distinguish them from each other. Interestingly, in a series of computer experiments we observed that a classifier trained on very noisy data performs still reasonably well on a test set with low levels of noise.

#### 4.4. Features importance

One of the benefits of feature-based methods is the relative ease of their interpretation. In particular, it is possible to assign a score to each of the features indicating how important these features are to the model's predictions. This importance aids to extract relevant knowledge concerning the relationships contained in the data or learned by the model. But it can also help in the selection of the meaningful features for a simplified version of the classifier.

**Table 6.** Gain feature importances for the XGB classifier.

Model characteristic	Feature importance
alpha	177.31
mean_squared_displacement_ratio	135.70
mw_x_mean10	92.59
mw_y_mean10	60.82
dagostino_y	60.21
max_ts	48.25
max_std_x	39.82
dagostino_x	38.16
max_std_y	27.55
M	24.56
fractal_dimension	23.44
p-variation	20.20
mean_gaussianity	13.93
mw_y_std10	13.27
p_var_1	13.02
mw_x_std10	12.84
max_std_change_x	11.01
L	10.73
max_std_change_y	9.88
mw_x_std20	9.75

One of the popular feature assessment techniques is the permutation feature importance [68]. It is defined as the decrease in a classifier's score when a single feature values are randomly shuffled. Since this procedure breaks the relationship between the feature and the target, the drop in the score indicates how much the classifier depends on the feature. In table 5, the 20 most important features in our model according to the permutation method are presented. It is worth to note that the features added in the extended version of the model play the key role among them. In particular, the average MW features have the highest score, followed by the D'Agostino–Pearson test statistic  $\kappa^2$ .

Since the permutation importance is known to have several drawbacks while working with correlated features [86] and this is the case in our model (e.g. we have several features based on MSD), we decided to check the gain importance as well [87]. In this approach, the importance of a feature represents the relative contribution of the feature to the model, calculated by taking each feature's contribution for each tree in the model. A higher value of this metric when compared to other features implies it is more important for generating predictions. The top 20 features according to gain are listed in table 6. Although the moving average window and D'Agostino–Pearson characteristics are still very important, the anomalous exponent  $\alpha$  and the MSD ratio get now the highest scores. But again, many of the most important features were added in the extended version of the model. Thus it seems we made some reasonable choices to improve the classification accuracy.

Last but not least, to further elaborate on that issue, we will employ another well known technique—the SHAP values based on feature attribution, i.e. the impact of having a certain value of a given feature in comparison to the prediction we would make if that feature took some baseline value [88, 89]. Compared the previous methods, SHAP values may be used to

**Table 7.** SHAP values for most important features for different diffusion models in the case of the extended XGB model.

ATTM		FBM		SBM	
Model characteristic	SHAP value	Model characteristic	SHAP value	Model characteristic	SHAP value
M	0.13	M	0.07	M	0.20
max_std_x	0.08	alpha	0.06	dagostino_y	0.05
max_std_y	0.08	dagostino_y	0.06	dagostino_x	0.04
dagostino_y	0.07	dagostino_x	0.06	alpha	0.03
mw_x_mean10	0.06	max_std_x	0.05	max_std_y	0.03
mw_y_mean10	0.06	max_std_y	0.05	max_std_x	0.03
mean_gaussianity	0.06	max_std_change_y	0.03	mw_y_mean10	0.02
dagostino_x	0.06	mean_gaussianity	0.03	ksstat_chi2	0.02
p_var_1	0.05	p_var_1	0.03	vac_lag_1	0.02
alpha	0.05	vac_lag_1	0.03	mean_gaussianity	0.02

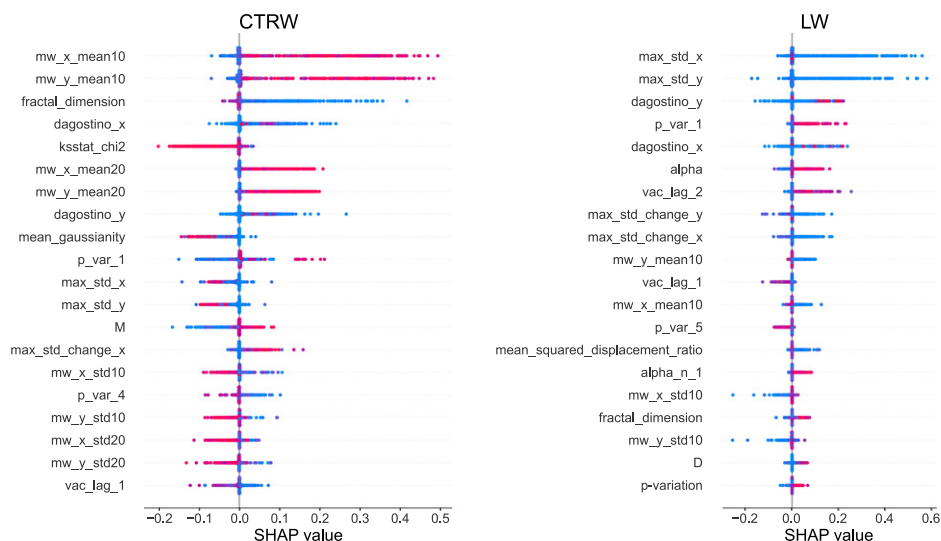
CTRW		LW	
Model characteristic	SHAP value	Model characteristic	SHAP value
mw_x_mean10	0.07	max_std_x	0.05
mw_y_mean10	0.07	max_std_y	0.05
fractal_dimension	0.04	dagostino_y	0.02
dagostino_x	0.03	p_var_1	0.02
ksstat_chi2	0.02	dagostino_x	0.02
mw_x_mean20	0.02	alpha	0.02
mw_y_mean20	0.02	vac_lag_2	0.01
dagostino_y	0.02	max_std_change_y	0.01
mean_gaussianity	0.02	max_std_change_x	0.01
p_var_1	0.01	mw_y_mean10	0.01

explain individual predictions. Moreover, they offer a more consistent approach to assess the overall contribution of features to the final outcome of the model.

In table 7, SHAP values for the most important features in different diffusion models are shown, in the case of the extended model. It should be noted that for every single diffusion model, the additional features turned out to be the most important ones, in agreement with the previous methods. In other words, the additional features prove to be an excellent choice. This also shows how critical the choice of features can be for a classifier (see reference [52] for further details).

The Moses exponent  $M$ , which relates to non-stationary increments, turned out to be the essential attribute for distinguishing ATTM and SBM from the rest of models; it seems to be quite important for FBM as well. However, in the latter case, the anomalous exponent is the most meaningful feature, followed by the D'Agostino–Pearson test statistic  $\kappa^2$ . Characteristics based on the average MW have the highest importance for CTRW. And finally, the maximum standard deviation feature is the top one for LW.

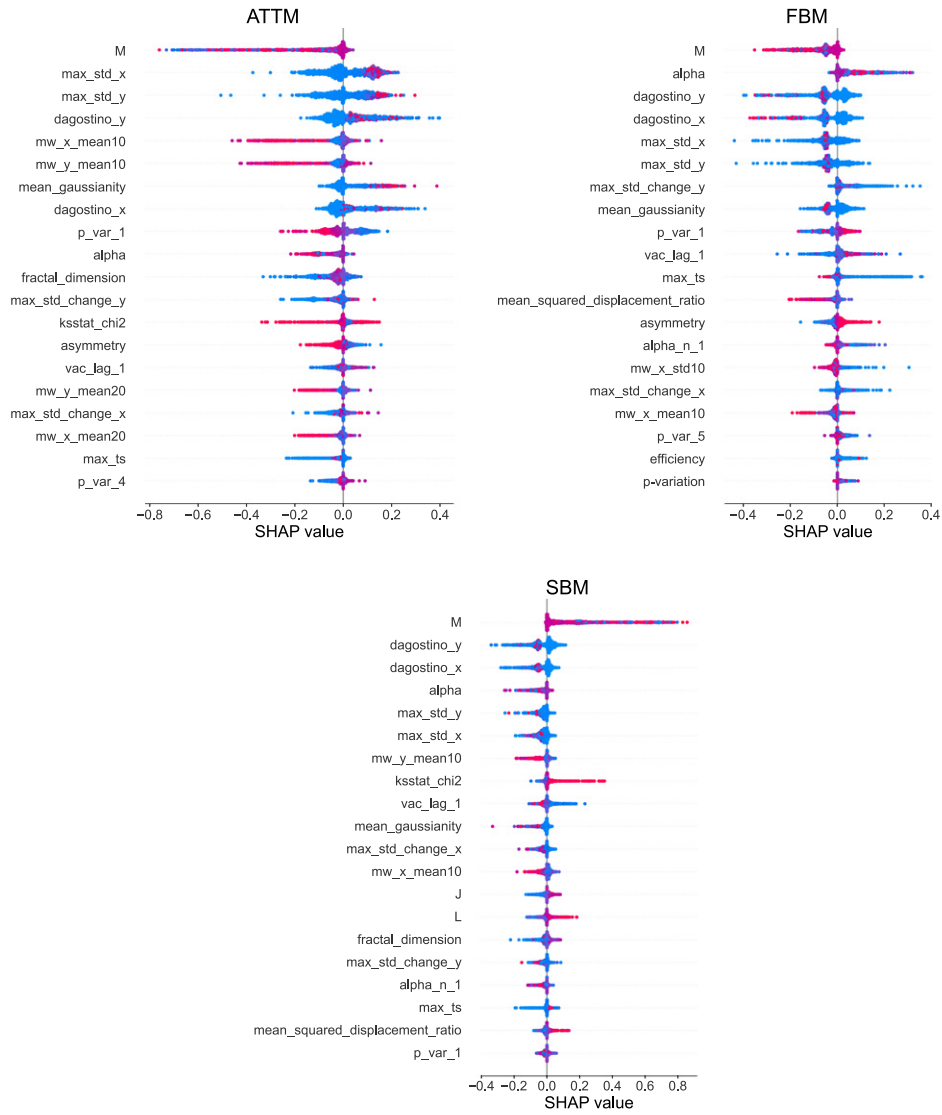
In figures 5 and 6, we may check how each of the 20 most relevant features impacts the decisions of the classifier for each diffusion model. A dot in the plot corresponds to a single instance of the explanation given by the classifier. The characteristics on the  $Y$  axis are sorted by their mean importance. The SHAP values on the horizontal axis reflect the positive or negative influence of each factor on the model prediction. A positive SHAP value increases the



**Figure 5.** SHAP values for 20 most relevant features in case of CTRW and LW. Dots represent the value for a single explanation (trajectory). The characteristics on the vertical axis are sorted by their mean importance. The SHAP values on the horizontal axis reflect the positive or negative influence of each factor on the model prediction. A positive SHAP value improves the likelihood that the sample will be classified as a specific model. A longer tail indicates a greater impact on the prediction. The dependent variable  $Y$  is positively correlated with high SHAP values on the positive side of the  $X$  axis.

likelihood that the sample will be classified as a specific model, whereas a negative SHAP value—that the observation belongs to one of the other classes. Larger values of SHAP (positive or negative) indicate a greater impact on the final prediction. Additionally, the colormap represents the feature values—red being high and blue being low (the color values are relative). The dependent variable  $Y$  is positively correlated with high SHAP values on the positive side of the  $X$  axis. In other words, those values have a beneficial effect on prediction. For instance, in the case of CTRW, the larger the value of  $mw\_y\_mean\_10$ , the higher SHAP attribution, thus the likelihood of classification as CTRW increases.

As can be seen, the SHAP values for CTRW and LW indicate that our classifier is more likely to distinguish these classes from the rest (some features have a long-tailed distribution of values and the colors indicating positive and negative impact are well separated from each other). In contrast, ATTM, FBM and SBM models reveal a high concentration of samples near 0 for many features. Moreover, the colors are mixed with each other indicating that there is no unique relationship between the SHAP value of a feature and its impact on the prediction. In this case, only a combination of all features can give us a reasonable outcome. In other words, it would be rather impossible to build a statistical test to distinguish these three models that operates with a single feature (among the ones present in our set).



**Figure 6.** SHAP values for the 20 most relevant features in the case of ATTM, FBM, and SBM. See caption of figure 5 for further details.

## 5. Conclusion

In this paper, we have introduced two sets of attributes that may be used for the classification of SPT data with help of feature-based machine learning methods. The first of these sets was proposed as our contribution to task 2 of the AnDi challenge [58, 59]. The latter one is the result of our attempt to improve the performance of the classifier presented in the challenge.

The XGB was used as the classification model. It is an ensemble method, which combines multiple decision trees to obtain better predictive performance. Although the deep-learning approach constitutes the state-of-the-art technology for data classification in many domains,

we deliberately decided to pick one of the traditional machine learning algorithms due to its superior interpretability.

Our original method turned out to be inferior to the top AnDi challenge teams, but still offered a reasonable performance of 73%. Moreover, further elaboration on the feature set after the challenge allowed us to significantly improve its accuracy by adding some new characteristics and to achieve performance similar to the winning teams (83% in 2D). Although the final accuracy is already quite good, there is still room for improvement, in particular in the case of short trajectories. Thus, more research into meaningful characteristics is encouraged.

Our results show how crucial the choice of features is for good classification performance. Moreover, we were able to identify the most important among the features for all diffusion types. It does not only contribute to our overall understanding of the decisions made by the classifier. The analysis of feature importance may also be helpful in the selection of meaningful features that can be then used to build simpler classifiers trained on subsets of the attributes.

## Acknowledgments

The work of PK, HL-O, and JS was supported by NCN-DFG Beethoven Grant No. 2016/23/G/ST1/04083. Calculations have been partially carried out using resources provided by Wrocław Centre for Networking and Supercomputing (<http://wcss.pl>).

## Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

## ORCID iDs

Patrycja Kowalek  <https://orcid.org/0000-0002-0743-5125>  
Hanna Loch-Olszewska  <https://orcid.org/0000-0002-3108-4839>  
Łukasz Łaszczuk  <https://orcid.org/0000-0002-6615-661X>  
Jarosław Opała  <https://orcid.org/0000-0002-5016-3997>  
Janusz Szwabiński  <https://orcid.org/0000-0002-6992-3634>

## References

- [1] Manzo C and Garcia-Parajo M F 2015 *Rep. Prog. Phys.* **78** 124601
- [2] Shen H, Tauzin L J, Baiyasi R, Wang W, Moringo N, Shuang B and Landes C F 2017 *Chem. Rev.* **117** 7331
- [3] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 *Phys. Chem. Chem. Phys.* **16** 24128
- [4] Mahowald J, Arcizet D and Heinrich D 2009 *ChemPhysChem* **10** 1559
- [5] Einstein A 1956 *Investigations on the Theory of the Brownian Movement* ed R Furth (New York: Dover)
- [6] Lemons D S and Gythiel A 1997 *Am. J. Phys.* **65** 1079
- [7] Klages R, Radons G and Sokolov I M 2008 *Anomalous Transport* (New York: Wiley)
- [8] Metzler R and Klafter J 2000 *Phys. Rep.* **339** 1



- [9] Hoze N, Nair D, Hosy E, Sieben C, Manley S, Herrmann A, Sibarita J-B, Choquet D and Holcman D 2012 *Proc. Natl Acad. Sci. USA* **109** 17052–7
- [10] Saxton M J 1994 *Biophys. J.* **67** 2110
- [11] Berry H and Chaté H 2014 *Phys. Rev. E* **89** 022708
- [12] Weiss M, Elsner M, Kartberg F and Nilsson T 2004 *Biophys. J.* **87** 3518
- [13] Höfling F and Franosch T 2013 *Rep. Prog. Phys.* **76** 046602
- [14] Golding I and Cox E C 2006 *Phys. Rev. Lett.* **96** 098102
- [15] Tolić-Nørrelykke I M, Munteanu E-L, Thon G, Oddershede L and Berg-Sørensen K 2004 *Phys. Rev. Lett.* **93** 078102
- [16] Jeon J-H, Tejedor V, Burov S, Barkai E, Selhuber-Unkel C, Berg-Sørensen K, Oddershede L and Metzler R 2011 *Phys. Rev. Lett.* **106** 048103
- [17] Bronstein I, Israel Y, Kepten E, Mai S, Shav-Tal Y, Barkai E and Garini Y 2009 *Phys. Rev. Lett.* **103** 018102
- [18] Weigel A V, Tamkun M M and Krapf D 2013 *Proc. Natl Acad. Sci. USA* **110** E4591
- [19] Heinemann F, Vogel S K and Schwille P 2013 *Biophys. J.* **104** 1465
- [20] Torreno-Pina J A, Castro B M, Manzo C, Buschow S I, Cambi A and Garcia-Parajo M F 2014 *Proc. Natl Acad. Sci. USA* **111** 11037
- [21] Bouchaud J-P and Georges A 1990 *Phys. Rep.* **195** 127
- [22] Ghosh S K, Cherstvy A G, Grebenkov D S and Metzler R 2016 *New J. Phys.* **18** 013027
- [23] Bursac P, Fabry B, Trepas X, Lenormand G, Butler J P, Wang N, Fredberg J J and An S S 2007 *Biochem. Biophys. Res. Commun.* **355** 324
- [24] Kahana A, Kenan G, Feingold M, Elbaum M and Granek R 2008 *Phys. Rev. E* **78** 051912
- [25] Richardson L F 1926 *Proc. R. Soc. A* **110** 709
- [26] Shlesinger M F, West B J and Klafter J 1987 *Phys. Rev. Lett.* **58** 1100
- [27] Viswanathan G M, da Luz M G E, Raposo E P and Stanley H E 2009 *The Physics of Foraging* (Cambridge: Cambridge University Press)
- [28] Zaburdaev V, Denisov S and Klafter J 2015 *Rev. Mod. Phys.* **87** 483
- [29] Michalet X 2010 *Phys. Rev. E* **82** 041914
- [30] Saxton M J and Jacobson K 1997 *Annu. Rev. Biophys. Biomol. Struct.* **26** 373–99
- [31] Kepten E, Weron A, Sikora G, Burnecki K and Garini Y 2015 *PLoS One* **10** e0117722
- [32] Saxton M J 1993 *Biophys. J.* **64** 1766
- [33] Grebenkov D S, Vahabi M, Bertseva E, Forró L and Jeney S 2013 *Phys. Rev. E* **88** 040701
- [34] Fuliński A 2017 *J. Phys. A: Math. Theor.* **50** 054002
- [35] Raupach C, Zitterbart D P, Mierke C T, Metzner C, Müller F A and Fabry B 2007 *Phys. Rev. E* **76** 011918
- [36] Burov S, Tabei S M A, Huynh T, Murrell M P, Philipson L H, Rice S A, Gardel M L, Scherer N F and Dinner A R 2013 *Proc. Natl Acad. Sci. USA* **110** 19689
- [37] Tejedor V, Bénichou O, Voituriez R, Jungmann R, Simmel F, Selhuber-Unkel C, Oddershede L B and Metzler R 2010 *Biophys. J.* **98** 1364
- [38] Burnecki K, Kepten E, Garini Y, Sikora G and Weron A 2015 *Sci. Rep.* **5** 11306
- [39] Schütz G J, Schindler H and Schmidt T 1997 *Biophys. J.* **73** 1073
- [40] Das R, Cairo C W and Coombs D 2009 *PLoS Comput. Biol.* **5** e1000556
- [41] Slatore P J, Cairo C W and Burroughs N J 2015 *PLoS One* **10** e0140759
- [42] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R and Fei-Fei L 2014 *Proc. Computer Vision and Pattern Recognition (CVPR) Conf.*
- [43] Deng L, Hinton G and Kingsbury B 2013 *2013 IEEE Int. Conf. Acoustics, Speech and Signal Processing* pp 8599–603
- [44] Collobert R and Weston J 2008 *Proc. 25th Int. Conf. Machine Learning*
- [45] Monnier N, Guo S-M, Mori M, He J, Lénárt P and Bathe M 2012 *Biophys. J.* **103** 616
- [46] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 *Phys. Chem. Chem. Phys.* **20** 29018
- [47] Cherstvy A G, Thapa S, Wagner C E and Metzler R 2019 *Soft Matter* **15** 2526
- [48] Wagner T, Kroll A, Haramagatti C R, Lipinski H-G and Wiemann M 2017 *PLoS One* **12** e0170165
- [49] Kowalek P, Loch-Olszewska H and Szwabiński J 2019 *Phys. Rev. E* **100** 032410
- [50] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 *New J. Phys.* **22** 013010
- [51] Janczura J, Kowalek P, Loch-Olszewska H, Szwabiński J and Weron A 2020 *Phys. Rev. E* **102** 032402



- [52] Loch-Olszewska H and Szwabiński J 2020 *Entropy* **22** 1436
- [53] Dosset P, Rassam P, Fernandez L, Espenel C, Rubinstein E, Margeat E and Milhiet P-E 2016 *BMC Bioinform.* **17** 197
- [54] Bo S, Schmidt F, Eichhorn R and Volpe G 2019 *Phys. Rev. E* **100** 010102
- [55] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 *Biophys. J.* **117** 185
- [56] Gentili A and Volpe G 2021 *J. Phys. A: Math. Theor.* **54** 314003
- [57] Gajowczyk M and Szwabiński J 2021 *Entropy* **23** 649
- [58] Muñoz-Gil G, Volpe G, García-March M A, Metzler R, Lewenstein M and Manzo C 2020 *Emerging Topics in Artificial Intelligence*
- [59] Muñoz-Gil G et al 2021 *Nat. Commun.* **12** 6253
- [60] Scher H and Montroll E W 1975 *Phys. Rev. B* **12** 2455
- [61] Mandelbrot B B and Van Ness J W 1968 *SIAM Rev.* **10** 422
- [62] Klafter J and Zumofen G 1994 *Phys. Rev. E* **49** 4873
- [63] Klafter J, Shlesinger M F and Zumofen G 1996 *Phys. Today* **49** 33
- [64] Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J 2014 *Phys. Rev. Lett.* **112** 150603
- [65] Lim S C and Muniandy S V 2002 *Phys. Rev. E* **66** 021114
- [66] Qian H, Sheetz M P and Elson E L 1991 *Biophys. J.* **60** 910
- [67] Weigel A V, Simon B, Tamkun M M and Krapf D 2011 *Proc. Natl Acad. Sci. USA* **108** 6438
- [68] Raschka S 2015 *Python Machine Learning* (Birmingham: Packt Publishing)
- [69] Song Y-Y and Ying L 2015 *Shanghai Arch. Psychiatry* **27** 130
- [70] Chen T and Guestrin C 2016 A scalable tree boosting system *Pro. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '16)* (San Francisco, CA: Association for Computing Machinery) pp 785–94
- [71] Lanoiselée Y, Sikora G, Grzesiek A, Grebenkov D S and Wyłomańska A 2018 *Phys. Rev. E* **98** 062139
- [72] Helmuth J A, Burckhardt C J, Koumoutsakos P, Greber U F and Sbalzarini I F 2007 *J. Struct. Biol.* **159** 347
- [73] Weber S C, Spakowitz A J and Theriot J A 2010 *Phys. Rev. Lett.* **104** 238102
- [74] Katz M and George E 1985 *Bull. Math. Biol.* **47** 273
- [75] Ernst D, Köhler J and Weiss M 2014 *Phys. Chem. Chem. Phys.* **16** 7686
- [76] Burnecki K and Weron A 2010 *Phys. Rev. E* **82** 021130
- [77] Magdziarz M, Weron A, Burnecki K and Klafter J 2009 *Phys. Rev. Lett.* **103** 180602
- [78] D'Agostino R B 1971 *Biometrika* **58** 341
- [79] D'Agostino R and Pearson E S 1973 *Biometrika* **60** 613
- [80] Aghion E, Meyer P G, Adlakha V, Kantz H and Bassler K E 2021 *New J. Phys.* **23** 023002
- [81] Sikora G 2018 *Chaos Solitons Fractals* **116** 54
- [82] Balcerek M, Burnecki K, Sikora G and Wyłomańska A 2021 *Chaos* **31** 063101
- [83] Hubicka K and Janczura J 2020 *Phys. Rev. E* **101** 022107
- [84] Muñoz-Gil G, Requena B, Volpe G and Garcia-March M A 2021 *AnDiChallenge/ANDI\_Datasets: Challenge 2020 Release (v.1.0)* (<https://doi.org/10.5281/zenodo.4775311>)
- [85] Pedregosa F et al 2011 *J. Mach. Learn. Res.* **12** 2825–30
- [86] Hooker G, Mentch L and Zhou S 2021 *Stat. Comput.* **31** 82
- [87] 2022 Introduction to boosted trees <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
- [88] Lundberg S and Lee S-I 2017 A unified approach to interpreting model predictions (arXiv:1705.07874 [cs.AI])
- [89] Lundberg S M and Lee S-I 2018 Consistent feature attribution for tree ensembles (arXiv:1706.06060 [cs.AI])