

Dead or alive

Citation for published version (APA):

Janzen, G., Smeets, X. L. J. A., Debets, V. E., Luo, C., Storm, C., Janssen, L. M. C., & Ciarella, S. (2023). Dead or alive: Distinguishing active from passive particles using supervised learning. *EPL*, 143(1), Article 17004. <https://doi.org/10.1209/0295-5075/acdf1b>

Document license:
CC BY

DOI:
[10.1209/0295-5075/acdf1b](https://doi.org/10.1209/0295-5075/acdf1b)

Document status and date:
Published: 01/07/2023

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.



LETTER • OPEN ACCESS

Dead or alive: Distinguishing active from passive particles using supervised learning^(a)

To cite this article: Giulia Janzen *et al* 2023 *EPL* **143** 17004

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

You may also like

- [Force-free and autonomous active Brownian ratchets](#)
Constantin Rein, Martin Kolá, Klaus Kroy et al.
- [Dosimetric comparison of autocontouring techniques for online adaptive proton therapy](#)
Andreas Johan Smolders, Evangelia Chouliilitsa, Katarzyna Czerska et al.
- [Tribological and microstructure studies of LM26/SiC metal matrix composite materials and structures for high temperature applications](#)
Suyash Y. Pawar, Soheil Gohari, Mizan Ahmed et al.

Focus Article

Dead or alive: Distinguishing active from passive particles using supervised learning^(a)

GIULIA JANZEN^{1,2(b)}, XANDER L. J. A. SMEETS^{1(b)}, VINCENT E. DEBETS^{1,2}, CHENGJIE LUO^{1,2}, CORNELIS STORM^{1,2}, LIESBETH M. C. JANSSEN^{1,2(c)} and SIMONE CIARELLA^{1,3(d)}

¹ Department of Applied Physics, Eindhoven University of Technology - P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

² Institute for Complex Molecular Systems, Eindhoven University of Technology - P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

³ Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris - F-75005 Paris, France

received 16 February 2023; accepted in final form 16 June 2023
published online 28 June 2023

Abstract – A longstanding open question in the field of dense disordered matter is how precisely structure and dynamics are related to each other. With the advent of machine learning, it has become possible to agnostically predict the dynamic propensity of a particle in a dense liquid based on its local structural environment. Thus far, however, these machine-learning studies have focused almost exclusively on simple liquids composed of passive particles. Here we consider a mixture of both passive and active (*i.e.*, self-propelled) Brownian particles, with the aim to identify the active particles from minimal local structural information. We compare a state-of-the-art machine learning approach for passive systems with a new method we develop based on Voronoi tessellation. Both methods accurately identify the active particles based on their structural properties at high activity and low concentrations of active particles. Our Voronoi method is, however, substantially faster to train and deploy because it requires fewer, and easy to compute, input features. Notably, both become ineffective when the activity is low, suggesting a fundamentally different structural signature for dynamic propensity and non-equilibrium activity. Ultimately, these efforts might also find relevance in the context of biological active glasses such as confluent cell layers, where subtle changes in the microstructure can hint at pathological changes in cell dynamics.

open access

focus article

Copyright © 2023 The author(s)

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

Introduction. – A central notion in the study of active particulate matter —systems of discrete entities which consume energy to perform work and move autonomously— is that the presence of activity can dramatically alter both spatial organization and (collective) dynamics [1]. The fact that active matter is intrinsically out of equilibrium renders the standard tools of statistical physics of limited use and leaves open the question of

which quantifiers most accurately characterize and predict the dynamics of active matter [2]. This is a profound issue, especially in densely disordered phases such as liquids and glasses, where the relation between spatial structure and emergent dynamics is notoriously obscure [1,3]. A better understanding of structure-dynamics relations in active matter would be highly desirable both from a fundamental and more applied perspective. Notably, in the context of biological tissues and confluent cell layers, subtle structural changes can correlate with the motile properties of the individual cells, with relevance in processes such as cancer metastasis and embryonic development [4,5].

Recent efforts have demonstrated that machine learning (ML) approaches are extremely effective in finding simple

^(a)Contribution to the Focus Issue *Statistical Physics of Self-Propelled Colloids* edited by Hartmut Löwen, Sabine Klapp and Holger Stark.

^(b)These authors contributed equally to this work.

^(c)E-mail: l.m.c.janssen@tue.nl (corresponding author)

^(d)E-mail: simoneciarella@gmail.com (corresponding author)

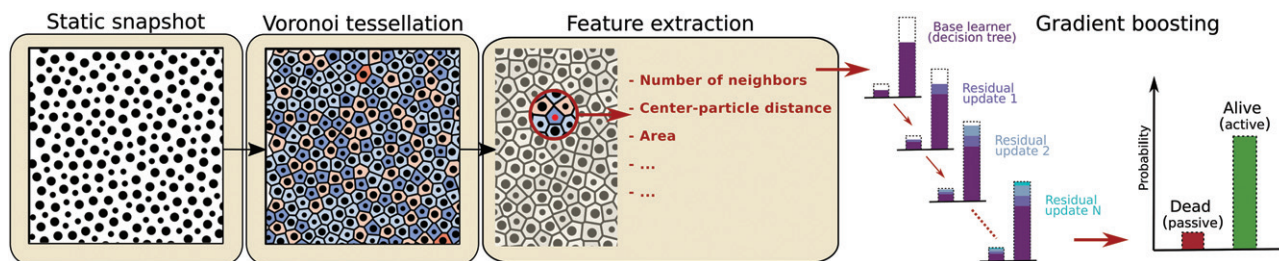


Fig. 1: Sketch of our Voronoi-based machine learning approach for the identification of active particles in an active/passive mixture. The snapshot is processed by constructing its Voronoi tessellation, then extracting particle-specific features. A gradient boosting model is trained to evaluate the probability that each particle is dead (passive) or alive (active), only using the static features extracted from the single snapshot.

structural indicators that predict dynamical properties in densely disordered (near-)equilibrium systems [6–25]. These findings firmly establish a correlation between local structure and the propensity of passive particles to move in a crowded environment. While ML has also been applied with considerable success in purely active systems [25–33], in practice and particularly in biological settings the system of interest will generally contain actors whose activity parameters are different, and distributed. In tumor tissue, for instance, there may be a distribution of epithelial (more stationary, passive) phenotypes and mesenchymal (more motile, active) ones, with the presence of mesenchymal cells being generally associated with greater metastatic potential [5,34]. Thus, developing the ability to reliably distinguish active from passive elements in dense collectives holds diagnostic and prognostic potential.

Inspired by this challenge, our work addresses a seemingly simple issue: Can we identify the active species in a binary model system of active and passive particles? Here we refer to passive entities as “dead” particles that do not possess any self-propulsion, while the active, self-propelled particles are “alive”. Naturally, the distinction would be most conveniently done based on dynamical (*i.e.*, time-resolved) information, but in biological settings such dynamical data may be difficult to obtain. This is why we complicate the challenge considerably, and demand that the identification is performed based solely on static images (“snapshots”) of the system.

Recent studies [35–39] have shown that the presence of a small number of active particles within a dense passive bath can induce specific structural inhomogeneities. Explicitly, bath particles tend to accumulate at the front of an active particle while leaving a depletion zone behind it. This spatial anisotropy is usually irrelevant in the absence of activity, and is therefore also frequently ignored in ML methods designed for purely passive systems [17,20,22,40]. However, it could be essential to distinguish dead particles from those that are alive.

In this work, we quantify the spatial anisotropy of each particle’s local environment through a simple Voronoi tessellation [40,41]; from this instantaneous structural information, we seek to predict the particle identity (dead or alive) using ML. We show that when the active particle

fraction is low and the activity level is high, the shape of the Voronoi polygons around active particles exhibits distinct characteristics compared to those around passive particles, thus providing sufficient static information to reliably distinguish the two species. This observation highlights the effectiveness of Voronoi tessellation as a suitable tool for quantifying unique structural signatures associated with active particles, at least within a suitably chosen parameter range.

We also compare our method with an existing state-of-the-art ML approach based on particle configurations, originally developed for purely passive simulated supercooled liquids [16,17,22]. We demonstrate that both approaches are only successful when the fraction of active particles is low and the activity is high. Furthermore, we show that the Voronoi method is substantially faster and easier to compute because it requires fewer and cheaper input features. This emphasizes the significance of anisotropy in telling apart dead from alive. However, when the active particle concentration is not dilute or the active force is too small, all these static approaches fail to provide accurate predictions, suggesting that the instantaneous local structure no longer fully encodes the particle identity. For this more challenging regime, we develop a pseudo-static approach, detailed in the Supplementary Material [SupplementaryMaterial.pdf](#) (SM), which also incorporates information on the averaged statistical *fluctuations* of the local structure.

Briefly, our Voronoi approach to distinguish between active and passive particles from a single snapshot is outlined in fig. 1. First, the snapshots are processed by performing a Voronoi tessellation [41], from which we compute 11 relevant features that effectively capture local anisotropies. At the core of our Voronoi approach is a gradient boosting model, discussed in the “Methods” section, that takes into account the local environment of each particle to predict if it is dead or alive. Furthermore, the predictions generated by our ML model are explainable, enhancing the interpretability of the results.

Ultimately, our findings lead us to conclude that there are two distinct regions within the parameter space defined by the fraction of active particles and the active force. In the first region, characterized by high activity and a low

fraction of active particles, active and passive particles can be distinguished effectively using a static approach. However, in the second region, active and passive particles become indistinguishable from a static point of view.

Methods. –

Simulation model. Our model system is a two-dimensional (2D) Kob-Andersen mixture [42–44] composed of 650 and 350 particles of type A and B, respectively. We extend the model by turning a fraction ϕ_a into active Brownian particles [45–48] (same proportion for each particle type). We have also conducted similar investigations in a three-dimensional (3D) mixture, comprising 800 particles of type A and 200 particles of type B, and obtained similar results.

The self-propelled motion is characterized by a constant active force of magnitude F_a . To develop a model that can be applied to study highly crowded biological systems, such as dense cellular collectives, we have chosen to focus on steady-state configurations with a density of $\rho = 1.2$ and temperature of $T = 0.5$ (in simulation units) and retrieve 2000 different independent configurations for each studied setting. Moreover, to effectively characterize the large activity regime, we consider activities ranging from 0.5 to 100, as cells primarily move due to active motion. Furthermore, we have verified that similar results are obtained at a temperature of $T = 0.6$. Finally, to confirm that the dynamics at high activities are not influenced by finite-size effects, we performed the same analysis on a larger system with 10000 particles, and we obtained similar results. All the details of the simulation model and the data collection are reported in the SM.

Gradient boosting model. We treat the identification of the active particles as a binary classification problem. For this we use LightGBM, which is a gradient-boosting machine-learning method based on decision trees [49]. This algorithm is widely used due to its efficiency, accuracy, and interpretability, and has been successfully employed, *e.g.*, in the context of low-temperature glassy materials [23]. We show in the SM that it outperforms neural networks for our task of distinguishing active from passive particles.

In general, gradient boosting decision trees [50] comprise an ensemble learning method that combines the power of decision trees [51] and gradient boosting. It involves building a sequence of decision trees, where each subsequent tree corrects the errors made by the previous ones, by performing a residual update step. By iteratively adding decision trees and adjusting their weights based on the residuals, the model gradually improves its ability to make accurate predictions.

LightGBM introduces gradient-based one-side sampling and exclusive feature bundling [49] in order to avoid the need to scan all the data to update the residuals at each step. As a result, gradient boosting can be used for large datasets, such as the collection of simulation snapshots used in this work. In the end, after calculating all the

relevant structural input features (discussed in the next section), the training of our model takes only several minutes on a standard laptop.

To quantify our ML performance we use the accuracy, defined as the number of correct predictions divided by the total number of predictions. When the fraction of active particles $\phi_a \neq 0.5$ we simply discard a subset of the particles from the data such that the number of active and passive species is equal. This allows us to work with a balanced dataset, which enhances the classifier’s generalizability. Since our input features are particle-resolved and additional data collection is quite inexpensive, this presents no practical challenges.

Structural input features. Our main goal is to distinguish the active particles from the passive ones using only instantaneous static information, *i.e.*, a single snapshot. For this, we employ two alternative sets of local, particle-resolved structural properties as input for our machine-learning model. We refer to these as i) the *shell-based approach*, developed earlier for purely passive systems [16], and ii) the *Voronoi approach*, which we introduce here to capitalize on the information contained in local anisotropy.

The features of the *shell-based approach* have been introduced to effectively predict the dynamic propensity of purely passive particles [16,17]. Since the dynamic propensity quantifies the average squared displacement of a particle from its initial condition [52,53], and active particles have an additional self-propelling force, it is natural to expect that active particles have a larger dynamic propensity. We also confirm this observation by measuring the mean-squared displacement [54,55], reported in the SM. Notice that similar indicators also predict localized plastic events [56,57]. This means that if the dynamical information remains encoded in the structure even for active systems, then this approach should be able to pick it up and identify the active particles from their enhanced propensity.

To implement the shell-based approach we follow the definitions of ref. [16]. Briefly, these features consist of n -th-order radial and angular descriptors, denoted as $G_i^{(n)}$ and $q_i^{(n)}$, respectively. The former measures for each particle i the local particle density within a radial shell, while the latter expands the local density within a shell in terms of spherical harmonics; the order n indicates the degree of averaging over the structural features of nearby particles. Here we consider $n = 0, 1, 2$. For $n = 0$, we compute 50 radial descriptors per species with radial distance $r \in [0, 5]$ and shell width $\delta = 0.1$, and 192 angular descriptors with $r \in [1, 2.5]$, $\delta = 0.1$, and spherical harmonics of order $l \in [1, 12]$. In sum, we use a total of 876 static quantities as input in the shell-based approach.

As an alternative to the shell-based features of ref. [16], we introduce the *Voronoi approach*, which naturally quantifies the local anisotropy created by the active particles. Briefly, we perform a Voronoi tessellation of all particle positions and compute the features based on the shape of

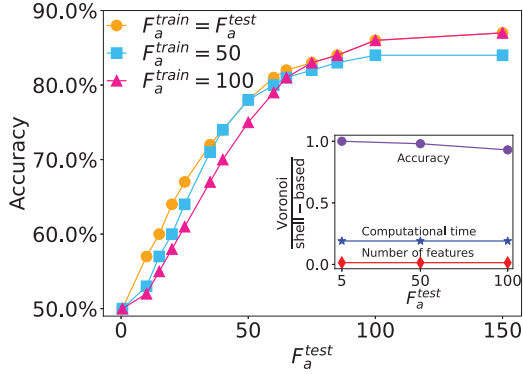


Fig. 2: Accuracy of the Voronoi ML approach as a function of the active force $F_a = F_a^{\text{test}}$, with $\phi_a = 0.008$. The orange circles represent scores obtained from separate models, where each one was trained using $F_a^{\text{train}} = F_a^{\text{test}}$. The blue squares and pink triangles represent scores obtained from a single global model trained with data only for $F_a = 50$ or $F_a = 100$, respectively. The inset shows the comparison between the Voronoi and the shell-based approaches for three different active forces $F_a^{\text{test}} = F_a^{\text{train}} = 5, 50, 100$. The purple circles, blue stars, and red diamonds represent the ratio between accuracy, computational times, and number of features, respectively.

the Voronoi polygon around each particle. Explicitly, we consider the number of neighbors, the polygon area and perimeter, the distance between the center of the polygon and the actual particle position, the maximum and minimum distances between nearest neighbors, the maximum and minimum distances between the vertices of a polygon, the maximum and minimum distances between the particle position and its polygon vertices, and finally we add the particle type (A or B). Overall this amounts to only 11 relatively simple input features. Compared to the shell-based approach, the Voronoi approach is thus significantly cheaper.

Results and discussion. –

Distinguishing dead from alive. We report in fig. 2 the accuracy achieved by our Voronoi ML model in the classification of an active/passive mixture with $\phi_a = 0.008$, for different values of the active force F_a . We compare the accuracy that we get when the model is trained and tested at a single specific value of F_a (orange circles), with the accuracy of a model trained when fixing $F_a = 50$ (blue squares) or $F_a = 100$ (pink triangles). All three curves produce rather similar accuracies, implying that a single ML model trained at a fixed F_a can also produce good predictions for unseen parameter regimes. This shows reasonable generalizability of the model. However, the accuracy in all three cases drops significantly for lower activity. We attribute this to the fact that small values of F_a render it extremely challenging to differentiate between structural signatures induced by either active forces or passive Brownian motion. Hence, any static approach based purely on instantaneous structural properties fails.

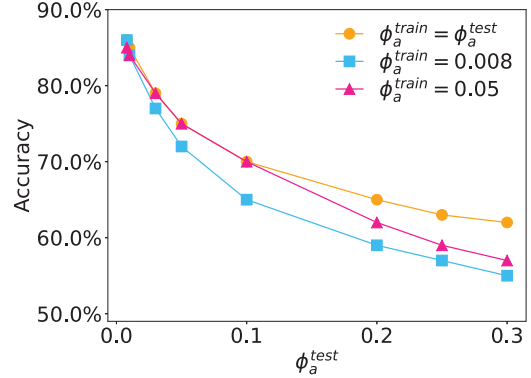


Fig. 3: Accuracy of the Voronoi ML approach as a function of the fraction of active particles $\phi_a = \phi_a^{\text{test}}$, with $F_a = 100$. The orange circles represent scores obtained from separate models, where each one was trained using $\phi_a^{\text{train}} = \phi_a^{\text{test}}$. The blue squares and pink triangles represent scores obtained from a single global model trained with data only for $\phi_a = 0.008$ or $\phi_a = 0.05$, respectively.

The inset of fig. 2 demonstrates that very similar accuracies can be achieved when training the ML algorithm with the more intricate structural properties $G_i^{(n)}$ and $q_i^{(n)}$ as input (shell-based approach). However, the shell-based approach is considerably more computationally expensive compared to the Voronoi method. The reason is that the shell-based approach employs almost two orders of magnitude more features. Consequently, we can conclude that, for this specific classification task, the Voronoi approach is more efficient. Moreover, we find similar results for both the Voronoi and the shell-based approach in 3D (see the SM).

Figure 3 reports the performance of the Voronoi active/passive classifiers as a function of the percentage of active particles ϕ_a , while keeping F_a fixed at 100. For $\phi_a \leq 0.1$, our approach demonstrates excellent predictive capability when the ML model is trained and tested at a specific ϕ_a value (represented by orange circles). Additionally, we compare the accuracy obtained when training the model with ϕ_a fixed at 0.008 (blue squares) or 0.05 (pink triangles). The latter case yields better transferability than when training only on $\phi_a = 0.008$, especially in the region $\phi_a \leq 0.1$, where it rivals the performance of the model trained at each ϕ_a value separately. Noticeably, however, when the fraction of active particles ϕ_a becomes large, the accuracy drops to below 70%, even for our best ML model. This indicates that the instantaneous structural input becomes less informative to differentiate between active and passive particles, which we attribute to the fact that the presence of more active particles increasingly affects the entire structure of the material and diminishes the front-wake asymmetry. Finally, we have also repeated the Voronoi tessellation analysis at a lower density of $\rho = 1.0$ with a fraction of active particles $\phi_a = 0.008$. Interestingly, the results are qualitatively similar to those obtained at $\rho = 1.2$, but with a

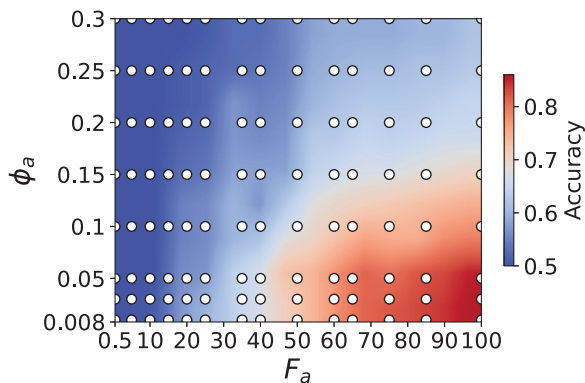


Fig. 4: Accuracy map of the Voronoi ML approach in the (F_a, ϕ_a) -plane for a 2D system at temperature $T = 0.5$ and density $\rho = 1.2$.

more pronounced anisotropy observed at a lower activity ($F_a = 50$) compared to the higher-density system (see the SM for more details).

A full overview of the Voronoi ML model accuracy in the (F_a, ϕ_a) -plane is reported in fig. 4. This figure reveals two distinct regions: a red region representing a “distinguishable” area and a blue region representing an “indistinguishable” area. In the red region, characterized by low ϕ_a and high F_a values, the algorithm achieves an accuracy greater than 0.75, enabling it to differentiate adequately between active and passive particles based on a single snapshot. In this region, the structure surrounding the active particle significantly differs from that of the passive particle. As a result, even a shell-based approach, which does not consider the directional dependence of passive particle density around the active particle, still yields accurate results (inset of fig. 2).

In the blue region, characterized by either high ϕ_a or low F_a , any static approach (either shell-based or Voronoi) is insufficient to effectively differentiate between active and passive particles. When $F_a < 40$, the activity is not high enough to induce noticeable spatial inhomogeneities. On the other hand, when ϕ_a is too high, the interaction among active particles complicates the classification task. Particularly, in cases where both F_a and ϕ_a are high, the system can undergo motility-induced phase separation (MIPS) [37]. In view of the poor overall performance of our ML method for both high ϕ_a and low F_a , we conclude that if a local structure-dynamics relationship exists in active systems, it must be significantly different from that observed in passive systems, as indicated by the failure of the shell-based approach. Additionally, this relationship is not hidden in the anisotropy, as indicated by the failure of the Voronoi approach.

In order to still achieve good predictability in the regime where a purely static ML approach fails (blue region in fig. 4), we have developed a so-called pseudo-static method. This approach, detailed in the SM, is able to identify active particles at the cost of more input information. Briefly, while not requiring a fully time-resolved

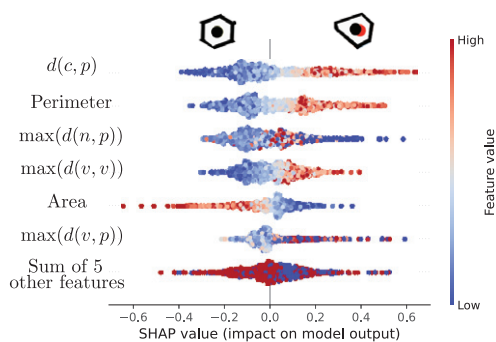


Fig. 5: The six most important features for the Voronoi model, represented as a SHAP beeswarm plot at $F_a = 100$ and $\phi_a = 0.008$. The position of the dots is determined by the SHAP values of the features, and the color is used to display the value of the features. The top two polygons in the figure are a detail of a snapshot that represents the typical shapes observed in the passive (left) and active (right) cases. In the passive scenario, the center of the polygon aligns with the particle position, whereas in the active case they do not coincide.

dynamical trajectory, it requires some knowledge of the averaged statistical *fluctuations* of the local structure. It is possible to obtain these statistics from a collection of snapshots that do not have to be time-ordered, but in which each particle has to be tracked. This tracking information makes the method naturally more computationally expensive. Overall, in the pseudo-static approach, the information about the local structure is complemented by statistical measurements of the same features at different times. We have verified that using this information on the structural fluctuations, an accuracy of 95%–100% can be achieved (see the SM).

Model explanation. Let us return to the parameter regime where our Voronoi ML model is most successful in distinguishing active from passive particles, *i.e.*, small ϕ_a and large F_a . In this regime we can interpret the decisions made by the ML model. To do so, we calculate the model explanations using SHapley Additive exPlanation (SHAP) analysis [58]. Briefly, this method, based on cooperative game theory, calculates Shapley values that distribute the model’s prediction among the features while considering their individual contributions to the output. We apply this analysis to the Voronoi model trained on $F_a = 100$ and $\phi_a = 0.008$.

Figure 5 shows the SHAP beeswarm plot, which indicates the six most important features and how the values of these features influence the model’s predictions. The first six features are, in order of importance: the distance between the center of the polygon and the particle ($d(c,p)$), the perimeter of the polygon, the maximum distance between the particle and its neighbors ($\max(d(n,p))$), the maximum distance between the vertices of the polygon ($\max(d(v,v))$), the area of the polygon, and the maximum distance between the particle and the vertices of the polygon ($\max(d(v,p))$). The colors show that the model interprets low values of $d(c,p)$ as

a passive particle and high values of $d(c, p)$ as an active particle. Considering that passive particles tend to accumulate in front of active particles while creating a void behind them, we anticipate the polygons around active particles to exhibit elongation. Consequently, due to this anisotropy, we expect the active particle to be situated at a different position compared to the center of the polygon, as we confirm in representative polygons shown at the top of fig. 5. The SHAP analysis clearly validates this expectation. Hence, rather than using the ML model solely as a black box, we infer that the model captures the correct physical picture and confirms the importance of spatial anisotropy induced by activity. Lastly, the SHAP analysis has also been performed in 3D, and the corresponding results can be found in the SM.

Conclusions. – The focus of this work is to identify active (alive) entities in a densely disordered mixture of active and passive (dead) particles, using only a single picture of the system as input. We compare two static ML approaches: one based on the Voronoi tessellation and another employing a shell-based approach that utilizes descriptors previously employed in passive systems to predict dynamic particle propensities [9–11,13,14,16].

We find that a static approach based on a single snapshot is able to predict whether a particle is dead or alive when the activity is high and the fraction of active particles is low, in both 2D and 3D. However, this approach fails when the activity is low or the fraction of active particles is high, implying a different structural signature for activity and dynamic propensity in passive systems. Moreover, we observe that while the Voronoi and shell-based approaches yield similar accuracies, the shell-based approach is computationally more expensive.

Within the region of the (F_a, ϕ_a) -plane where the static approach is effective for particle classification, our model exhibits reasonably good performance when extrapolating to active forces and compositions outside its training range, demonstrating its robustness. This transferability can be particularly useful in experiments, where the active force or the precise number of active species may be difficult to quantify.

Additionally, we explain the decisions of our ML model in order to understand how it is able to identify the active entities from the Voronoi input. We find that the distance between a particle and the center of its corresponding polygon is the most significant feature for this classification task, serving as a simple and reliable measure of the distinct spatial anisotropy surrounding an active particle.

We hypothesize that the region of the (F_a, ϕ_a) -plane where a purely static approach fails depends on system parameters such as density and temperature. Indeed, when considering a lower density, we have found that our classifier can perform better for certain levels of activity (see the SM). However, while the boundary of the distinguishable region may thus be varied with external control

parameters, we still expect a region where active and passive particles remain indistinguishable from a single snapshot. In this regime more information is inevitably needed, as confirmed by our pseudo-static approach that incorporates additional knowledge of structural *fluctuations*.

In summary, this work demonstrates that the shape of the polygon is sufficient to differentiate between active and passive particles in a specific region of the (F_a, ϕ_a) -plane. This work serves as a step to better understand the elusive structure-dynamics relation in densely disordered non-equilibrium systems. We believe that our approach can also be a valuable tool for investigating experimental systems such as biological cells, where identifying the most active entities visually can be challenging. It is thus our hope that the here presented approach can help to process large experimental datasets and contribute to the discovery of new connections between structural and dynamical properties in complex non-equilibrium materials.

This work has been financially supported by the Dutch Research Council (NWO) through a START-UP grant (VED, CL, and LMCJ), Physics Projectruimte grant (GJ and LMCJ), ENW-XL grant (CS and LMCJ), and Vidi grant (LMCJ).

Data availability statement: The data that support the findings of this study are available upon reasonable request from the authors.

REFERENCES

- [1] BECHINGER C., LEONARDO R. D., LÖWEN H., REICHHARDT C., VOLPE G. and VOLPE G., *Rev. Mod. Phys.*, **88** (2016) 45006.
- [2] MARCHETTI M. C., JOANNY J. F., RAMASWAMY S., LIVERPOOL T. B., PROST J., RAO M. and SIMHA R. A., *Rev. Mod. Phys.*, **85** (2013) 1143.
- [3] JANSSEN L. M. C., *J. Phys.: Condens. Matter*, **31** (2019) 503002.
- [4] BLAUTH E., KUBITSCHKE H., GOTTHEIL P., GROSSER S. and KÄS J. A., *Front. Phys.*, **9** (2021) 666709.
- [5] GROSSER S., LIPPOLDT J., OSWALD L., MERKEL M., SUSSMAN D. M., RENNER F., GOTTHEIL P., MORAWETZ E. W., FUHS T., XIE X., PAWLIZAK S., FRITSCH A. W., WOLF B., HORN L.-C., BRIEST S., AKTAS B., MANNING M. L. and KÄS J. A., *Phys. Rev. X*, **11** (2021) 11033.
- [6] CUBUK E. D., SCHOENHOLZ S. S., RIESER J. M., MALONE B. D., RÖTTLER J., DURIAN D. J., KAXIRAS E. and LIU A. J., *Phys. Rev. Lett.*, **114** (2015) 108001.
- [7] CUBUK E. D., IVANCIC R. J. S., SCHOENHOLZ S. S., STRICKLAND D. J., BASU A., DAVIDSON Z. S., FONTAINE J., HOR J. L., HUANG Y.-R., JIANG Y., KEIM N. C., KOSHIGAN K. D., LEFEVER J. A., LIU T., MA X.-G., MAGAGNOSC D. J., MORROW E., ORTIZ C. P., RIESER J. M., SHAVIT A., STILL T., XU Y., ZHANG Y., NORDSTROM K. N., ARRATIA P. E., CARPICK R. W., DURIAN D. J., FAKHRAAI Z., JEROLMACK D. J., LEE D., LI J., RIGGLEMAN R., TURNER K. T., YODH A. G., GIANOLA D. S. and LIU A. J., *Science*, **358** (2017) 1033.

- [8] CUBUK E. D., SCHOENHOLZ S. S., KAXIRAS E. and LIU A. J., *J. Phys. Chem. B*, **120** (2016) 6139.
- [9] SUSSMAN D. M., PAOLUZZI M., MARCHETTI M. C. and MANNING M. L., *EPL*, **121** (2018) 36001.
- [10] BOATTINI E., RAM M., SMALLENBURG F. and FILION L., *Mol. Phys.*, **116** (2018) 3066.
- [11] BOATTINI E., DIJKSTRA M. and FILION L., *J. Chem. Phys.*, **151** (2019) 154901.
- [12] SCHOENHOLZ S. S., CUBUK E. D., SUSSMAN D. M., KAXIRAS E. and LIU A. J., *Nat. Phys.*, **12** (2016) 469.
- [13] BAPST V., KECK T., GRABSKA-BARWIŃSKA A., DONNER C., CUBUK E. D., SCHOENHOLZ S. S., OBIKA A., NELSON A. W. R., BACK T., HASSABIS D. and KOHLI P., *Nat. Phys.*, **16** (2020) 448.
- [14] PARET J., JACK R. L. and COSLOVICH D., *J. Chem. Phys.*, **152** (2020) 144502.
- [15] LANDES F. M. C. P., BIROLI G., DAUCHOT O., LIU A. J. and REICHMAN D. R., *Phys. Rev. E*, **101** (2020) 010602.
- [16] BOATTINI E., SMALLENBURG F. and FILION L., *Phys. Rev. Lett.*, **127** (2021) 88007.
- [17] ALKEMADE R. M., BOATTINI E., FILION L. and SMALLENBURG F., *J. Chem. Phys.*, **156** (2022) 204503.
- [18] OYAMA N., KOYAMA S. and KAWASAKI T., *Front. Phys.*, **10** (2023) 1007861.
- [19] TAH I., RIDOUT S. A. and LIU A. J., *J. Chem. Phys.*, **157** (2022) 124501.
- [20] JUNG G., BIROLI G. and BERTHIER L., *Phys. Rev. Lett.*, **130** (2023) 238202.
- [21] CIARELLA S., CHIAPPINI M., BOATTINI E., DIJKSTRA M. and JANSSEN L. M. C., *Mach. Learn.: Sci. Technol.*, **4** (2023) 025010.
- [22] COSLOVICH D., JACK R. L. and PARET J., *J. Chem. Phys.*, **157** (2022) 204503.
- [23] CIARELLA S., KHOMENKO D., BERTHIER L., MOCANU F. C., REICHMAN D. R., SCALLIET C. and ZAMPONI F., arXiv preprint, arXiv:2212.05582 (2022).
- [24] ALKEMADE R. M., SMALLENBURG F. and FILION L., *J. Chem. Phys.*, **158** (2023) 134512.
- [25] JANZEN G., SMIT C., VISBEEK S., DEBETS V. E., LUO C., STORM C., CIARELLA S. and JANSSEN L. M. C., arXiv preprint, arXiv:2303.00636 (2023).
- [26] CICHOS F., GUSTAVSSON K., MEHLIG B. and VOLPE G., *Nat. Mach. Intell.*, **2** (2020) 94.
- [27] NEWBY J. M., SCHAEFER A. M., LEE P. T., FOREST M. G. and LAI S. K., *Proc. Natl. Acad. Sci. U.S.A.*, **115** (2018) 9026.
- [28] JECKEL H., JELLI E., HARTMANN R., SINGH P. K., MOK R., TOTZ J. F., VIDAKOVIC L., ECKHARDT B., DUNKEL J. and DRESCHER K., *Proc. Natl. Acad. Sci. U.S.A.*, **116** (2019) 1489.
- [29] BO S., SCHMIDT F., EICHHORN R. and VOLPE G., *Phys. Rev. E*, **100** (2019) 10102.
- [30] NOZ GIL G. M., GARCIA-MARCH M. A., MANZO C., MARTÍN-GUERRERO J. D. and LEWENSTEIN M., *New J. Phys.*, **22** (2020) 013010.
- [31] TAH I., SHARP T. A., LIU A. J. and SUSSMAN D. M., *Soft Matter*, **17** (2021) 10242.
- [32] BAG S. and MANDAL R., *Soft Matter*, **17** (2021) 8322.
- [33] RUIZ-GARCIA M., GUTIERREZ C. M. B., ALEXANDER L. C., AARTS D. G. A. L., GHIRINGHELLI L. and VALERIANI C., arXiv preprint, arXiv:2203.14846 (2022).
- [34] COBAN B., BERGONZINI C., ZWEEMER A. J. M. and DANNEN E. H. J., *Br. J. Cancer*, **124** (2021) 49.
- [35] MEJÍA-MONASTERIO C. and OSHANIN G., *Soft Matter*, **7** (2011) 993.
- [36] BÉNICHOU O., ILLIEN P., MEJIA-MONASTERIO C. and OSHANIN G., *J. Stat. Mech.: Theory Exp.*, **2013** (2013) P05008.
- [37] WITTKOWSKI R., STENHAMMAR J. and CATES M. E., *New J. Phys.*, **19** (2017) 105003.
- [38] VASILYEV O. A., BÉNICHOU O., MEJÍA-MONASTERIO C., WEEKS E. R. and OSHANIN G., *Soft Matter*, **13** (2017) 7617.
- [39] BANERJEE J. P., MANDAL R., BANERJEE D. S., THUTUPALLI S. and RAO M., *Nat. Commun.*, **13** (2022) 4533.
- [40] ZHANG H., LIU F. and HAN Y., arXiv preprint, arXiv:2305.04179 (2023).
- [41] AURENHAMMER F. and KLEIN R., *Handbook of Computational Geometry* (North-Holland, Amsterdam) 2000, pp. 201–290.
- [42] KOB W. and ANDERSEN H. C., *Phys. Rev. Lett.*, **73** (1994) 1376.
- [43] MICHELE C. D., SCIORTINO F. and CONIGLIO A., *J. Phys.: Condens. Matter*, **16** (2004) L489.
- [44] FLENNER E. and SZAMEL G., *Phys. Rev. E*, **72** (2005) 1.
- [45] ROMANCZUK P., BÄR M., EBELING W., LINDNER B. and SCHIMANSKY-GEIER L., *Eur. Phys. J. ST*, **202** (2012) 1.
- [46] RAMASWAMY S., *J. Stat. Mech.: Theory Exp.*, **2017** (2017) 054002.
- [47] LÖWEN H., *J. Chem. Phys.*, **152** (2020) 040901.
- [48] TEN HAGEN B., VAN TEEFFELEN S. and LÖWEN H., *J. Phys.: Condens. Matter*, **23** (2011) 194119.
- [49] KE G., MENG Q., FINLEY T., WANG T., CHEN W., MA W., YE Q. and LIU T.-Y., *LightGBM: A highly efficient gradient boosting decision tree*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems NIPS'17* (Curran Associates Inc., Red Hook, NY, USA) 2017, pp. 3149–3157.
- [50] FRIEDMAN J. H., *Ann. Stat.*, **29** (2001) 1189.
- [51] KOTSIAANTIS S. B., *Artif. Intell. Rev.*, **39** (2013) 261.
- [52] BERTHIER L. and JACK R. L., *Phys. Rev. E*, **76** (2007) 41509.
- [53] WIDMER-COOPER A. and HARROWELL P., *J. Chem. Phys.*, **126** (2007) 154503.
- [54] LILUASHVILI A., ÓNODY J. and VOIGTMANN T., *Phys. Rev. E*, **96** (2017) 062608.
- [55] BECHINGER C., DI LEONARDO R., LÖWEN H., REICHARDT C., VOLPE G. and VOLPE G., *Rev. Mod. Phys.*, **88** (2016) 045006.
- [56] RICHARD D., OZAWA M., PATINET S., STANIFER E., SHANG B., RIDOUT S. A., XU B., ZHANG G., MORSE P. K., BARRAT J.-L., BERTHIER L., FALK M. L., GUAN P., LIU A. J., MARTENS K., SASTRY S., VANDEMBROUCQ D., LERNER E. and MANNING M. L., *Phys. Rev. Mater.*, **4** (2020) 113609.
- [57] RIDOUT S. A., ROCKS J. W. and LIU A. J., *Proc. Natl. Acad. Sci. U.S.A.*, **119** (2022) e2119006119.
- [58] LUNDBERG S. M. and LEE S.-I., *A unified approach to interpreting model predictions*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems NIPS'17* (Curran Associates Inc., Red Hook, NY, USA) 2017, pp. 4768–4777.