Classifying the age of a glass based on structural properties: A machine learning approach

Giulia Janzen,[1,2] Casper Smit *,[1,3] Samantha Visbeek,[1,3,*] Vincent E. Debets,[1,2]
Chengjie Luo,[1,2] Cornelis Storm,[1,2] Simone Ciarella,[1,2,4,5,†] and Liesbeth M.C. Janssen[1,2,‡]

[1] *Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
[2] *Institute for Complex Molecular Systems, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
[3] *Institute of Physics, University of Amsterdam, Science Park 904, Amsterdam, 1098 XH, The Netherlands*
[4] *Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France*
[5] *Netherlands eScience Center, Amsterdam 1098 XG, The Netherlands*

(Dated: February 26, 2024)

It is well established that physical aging of amorphous solids is governed by a marked change in dynamical properties as the material becomes older. Conversely, structural properties such as the radial distribution function exhibit only a very weak age dependence, usually deemed negligible with respect to the numerical noise. Here we demonstrate that the extremely weak age-dependent changes in structure are in fact sufficient to reliably assess the age of a glass with the support of machine learning. We employ a supervised learning method to predict the age of a glass based on the system's instantaneous radial distribution function. Specifically, we train a multilayer perceptron for a model glassformer quenched to different temperatures, and find that this neural network can accurately classify the age of our system across at least four orders of magnitude in time. Our analysis also reveals which structural features encode the most useful information. Overall, this work shows that through the aid of machine learning, a simple structure-dynamics link can indeed be established for physically aged glasses.

arXiv:2303.00636v2 [cond-mat.soft] 23 Feb 2024

I. INTRODUCTION

The structural, dynamical and mechanical properties of a material change as it gets older, i.e. it ages [1–9]. Physical aging is particularly well studied for glasses due to their slow relaxation dynamics [10–14]. One of the most common methods to study the aging dynamics of a glass consists of a temperature quench toward a lower temperature [15–18]. After the quench, as the material seeks to recover equilibrium at the new temperature, the relaxation time of the system will increase with its age [10, 19–21]. The physical aging in glassy systems can thus be understood as a gradual approach towards increasingly lower-energy equilibrium states [13]. It is also well known that, besides a rapid short-time change, the structural properties change only extremely weakly with time [22–27]. In contrast, the dynamical properties exhibit significant changes over multiple orders of magnitude in time as shown in Fig. 1 and the Supplementary Material [28]. It is therefore customary to characterize the aging behavior of a system by means of its dynamical properties. At the same time, it remains unclear how these strong dynamical changes of an aging glass are connected to its almost constant structure [24].

To bridge this gap, Cubuck *et al.* [29] have recently developed a pioneering approach which demonstrates that machine learning techniques can in fact successfully correlate structure and dynamics in glassy systems. Cubuck *et al.* have introduced a machine learning microscopic structural quantity, so-called softness, which characterizes the local structure around each particle. Based on this approach, several recent works [30–41] have extended our conceptual understanding of glassy liquids by convincingly demonstrating that machine learning is able to accurately connect structural properties with the corresponding dynamics. In particular, standard machine learning tools like support vector machines have been able to compute the relaxation time through softness [42] and collective effects like fragility [36] and low-temperature defects [43]. More sophisticated models like graph neural networks [44] give accurate predictions of dynamic propensity, but similar results can be achieved by simpler models with accurate structural indicators [45]. It is thus evident that machine learning is a powerful tool to study glassy systems and, as suggested by Schoenholz *et al.* [42], it is plausible that it could also be used to shed new light on aging behavior.

Still, it is a priori not clear whether this level of complexity, both in the machine-learning model and in the input set, is strictly necessary to predict the age of a system from structural properties. Indeed, an analysis of the softness suggests that the radial distribution function's first peak contributes the most to predicting rearrangements [30]. Since the radial distribution function does change weakly with age, one could argue that a traditional approach, which could consist of selecting the radial distribution function's values that change the most

* These authors contributed equally to this work.
† simoneciarella@gmail.com
‡ l.m.c.janssen@tue.nl

with age and applying linear regression [46–48], might already be sufficient to extract the age of a system. However, such a traditional approach would only be expected to work if the uncertainty in the data is sufficiently small, e.g. in the thermodynamic limit, while in reality a system is typically finite-sized and thus susceptible to noise.

Here our goal is to classify the age of a glassy system based solely on a snapshot, i.e. an instantaneous particle configuration of a finite-sized system. In particular, we compute the radial distribution function at every age and we use this simple feature as input for a neural network. We find that a neural network that is trained and tested at a fixed quenching temperature can distinguish between a young and an old glass with $94\% - 97\%$ accuracy. We compare our machine learning method with a traditional approach, confirming the superiority of the first. The traditional approach involves manual selection of features that–on average–exhibit the most significant changes with age. However, due to noise, these features yields significantly less reliable age predictions compared to those obtained using features selected through a machine learning approach.

In order to analyze our machine learning results in more detail, we perform both a Principal Component Analysis (PCA) and a Shapley Additive Explanation (SHAP). These methods reveal the principal components or the structural features that most strongly encode the age, also allowing us to infer the age of the system from only a subset of the structural data. Finally, we explore the role of the quenching temperature, also proving that a neural network trained with a set of multiple quenching temperatures generalizes well when tested at a new temperature. Though we primarily focus on passive systems, we also verify our model for active systems. Ultimately, we conclude that a machine learning approach purely based on simple structural properties can reliably infer the age of a glassy system.

II. METHODS

A. Simulation model

We study a two-dimensional (2D) binary mixture of Brownian particles. The overdamped equations of motion for each particle i are given by

$$\gamma\, \dot{\boldsymbol{r}}_i = \sum_{j \neq i} \boldsymbol{f}_{ij} + \sqrt{2D_T}\, \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{r}_i = (x_i, y_i)$ represents the particle's spatial coordinates and the dot denotes the time derivative. The translational diffusion constant is denoted as $D_T = k_B T/\gamma$ and the thermal noise is represented by independent Gaussian stochastic processes $\boldsymbol{\eta} = (\eta_x, \eta_y)$ with zero mean and variance $\delta(t - t')$, where k_B is the Boltzmann constant, T the temperature, and γ the friction coefficient. Lastly, $\boldsymbol{f}_{ij} = -\nabla_i V(r_{ij})$ is the interaction force

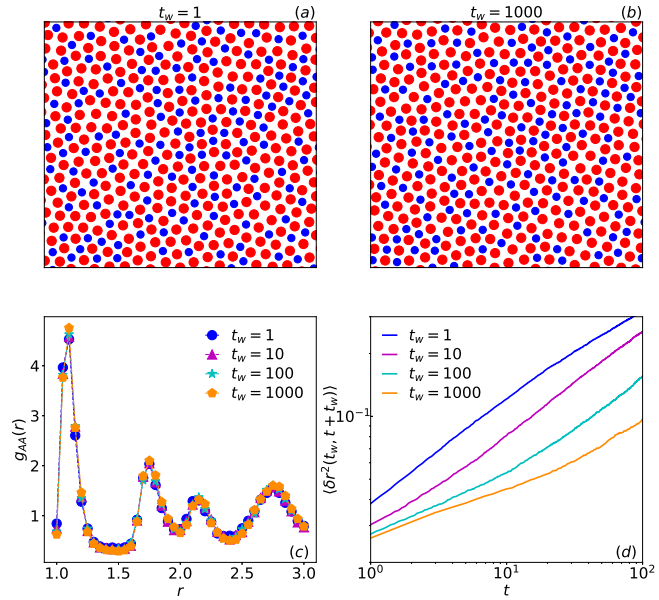

FIG. 1. Age dependence in structural and dynamical properties. Snapshots after a temperature quench for waiting time (a) $t_w = 1$ and (b) $t_w = 1000$. (c) Radial distribution function for A particles $g_{AA}(r)$ and (d) mean-square displacement $\langle \delta r^2(t_w, t + t_w) \rangle$ at fixed quenching temperature $T_q = 0.375$ for waiting times $t_w = 1, 10, 100, 1000$.

between particles i and j, where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and V is a Lennard-Jones potential [49] with a cutoff distance $r_{ij} = 2.5\sigma_{ij}$. In order to prevent crystallization we use the 2D binary Kob-Andersen mixture [50]: $A = 65\%$, $B = 35\%$, $\epsilon_{AA} = 1$, $\epsilon_{BB} = 0.5\epsilon_{AA}$, $\epsilon_{AB} = 1.5\epsilon_{AA}$, $\sigma_{AA} = 1$, $\sigma_{BB} = 0.88\sigma_{AA}$ and $\sigma_{AB} = 0.8\sigma_{AA}$. We set the number density to $\rho = 1.2$, the number of particles to $N = 10\,000$ and $D_T = 1$. Results are in reduced units, where σ_{AA}, ϵ_{AA}, $\frac{\sigma_{AA}^2 \gamma}{\epsilon_{AA}}$, and $\frac{\epsilon_{AA}}{k_B}$ are the units of length, energy, time, and temperature, respectively. Simulations have been performed using LAMMPS [51] by solving Eq. 1 via the Euler-Maruyama method [52] with a step size $\delta t = 10^{-4}$.

As additional verification of our method we also study the aging behavior of an active glass. For this we use the active Brownian particle (ABP) model, which combines thermal motion with a constant self-propulsion speed [53–58]. To obtain the equation of motion for ABPs, in Eq. 1 we need to add the self-propulsion term. This term is defined as $f\, \boldsymbol{n}_i$, where f/γ is the constant self-propulsion speed along a direction $\boldsymbol{n}_i = (\cos\theta_i, \sin\theta_i)$, θ_i is the rotational coordinate, and f is the magnitude of the active force. The rotational coordinate obeys $\dot{\theta}_i = \sqrt{2D_r}\, \eta_\theta$, where D_r is the rotational diffusion coefficient and η_θ is a Gaussian stochastic process. The persistence time, τ_r, is defined as the inverse of the rotational diffusion coefficient and determines the decay time of a particle's orientation [59]. Finally, we choose to focus on a relatively large system with $N = 10\,000$ particles, but we verified that our machine learning approach also

performs well for a smaller system with $N = 1\,000$ (Supplementary material [28]).

B. Aging

For our data set, we prepare 20 independent configurations and let each of them equilibrate at the initial temperature T_i. In this work we consider $T_i = 1$, which corresponds to the liquid phase, but similar results can in principle be obtained for other initial temperatures. Moreover, the dataset consists of 20 independent configurations since these are sufficient to obtain good performance. After the equilibration process we apply a quench to the final temperature T_q that is lower than the glass transition temperature T_g (for this system $T_g \approx 0.4$ [60–62]). We use quenching temperatures between $T_q = 0.1$ and $T_q = 0.375$ and collect data for waiting times between 0 and 10^4. It is well known that the relaxation time as a function of the waiting time follows a power law [15]. We therefore split the data into 5 different classes following a logarithmic scale, as shown in Tab. I. Each class consists of 900 different waiting times t_w that we also refer to as ages, except for class 0 which consists of only the single age $t_w = 0$. In order to have the same amount of data in each class, we save the particle's configurations every t_s time units, with t_s specified in Table I.

For each age, we compute the radial distribution function $g_i(r)$ averaged over the number of particles, where $i \in \{AA, BB, AB\}$ indicates the interaction pairs. It has been shown that the radial distribution function's first peak is one of the most important features to predict rearrangements [30]. To verify whether this also applies to the age classification, we compare the results when the radial distribution function includes or excludes the first peak, corresponding to $g_i(r)$ with $\sigma_i \leq r < 3\sigma_i$ and $g_i(r')$ with $\sigma_i + 0.15 \leq r' < 3\sigma_i + 0.15$, respectively. In this paper, we will refer to the radial distribution function without the first peak as $\hat{g}_i(r)$. To compute $\hat{g}_i(r)$ or $g_i(r)$ we use a bin width of $\delta r = 0.05$, resulting in 40 data points for each of the three partial radial distribution functions $g_i(r)$. These 120 structural properties will be used as an input for our machine learning model. The dataset is randomly divided into a training and a test set that includes 70% and 30% of the data, respectively. To verify that this model also works for an active particle system, we study ABPs with an active force $f = 0.5$, a persistence time $\tau_r = 1$, and a quenching temperature $T_q = 0.25$. We chose these parameters such that the relaxation times of the active and passive systems are of the same order of magnitude [62].

C. Classification model

To carry out the age classification task we use a multilayer perceptron [63, 64] as implemented in Scikit-learn [65]. This neural network (NN) is composed of multiple layers of interconnected neurons. In the first layer, i.e. the input layer, the neurons receive the input vector, while the output layer yields the output signals or classifications with an assigned weight. The hidden layers optimize the weights until the neural network's margin of error is minimal [66].

class	age	t_s
0	$t_w = 0$	10^{-4}
1	$10^0 \leq t_w < 10^1$	10^{-2}
2	$10^1 \leq t_w < 10^2$	10^{-1}
3	$10^2 \leq t_w < 10^3$	10^0
4	$10^3 \leq t_w < 10^4$	10^1

TABLE I. Description of the dataset. Data are collected from age 0 to age 10^4 and then divided in 5 different classes. The left column shows the label given to each class, the central column the ages that belong to each class, and the right column the time interval t_s used to collect data.

In this work we will use two different NN architectures consisting of either four or twelve hidden layers. In both cases, all hidden layers have 100 nodes except for the last two which have 50 and 30 nodes, respectively. The ADAM algorithm has been used to update the weights [67].

To evaluate the model we compute the f1-score

$$\text{f1-score} = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

where the precision is the sum of true positives across all classes divided by the sum of both true and false positives over all classes, and the recall is the sum of true positives across all classes divided by the sum of true positives and false negatives across all classes. The f1-score reaches its largest value of 1 when the model has perfect precision and recall and its lowest value of 0 if either the precision or the recall is equal to zero. The list of hyperparameters used for the multilayer perceptron is reported in the supplementary material [28].

D. Feature selection: traditional approach, SHAP analysis, and PCA

A key aspect of our work is to establish whether machine learning is truly of added value when inferring the age of a finite system, as opposed to a more traditional approach. Some signatures of aging in the $g(r)$ have already been observed [22–25], and a traditional approach would focus on the features that change the most with age, that usually include the first peak. If the age dependence of these features is linear or polynomial, we could apply a simple algorithm, e.g., linear regression, to make predictions. To verify if this approach is efficient,

we compute

$$\langle \delta g_i(t_w, r) \rangle = \left\langle \frac{g_i(t_w, r) - g_i(t_w = 1, r)}{g_i(t_w = 1, r)} \right\rangle$$

where $i \in \{AA, BB, AB\}$, $\sigma_i \leq r < 3\sigma_i$ and $\langle \dots \rangle$ denotes an average over twenty independent configurations. The variation $\langle \delta g_i(t_w, r) \rangle$ tells us how much the radial distribution function at age t_w changes compared to $g(r)$ obtained for a very young glass, i.e. $t_w = 1$. To select the features that are changing the most, we measure $\delta = \max(\langle \delta g_i(t_w, r) \rangle) - \min(\langle \delta g_i(t_w, r) \rangle)$ and select those with the higher δ.

We then compare this traditional approach to our machine learning strategy. The machine-learning model calculates its predictions using all the available data, but we can also identify which features have a stronger influence on the neural network's prediction. We can then verify if these features correspond to those selected with the traditional approach. Moreover, in order to gain more insight into the machine learning model's prediction, we perform a SHAP analysis [68] that calculates the relative contribution of each feature to the prediction. Briefly, the SHAP explanation method computes Shapley values incorporating concepts from cooperative game theory. The goal of this analysis is to distribute the total payoff among players taking into account the importance of their contribution to the final outcome. In this context, the feature values are the players, the model is the coalition, and the payoff is the model's prediction.

Finally, we perform a PCA analysis [69]. PCA is a valuable tool for condensing multidimensional data with correlated variables into new variables that represent linear combinations of the original ones. Essentially, PCA serves as a means to reduce the dimensionality of high-dimensional data. Through the identification of variables exhibiting significant variances, we can uncover the inherent characteristics within the data. The first component corresponds to the projection axis that maximizes the variance in a certain direction, while the second principal component is an orthogonal projection axis that maximizes the variance along the next-leading direction. This process can be iterated to identify additional components.

We will explain in the next sections how machine learning approaches outperform the traditional approach for the system under study, demonstrating that machine learning can be more efficient in inferring the age of a material from simple static properties when noise is inherently present in the data.

III. RESULTS AND DISCUSSION

A. Fixed quenching temperature

Let us first focus on the situation where both the training and prediction have been carried out for a single quenching temperature T_q. This allows us to finely tune the machine-learning model. In the following, networks that are trained with a single quenching temperature will be referred to as '$\mathcal{S}$'. In Sec. III B we compare these models with a more generalized machine-learning model that is trained for a broad range of quenching temperatures.

T_q	class	f1-score	score
Passive system			
0.1	0	1	0.97
	1	0.99	
	2	0.97	
	3	0.96	
	4	0.98	
0.25	0	1	0.94
	1	0.98	
	2	0.92	
	3	0.91	
	4	0.97	
0.375	0	1	0.95
	1	0.97	
	2	0.93	
	3	0.94	
	4	0.96	
Active system			
0.25	0	1	0.93
	1	0.96	
	2	0.92	
	3	0.91	
	4	0.92	

TABLE II. Classification performance of the neural networks $\mathcal{S}$ in the passive and active case. The passive neural networks are trained and tested with $T_q = \{0.1, 0.25, 0.375\}$, while the active NN is trained and tested with $T_q = 0.25$, an active force $f = 0.5$, and a persistence time $\tau_r = 1$. The model has $g_i(r)$ as input, with $\sigma_i \leq r < 3\sigma_i$. In the left column we show the T_q at which each $\mathcal{S}$ is trained, then we show the class label and its corresponding f1-score, and finally in the last column we provide the overall score obtained in the test set.

1. *Age prediction*

To infer the age of a glass, we use a NN that only uses the instantaneous radial distribution functions $g_i(r)$ (120 features in total) as input. We train three different neural networks $\mathcal{S}$, composed by four hidden layers, and trained and tested at quenching temperatures $T_q = 0.1, 0.25$ and 0.375, respectively. We have verified that our bigger alternative NN with twelve hidden layers does not improve the performances (see supplementary material [28]). In Table II, we show the f1-score for each class and the overall score computed in the test set. Table II shows that the f1-score for each class is always higher than 0.9, regardless of the quenching temperature. From this excellent score across all age categories it is clear that, even if the waiting time dependence of the radial distribution function is considered weak, a NN trained exclusively on

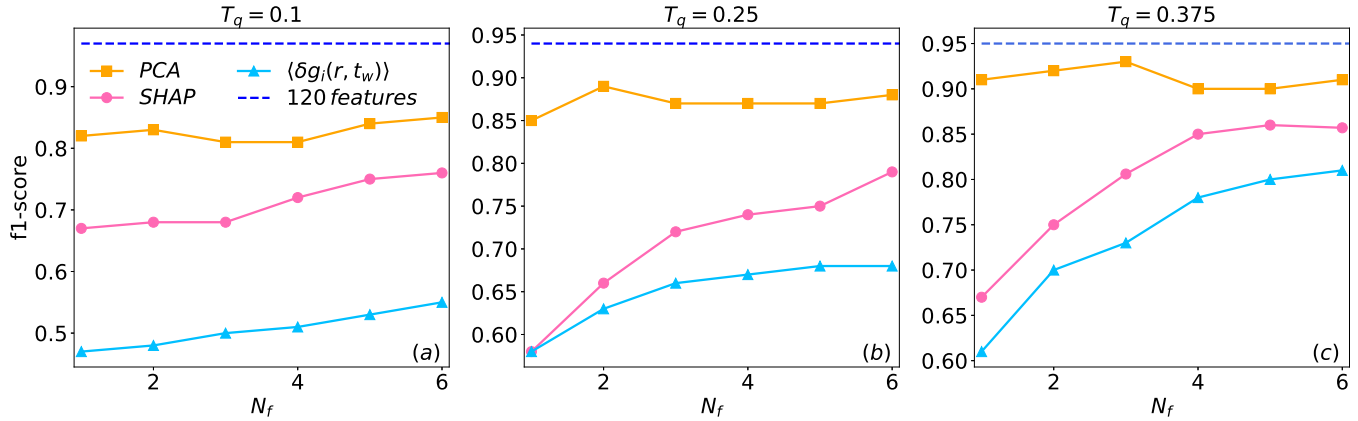

FIG. 2. The f1-score as a function of the number of features N_f used in the machine-learning model. The blue dashed line corresponds to the f1-score showed in Table II obtained with the full dataset (120 features). The orange squares represent the f1-score obtained with the principal components selected by PCA, the pink dots correspond to the f1-score obtained with the most important features selected with a SHAP analysis, and finally the light blue triangles represent the f1-score gained using the features that change the most on average. (a) Quenching temperature $T_q = 0.1$. The SHAP analysis shows that the six most important features are: $g_{AA}(1.05)$, $g_{AA}(1.65)$, $g_{AB}(1.35)$, $g_{AA}(1.80)$, $g_{AA}(1.75)$, $g_{AB}(1.30)$. The six features that are changing the most on average are: $g_{AA}(1.00)$, $g_{BB}(2.00)$, $g_{AB}(1.25)$, $g_{BB}(2.05)$, $g_{AA}(1.60)$, $g_{AB}(1.40)$. (b) Quenching temperature $T_q = 0.25$. The SHAP analysis shows that the six most important features are: $g_{AA}(1.00)$, $g_{AA}(1.65)$, $g_{AB}(1.60)$, $g_{AB}(1.55)$, $g_{BB}(1.80)$, $g_{AA}(1.05)$. The six features that are changing the most on average are: $g_{AA}(1.00)$, $g_{AB}(1.55)$, $g_{BB}(1.00)$, $g_{BB}(2.25)$, $g_{AB}(1.40)$, $g_{BB}(1.05)$.(c) Quenching temperature $T_q = 0.375$. The SHAP analysis shows that the six most important features are: $g_{BB}(1.50)$, $g_{AA}(1.00)$, $g_{AA}(2.45)$, $g_{AB}(1.60)$, $g_{BB}(1.55)$, $g_{BB}(1.45)$. The six features that are changing the most on average are: $g_{BB}(1.00)$, $g_{BB}(1.55)$, $g_{BB}(1.05)$, $g_{BB}(1.50)$, $g_{BB}(2.35)$, $g_{BB}(2.30)$.

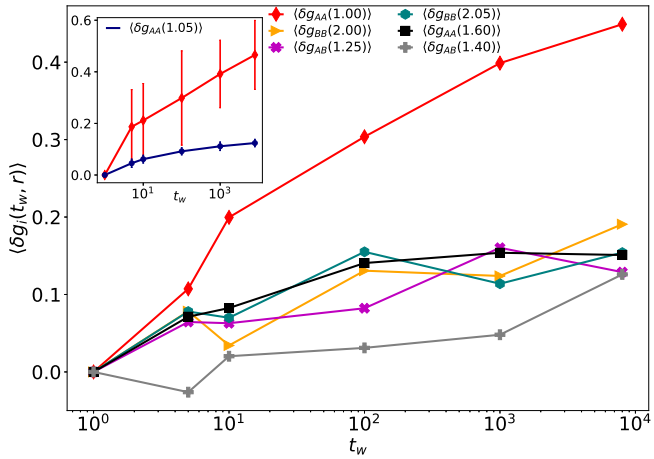

FIG. 3. Variation of the radial distribution function, $\langle \delta g_i(t_w, r) \rangle$, as a function of age t_w. The symbols represent the six features that change the most on average for a system at $T_q = 0.1$. In the inset the red diamonds show the feature that changes the most, $\langle \delta g_{AA}(1.00) \rangle$, with its corresponding standard deviation, and the blue dots correspond to the most important feature according to the SHAP analysis, $\langle \delta g_{AA}(1.05) \rangle$, and its standard deviation.

this structural property is able to distinguish between a young and an old glass with remarkable accuracy.

To verify whether our machine learning approach can also classify the age of an active system, we train and test a model $\mathcal{S}$ with the data of dense ABPs. In Tab. II we show the f1-score corresponding to an aging active system ($f = 0.5$, $\tau_r = 1$) at quenching temperature $T_q = 0.25$. As in the passive case, the f1-scores exceed 0.9 for all age categories across four decades in time. Thus, the neural network also performs well for active glasses when trained and tested at the same temperature. This is consistent with recent works [62, 70] demonstrating that an active system's aging behavior shares several similarities with a passive glass, notably the power-law growth of the alpha relaxation time as a function of the waiting time. In particular, this explains why our machine-learning models for passive and active systems have a similar predictive performance. Finally, while we focused on a two-dimensional system, we have verified that this model also works for a three-dimensional system [28].

2. *Traditional approach versus machine learning*

In the previous section we have shown that a NN trained with 120 static features can reliably predict the age of the system at a given temperature. Here we explore whether all these features are necessary to train a well-performing model, since a subset of features might already efficiently encode the age of the material. To this end, we sort all $g_i(r)$ features in order of importance; The order is determined either from a traditional approach that simply looks for the values of $g_i(r)$ changing the most with age, a machine-learning-based SHAP analysis, or a PCA analysis which extracts the most important

components (see Sec. II D). For these three sortings, we can then train a NN with only the most important features and establish how the age can be most efficiently predicted from minimal structural information.

To compare the traditional approach with machine learning, we train neural networks $\mathcal{S}$ with a different number of features N_f, where $N_f \in \{1, 2, 3, 4, 5, 6\}$. In Fig. 2 we show the f1-score as a function of N_f for both the traditional and machine techniques, namely SHAP-based feature selection and PCA. Each panel corresponds to a different quenching temperature. It can be seen that for all considered temperatures, the predictions restricted to the features selected by SHAP are better than the traditional approach, demonstrating that machine-learning-based feature selection is superior to the traditional 'human learning' approach in this case. Furthermore, PCA outperforms both the conventional method and the SHAP analysis. Importantly, however, the variance associated with the number of principal components ranging from one to six is below 0.5, meaning that the first six components do not contain all the information within the dataset (see Supplementary Material [28]). Consequently, this suggests that the dimensionality-reduced dataset from PCA does not comprehensively represent the entire dataset.

Even though the f1-score for a restricted model is always lower than that for the full model with 120 features (see Fig. 2), both PCA and SHAP restricted to $N_f = 6$ can be considered good classifiers, since their f1-scores are always greater than 0.76 and 0.80, respectively. However, the list of the six optimal features with both SHAP and PCA changes with the quenching temperature, while the full model leads to a f1-score higher than 0.9 regardless of the quenching temperature. Thus, while fewer features can indeed be used to obtain good predictions, this comes at the price of performing a new PCA or SHAP analysis for the full model at each temperature, and hence the full model can be deemed more efficient overall.

Let us now inspect the feature selection more closely to determine why the machine-learning-based selection outperforms the traditional approach. To compare the most important features selected by these two approaches, the focus of the remainder of this section will be on the SHAP analysis and the traditional approach. We note that one could also perform a more in-depth analysis of the main PCA components, but due to their relatively small variance, we prefer to focus on SHAP instead.

A key point in support of machine learning is its ability to perform well for noisy data, i.e. in the presence of fluctuations that are inevitable in experimental or simulation data of finite-sized systems. Figure 3 shows the six features that change the most *on average* for a system with quenching temperature $T_q = 0.1$. From this plot it is clear that $g_{AA}(1.00)$ is the feature that varies the most with age. In particular, an older system corresponds to a larger value of $\langle \delta g_{AA}(1.00) \rangle$. Therefore, one could argue that this feature alone should be sufficient to predict the age of the system. However, Fig. 2(a) shows that the f1-score obtained from a NN trained with only $g_{AA}(1.00)$ (light blue point at $N_f = 1$) is lower than 0.5, while the one corresponding to a single SHAP-selected feature (pink point at $N_f = 1$) is greater than 0.6. This single most important feature according to SHAP is $g_{AA}(1.05)$, which the machine-learning model selects even if it does not change much with age (inset Fig. 3). The reason for this choice, also highlighted in the inset of Fig. 3, is that the *standard deviation* associated to $\langle \delta g_{AA}(1.00) \rangle$ is much larger than the one obtained for $\langle \delta g_{AA}(1.05) \rangle$. Furthermore, in the supplementary material [28], we present the results achieved by subtracting the mean radial distribution function of the complete dataset from the radial distribution function at different ages, denoted as $g_i(r) - \langle g_i(r) \rangle$ with $i \in AA, BB, AB$, at $T_q = 0.375$ for a single snapshot. This analysis confirms results similar to those reported in Fig. 3, and we do not observe any additional significant effects.

From this analysis we can conclude that, according to SHAP, the features that have the biggest influence on the model's prediction are not necessarily those that change the most with age, but rather features that change monotonically with age and have a relatively small standard deviation. Overall, we see that the noise associated with an instantaneous configuration is usually too large to make reliable age predictions based on features selected with the traditional approach. Therefore, we conclude that in order to properly classify the age of a glass from a single snapshot, a machine learning approach is preferred, since it is better equipped to handle noise.

B. Quenching temperature dependence

1. *Age prediction with a generalized model*

We now aim to build a general model that is able to classify the age of the system at any quenching temperature regardless of the T_q used in the training. The first attempt to achieve this goal consists of determining whether the model $\mathcal{S}$, introduced in the previous section III A, can correctly classify unseen data at different temperatures. Therefore, we test each neural network $\mathcal{S}$ with $T_{q_{test}} = 0.11, 0.12, 0.15, 0.17, 0.2, 0.23, 0.3, 0.32, 0.35$. Our results show that the model $\mathcal{S}$ trained with the partial radial distribution functions without the first peaks, $\hat{g}_i(r)$, generalizes better than when trained with the full radial distribution functions $g_i(r)$ [28]. This is not only due to the strong temperature dependence of the main peaks, but also to the fact that those data points are extremely noisy (as shown in Sec. III A 2). For this reason, in this section we will focus on the results corresponding to neural networks trained with $\hat{g}_i(r)$. Moreover, we have found that this model can extrapolate reasonably well only when the difference between $T_{q_{train}}$ and $T_{q_{test}}$ remains sufficiently small. Since this model cannot be used to predict the age of the system at an arbitrary quenching temperature, we examine whether the perfor-

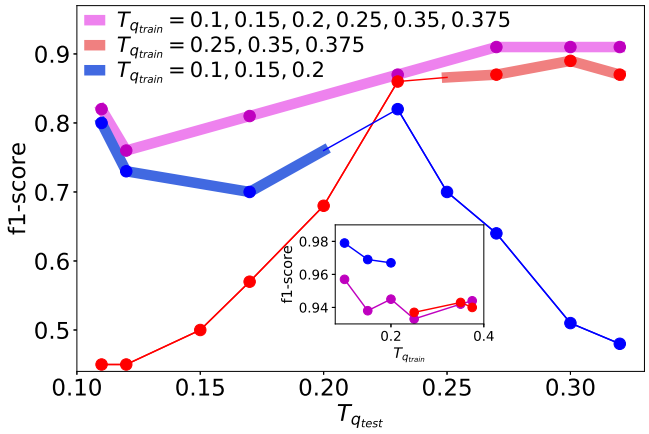

FIG. 4. The f1-score as a function of the quenching temperature used to test the model $T_{q_{test}}$. The purple line shows the neural network $\mathcal{M}$ trained with $T_{q_{train}} = 0.1, 0.15, 0.2, 0.25, 0.35, 0.375$, the red line represents $\mathcal{M}_{high}$ trained with $T_{q_{train}} = 0.25, 0.35,$ 0.375 and the blue line indicates $\mathcal{M}_{low}$ trained for $T_{q_{train}} = 0.1, 0.15, 0.2$. Each dot corresponds to the f1-score obtained in the test set when $T_{q_{test}} \neq T_{q_{train}}$. The temperatures used in the test set are $T_{q_{test}} = 0.1, 0.11, 0.12, 0.15, 0.17, 0.2, 0.23, 0.25, 0.27, 0.3, 0.32$. The inset shows the f1-score when the networks $\mathcal{M}$, $\mathcal{M}_{high}$ and $\mathcal{M}_{low}$ are tested with $T_{q_{test}} = T_{q_{train}}$.

mance of our model further improves when it is trained with a set of multiple (in our case three or six) different quenching temperatures.

To this end we use a new neural network with twelve hidden layers, referred to as '$\mathcal{M}$', that is trained with $T_{q_{train}} = 0.1, 0.15, 0.2, 0.25, 0.35, 0.375$, and subsequently tested with $T_{q_{test}} = 0.11, 0.12, 0.17, 0.23, 0.27, 0.3, 0.32$. We have also verified that for this dataset a NN with twelve hidden layers generalizes better compared to a smaller network (see supplementary material [28]), and that using T_q and $\hat{g}_i(r)$ as input yields the best performance. The purple line in Fig. 4 shows the f1-score of our most general model $\mathcal{M}$ as a function of $T_{q_{test}}$. It can be seen that the f1-score in the test set is always higher than 0.76. Therefore, this model is able to interpolate reasonably well for unseen data. Specifically, for $0.11 \leq T_{q_{test}} \leq 0.17$ we find that $0.76 \leq$ f1-score ≤ 0.82, while, when $0.23 \leq T_{q_{test}} \leq 0.32$ the model has $0.87 \leq$ f1-score ≤ 0.91. Our neural network $\mathcal{M}$ thus performs better for the higher quenching temperatures, i.e. when $T_{q_{test}} \geq 0.23$.

To better understand this behavior and to test if different aging regimes exist, we split the training set into two parts: One for low temperatures $\mathcal{M}_{low}$, with $T_{q_{train}} = 0.1, 0.15, 0.2$, and one for higher quenching temperatures $\mathcal{M}_{high}$, with $T_{q_{train}} = 0.25, 0.35, 0.375$. In both cases we use a NN with four hidden layers, because for these two datasets this performs better than a larger NN. From Fig. 4 we can see that $\mathcal{M}_{high}$ (red curve) performs well (f1-score ≥ 0.86) for $T_{q_{test}} \geq 0.23$. For these $T_{q_{test}}$ values the f1-score is very similar to the one obtained with $\mathcal{M}$. This means that for high temperatures even a small network trained with a smaller set of quenching temperatures is able to generalize to quenching temperatures close to those used in the training set. However, when $\mathcal{M}_{high}$ is tested with $T_{q_{test}} < 0.23$ the corresponding f1-score is lower than 0.7. For these temperatures $\mathcal{M}_{high}$ systematically overestimates the age of the system (see supplementary material [28]). For lower quenching temperatures, instead, $\mathcal{M}_{low}$ (blue curve) has a f1-score higher than 0.7 when it is tested with $T_{q_{test}} \leq 0.25$. In this case this NN performs worse compared to $\mathcal{M}$, but $\mathcal{M}_{low}$ is able to generalize in a larger range of $T_{q_{test}}$ compared to $\mathcal{S}$ trained with $T_{q_{train}} = 0.1$ (reported in the Supplementary material [28]). In order to have higher performances, the low temperature regime needs a bigger set of $T_{q_{train}}$ and a bigger NN. Moreover, similarly to $\mathcal{M}_{high}$, $\mathcal{M}_{low}$ has a f1-score lower than 0.7 when tested with $T_{q_{test}} > 0.23$. In this case, $\mathcal{M}_{low}$ underestimates the age of the system (see supplementary material [28]). As we shall discuss in the following section, the over- or underestimation of $\mathcal{M}_{high}$ and $\mathcal{M}_{low}$ in unseen temperature ranges might be related to the true underlying physics, as the rate of aging depends on the quenching temperature.

The inset of Fig. 4 shows that when we test the three neural networks ($\mathcal{M}$, $\mathcal{M}_{high}$, and $\mathcal{M}_{low}$) with $T_{q_{test}} = T_{q_{train}}$, the f1-score is always higher than 0.93, i.e., all models yield excellent predictions when tested for the temperatures they were trained for. From Fig. 4, we can conclude that a NN trained with quenching temperatures $T_{q_{train}} = T_{q_1}, \ldots, T_{q_m}$ performs well when tested with $T_{q_1} \leq T_{q_{test}} \leq T_{q_m}$. Schoenholtz *et al.* [42] have shown that the history-dependent dynamics in glassy systems can be quantified by the softness and that this property can be used to predict t_w even for systems at different temperatures. Our results show that a simpler model, based only on the radial distribution function, can predict the age of a system at any temperature if the NN is trained on a set of multiple quenching temperatures.

Finally, we have also verified that the model $\mathcal{M}$ trained with passive data can correctly classify the age of an unseen active system ($f = 0.5$ and $T_q = 0.25$) with an f1-score equal to 0.85. This remarkably good performance can be rationalized as follows. In the steady state, an active system can be mapped onto a passive system using an effective temperature, while during aging the effective temperature will change with the age of the system [62]. In this context each class will correspond to a different effective temperature, and for this reason the NN trained on a passive system with multiple quenching temperatures has high performances when tested on an active system.

2. *Physical interpretation of the most important features*

Lastly, we aim to identify which features have a bigger impact on model $\mathcal{M}$'s predictions and how to in-

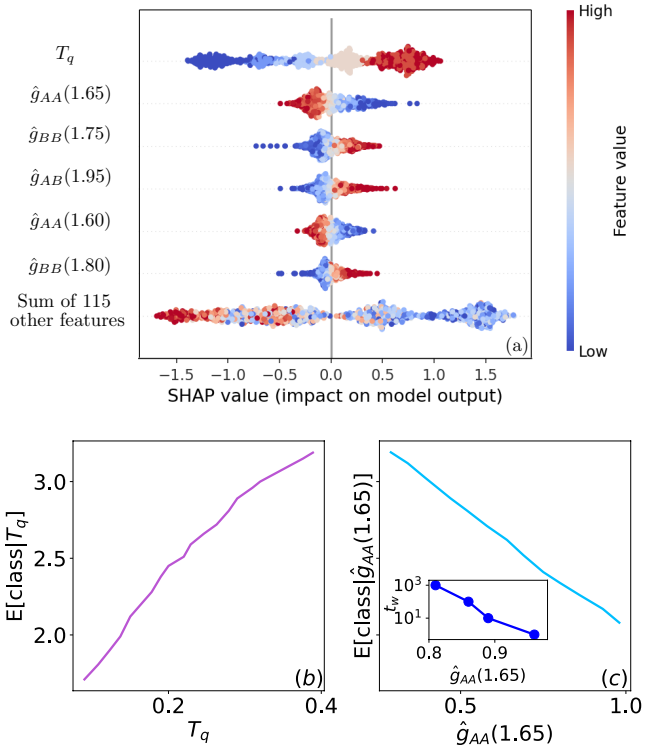

FIG. 5. SHAP-based interpretation of the multilayer perceptron predictions. Here we analyze the neural network $\mathcal{M}$ with twelve hidden layers that has T_q and $\hat{g}_i(r)$ as input and is trained with $T_{q_{train}} = 0.1, 0.15, 0.2, 0.25, 0.35, 0.375$. (a) SHAP beeswarm plot that shows how the most important features impact the model's output. The x position of the dots is determined by the SHAP values of the features and color is used to display the original value of the features. Partial dependence plot for (b) T_q and (c) $\hat{g}_{AA}(1.65)$. The x-axis is the value of the feature and the y-axis is the average value of the model output when we fix T_q or $\hat{g}_{AA}(1.65)$ to a given value. Each class has a label that goes from 0, young glass with $t_w = 0$, to 4, old glass with $10^3 \leq t_w \leq 10^4$. The inset of panel (c) shows the waiting time t_w as a function of $\hat{g}_{AA}(1.65)$. Here we show the actual data for a passive system quenched at $T_q = 0.35$.

terpret the machine learning approach from a physical point of view. For this identification, we choose to employ a SHAP analysis, but it is important to mention that a comparable analysis could also be carried out using methods like PCA. The results of this analysis for our most general model trained on all quenching temperatures (purple line in Fig. 4) are presented in Fig. 5. In particular, Fig. 5(a) shows the SHAP beeswarm plots which indicate the six most important features and how the values of these features influence the model's predictions. The quenching temperature T_q is seen to be the most important feature and the colors in Figure 5(a) show that the model interprets low values of T_q as a young glass and high values of T_q as an old glass.

To better understand this behavior, we have also plotted the partial dependence of T_q in Fig. 5(b). In this plot the quenching temperature is handled independently from the other features, allowing us to precisely pinpoint how changing T_q impacts the model's predictions. In agreement with Fig. 5(a), this plot shows that according to the NN a low T_q is more likely to correspond to a young glass. At first glance this interpretation may look incorrect since the dataset consists of the same amount of ages for each temperature. However, at any fixed waiting time t_w, a system quenched to a higher T_q is always closer to its steady state compared to a system at a lower quenching temperature, because its temperature jump is smaller. Therefore, for any given t_w, the system at a higher T_q is effectively older than the one quenched to a lower T_q. This analysis shows that the NN understands that glasses quenched at higher temperatures age faster. Therefore, the misclassification of $\mathcal{M}_{high}$ and $\mathcal{M}_{low}$ at low and high temperatures (as shown in Sec. III B 1), respectively, might be due to the model's ability to learn that the rate of aging depends on the quenching temperature.

Finally, let us look at the most important structural feature for model $\mathcal{M}$'s predictions. In Fig. 5(a) it is shown that the most important structural feature is $\hat{g}_{AA}(1.65)$, i.e., the point just before the second peak of $\hat{g}_{AA}$. As discussed in Sec. III B 1, the main peak of the radial distribution function strongly depends on temperature and is affected by noise. Therefore, we excluded the first peak from the dataset. Our work does not necessarily imply that the main peak is unimportant, and indeed Schoenholz *et al.* [30] have shown that the radial distribution function's first peak gives 77% accuracy to predict rearrangements. Rather, our work shows that even without the main peak, and focusing only on a seemingly small feature as $\hat{g}_{AA}(1.65)$, we can reliably classify the age. Thus, even a region where the correlation between particles is low contains enough information to classify the system's age. Moreover, in Fig. 5(c) we show that the NN interprets large values of $\hat{g}_{AA}(1.65)$ as an old glass. This feature interpretation is in agreement with the data, as shown in the inset of Fig. 5(c).

IV. CONCLUSIONS

In summary, this proof-of-principle study demonstrates that a simple supervised machine learning method can accurately classify the age of a glass undergoing a temperature quench, relying only on partial radial distribution functions (obtained from an instantaneous configuration, averaged over all particles). The performance of our machine learning algorithm is extremely accurate when the quenching temperature T_q used during training is equal to the one used in the test set (model $\mathcal{S}$), and the model also generalizes well to datasets consisting of multiple quenching temperatures (model $\mathcal{M}$). This good performance for various temperatures indicates the robustness of our method. Extrapolation to unseen temperatures outside the training window is also

reasonable, provided that the temperature difference is not too large. When extrapolating to significantly lower or higher temperatures, however, we find that our neural network tends to systematically under- or overestimate the age of the glass, respectively. This breakdown of the model extrapolation could be ultimately driven by a different physical behavior, as it is well known that higher-temperature glasses effectively age faster.

To establish which features in the radial distribution functions best encode the age of a glassy configuration, we have compared a traditional approach based on physical intuition with a machine learning-based analysis employing SHAP or PCA. The traditional approach manually seeks the values of the radial distribution functions that–on average–change the most with age, while the SHAP method extracts the most important features from a trained neural network. This comparison reveals that machine learning methods strongly outperform the more traditional one. The reason for this is the inevitable statistical noise in the data. Indeed, the fluctuations in the radial distribution functions can vary significantly among different configurations, and the machine-learning model is able to adequately filter out these statistical fluctuations. However, the list of key features selected by SHAP or the principal components selected by PCA changes with the quenching temperature (see Supplementary Material [28]). It follows that in order to identify the most important structural features, one should in principle train a neural network at each T_q with the full dataset and later perform a SHAP analysis or PCA to identify the key features. Since there is usually no cost associated to using a larger number of features, overall we conclude that a model trained with the full data set (120 features) is the most efficient approach.

For our most general machine-learning model (model $\mathcal{M}$), we have also employed SHAP to explain the predictions. This analysis shows that the two most important features are the quenching temperature T_q and the partial radial distribution function $\hat{g}_{AA}(1.65)$. Interestingly, the model is thus able to learn that the rate of aging depends on the quenching temperature and, surprisingly, that $\hat{g}_{AA}(1.65)$, the point just before the radial distribution function's second peak, contains enough information to predict the system's age.

While we have focused on the age classification of a passive glass, we have verified that this machine-learning model works remarkably well even for an active glass composed of active Brownian particles. Our results show that model $\mathcal{M}$ trained with passive data can correctly classify the age of an active system. Therefore, this method could also be used to map the aging behavior of an active glass onto a passive glass at different quenching temperatures [62].

A potential next step of this work could involve incorporating additional structural descriptors to further investigate the relationship between structure and dynamics in aged glasses. Since in recent years, Smooth Overlap of Atomic Positions (SOAP) parameters have proven to be effective in encoding atomic structures [38, 71, 72], one could explore training a machine learning algorithm using these parameters as input.

Our work demonstrates that, even though the radial distribution function of an aging glass is usually considered to remain constant with age, the age dependence, albeit subtle, is already fully encoded in this simple structural property. We thus argue that machine learning methods can be of true added value compared to traditional physical approaches, since they can uncover previously unseen correlations that would be difficult if not impossible to detect with the human eye. Owing to the simplicity and computational efficiency of our approach, we envision that our machine-learning method can be used in a variety of applications, e.g. to quickly distinguish a system that has already reached its steady state from a system that is still aging. This could be particularly attractive for studies in which physical aging is an undesirable and difficult problem, such as equilibration of deeply supercooled liquids; With our model, it would be possible to verify whether a supercooled liquid has reached equilibrium from a single snapshot.

ACKNOWLEDGMENTS

It is a pleasure to thank Robert Jack for stimulating discussions. 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), and Vidi grant (LMCJ).

[1] I. M. Hodge, Science **267**, 1945 (1995).
[2] L. Berthier and G. Biroli, Encyclopedia of Complexity and Systems Science , 4209 (2009).
[3] P. Lunkenheimer, R. Wehn, U. Schneider, and A. Loidl, Physical Review Letters **95**, 055702 (2005).
[4] J. Zhao, S. L. Simon, and G. B. McKenna, Nature Communications **4**, 1783 (2013).
[5] J. Y. Raty, W. Zhang, J. Luckas, C. Chen, R. Mazzarello, C. Bichara, and M. Wuttig, Nature Communications **6**, 2041 (2015).
[6] P. Wang, C. Song, and H. A. Makse, Nature Physics **2**, 526 (2006).
[7] G. Odegard and A. Bandyopadhyay, Journal of polymer science Part B: Polymer physics **49**, 1695 (2011).
[8] B. Martin, Calcified Tissue International **53**, S34 (1993).
[9] G. B. McKenna, Y. Leterrier, and C. R. Schultheisz, Polymer Engineering & Science **35**, 403 (1995).
[10] L. C. E. Struik, Polymer Engineering & Science **17**, 165 (1977).
[11] K. Binder and W. Kob, *Glassy materials and disor-*

dered solids: An introduction to their statistical mechanics (World Scientific, 2011).
[12] G. Biroli and J. P. Garrahan, The Journal of Chemical Physics **138**, 12A301 (2013).
[13] P. G. Debenedetti and F. H. Stillinger, Nature **410**, 259 (2001).
[14] F. Turci, C. P. Royall, and T. Speck, Phys. Rev. X **7**, 031028 (2017).
[15] W. Kob and J. L. Barrat, Physical Review Letters **78**, 4581 (1997).
[16] G. Foffi, E. Zaccarelli, S. Buldyrev, F. Sciortino, and P. Tartaglia, The Journal of Chemical Physics **120**, 8824 (2004).
[17] M. Warren and J. Rottler, Europhysics Letters **88**, 58005 (2009).
[18] M. Warren and J. Rottler, Phys. Rev. Lett. **110**, 025501 (2013).
[19] J. M. Hutchinson, Progress in Polymer Science **20**, 703 (1995).
[20] J. L. Barrat and W. Kob, EPL (Europhysics Letters) **46**, 637 (1999).
[21] W. Kob and J. L. Barrat, The European Physical Journal B: Condensed Matter and Complex Systems **13**, 319 (2000).
[22] W. Kob, J.-L. Barrat, F. Sciortino, and P. Tartaglia, Journal of Physics: Condensed Matter **12**, 6385 (2000).
[23] T. Kawasaki and H. Tanaka, Phys. Rev. E **89**, 062315 (2014).
[24] M. Warren and J. Rottler, Phys. Rev. E **76**, 031802 (2007).
[25] Y. Waseda and T. Egami, J. Mater. Sci. **14**, 1249 (1979).
[26] M. A. Popescu, Journal of non-crystalline solids **169**, 155 (1994).
[27] Y. Fan, T. Iwashita, and T. Egami, Phys. Rev. E **89**, 062313 (2014).
[28] *See Supplemental Material at [URL] for more details on the hyperparameters used for the machine-learning model, principal component analysis, and Shap analysis for different temperatures and the active system.*
[29] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Phys. Rev. Lett. **114**, 108001 (2015).
[30] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, Nature Physics **12**, 469 (2016).
[31] E. D. Cubuk, R. J. S. Ivancic, S. S. Schoenholz, D. J. Strickland, A. Basu, Z. S. Davidson, J. Fontaine, J. L. Hor, Y.-R. Huang, Y. Jiang, N. C. Keim, K. D. Koshigan, J. A. Lefever, T. Liu, X.-G. Ma, D. J. Magagnosc, E. Morrow, C. P. Ortiz, J. M. Rieser, A. Shavit, T. Still, Y. Xu, Y. Zhang, K. N. Nordstrom, P. E. Arratia, R. W. Carpick, D. J. Durian, Z. Fakhraai, D. J. Jerolmack, D. Lee, J. Li, R. Riggleman, K. T. Turner, A. G. Yodh, D. S. Gianola, and A. J. Liu, Science **358**, 1033 (2017), https://www.science.org/doi/pdf/10.1126/science.aai8830.
[32] F. m. c. P. Landes, G. Biroli, O. Dauchot, A. J. Liu, and D. R. Reichman, Phys. Rev. E **101**, 010602 (2020).
[33] E. D. Cubuk, S. S. Schoenholz, E. Kaxiras, and A. J. Liu, The Journal of Physical Chemistry B **120**, 6139 (2016), pMID: 27092716, https://doi.org/10.1021/acs.jpcb.6b02144.
[34] D. Richard, M. Ozawa, S. Patinet, E. Stanifer, B. Shang, S. A. Ridout, B. Xu, G. Zhang, P. K. Morse, J.-L. Barrat, L. Berthier, M. L. Falk, P. Guan, A. J. Liu, K. Martens, S. Sastry, D. Vandembroucq, E. Lerner, and M. L. Manning, Phys. Rev. Mater. **4**, 113609 (2020).
[35] N. Oyama, S. Koyama, and T. Kawasaki, Frontiers in Physics **10** (2023), 10.3389/fphy.2022.1007861.
[36] I. Tah, S. A. Ridout, and A. J. Liu, The Journal of Chemical Physics **157**, 124501 (2022).
[37] G. Jung, G. Biroli, and L. Berthier, Phys. Rev. Lett. **130**, 238202 (2023).
[38] D. Coslovich, R. L. Jack, and J. Paret, The Journal of Chemical Physics **157**, 204503 (2022).
[39] S. Ciarella, M. Chiappini, E. Boattini, M. Dijkstra, and L. M. C. Janssen, Machine Learning: Science and Technology **4**, 025010 (2023).
[40] R. M. Alkemade, F. Smallenburg, and L. Filion, The Journal of Chemical Physics **158** (2023), 10.1063/5.0144822, 134512.
[41] G. Janzen, X. L. J. A. Smeets, V. E. Debets, C. Luo, C. Storm, L. M. C. Janssen, and S. Ciarella, Europhysics Letters **143**, 17004 (2023).
[42] S. S. Schoenholz, E. D. Cubuk, E. Kaxiras, and A. J. Liu, Proceedings of the National Academy of Sciences **114**, 263 (2017), https://www.pnas.org/content/114/2/263.full.pdf.
[43] S. Ciarella, D. Khomenko, L. Berthier, F. C. Mocanu, D. R. Reichman, C. Scalliet, and F. Zamponi, Nat. Commun. **14**, 4229 (2023).
[44] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. R. Nelson, T. Back, D. Hassabis, and P. Kohli, Nature Physics **16**, 448 (2020).
[45] E. Boattini, F. Smallenburg, and L. Filion, Physical Review Letters **127**, 88007 (2021).
[46] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).
[47] H. F. Mahmoud, International Journal of Statistics and Probability **10** (2021).
[48] Y. Tokuda, M. Fujisawa, D. M. Packwood, M. Kambayashi, and Y. Ueda, AIP Advances **10**, 105110 (2020), https://doi.org/10.1063/5.0022451.
[49] J. E. Lennard-Jones, Proceedings of the Physical Society **43**, 461 (1931).
[50] W. Kob and H. C. Andersen, Physical Review E **51**, 4626 (1995).
[51] S. Plimpton, Journal of Computational Physics **117**, 1 (1995).
[52] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability (Springer Berlin Heidelberg, 2011).
[53] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, The European Physical Journal Special Topics **202**, 1 (2012).
[54] S. Ramaswamy, Annual Review of Condensed Matter Physics **1**, 323 (2010).
[55] B. Lindner and E. Nicola, The European Physical Journal Special Topics **157**, 43 (2008).
[56] H. Löwen, The Journal of Chemical Physics **152**, 040901 (2020).
[57] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger, Nature Reviews Physics **2**, 181 (2020).
[58] L. Dabelow, S. Bo, and R. Eichhorn, Phys. Rev. X **9**, 021009 (2019).
[59] A. Zöttl and H. Stark, Journal of Physics: Condensed Matter **28**, 253001 (2016).

[60] E. Flenner and G. Szamel, Nature Communications **6**, 7392 (2015).
[61] D. Li, H. Xu, and J. P. Wittmer, Journal of Physics: Condensed Matter **28**, 045101 (2016).
[62] G. Janzen and L. M. C. Janssen, Phys. Rev. Research **4**, L012038 (2022).
[63] M. Gardner and S. Dorling, Atmospheric Environment **32**, 2627 (1998).
[64] S. Pal and S. Mitra, IEEE Transactions on Neural Networks **3**, 683 (1992).
[65] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, the Journal of machine Learning research **12**, 2825 (2011).
[66] S. Haykin, *Neural Networks and Learning Machines*, Pearson International Edition (Pearson, 2009).
[67] D. P. Kingma and J. Ba, arXiv preprint arXiv:1412.6980 (2014).
[68] S. M. Lundberg and S.-I. Lee, in *Proceedings of the 31st international conference on neural information processing systems* (2017) pp. 4768–4777.
[69] I. T. Jolliffe, *Principal component analysis for special types of data* (Springer, 2002).
[70] R. Mandal and P. Sollich, Physical Review Letters **125**, 218001 (2020).
[71] A. P. Bartók, R. Kondor, and G. Csányi, Physical Review B **87**, 184115 (2013).
[72] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).