Predicting creep failure by machine learning - which features matter?

Stefan Hiemer [a], Paolo Moretti [a], Stefano Zapperi [a,b,c], Michael Zaiser [*,a]

[a] Institute of Materials Simulation, Department of Materials Science Science and Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Dr.-Mack-Str. 77, Fürth 90762, Germany
[b] Center for Complexity and Biosystems, Department of Physics, University of Milan, via Celoria 16, Milan 20133, Italy
[c] CNR - Consiglio Nazionale delle Ricerche, Istituto di Chimica della Materia Condensata e di Tecnologie per l'Energia, Via R. Cozzi 53, Milan 20125, Italy

ARTICLE INFO

ABSTRACT

Spatial and temporal features are studied with respect to their predictive value for failure time prediction in subcritical failure with machine learning (ML). Data are generated from simulations of a novel, brittle random fuse model (RFM), as well as elasto-plastic finite element simulations (FEM) of a stochastic plasticity model with damage, both models considering stochastic thermally activated damage/failure processes in disordered materials. Fuse networks are generated with hierarchical and nonhierarchical architectures. Random forests - a specific ML algorithm - allow us to measure the feature importance through a feature's average error reduction. RFM simulation data are found to become more predictable with increasing system size and temperature. Increasing the load or the scatter in local materials properties has the opposite effect. Damage accumulation in these models proceeds in stochastic avalanches, and statistical signatures such as avalanche rate or magnitude have been discussed in the literature as predictors of incipient failure. However, in the present study such features proved of no measurable use to the ML models, which mostly rely on global or local strain for prediction. This suggests the strain as viable quantity to monitor in future experimental studies as it is accessible via digital image correlation.

1. Introduction

Creep failure is an example of a subcritical failure process, where an applied load which is insufficient to instantaneously break the sample drives time dependent damage accumulation. This gradual accumulation of damage deteriorates the strength of the material and ultimately results in delayed failure [1].

It is in general unfeasible to design structures in such a manner as to avoid the damage processes that lead to subcritical failure. Predicting the residual lifetime of a structure under subcritical load is therefore an important issue that is actively investigated by both physicists and engineers. Reliable lifetime predictions may help to avoid catastrophic in-service failure of components and systems and harness substantial economic benefits by adapting and where possible extending replacement cycles.

To assist prediction, it is desirable to obtain sample specific information about the damage accumulation process through non destructive means. Such information can be obtained from the macroscopic sample response, i.e., the time dependent creep strain or strain rate as accessible by surface monitoring. Additional and more detailed information can be drawn from analysis of the spatio-temporal pattern of energy releases as local creep damage accumulates, as microcrack formation is accompanied by elastic energy release which can be recorded by monitoring the acoustic emission (AE) of the sample.

Several empirical approaches have been proposed to predict sample specific failure times from macroscopic creep strain data. The simplest possible approach is to correlate the time t_m of minimum strain rate with the catastrophic failure time t_f, in the simplest case by assuming a linear relationship between both [2,3]. A variant consists in relating the failure time to the duration of the primary (decelerating) creep stage [4].

A slightly different approach towards failure time prediction based on macroscopic strain (strain rate) focuses, instead, on the rapid increase of creep strain and strain rate in the run-up to failure, which typically is characterized by a creep strain (strain rate) that increases like an inverse power of the time-to-failure. Fitting such a power law to the data recorded until a given moment implies a prediction of the residual lifetime – an approach which has been promoted by D. Sornette and applied, in different variations, to catastrophic phenomena from material rupture over financial crises to childbirth [5] to the catastrophic breakdown of civilization as we know it [6].

A related prediction approach focuses on temporal statistics of precursor events, whose magnitudes and rates also may develop characteristic singularities in the approach to failure. For instance, one may exploit the observation made both in simulations [7] and experiments [8] that the AE event rate ν_{AE} accelerates towards failure according to a reverse Omori law, $\nu_{AE} \propto (t - t_f)^{-p}$ with $p \approx 1$. This behavior is also found in mean-field models of thermally activated rupture [9] and allows to obtain the failure time by fitting the Omori law to the previous AE record and extracting the expected failure time t_f, which is one of the fit parameters. At the same time, the approach to failure may be accompanied with other characteristic changes in the AE burst statistics, such as an increase in the AE event size or characteristic changes in the Gutenberg-Richter exponent of the power law type energy statistics [7], which may also be used for monitoring and prediction.

Beyond the temporal record of strain, strain rate and AE, additional information can be obtained by simultaneously monitoring the spatial pattern of local strains, or more generally of damage accumulation. Materials failure is associated with localization of damage [7,8], and signatures of damage or strain localization may provide additional features that assist failure forecasting. To accurately describe the material response, machine learning (ML) has arisen as a new tool in recent years [10]. Among the goals are finding new models [11,12], bypassing expensive solvers with ML as surrogate [13,14] and reducing models with a large number of parameters with the few necessary by inclusion of the proper symmetries and regularization [15]. For the special case of creep, research has mainly focused on making average lifetime predictions based on composition, microstructure and processing based features [16–18].

In the present investigation, we use ML approaches in a different sense, namely to make *sample specific* predictions of deformation behavior. In the context of plasticity, machine learning has been used to predict sample specific stress-strain curves in dislocation plasticity based on initial dislocation microstructure [19] and based on surface strain patterns determined in the microplastic regime [20]. A more general approach to predict the onset of plasticity or damage based on generic microstructural signatures was proposed in Ref.[21]. Regarding fracture, machine learning has been used to identify critical fracture loads together with crack nucleation sites and propagation pathways based on the atomic configuration of individual samples of amorphous silica [22], and in the context of subcritical failure, ML has been used to predict sample specific failure times in disordered solids by means of random forest regression [23] from characteristics of local deformation events recorded up to a given moment in time. Even though the event statistics for the data used in this model is characterized by a very clear reverse Omori law [7], the ML approach clearly outperformed lifetime predictions based on Omori law fits. Random forests were also used to forecast laboratory earthquakes from AE time series records [24]. In the latter context, [25] have issued a Kaggle challenge to benchmark the performance of different ML algorithms with a prize money of 25,000 $ which was won by decision tree based method.

In the present study, we use ML to assess the usefulness of different spatial and temporal features for predicting creep failure times. As data base we consider simulation data from two different types of models, namely a highly simplified model (random fuse model, RFM) of thermally activated damage accumulation, as well as the stochastic FEM model of [7] that was used in the previous work by Biswas et al. [23]. These models are introduced in Section 2 together with the respective feature sets extracted from the simulations for sample specific failure time prediction. The behavior of the RFM in the run-up to failure is discussed in Section 3.1, while results of the ML analysis of the data sets are given in Section 3.2. Section 4 concludes with a critical discussion of the usefulness of different features, and the possibilities and limitations of sample specific lifetime prediction.

2. Methods

2.1. RFM simulations

We consider simplified models of failure, so called random fuse models, which provide a scalar caricature of elastic-brittle behavior by modelling materials as networks of scalar load carrying elements. We envisage two-dimensional structures as shown in Fig. 1. In mechanical terms, these structures can be envisaged as modelling sheet-like materials deformed in plane stress conditions. The structures consist of beam-like load carrying elements of unit length, which form a two-dimensional lattice of junction points, or *nodes*. A tensile-like load is applied along a statistical symmetry axis of the lattice structure, corresponding to the vertical direction in Fig. 1. In the continuum limit of infinitely many load carrying elements, the scalar character of load and displacement corresponds to a material of zero Poisson ratio loaded in uni-axial tension, for which the equations of elastostatics reduce to the Laplace equation.

The elements are arranged using the following architecture: For a given system size L, L fibers consisting each of L vertically adjacent, vertically oriented elements transmit load across the system in the load-parallel direction, while a fixed number $C < L$ of horizontal cross-linking elements is responsible for load redistribution in the load-perpendicular direction. Fig. 1 shows three possible variants of this construction. In the Random Fuse Network (RFN), horizontal cross-linking elements are distributed randomly. This arrangement results in the formation of vertical *gaps*, which interrupt load redistribution and exhibit an exponential length distribution [26]. A deterministic hierarchical fuse network (D-HFN) refers to a similar system, where however the cross links are distributed hierarchically, and the resulting gaps have a heavy tailed, power-law size distribution [26]. Finally, a shuffled hierarchical fuse network (S-HFN) is constructed from a D-HFN by randomly shuffling rows and columns of the network adjacency matrix. This construction maintains the power-law gap-size distribution and shows the same failure phenomenology as exhibited by the D-HFN, but at the same time allows for averaging over different network realizations. In all systems, periodic boundary conditions are imposed in the load perpendicular direction. In the actual simulations a D-HFN pattern for a prescribed size L is constructed first, and RFN and S-HFN variants are generated from that, making sure that the number of horizontal fuses is the same. In the following, we restrict our study to RFN and S-HFN, and we refer to S-HFN simply as HFN from now on.

The term 'random fuse model' is used because the dynamics of damage accumulation and fracture can be described using an electrical analogue [27,28]: Owing to the scalar nature of the load variable, the load carrying elements can be envisaged as fuses carrying currents according to scalar constitutive equations. Identifying a fuse ij by the indices of its two end-nodes i and j, we use I_{ij} to indicate the scalar force (current) acting on ij, and we relate it to the scalar displacements (voltages) V_i and V_j of the fuse endpoints through the scalar Hooke's law (Ohm's law)

$$I_{ij} = V_i - V_j, \tag{1}$$

where we have assumed a unit elastic modulus (resistance), while the balance equation for each node i can be computed imposing that the algebraic sum of the scalar forces acting on i is zero (in the electrical analogue: Kirchhoff's law). An externally imposed displacement is simulated by applying a prescribed voltage difference between the upper and lower boundaries, while an imposed external load is represented by an imposed global current across the network. The response of the system is then captured by monitoring the evolution of the complementary variable (applied voltage V for load control, global current I for displacement control). The RFM implies a simplification of the elastic problem which however preserves, in the continuum limit, essential aspects of fracture mechanics such as load re-distribution and stress
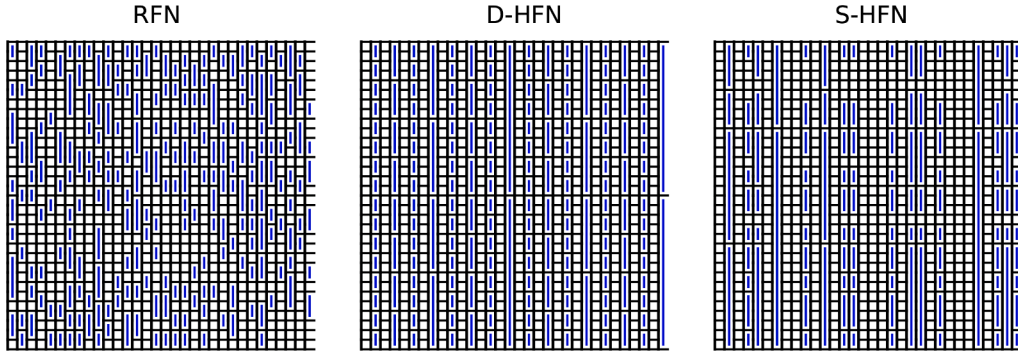
RFN
D-HFN
S-HFN

Fig. 1. Two-dimensional fuse networks of size $L = 32$. External uni-axial loads are applied in form of fixed global currents flowing through the networks from top to bottom. All systems share the same number of horizontal cross-linking fuses. Gaps are emphasized in blue. Gap sizes are exponentially distributed in the non-hierarchical case (RFN) and power-law distributed in the hierarchical case (D- and S-HFN). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

singularities at crack tips. RFM models have for this reason been widely used in statistical modelling of fracture [29].

In order to account for local strength fluctuations, each fuse ij is assigned a threshold θ_{ij}, corresponding to a critical current at which it fails irreversibly (i.e., the fuse conductivity is set to zero). Thresholds are distributed following a Weibull cumulative distribution function of the form

$$C\left(t_{ij}\right) = 1 - \exp\left[- \left(\frac{\theta_{ij}}{\theta_0}\right)^k \right], \tag{2}$$

where we choose $\theta_0 = 1/\Gamma(1 + 1/k)$ to set the mean value $\langle \theta_{ij} \rangle = 1$, and we vary the shape parameter k to control the degree of heterogeneity in the material model, lower k leading to larger fluctuations [30].

To simulate thermally activated sub-critical creep failure, we tweak the voltage difference between top and bottom boundary such as to ensure that the total current I remains at a fixed, prescribed level. Fuses may in this model fail either instantaneously by overloading ($\theta_{ij} - I_{ij} < 0$) or by thermal activation, which is considered in terms of crossing an energy barrier

$$\Delta U = \theta_{ij} - I_{ij}. \tag{3}$$

Whenever $\theta_{ij} \leq I_{ij}$ holds for any number of fuses, the corresponding fuses ij are removed instantaneously, equilibrium equations are solved again and the process is repeated until no instantaneous removals occur. When $\theta_{ij} > I_{ij}$ for every fuse ij, a Kinetic Monte Carlo step is performed to identify the fuse which fails next because of thermally assisted energy barrier crossing, and to determine the time interval Δt after which this happens. The attempts to cross barriers are assumed to be stochastically independent, and thus described by Poisson processes with the transition rates

$$\nu_{ij} = \nu_0 \exp\left(- \frac{\theta_{ij} - I_{ij}}{T} \right) \tag{4}$$

with the characteristic frequency ν_0 and the temperature T scaled with appropriate constants (e. g. Debye frequency, Boltzmann's constant). The unit of time of the simulations is ν_0^{-1}, thus we set $\nu_0 = 1$ without loss of generality. A failing fuse is selected and removed accordingly, and time is increased by an interval $\Delta \tau$ extracted from an exponential distribution with mean value ν^{-1} where the total event rate is

$$\nu = \sum_{ij \in \mathscr{F}} \nu_{ij}. \tag{5}$$

Here $\mathscr{F}$ refers to the set of surviving fuses at a given time. Note that thermally activated failure of one fuse may, due to load re-distribution, lead to overloading of other fuses which then fail instantaneously. In this case we speak of an *avalanche* of size s where s is the number of fuses that fail in a correlated manner until the stability condition $\theta_{ij} > I_{ij} \forall ij$ is met. The next event is then again thermally activated.

A subset of simulations is also run resorting to an extremal (i.e. non-stochastic) failure criterion (see e.g. Alava et al. for details [29]) which is equivalent to the zero-temperature limit of the protocol outlined above. In extremal simulations, due to the absence of thermal activation all fuses fail by overloading – either by an increase in applied load or at constant load in an avalanche. Damage accumulation and failure are driven by increasing the applied load, and extremal simulations thus allow us to identify the global peak load (peak current) I_p which can be supported by the system of fuses before it undergoes instantaneous failure.

Subcritical creep loading, on the other hand, is performed by applying a constant load I in the interval $]0, I_p[$ at finite temperature, such that fuses can fail by thermal activation. For all considered combinations of parameter values (Table 1) we gathered data from multiple samples, performing for each parameter combination 1200 simulations with different sets of random thresholds assigned from the same statistical distribution.

2.2. Finite element data

As an alternative source of information, and to ensure that our findings are not contingent on peculiarities of the creep model we use, data created in a previous study [23] from a two-dimensional elasto-plastic creep finite element (FEM) model were also considered. This model considers a 2D square block made of a stochastically heterogeneous material where each element has a different, randomly assigned local yield stress. The block is loaded in simple plane shear. For details of the simulation we refer to [7,31], here we only summarize the main analogies and differences between FEM and RFM creep models: i) The RFM model is equivalent to a periodically continued domain loaded in pure tension without cross contraction, where the axial stress corresponds to the scalar current of the electrical analogue. The FEM model, on the other hand, considers plane strain deformation which makes a tensorial treatment mandatory. The current is now replaced by the von Mises equivalent stress calculated from the deviatoric stress, the local threshold is the counterpart of a local yield stress, and the role of fuses is taken over by the elements. ii) Once the threshold of an element is exceeded by the local stress, the element deforms by a plastic incremental strain $\Delta \epsilon$. Unlike the RFM, elements can deform repeatedly, however, with each threshold crossing damage is added to the element.

Table 1
Choice of parameters investigated for the RFM. All combinations of these parameters are investigated with machine learning except current 0.1 for size 128.

parameter	parameter values
k	1,2,4,8,16
T	0.01,0.03,0.05,0.1
I/I_p	(0.1),0.3,0.5,0.7,0.9
L	16,32,64,128

After the threshold crossing, a new threshold is re-drawn from a Weibull distribution with shape parameter k whose mean value decreases with the accumulated element damage. Again, deformation proceeds either by thermally activated events or by stress re-distribution leading to overloading, causing avalanches. The avalanche size is here defined as the number of strain increments caused by correlated overloading between two consecutive thermally activated barrier crossing events. The simulation ends once the system enters a never-ending avalanche, pragmatically defined as an avalanche size that exceeds the total number of elements.

The creep load is again measured in units of the zero-temperature failure stress of the specific system. For stress 0.7, systems of varying disorder ($k \in \{1, 2, 4, 8, 16\}$) are investigated while for stress 0.9 just $k = 4$ is explored. The recorded data are the times and locations of thermally activated events as well as the sizes of the ensuing avalanches.

2.3. Machine learning methodology

When deciding which machine learning method should be used for our prediction problem, there are roughly three groups of methods to think about: Neural networks, kernel methods like the support vector machine, and decision tree based methods. The chosen method should ideally meet the following criteria: i) it should be applicable to data from both sources (FEM and RFM) to allow for comparison; ii) it should scale computationally well enough to handle tens of millions of samples as the average number of events per FEM trajectory is of the order of 10^4; iii) it should offer the possibility of obtaining information regarding feature importance; iv) it should be able to yield a time to failure distribution (such as to be able to assess prediction scatter) as our data stem from stochastic processes.

Convolutional neural networks (CNN) - the backbone of almost all modern neural network architectures for image recognition [32–34] - are applicable to the FEM data, as these data can be conveniently transformed to a time series of images. For RFM this is not possible as there is no convenient way to convert the adjacency and edge threshold information into an image without either losing key information regarding adjacency or creating an image of very large size. The alternative to CNNs would be graph neural networks (GNN) which are capable of handling both data types. Modern GNNs typically operate by converting each sample into a graph structure in which each node draws information from the neighbor nodes and the edges connecting them. The process of drawing information is called graph convolution. The final prediction is made by summing over all nodes. Thus the number of convolutions determines the amount of nonlocal information that the network can analyze. Unfortunately, today's GNNs cannot perform nearly the same number of convolutions as CNN based architectures because computational cost and memory requirements increase rapidly with increasing graph size [35,36]. Building deep GNNs is still an open problem and interpretation is also not straightforward. Additionally, when looking at a well-established area of GNN application such as molecular machine learning, one has to realize that the RFMs used here define much larger graphs than their molecular counterparts. As example, take the QM9 dataset [37] whose largest molecules have 29 atoms/nodes while our largest graphs have 2^{14} nodes. To summarize: Point i) is not met for CNNs, for GNNs point iii) is not met and point iv) is questionable. Given the fact that each node has almost no information about the state of the entire network due to the limited number of convolutions, chances of success for GNNs are slim in this setting.

For kernel methods like kernel ridge regression, the computational cost of training scales in proportion with the third power of the size of the training set and the memory requirement with the second power due to the kernel matrix [38]. Given the fact that the training set size here can reach well above a few millions, kernel methods can be ruled out for reasons of numerical cost. As kernel methods belong to the class of instance based models, meaning they make predictions based on similarity/distance to samples in the training set, they do not make an assessment of individual features, since they only care about the total distance.

Decision tree based regressors meet requirements (i), (ii) and (iii), while point (iv) is met by random forests. Given the success of random forests in laboratory-scale earthquake prediction in a past publication [24], the fact that a recent Kaggle challenge for laboratory-scale earthquake prediction was also won by a tree based algorithm [25], and the promising results obtained with the random forest method in a previous publication [23], we consider random forests as the regressor of choice for this study.

2.4. Random forests

Random forests are constructed by bootstrapping of decision trees, which means that, instead of considering the predictions of a single tree, one takes the average over the predictions of an ensemble of regressors that have been trained on different random subsamples of the training data. This mitigates against overfitting (for which single decision trees are notorious) and provides additional probabilistic information in terms of the statistical distribution of predictions. Statistical signatures of this distribution, such as the variance of predicted lifetimes, provide a means not only to forecast a single lifetime value but also to estimate the reliability of the predicted lifetime in terms of the expected scatter of the predictions. Single decision trees divide the feature space recursively into partitions. The predicted value of a variable – here: the residual lifetime – in each partition is the average of the actual values of that variable for the training samples that fall into the partition. The partitioning is done on form of binary decisions ('splits') based on single features, taking in each recursive step the decision which minimizes the mean squared error of the ensuing predictions. Feature importance is determined in terms of the (normalized) average error reduction that is achieved by making splits based on a given feature [39]. For details we refer to [40–42].

Scikit version 0.23.2 is used for implementation of the machine learning scheme [43]. All machine learning algorithms rely on sets of hyperparameters which control the working of the algorithms. In the context of random forests, these hyperparameters describe the manner in which random subsamples of the training data are constructed (size of the subsamples, size of the forest) and how the trees are organized (number of binary decisions made). Apart from the number of trees (1000), we use default parameters for the random forest (which can be found in the manual of version 0.23.2), because preliminary studies as well as the previous work of Biswas et al. [23] indicate that the results are insensitive to the choice of hyperparameters. To measure prediction performance, besides summary statistical signatures like the mean squared error, the machine learning prediction (ML) at time t is compared to a set of four simple baseline predictions t_{BA}

$$t_{\text{BA}}(t) = (t_{\text{ref}} - t)H(t_{\text{ref}} - t), \tag{6}$$

where t_{ref} is one of four reference times. H is the Haeviside function to avoid prediction of a negative remaining time to failure. Two simple reference times are the average and median lifetime of the entire training set. For the other two, one compares t with the lifetimes t_{f} in the training set, ignores all lifetimes smaller than t and takes the mean or median lifetime of the remaining samples. We refer to the first two reference as static mean and median whereas the last two as dynamic mean and median.

2.5. Feature selection

For the FEM data, two different feature sets have been used: The features described the previous work of Biswas et al. [23] and a new set consisting of time, total damage as proxy for the total strain, thermally activated event rate, the average avalanche amplitude over a finite time window, and a spatial damage localization parameter. The latter

quantity is defined resorting to domain knowledge, which tells us that for the imposed loading conditions (simple shear imposed on the boundaries of a quadratic domain), an incipient shear band must be aligned with the domain boundaries. Accordingly, we determine the maximum value of strain (or equivalently damage) within a rectangle of predefined width that spans the system parallel to one of its boundaries, i.e. in one of the two directions of a potential shear band. The rectangle width is varied (2,4,8,32 elements) such as to achieve optimum prediction performance. To calculate the thermally activated event rate, a running average is used over $n_{av} = \text{floor}\left(\frac{\overline{N}_{av}}{N}\right)$ events where $\overline{N}_{av}$ is the average number of thermally activated events in a simulation and N an integer value that needs to be chosen. We refer to N as rate parameter. In this study the values $50, 100, 200$ are compared. The event rate r at the ith event then is defined as

$$r_i = \frac{n_{av}}{\left(t_i - t_{(i-n_{av})}\right)} \tag{7}$$

The average avalanche amplitude is calculated over the same running window.

For the RFM data, also two feature sets are used. For the first set, the features are time, voltage, the total damage measured as the number of failed fuses, as well as the maximum number of failed fuses aligned in the direction perpendicular to the applied load, which corresponds to the propagation direction of an emergent tensile crack. The second feature set consists of time, voltage, the total damage measured as the number of failed fuses and the maximum crack size. The latter is obtained by determining all connected clusters of broken fuses. The size of such a cluster is defined as the number of broken fuses. The largest crack, i.e. the largest connected component, is found by using the graph implementation of scipy [44]. No rate based features were calculated for the RFM data as this implies a smoothing/averaging step which can only be meaningfully performed if the time series is of sufficient length. Typical RFM parameter combinations exhibit (on average) $10^1 - 10^2$ thermal events until failure, whereas the FEM model has event numbers in the order $10^4 - 10^5$. This makes rate features unfeasible for the RFM model (Table 2).

3. Results and discussion

3.1. Macroscopic RFM model behaviour

Fig. 2 shows average time-voltage curves, for a relative applied load $I/I_p = 0.1$ and temperature $T = 0.1$, and for varying Weibull shape parameters k, in hierarchical and non-hierarchical systems of size $L = 128$. In a typical creep strain vs time curve, one can identify three regimes: a strain hardening regime at small t/t_f during which the strain rate decreases, a stationary regime with almost constant strain rate for intermediate t/t_f, and strain softening (accelerating creep rate) towards the end of the sample lifetime t_f. Fig. 2 confirms this picture, with a deviation in the case of very low disorder (high k) where the initial strain hardening part becomes less and less prominent. A more complete picture of the influence of simulation parameters on the behavior of the system is presented in the Appendix (Figs. S1–S5), where results for different temperatures and applied loads are compiled.

Table 2
Overview over feature sets used for the RFM data.

	FEM features
Set 1	time, avalanche size, min. line damage, max. line damage
Set 2	time, avalanche rate, average avalanche amplitude total damage
	RFM features
Set 1	time, voltage, max. line damage vertical max. line damage horizontal, total damage
Set 2	time, voltage, total damage, max. crack size

3.2. Machine learning

To measure the performance of machine learning algorithms, usually statistical signatures like the coefficient of determination

$$R_X^2(t) = 1 - \frac{\sum_i \left(t_{X,i} - t_{a,i}\right)^2}{\sum_i \left(\langle t_{a,i} \rangle - t_{a,i}\right)^2} \tag{8}$$

are used. Here the subscript X = ML refers to a machine learning prediction and X=BA to a baseline prediction. $t_{X,i}(t)$ is the predicted lifetime and $t_{a,i}$ the actual lifetime. The average $\langle .. \rangle$ is, in the following, always evaluated over the entire data set.

For failure time prediction it is important to know how the prediction quality evolves over time. Thus, we define time dependent prediction error and score:

$$e_X(t) = \frac{|t_X(t) - t_a(t)|}{t_f} \tag{9}$$

$$s_{ML} = 1 - \frac{e_{ML}}{e_{BA}} \tag{10}$$

were t is the time at which the prediction is made, and times are normalized by the sample lifetime as t_f. The ML score should ideally be one or at least above zero which indicates superiority of the forest regression compared to the baseline estimate. These quantities are collected for each individual sample and sorted into bins according to the value of $\frac{t}{t_f}$. For binning we use equi-spaced partitions which divide the unit interval into 100 bins for the FEM and 20 bins for the RFM data, thus accounting for the fact that the RFM produce fewer data points in a simulation.

3.2.1. Baseline evaluation

For the FEM data (Fig. 3) in the limit of high disorder, the mean lifetime performs very badly as a prediction of the lifetime of individual samples, irrespective whether the mean is taken over the total or the surviving population. The reason is the extremely high scatter of the lifetime distribution which exhibits a coefficient of variation much larger than one. The median lifetime, which is less dominated by extremely long-lived samples, performs better but still badly (Fig. 3, top left).

When considering more ordered systems, the difference between median-based baselines and mean-based baselines vanishes especially at the start of the simulation up roughly half of the lifetime. This is expected as with decreasing disorder the lifetime distributions become less skewed, thus median and mean start to coincide. When close to failure, paradoxically, the dynamic baselines that consider only the surviving population perform worse than the static baselines that consider the whole initial population. This is the case because the surviving population exhibits a stronger outlier sensitivity: Its mean and median are increasingly dominated by the fittest samples which may show atypically high lifetimes.

To judge baseline performance across parameters, a box whisker plot (Fig. 4) of R^2 is used. Each data point in the plot represents the baseline performance for one combination of parameters (e. g. $k = 1$, $T = 0.01$, etc.). All baseline medians and averages (green line and dot) are above $R^2 = 0$. The worst case performance assessed by the farthest outlier in the negative domain is for all four baselines of the same magnitude. The dynamically updated median has the highest average (green dot) and highest median performance (green line). The spread in performance between the dynamic baselines is similar, whereas the static baselines show a larger spread in performance (box length). To explain the negative R^2 values in Fig. 4, we show in Figs. S6 and S7 the empirical complementary cumulative lifetime distribution (ccdf) of normalized lifetimes $t_N = \frac{t_f - \langle t_f \rangle}{\sigma(t_f)}$ where $\sigma(t_f)$ is the standard deviation. The distributions show a spread between 5 and 20 times the standard deviation and
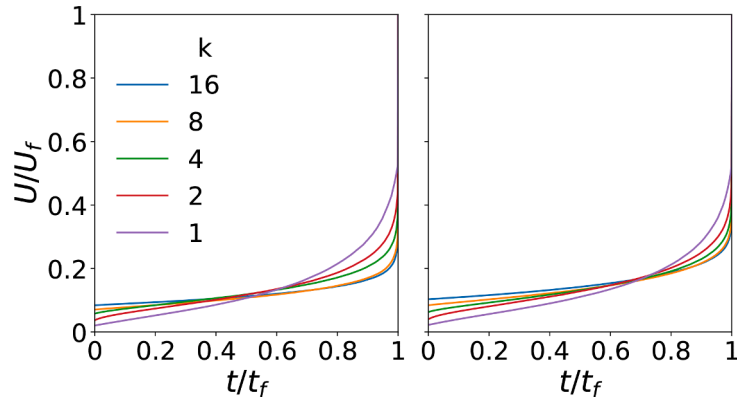
Fig. 2. Normalized average time-voltage curve for $L = 128, \frac{I}{I_p} = 0.1$ and $T = 0.1$. Data of nonhierarchical networks are on the right and of hierarchical networks on the left.
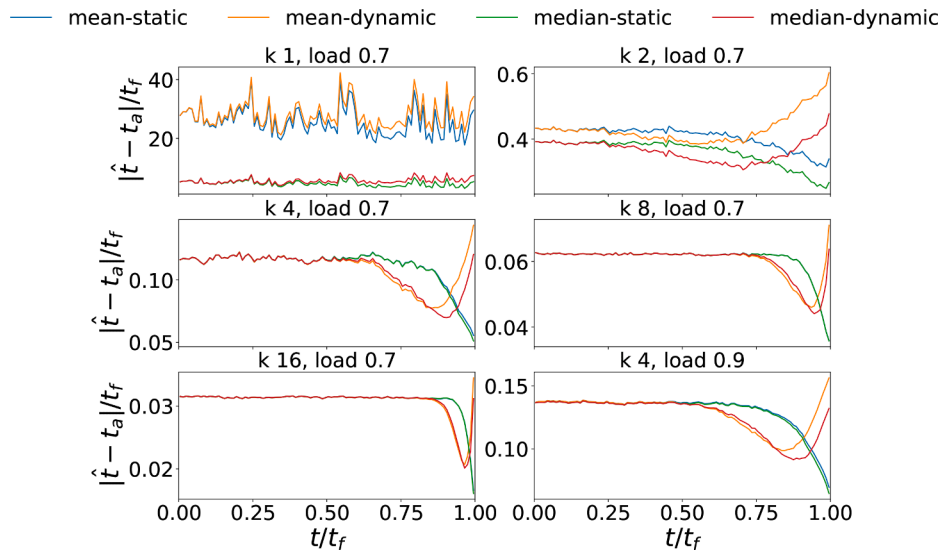

Fig. 3. Prediction error of different baselines, as function of time-to-failure in the FEM simulated data. The dynamic baselines calculate the mean or median lifetimes from the sub-population of the training set that still survives at time t, whereas static baselines consider the entire population.

with lower temperature the appearance of a tail. Few long living samples cause large errors as (i) they heavily distort the baseline predictions for the mean baseline – thus, shorter living samples are badly predicted; (ii) they are more or less ignored by the median baselines, thus give bad predictions for the "Methusalem" samples in the test set. In addition to these points, we note that it is actually surprising that these very simple baselines show reasonable performance as they incorporate neither domain knowledge nor a sophisticated statistical approach.

The static median is the best performing baseline for FEM, thus we choose it as baseline there. For RFM, the dynamic median is chosen as it has the highest average and median performance, although it is very similar to the dynamic mean.

3.2.2. Performance

When considering the score-time curve for the FEM data in Fig. 5, the new feature set 2 (full lines) performs slightly better than the feature set used by Biswas et al. [23], but for the highly disordered samples, both feature sets fail to outperform the baseline until just before the failure event. Even then, the mean error is still of the order of five times the mean residual lifetime, thus the prediction cannot be called successful. For less disordered samples, the ML algorithm outperforms the baseline across the entire specimen lifetime. The time series parameter N does not affect the performance while the shear zone width shows a small

peak performance at a width of four elements, but not a game changing improvement (Fig. 6). As established in a previous work [23], different applied loads also do not strongly affect the quality of the ML model predictions (Fig. 5), even though the absolute lifetimes may change by many orders of magnitude. Higher disorder generally leads to a decrease in prediction performance.

Random forests trained on RFM data can outperform the corresponding dynamic median baseline, but it depends strongly on the system parameters to which extent this is the case. This can be seen when comparing Figs. 7 and 8. Small systems show a less stable trend as these systems produce only few thermally activated events before failure and thus deliver less reliable statistics for the plot, but also make the task harder for the random forest. Similar problems are encountered in ordered systems where, owing to the small scatter in strength, load redistribution produces large avalanches, resulting again in a small number of large events (Fig. S5) and poor statistics.

To gain more systematic insight on how ML performance depends on the simulation parameters, we calculate the average rank correlation as illustrated in the following for the system size: We iterate over all combinations of k, T and $\frac{I}{I_p}$, calculate for each combination the rank correlation between system size L and R^2, and take the average over all combinations. The results are compiled in Table 3.

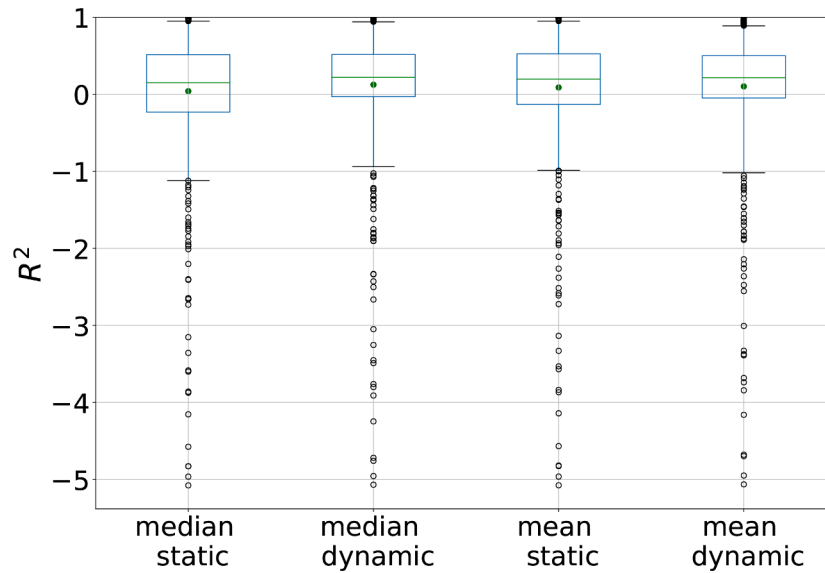Increasing temperature and system size, and reducing material

Fig. 4. Box and whisker plot of R^2 for RFM data with respect to different baseline choices. The data points correspond to different sets of simulation parameters, the green line indicates the median R^2, the green dot the average, the box bounds are located at the 25th and 75th percentile respectively and the whisker at the 5th and 95th percentile. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).
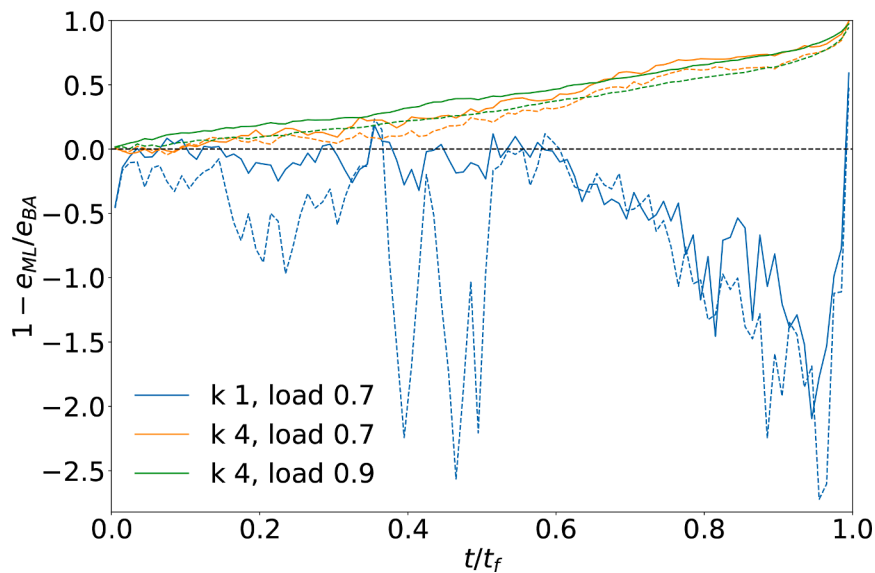

Fig. 5. Prediction score for the FEM data with the dotted lines representing feature set 1 already used in a previous publication [23] and the full lines the new set 2. The baseline is the static median.

disorder, lead to increased prediction performance while increasing the load (applied current) does the opposite. For disorder and temperature the correlation depends strongly on network architecture. The trend with regards to temperature is most pronounced for the nonhierarchical networks. This can be understood as follows: In the low-temperature limit, the crack path is identical to the quasistatic extremal fracture path which is controlled by the intrinsic fluctuations in material strength and local network geometry. Thus, the prediction problem is tantamount to identifying the extremal path and assessing its strength – an inherently difficult problem since the algorithm relies on the activity record, but has no direct 'structural' information about local strength or network morphology. Increasing temperature leads to increased activity at other sites of the network and to more widespread damage, whereas 'quenched' structural features such as strength or morphology variations become less important in determining the fracture scenario. As predictions use the activity record, this leads to better predictions at higher

temperature. In this context it is important to note that the hierarchical network architecture reduces the importance of stress concentrations and facilitates activity spread across the system [26]; because activity is diffusely spread even at zero temperature, the effect of temperature on predictability is reduced in hierarchical systems. High currents and small systems lead to quick failure, thus to shorter activity records which provide the prediction algorithm with less information to work with. The effect of disorder is ambiguous as high disorder leads to an increasing range of possible fracture paths, thus a larger feature space, but also to more activity as a source of information, while low disorder can lead to fracture after few thermally activated events, creating little information about fracture precursors that might be used to improve predictions above the baseline.

3.2.3. Feature selection

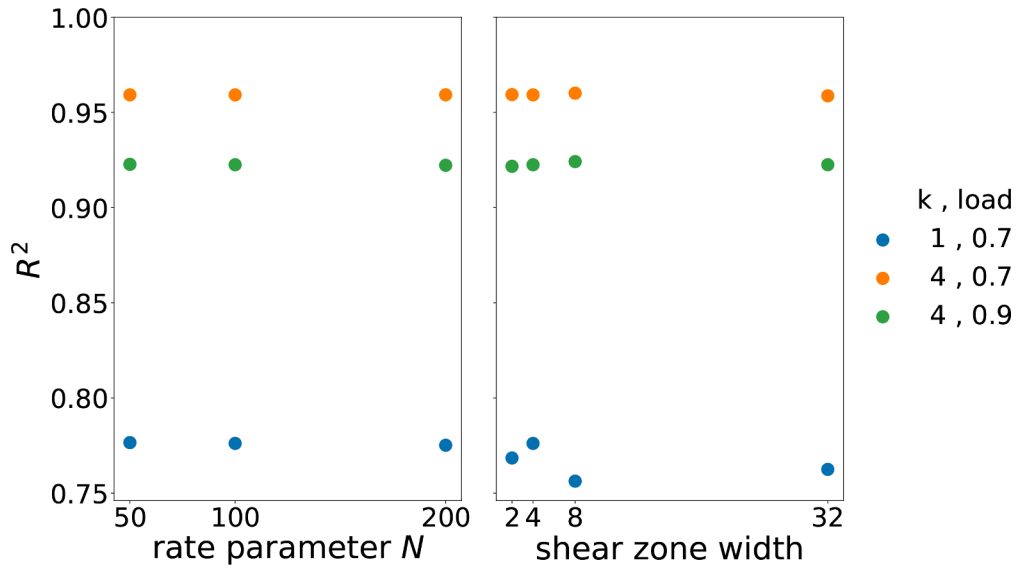Regarding feature selection, two main questions should be answered:

Fig. 6. Hyperparameters of feature set 2 for FEM data which show no significant impact on the prediction performance as measured by R^2. The rate parameter N is used to calculate event rates according to Eq. (7) in Section 2.3.
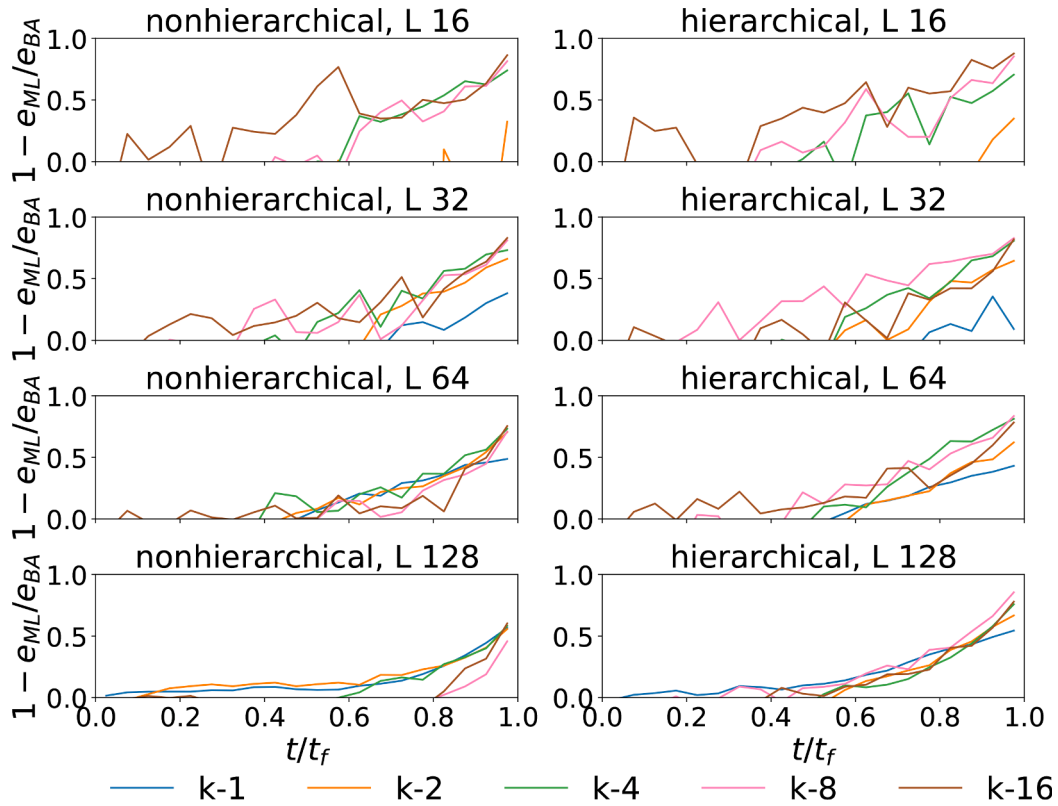

Fig. 7. Prediction score calculated for system with $T = 0.03$ and $\frac{L}{l_p} = 0.7$ for different values of k. The ML models were trained on feature set 1. The baseline for comparison is the dynamic median.

What are the most important features and how focused is the random forest on the top features? If the model attention is spread equally among all features, then no reduction in dimensionality is possible. For each feature set, we count how often a specific feature is deemed the most important according to the normalized average error reduction criterion (Fig. 9). To measure the attention spread of the model, we calculate the Gini Impurity [40]

$$1 - \sum_{i=1}^{N_{\text{feat}}} p_i^2 \tag{11}$$

from the normalized average error reduction (aer) p_i ($\sum_{i=1}^{N_{\text{feat}}} p_i = 1$) of feature i. N_{feat} is the number of features of the model. Close to zero Gini impurity indicates a model focused on a single feature.

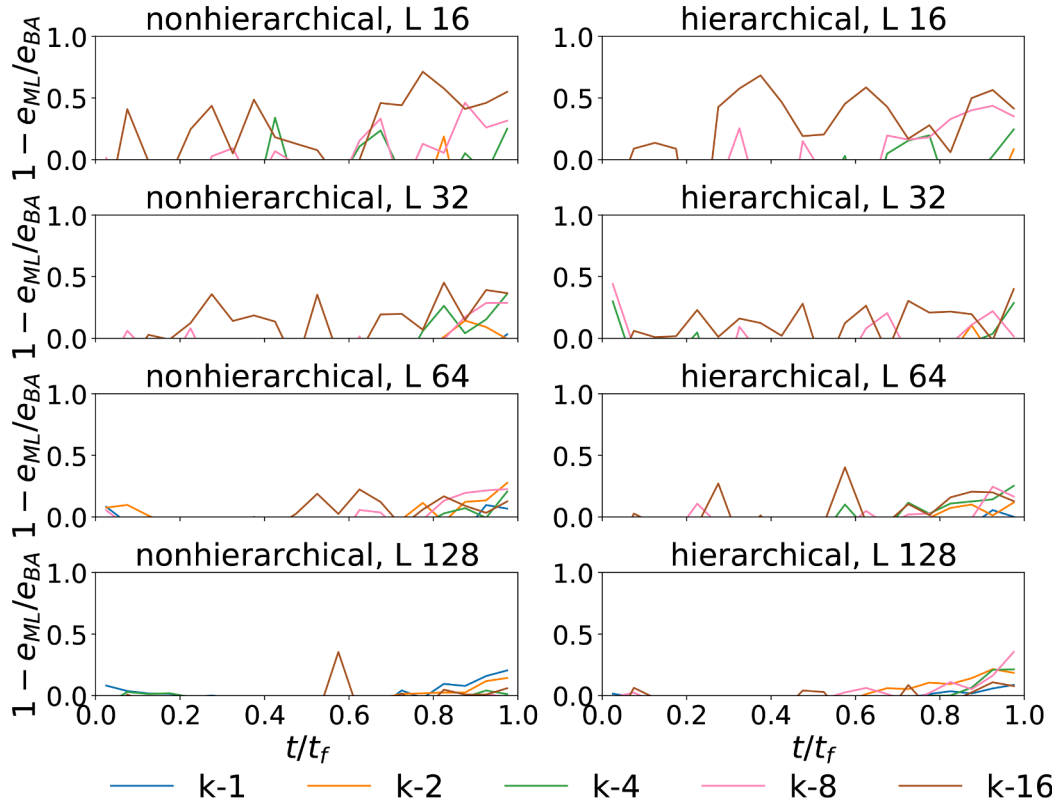Fig. 9 shows that damage based features dominate the partitioning

Fig. 8. Prediction score calculated for system with $T = 0.03$ and $\frac{I}{I_p} = 0.9$. The ML models were trained on feature set 2. The baseline for comparison is the dynamic median.

Table 3

Average rank correlation between the prediction performance R^2 and the simulation parameters for RFM data. Correlations obtained from random forests trained on feature set 1 are shown in the upper part of the table, those for feature set 2 on the lower part.

parameter	nonh. kendall	nonh. spearman	hier. kendall	hier. spearman
size	0.247	0.234	0.283	0.267
k	0.088	0.062	0.412	0.469
temperature	0.85	0.88	0.35	0.33
current	-0.683	-0.72	-0.567	-0.59
size	0.247	0.226	0.27	0.245
k	0.062	0.05	0.425	0.444
temperature	0.8	0.83	0.433	0.43
current	-0.667	-0.71	-0.6	-0.6

process for FEM data whereas models for predicting RFM behavior mainly use voltage (i.e., strain) as the source of their predictions. The importance of damage based features for the FEM model is easily rationalized, as this model shows shear banding behavior which leads to a large total damage as well as high damage concentration in the incipient shear band, a feature which can be easily detected since the direction of the incipient shear band is known from continuum mechanics (thus using domain knowledge). In the RFM, on the other hand, failure is controlled by cracks which comprise only a small number of all fuses; moreover, the cracks meander across the system in a self-affine manner (non-hierarchical systems) or via system-wide jumps (hierarchical systems), such that the crack direction undergoes random variations. This makes it difficult to identify the critical crack in terms of global damage or simple damage localization parameters.

For feature set 1 the avalanche amplitude is the least important feature regardless of investigated system and with a normalized aer ranging from zero to $p = 5.71 \times 10^{-6}$. In feature set 2 the event rate and

averaged avalanche amplitude are the least important features which never exceed an aer of $p = 1.29 \times 10^{-4}$. In fact, if the event rate is the least important feature, the second least important feature is always the averaged avalanche amplitude which never exceeds an aer of 2.56×10^{-2}. It is thus safe to say that, for our FEM data and feature sets, the avalanche amplitude and avalanche rate are not very useful for failure time prediction.

For the RFM data, models with high attention spread can occur where even the least important feature has an aer of $p \approx 0.17$. To investigate the consequences, we compare the model performance as measured by R^2 with the Gini impurity in Fig. 10. Every data point represents one combination of model parameters and the color represents the most important feature of the model trained on the specific combination. Attention focus and model performance correlate positively, although this correlation is less pronounced for hierarchical networks where in a number of cases, attention diversified models achieve comparable performance to focused models. When we calculate average rank correlations between the Gini impurity and the model parameters in the same ways as we have done previously for R^2 (Table 4), we can see that the numerical values of temperature and system size correlate positively with attention concentration, while higher current/load and higher disorder have the opposite effect. This trend is most stable for temperature with some deviations in the case of feature set 2 and hierarchical networks. While the trends for system size, current and disorder are not surprising, one might argue that higher temperature should cause a more stochastic/chaotic trajectory. However, this is only true on the micro scale of single elements but not on the system scale. To understand why, we observe that the RFMs considered here are disordered systems where stochastic time evolution has two different origins, namely (a) thermal fluctuations ('annealed disorder') and (b) structural disorder which is reflected by fluctuating thresholds/energy barriers and differing local network morphologies ('quenched disorder'). Quenched disorder leads to an unpredictable avalanche
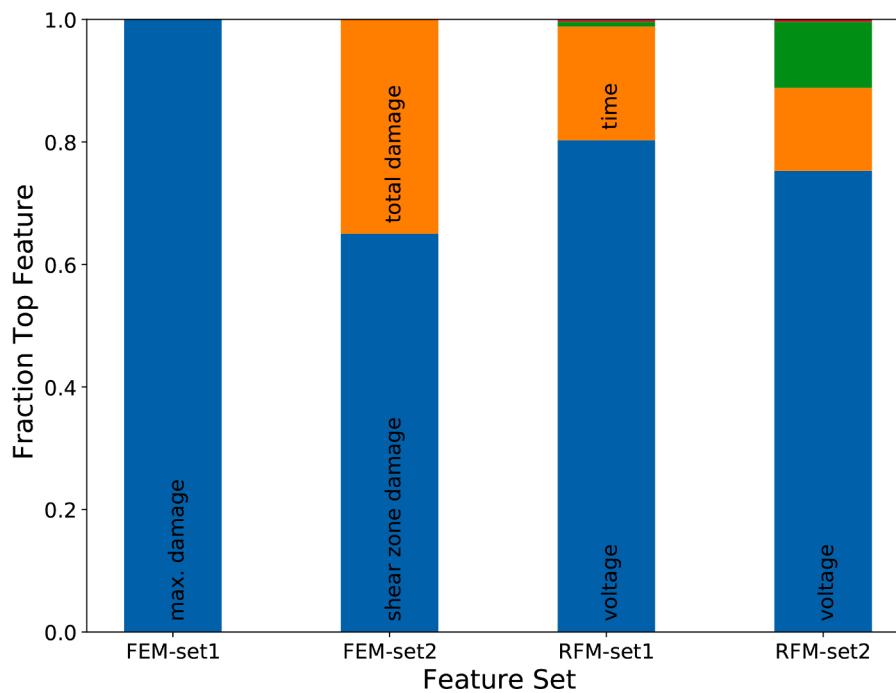
Fig. 9. Fraction of times a feature has been the most important one in a model trained on the specified feature set.

dynamics with scale-free statistical signatures even at zero temperature, and studies by other authors [3] show that ML methods cannot well predict this type of avalanche dynamics. On the other hand, at very high temperature the barrier fluctuations become asymptotically irrelevant (temperature 'irons out' the structural fluctuations) and as a consequence the material becomes indeed more predictable. This interpretation can be supported by considering the normalized lifetimes $t_N = \frac{t_f - \langle t_f \rangle}{\sigma(t_f)}$, where $\sigma(t_f)$ is the life time standard deviation. The empirical complementary cumulative lifetime distribution (Figs. S6 and S7) shows that higher temperature leads to a relatively more centralized lifetime distribution, whereas low temperature leads to a more pronounced tail. When comparing Tables 3 and 4, we notice that the different simulation parameters tend to have opposite-sign correlations with R^2 and Gini impurity, further supporting the connection between R^2 and ML model attention.

4. Conclusion

In this work, we determine the usefulness of a set of spatial and temporal features for predicting the time to failure in creep simulations by machine learning. The data for this study stem from a random fuse model (RFM) describing brittle, thermally activated statistical failure and an elasto-plastic finite element method (FEM) approach [7] which has been used in a previous study [23]. Fuse networks are created with two different morphologies (hierarchical vs nonhierarchical) as hierarchical architecture was shown to modify systems behavior in the approach to failure [26]. To measure performance, we compare the trained models with a simple statistical baseline. We make use of the ability of a special method of data science (random forest/decision trees) to measure the importance of a single feature for predictions via the average error reduction. Predictability of RFM systems is found to increase with increasing size, temperature and order, whereas increasing load/current has the opposite effect. Increasing size leads to smaller sample-by-sample variation, higher current to fewer events until failure thus "emitting" less information, while high disorder increases structural fluctuations leading to broader lifetime distributions. Increasing temperature centralizes the lifetime distribution by smoothing out threshold fluctuations, thus increasing predictability. The features

deemed most important by the random forest algorithm for prediction of FEM simulation data are the damage localized in a shear band and the total damage, whereas predictions for RFM data rely mainly on voltage. Note that voltage is the equivalent of strain in the RFM model and that damage in the FEM model is a monotonically increasing function of strain, hence, it is fair to say that in both cases the algorithms rely mostly on strain-like features. Avalanche-related features and event rate were found to be of little predictive power for the ML models in the case of FEM, whereas RFMs are too brittle – there are not enough events to calculate a sensible time dependent event rate. This suggests that future experimental studies - contrary to existing literature suggesting strain rate [2–4], event rate and amplitude [7–9] - should rely more strongly on strain-based features for predictions of the time to failure in a creep setting. In real life situations, strain patterns might be accessed nondestructively by surface monitoring, and the possibility of using such patterns for life time prediction was explored by Koivisto et al. [3] who considered the coefficient of spatial variation of strain (or incremental strain / strain rate) alongside spatial signatures of catastrophic failure accessible by thermography. These authors found that heterogeneities of strain correlate with sample lifetime even at early stages of creep life, and that this correlation becomes stronger after the strain-rate minimum is passed and the sample enters the tertiary creep stage. While such spatial features, considered in isolation, may be insufficient for reliable prediction [3], they may be an important source of information in ML based approaches. Spatial information is particularly useful in situations where (as in case of emergent shear bands in our FEM model, whose direction is for a given loading known from continuum mechanics) domain knowledge can be used to *a priori* design spatial features in such a manner that they capture the relevant deformation modes.

Data availability

The average, median and standard deviation of the peak current and voltage at peak current of the extremal simulations can be found here https://zenodo.org/record/7131727#.YzfJObTP1PY and the FEM raw data can be found here https://zenodo.org/record/7131771#.Yzgh RbTP1PY.
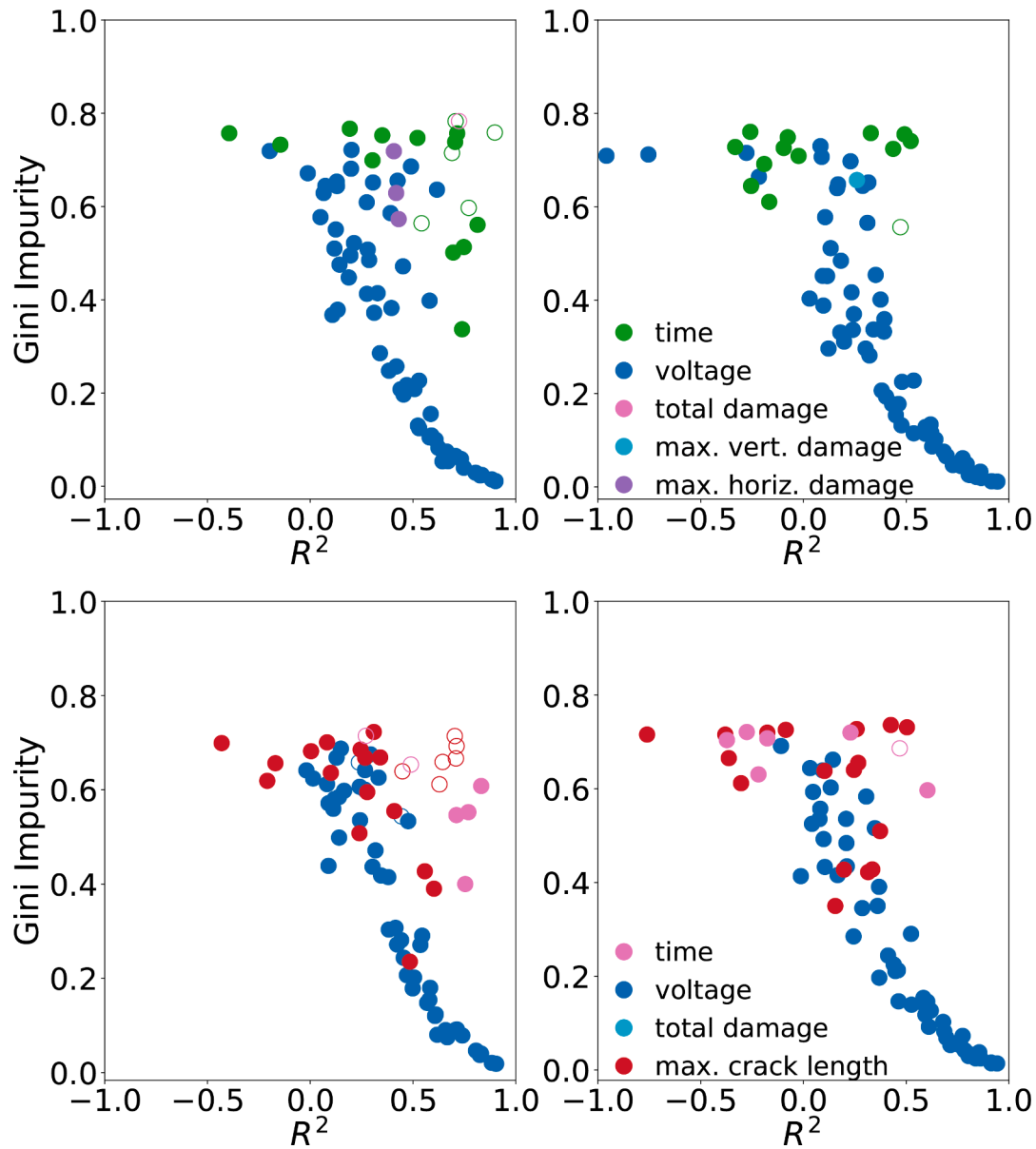
Fig. 10. Gini Impurity as a measure of attention spread versus model performance of the RFM data with hierarchical network data on the left and nonhierarchical on the right. The top row stem from ML models trained on feature set 1 and the bottom row on feature set 2. The color shows the top feature for the specific data point. Hollow dots indicate that for this data point the static median baseline's R^2 is higher than the random forest model.

Table 4
Average rank correlation between different simulation parameters and the Gini impurity of the ML model, for RFM data. Correlations obtained from random forests trained on feature set 1 are shown in the upper part of the table, those for feature set 2 on the lower part.

parameter	nonh. kendall	nonh. spearman	hier. kendall	hier. speaman
size	-0.53	-0.536	-0.423	-0.419
k	-0.338	-0.375	-0.587	-0.706
temperature	-0.95	-0.97	-0.9	-0.94
current	0.75	0.74	0.45	0.44
size	-0.507	-0.519	-0.52	-0.524
k	-0.55	-0.625	-0.488	-0.569
temperature	-0.967	-0.98	-0.717	-0.74
current	0.667	0.67	0.583	0.62

Code availability

The codes used in this paper are available at https://simlab.ww.uni-erlangen.de/publications/predicting-creep-failure-by-machine-learning/.

CRediT authorship contribution statement

Stefan Hiemer: Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. **Paolo Moretti:** Software, Visualization, Writing – original draft, Writing – review & editing. **Stefano Zapperi:** Visualization, Writing – original draft, Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Michael Zaiser:** Visualization, Writing – original draft, Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.finmec.2022.100141

References

[1] E.N.D.C. Andrade, On the viscous flow in metals, and allied phenomena, Proc. R. Soc. Lond. Ser. A 84 (567) (1910) 1–12.

[2] S.-W. Hao, B.-J. Zhang, J.-F. Tian, D. Elsworth, Predicting time-to-failure in rock extrapolated from secondary creep, J. Geophys. Res. Solid Earth 119 (3) (2014) 1942–1953.

[3] J. Koivisto, M. Ovaska, A. Miksic, L. Laurson, M.J. Alava, Predicting sample lifetimes in creep fracture of heterogeneous materials, Phys. Rev. E 94 (2) (2016) 023002.

[4] H. Nechad, A. Helmstetter, R. El Guerjouma, D. Sornette, Creep ruptures in heterogeneous materials, Phys. Rev. Lett. 94 (4) (2005) 045501.

[5] D. Sornette, Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes, and human birth, Proc. Natl. Acad. Sci. 99 (suppl_1) (2002) 2522–2529.

[6] A. Johansen, D. Sornette, Finite-time singularity in the dynamics of the world population, economic and financial indices, Physica A 294 (3–4) (2001) 465–502.

[7] D.F. Castellanos, M. Zaiser, Avalanche behavior in creep failure of disordered materials, Phys. Rev. Lett. 121 (12) (2018) 125501.

[8] S. Lennartz-Sassinek, I. Main, M. Zaiser, C. Graham, Acceleration and localization of subcritical crack growth in a natural composite material, Phys. Rev. E 90 (5) (2014) 052401.

[9] A. Saichev, D. Sornette, Andrade, omori, and time-to-failure laws from thermal noise in material rupture, Phys. Rev. E 71 (1) (2005) 016608.

[10] K. Frydrych, K. Karimi, M. Pecelerowicz, R. Alvarez, F.J. Dominguez-Gutiérrez, F. Rovaris, S. Papanikolaou, Materials informatics for mechanical deformation: areview of applications and challenges, Materials 14 (19) (2021) 5764.

[11] K. Linka, M. Hillgärtner, K.P. Abdolazizi, R.C. Aydin, M. Itskov, C.J. Cyron, Constitutive artificial neural networks: a fast and general approach to predictive data-driven constitutive modeling by deep learning, J. Comput. Phys. 429 (2021) 110010.

[12] F. Masi, I. Stefanou, Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (TANN), Comput. Methods Appl. Mech. Eng. 398 (2022) 115190.

[13] J.R. Mianroodi, N. H Siboni, D. Raabe, Teaching solid mechanics to artificial intelligence–a fast solver for heterogeneous materials, npj Comput. Mater. 7 (1) (2021) 1–10.

[14] A. Pandey, R. Pokharel, Machine learning enabled surrogate crystal plasticity model for spatially resolved 3D orientation evolution under uniaxial tension, arXiv preprint arXiv:2005.00951(2020).

[15] M.C. Messner, M. Rhee, A. Arsenlis, N.R. Barton, A crystal plasticity model for slip in hexagonal close packed metals based on discrete dislocation simulations, Model. Simul. Mater. Sci. Eng. 25 (4) (2017) 044001.

[16] D. Shin, Y. Yamamoto, M.P. Brady, S. Lee, J.A. Haynes, Modern data analytics approach to predict creep of high-temperature alloys, Acta Mater. 168 (2019) 321–330.

[17] Y. Liu, J. Wu, Z. Wang, X.-G. Lu, M. Avdeev, S. Shi, C. Wang, T. Yu, Predicting creep rupture life of ni-based single crystal superalloys using divide-and-conquer approach based machine learning, Acta Mater. 195 (2020) 454–467.

[18] X.-C. Zhang, J.-G. Gong, F.-Z. Xuan, A physics-informed neural network for creep-fatigue life prediction of components at elevated temperatures, Eng. Fract. Mech. 258 (2021) 108130.

[19] H. Salmenjoki, M.J. Alava, L. Laurson, Machine learning plastic deformation of crystals, Nat. Commun. 9 (1) (2018) 1–7.

[20] Z. Yang, S. Papanikolaou, A.C. Reid, W.-k. Liao, A.N. Choudhary, C. Campbell, A. Agrawal, Learning to predict crystal plasticity at the nanoscale: deep residual networks and size effects in uniaxial compression discrete dislocation simulations, Sci. Rep. 10 (1) (2020) 1–14.

[21] S. Papanikolaou, Microstructural inelastic fingerprints and data-rich predictions of plasticity and damage in solids, Comput. Mech. 66 (1) (2020) 141–154.

[22] F. Font-Clos, M. Zanchi, S. Hiemer, S. Bonfanti, R. Guerra, M. Zaiser, S. Zapperi, Predicting the failure of two-dimensional silica glasses, Nat. Commun. 13 (1) (2022) 1–11.

[23] S. Biswas, D. Fernandez Castellanos, M. Zaiser, Prediction of creep failure time using machine learning, Sci. Rep. 10 (1) (2020) 1–11.

[24] B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C.J. Humphreys, P.A. Johnson, Machine learning predicts laboratory earthquakes, Geophys. Res. Lett. 44 (18) (2017) 9276–9282.

[25] P.A. Johnson, B. Rouet-Leduc, L.J. Pyrak-Nolte, G.C. Beroza, C.J. Marone, C. Hulbert, A. Howard, P. Singer, D. Gordeev, D. Karaflos, et al., Laboratory earthquake forecasting: a machine learning competition, Proc. Natl. Acad. Sci. 118 (5) (2021).e2011362118

[26] P. Moretti, B. Dietemann, N. Esfandiary, M. Zaiser, Avalanche precursors of failure in hierarchical fuse networks, Sci. Rep. 8 (1) (2018) 1–7.

[27] L. de Arcangelis, S. Redner, H.J. Herrmann, A random fuse model for breaking processes, J. Phys. Lett. 46 (1985) L585.

[28] S. Zapperi, P. Ray, H. Stanley, A. Vespignani, Avalanches in breakdown and fracture processes, Phys. Rev. E 59 (1999) 5049–5057.

[29] M.J. Alava, P.K. Nukala, S. Zapperi, Statistical models of fracture, Adv. Phys. 55 (3–4) (2006) 349–476.

[30] W. Weibull, A statistical distribution function of wide applicability, J. Appl. Mech. (1951).

[31] D.F. Castellanos, M. Zaiser, Statistical dynamics of early creep stages in disordered materials, Eur. Phys. J. B 92 (7) (2019) 1–12.

[32] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[33] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks. European Conference on Computer Vision, Springer, 2016, pp. 630–645.

[34] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[35] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907(2016).

[36] Z. Liu, K. Zhou, F. Yang, L. Li, R. Chen, X. Hu, Exact: scalable graph neural networks training via extreme activation compression. International Conference on Learning Representations, 2021.–

[37] R. Ramakrishnan, P.O. Dral, M. Rupp, O.A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Sci. Data 1 (1) (2014) 1–7.

[38] C.M. Bishop, N.M. Nasrabadi, Pattern Recognition and Machine Learning, Springer, 2006.

[39] G. Louppe, Understanding random forests: from theory to practice, arXiv preprint arXiv:1407.7502(2014).

[40] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[41] T.K. Ho, Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition vol. 1, IEEE, 1995, pp. 278–282.

[42] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 832–844.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[44] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., Scipy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods 17 (3) (2020) 261–272.