On Predicting Material Fracture from Persistent Homology: Or, Which Topological Features are Informative Covariates?

James Amarel
Los Alamos National Laboratory
Los Alamos, NM 87545
jlamarel@lanl.gov

Nicolas Hengartner
Los Alamos National Laboratory
Los Alamos, NM 87545
nickh@lanl.gov

Robyn Miller
Los Alamos National Laboratory
Los Alamos, NM 87545
robynm@lanl.gov

Benjamin Migliori
Los Alamos National Laboratory
Los Alamos, NM 87545
ben.migliori@lanl.gov

Daniel Hope
Los Alamos National Laboratory
Los Alamos, NM 87545
hope@lanl.gov

Emily Casleton
Los Alamos National Laboratory
Los Alamos, NM 87545
ecasleton@lanl.gov

Alexei Skurikhin
Los Alamos National Laboratory
Los Alamos, NM 87545
alexei@lanl.gov

Earl Lawrence
Los Alamos National Laboratory
Los Alamos, NM 87545
earl@lanl.gov

Gerd J Kunde
Los Alamos National Laboratory
Los Alamos, NM 87545
g.j.kunde@lanl.gov

Abstract

We apply topological data analysis to characterize the simulated evolution of cracks in heterogeneous materials. Using persistent homology, we derive covariates for survival analysis, enabling lifetime prediction within a generalized linear modeling framework. Zeroth-homology features alone reproduce the ensemble survival curves of distinct materials, revealing that coarse topological statistics retain predictive signal even when important geometric details are abstracted away. We further compare the predictive capability of neural networks trained directly on damage fields with those trained on persistence-homology-derived representations, finding that the latter achieve superior accuracy. Finally, we investigate patched persistent homology, which encodes local topological information by computing persistence within spatial subdomains. This localized variant bridges global and geometric perspectives, capturing the collective mechanisms that govern fracture and may eventually yield representations better suited to the design and evaluation of fracture emulators.

1st Conference on Topology, Algebra, and Geometry in Data Science (TAG-DS 2025).

1 Introduction

Fracture in structural materials arises from the collective dynamics of defects across multiple scales, and its sudden onset makes lifetime prediction a central scientific and engineering challenge. Anticipating failure is critical for ensuring safety and reliability, as well as for accelerating the design of more durable materials. We study mesoscopic descriptions of fracture provided by phase-field [1–3] simulations (Figure A1), which coarse grain detailed atomistic dynamics into continuous fields. Cracks are defined as regions where the damage variable approaches a critical threshold, and structural failure is said to occur when a crack connects two opposing boundaries of the material specimen. Extended structures are not manifest in the canonical pixel-based representation of damage; their existence is exhibited only through correlations among local field values. persistent homology (PH) addresses this limitation by describing the evolving topology of the damage field, explicitly exposing the collective organization responsible for fracture.

PH provides a description of evolving structures that captures how extended features emerge and disappear across scales, without reference to microscopic coordinates, making it a promising tool for studying material degradation and failure. Yet, because geometric information enters only indirectly through the filtration scale, this abstraction entails a loss of spatial detail, which raises a key question: how much predictive signal do such topological summaries retain for fracture? Our aim is to assess the extent to which persistent homology captures features relevant to lifetime prediction.

2 Related Work

persistent homology provides a mathematical framework for tracking the birth and death of k-dimensional holes across scales in a filtration of topological spaces, quantifying how connectivity and structure evolve with resolution. Over the past three decades, PH has developed from an algebraic construct in computational topology into a core tool for data analysis [4].

Within materials research, PH has proven effective for describing the evolution of defect networks and fracture pathways in heterogeneous systems. In polymers, [5] demonstrated that PH together with principal component analysis can successfully reproduce stress–strain curves across systems. In metallic glasses under shear, [6] showed that PH provides a means to quantify structural changes accompanying yielding. In fracture-network flow, Suzuki *et al.* [7] used persistence to detect connected flow channels and estimate their effective apertures directly from images, yielding permeability predictions that correlate closely with direct simulations. [8] used coarse-grained molecular dynamics with persistence to show that craze formation in glassy polymers arises from the percolation of nanovoids during yielding. Relative to the above works, we employ a comparatively large synthetic dataset containing thousands of simulated fracture trajectories on a uniform grid, enabling statistically robust evaluation of topological predictors across materials with diverse constitutive behavior.

A central challenge in applying PH to learning is that persistence diagrams are multisets without canonical coordinates [9]. To address this, numerous embedding techniques have been proposed, including persistence landscapes [10–12], persistence images [13, 14], and kernel methods [15, 16]. These vectorizations embed persistence diagrams into Euclidean or reproducing-kernel Hilbert spaces, enabling standard machine-learning models to operate on topological data. Functional representations such as the accumulated persistence function [17], and persistence intensity or density functions [18–20] integrate over the multiset of birth-death pairs to yield smooth, interpretable summaries. Topological approaches to time-dependent and sequential data have explored extensions of persistence and clustering summaries to capture temporal evolution, including crocker plots and stacks [21] that summarize the evolving topology of dynamic metric spaces [22], formigrams that generalize dendrograms for dynamic clustering [23], and latent-ODE frameworks that learn the dynamics of vectorized persistence diagrams over time [24]. In contrast, we interpret the decay of topological features through the lens of survival analysis, defining conditional survival and hazard curves over persistence thresholds that reflect topological lifetimes.

In machine learning approaches to fracture modeling, evaluation metrics have increasingly shifted from pointwise losses to measures that capture geometric and structural fidelity. Studies on crack and fracture segmentation in CT and X-ray imagery commonly report overlap and distance-based metrics such as the Dice coefficient [25], structural similarity index (SSIM) [26] and mean absolute percentage error (MAPE) [27] to quantify the morphological accuracy of predicted fracture patterns.

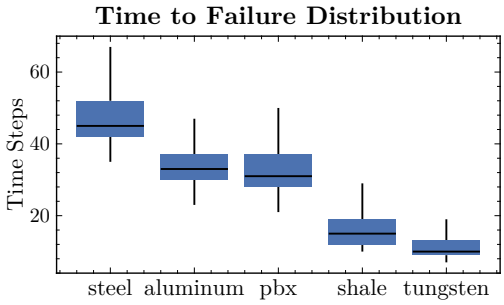


Figure 1: Lifetime distribution Tukey plot for each of the five materials under consideration.

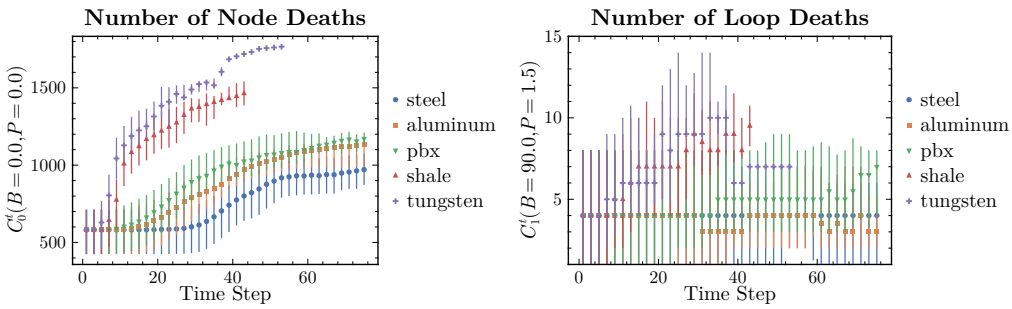
These metrics provide a more faithful assessment of physical and topological coherence than mean squared error (MSE), which penalizes local deviations but neglects spatial connectivity. Collectively, they underscore a broader trend toward topology-aware evaluation. Our work examines fracture metrics through the lens of persistence, quantifying changes in connectivity and structure directly from topological invariants.

3 Data

We use a subset of the `smartFRACs` dataset [28], specifically the first five batches for each material, comprised of $5,000$ two-dimensional fracture trajectories generated by a dynamic phase-field model for five representative materials: steel, aluminum, pbx, shale, and tungsten. Each trajectory records the temporal evolution of a scalar damage field, seeded from a stochastically generated initial condition, that monotonically accumulates from zero at pristine locations to unity in fully fractured regions. The stochastic initialization introduces substantial variation in crack geometry, leading to pronounced heterogeneity in both the complexity of fracture evolution and the resulting time-to-failure behavior. This variability, rooted in physically distinct crack-interaction pathways, creates a challenging testbed for evaluating the generalization capabilities of models across heterogeneous initial states and fracture morphologies. Methods capable of succintly capturing the relevant features of these extended structures remain to be developed, and how best to leverage persistent homology in systems with strong geometric and locality dependence remains an open methodological question.

The simulations follow a standard formulation that regularizes crack discontinuities over a finite length scale, coupling the damage field to the elastic and plastic response of the solid through strain-energy degradation. Materials include both brittle (shale, tungsten) and ductile (steel, aluminum, pbx) regimes (Figure 1), with elastoplastic response governed by a rate-independent von Mises yield criterion and isotropic hardening [28]. Each domain is discretized on a 128×128 Cartesian grid and subjected to tensile strain applied at a linearly ramping velocity on two opposing faces, identical across all simulations, with the remaining boundaries left free. Loading continues until a crack connects the two free faces, defining the failure event. For every material and loading configuration, the dataset provides the initial defect geometry, the full time-resolved damage evolution, and the corresponding failure time. This structured format enables both trajectory-level prediction and ensemble-level survival analysis across materials with contrasting constitutive behavior.

The resulting dataset provides a structured view of fracture evolution as a sequence of spatial fields with temporal dynamics. Each trajectory captures not only the onset and propagation of damage but also the formation, interaction, and coalescence of crack networks leading to system-wide failure. These evolving geometries exhibit rich topological signatures, such as the birth, merging, and annihilation of connected components, that are naturally suited for analysis via persistence. In the following section, we quantify these evolving structures using topological descriptors derived from the damage field, enabling a compact and interpretable representation of fracture dynamics across materials.



(a) Total number of zeroth homology birth-death pairs. (b) Total number of first homology birth-death pairs.

Figure 2: Number of feature events as a function of physical time; range bars capture the second and fourth quantiles, while markers represent the median.

3.1 Persistent Homology

We approach both persistent homology and material failure through the lens of survival statistics, a framework often applied to problems in biostatistics and reliability. Statistical structures, such as survival functions, hazards, and cumulative deaths, are employed to analyze both material failure rates and the lifetimes of topological features in persistence diagrams.

Let $\mathcal{D}_I$ be a persistence diagram with finite index set I, i.e., the multiset

$$\mathcal{D}_I = \{(b_i, p_i) : i \in I\}, \tag{1}$$

where each pair (b_i, p_i) corresponds to an event of birth time $b_i \in \mathbb{R} \geq 0$, death time $d_i \in \mathbb{R} > 0$, and persistence $p_i = d_i - b_i > 0$. For threshold pair (B, P) in the first quadrant of $\mathbb{R}^2$, the tail index

$$C(B, P) = \sum_{(b,p)\in\mathcal{D}_I} \mathbf{1}\{\, b \leq B \text{ and } p \geq P \,\} \tag{2}$$

counts the number of indices $i \in I$ born no later than B and of persistence at least P. The normalized empirical survival function, conditional on birth threshold B, is

$$S(P|B) = \frac{C(B, P)}{C(B, 0)}; \tag{3}$$

$S(P|B)$ is interpretable as the probability that a uniformly sampled feature with birth time $b \leq B$ has persistence at least P [29, 30]. This conditional formulation isolates persistence survival from birth abundance, rendering S a survival function in P for each fixed B. The cumulative hazard function

$$H(P|B) = -\log S(P|B) \tag{4}$$

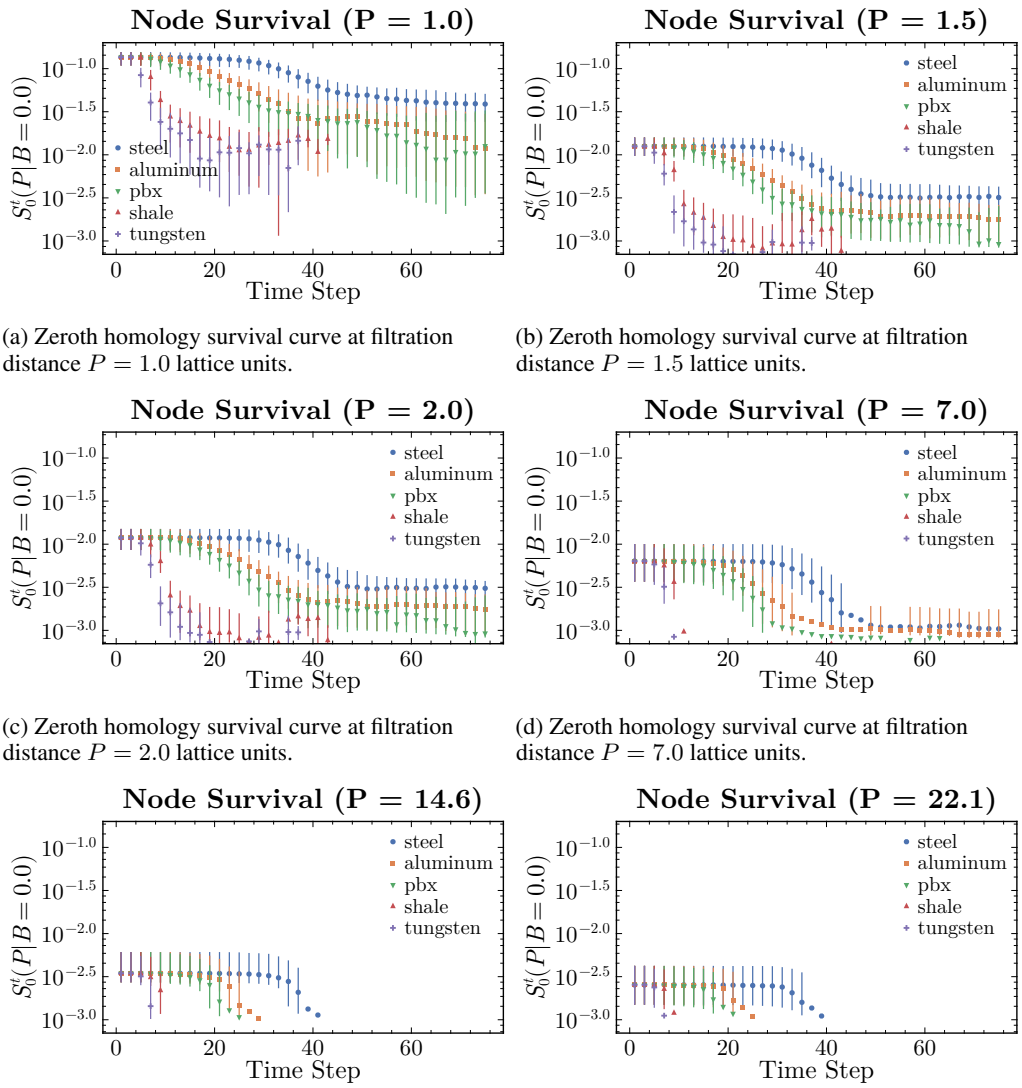
measures the total accumulated risk of disappearance up to persistence level P, for each fixed birth threshold B and the corresponding persistence hazard

$$h(P|B) = \partial_P H(P|B). \tag{5}$$

quantifies the instantaneous rate at which survival decays as persistence increases. The cumulative hazard and its derivative provide the standard survival-analysis interpretation of persistence lifetimes, conditional on birth time $b \leq B$.

Denoting by C_0^t and C_1^t the time-dependent tail indices for the zeroth and first homology, respectively, we track the total number of feature death events over time (Figure 2). Throughout, we work with natural units in which both the simulation time-step and unit-cell length are set to unity. The cumulative number of node deaths, corresponding to the integrated damage field, serves as a clear indicator of material susceptibility to failure [31]. In contrast, the evolution of loop deaths provides weaker separation across materials, likely reflecting differences in initial microstructural configuration rather than intrinsic fracture behavior.

Restricting attention to zeroth homology, we examine the conditional survival function S_0^t, evaluated at zero birth threshold without loss of generality. As shown in Figure 3, S_0^t decays monotonically with



(a) Zeroth homology survival curve at filtration distance $P = 1.0$ lattice units.

(b) Zeroth homology survival curve at filtration distance $P = 1.5$ lattice units.

(c) Zeroth homology survival curve at filtration distance $P = 2.0$ lattice units.

(d) Zeroth homology survival curve at filtration distance $P = 7.0$ lattice units.

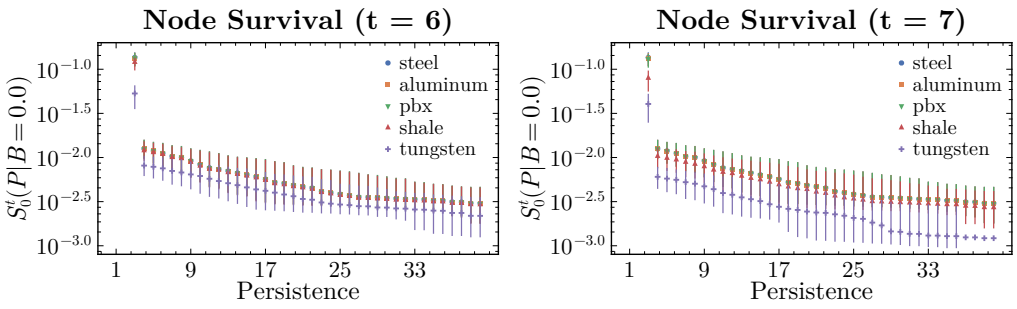
(e) Zeroth homology survival curve at filtration distance $P = 14.6$ lattice units.

(f) Zeroth homology survival curve at filtration distance $P = 22.1$ lattice units.

Figure 3: Survival curves for zeroth homology features across filtration levels P as a function of physical time; range bars capture the second and fourth quantiles, while markers represent the median. An animation is available upon reasonable request.

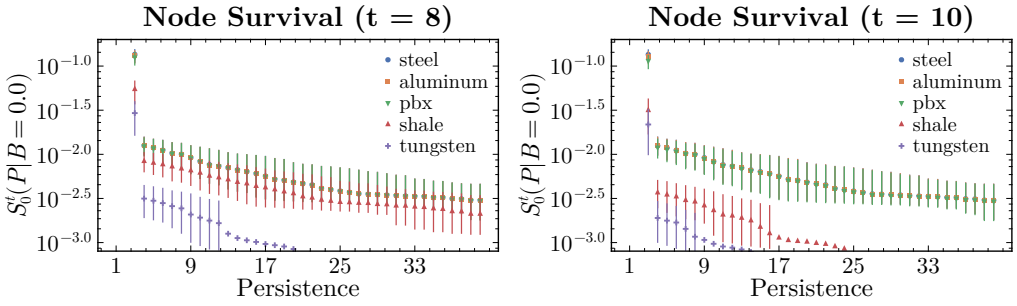
time, reflecting the progressive accumulation of damage under ramping tension. This representation cleanly separates brittle and ductile materials: brittle systems exhibit rapid early decay, while ductile specimen admit features that vanish relatively gradually. On increasing the persistence threshold P there develops pronounced collapse of the feature survival curves, indicating that highly persistent features are rare and that connectivity decays more abruptly in brittle materials. Overall, the odds of feature survival diminish jointly with time and scale, capturing the evolving connectivity of the damage field as fracture progresses.

Directly visualizing the filtration level dependence of the survival function provides a complementary perspective on the evolving topology of the damage field. As shown in Figure 4, materials that appear comparable under aggregate statistics, such as aluminum and pbx, whose time-to-failure and node-death distributions largely overlap, exhibit a modest separation of medians in this representation. Features persisting to large filtration thresholds P correspond to coherent structures of intermediate extent within the damage network, whose evolution reflects the formation and dissolution of medium-



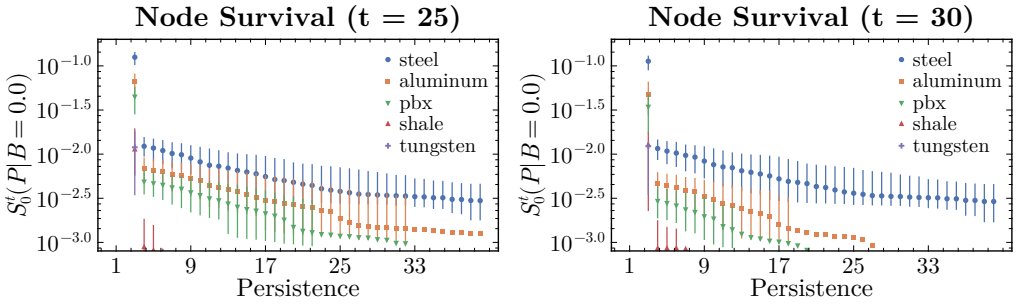
(a) Zeroth homology survival curve at time step 6.

(b) Zeroth homology survival curve at time step 7.



(c) Zeroth homology survival curve at time step 8.

(d) Zeroth homology survival curve at time step 10.



(e) Zeroth homology survival curve at time step 25.

(f) Zeroth homology survival curve at time step 30.

Figure 4: Survival curves for zeroth homology features across time t as a function of filtration level; range bars capture the second and fourth quantiles, while markers represent the median. An animation is available upon reasonable request.

range order as fracture progresses. The observed shift of the survival curves toward smaller P with increasing time reflects the dearth of such intermediate-scale structures, suggesting that the topology of the damage field becomes dominated by short-range connectivity near failure. The sensitivity of this medium-range organization to loading history highlights its potential as a topological indicator of material susceptibility to failure.

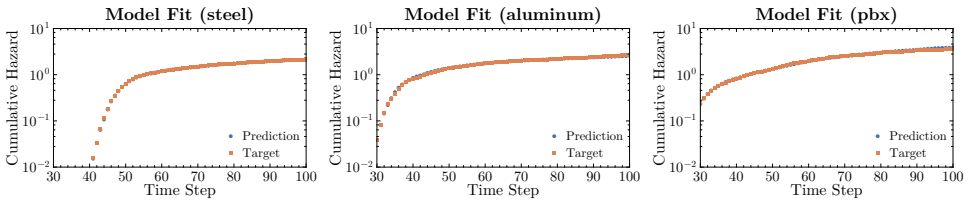
The persistence framework yields topological descriptors that summarize the structural organization of crack patterns and their variability across materials. We use these descriptors to investigate the link between damage field topology and specimen susceptibility to failure.

3.2 Survival Analysis

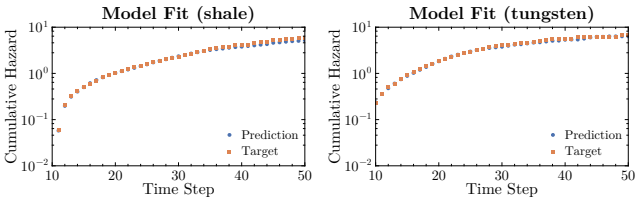
Each observed damage field trajectory is reduced to a time series of persistence diagrams

$$\mathcal{O}_n = \{\mathcal{D}_{I_n}^t : 1 \leq t \leq T_n\}, \tag{6}$$

where T_n denotes the failure time of observation n. From this sequence, we introduce a to be specified map Ψ that is responsible for selecting those topological features to be retained as covariates. In



(a) Steel cumulative hazard. (b) Aluminum cumulative hazard. (c) PBX cumulative hazard.



(d) Shale cumulative hazard. (e) Tungsten cumulative hazard.

Figure 5: Material failure cumulative hazard curve comparison between our generalized linear model (Prediction) prediction and empirical ground truth (Target).

principle, the only requirement on Ψ is that the resulting covariates encode physically admissible aspects of the progression toward fracture.

The material survival function S is determined by a causal hazard function λ, parameterized by a Cox-type log-linear model, i.e., the generalized linear model (GLM) [32, 33]

$$S^t[\Psi] = \exp\left[-\int_0^t d\tau\, \lambda^\tau[\Psi]\right], \tag{7}$$

with instantaneous hazard

$$\lambda^t[\Psi] = \exp\left\{\beta_0^t + \langle \beta_1, \Psi\rangle^t\right\}; \tag{8}$$

here, β_0^t specifies a time-dependent baseline hazard and β_1 is the collection of regression coefficients associated with our topological covariates, both of which are obtained via maximum-likelihood estimation under standard survival-model fitting procedures [34].

In selecting covariates Ψ, we seek a minimal model that permits direct interpretation of its coefficients. For simplicity, we include only data from the initial time: the zeroth-homology tail index at persistence threshold $P = 0$, which captures the integrated initial damage, together with the persistence hazards defined in Equation 5. We varied the maximum persistence level of the retained covariates and performed parameter inference on randomly selected 80% subsets of the dataset. The resulting holdout Brier scores, of order 0.125 across all materials, changed negligibly as hazard values at larger persistence thresholds were added (Figure A2). The mean time-to-failure obtained from survival analysis using initial-state persistence-homology features does not sharply resemble the empirical time-to-failure on an individual basis (Figure A3a). However, the ensemble survival curve is reproduced with high fidelity (Figure 5). This contrast between limited individual predictability and accurate ensemble calibration indicate that persistent homology may suffice only to capture the coarse statistical organization underlying macroscopic failure, in which geometric detail is effectively washed out through averaging over many heterogeneous cracked regions.

Our survival-analysis results suggest that persistence-based covariates extracted from the initial state capture ensemble-level failure statistics but do not sharply resolve individual failure times. To better probe per-example predictability, we now shift from semi-parametric survival models to data-driven predictors.

3.3 Machine Learning

We compare neural networks trained directly on damage fields with neural networks trained on persistence-homology-derived representations. This comparison isolates the role of representation, contrasting geometric with topological and global with localized features, in determining out-of-sample accuracy, calibration, and robustness across materials. It also highlights the difficulties

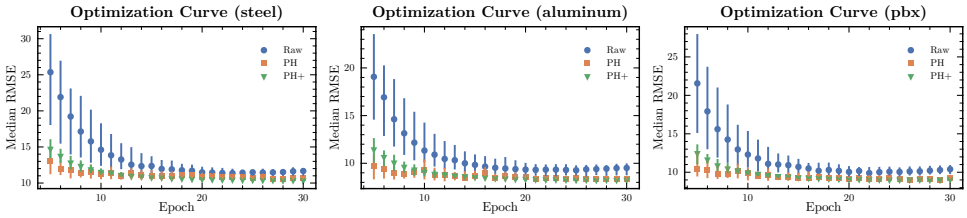
inherent in training neural networks on thresholded damage fields, whose sharp interfaces, sparsity, and binary character pose challenges for gradient-based optimization.

In addition to the raw-crack fields and zeroth homology tail index, we consider as candidate input features a patched variant of persistent homology (PH+), restricting attention to local neighborhoods of the damage field rather than deriving tail indices from all of space. Specifically, we evaluate PH on cracks within square patches extending twelve lattice units on each side and slide these patches across the domain with a stride of eight. Each patch yields a local persistence summary, and the collection of these summaries forms a $15 \times 15 \times N_P$ representation, where $N_P = 24$ corresponds to the chosen persistence threshold $P = 12$ sampled at half-integer intervals. This construction preserves coarse global structure while retaining sensitivity to local variations in crack connectivity. Overlapping spatial neighborhoods in Patched Persistence are intentionally introduced to ensure continuity of representation, permitting adjacent patches to share local topological structure. This redundancy is expected to promote smoother, translation-consistent feature fields that enhance convolution-based learning. Patch size and stride were empirically determined from survival-analysis feature-importance trends, indicating that the most informative persistence features emerge at intermediate scales.

Visualizations of PH+, averaged across failure configurations, show that this representation captures the phenomenological distinction between brittle and ductile behavior: brittle materials such as shale and tungsten are unable to redistribute stress over regions extending away from the central necking line and therefore fracture earlier. Beyond this physical correspondence, the PH+ representation is notable in that it embeds local topological structure into a spatially organized feature field, allowing patterns of connectivity and their evolution to be analyzed directly within the material domain. This fusion of topological abstraction with spatial locality suggests a promising direction for representing complex failure processes in forms amenable to learning and statistical modeling. Localizing the persistence computation in this way allows the representation to retain information about spatial heterogeneity that would otherwise be lost in a global summary. Crack evolution and coalescence are inherently local processes shaped by neighboring defects and stress concentrations. By preserving locality, patched persistent homology can, in principle, capture how mesoscale connectivity patterns contribute to failure initiation, providing a bridge between microscopic crack geometry and macroscopic response.

We implement three comparatively sized architectures aligned with the structure of their respective inputs. The raw crack-field and PH$^+$ representations, both spatially organized, are modeled using convolutional networks (CNN), while the zeroth homology PH representation, an ordered sequence of persistence-derived covariates, is modeled with a fully connected network (MLP). The crack-field CNN comprises five convolutional layers with 16 feature channels, beginning with a broad receptive kernel (12×12) and stride 8, followed by successively smaller 5×5 and 3×3 filters. The PH$^+$ network has similar depth but operates over $\sim$ 24 input channels with 5×5 and 3×3 kernels suited to its 15×15 patch grid. The PH network consists of four dense layers with 70 hidden units each. All models employ batch normalization with swish activation [35] and produce a single scalar output corresponding to the normalized failure time. Training targets were the z-score shifted logarithm of time-to-failure, which helps with training stability in the presence of outlier failure times. To prevent skew from rare but extreme cases, we excluded trajectories with lifetimes outside the 5th-85th percentile range prior to training.

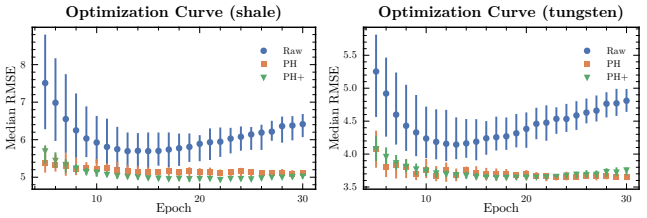
Neural networks trained across thirty-two random seeds using the Adam optimizer and data splits using a mean squared error objective reveal consistent underperformance of models trained directly on crack fields relative to their topological counterparts (Figure 6). Simple convolutional architectures ($\sim 2 \times 10^4$ parameters) trained on thresholded damage maps fail to match the predictive accuracy of models trained on persistence-homology representations. Both persistence-based representations, PH and PH$^+$, outperform the raw crack-field model. Notably, this comparison evaluates both an MLP and a CNN trained on PH-derived features against a CNN baseline using raw crack fields, providing a fair assessment that isolates representational differences from architectural ones. These findings suggest that while crack fields encode fine geometric detail, they constitute an inefficient representation of the physical mechanisms governing failure, whereas persistence-based features render manifest the relevant degrees of freedom. PH based neural networks also surpass the survival-model baseline (Figure A3b). All models were trained on an Ampere-based ARM Q80–30 node with an NVIDIA A100 40 GB GPU, totalling less than a day of wall time. We used the Julia library Lux.jl [36] for training and Makie.jl [37] for plotting.



(a) Validation performance over training epochs, steel.

(b) Validation performance over training epochs, aluminum.

(c) Validation performance over training epochs, pBX.



(d) Validation performance over training epochs, shale.

(e) Validation performance over training epochs, tungsten.

Figure 6: Model optimization and convergence behavior of the median root mean square (RMS) validation loss, evaluated over thirty-two independent training seeds for each material of interest. The curves depict the median validation loss across seeds, while range bars indicate one interquartile range. Raw crack field models yield inferior results to PH based training.

The present results remain preliminary: the models are small and have not undergone systematic hyperparameter optimization, architecture search, or large-scale data augmentation. Broader datasets and comprehensive tuning will be required for definitive assessment of representational performance. Furthermore, our models remain poor predictors on a per-example basis; the reported losses correspond to the optimization error computed after z-score normalization of the failure times.

4 Conclusion

This work advances topological perspectives on fracture prediction by analyzing a dataset spanning multiple materials and large geometric variability. We examine the difficulty of learning directly from raw geometric fields and motivate representations that encode the structural organization of fracture across heterogeneous systems. Within a survival-analysis formulation, persistence-based covariates derived from the initial configuration reproduce ensemble-level failure statistics with high fidelity, even when geometric detail is suppressed. Extending this framework to machine learning, we find that persistence-homology inputs facilitate the training of models that outperform networks trained directly on raw crack fields. We also introduce a locally patched variant of persistent homology that enriches this representation by embedding local topology within a spatially resolved feature field, yielding visual patterns consistent with phenomenological distinctions between brittle and ductile behavior.

Our results point to a broader role for topology-informed learning and evaluation in material modeling, capturing the essential structure of complex, multiscale processes without relying on pixel-wise representations.

5 Acknowledgements

Research presented in this report was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number(s) 20250637DI, 20250638DI, and 20250639DI. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. 89233218CNA000001. It is published under LA-UR-25-29771.

References

[1] J.-Y. Wu, V. P. Nguyen, C. T. Nguyen, D. Sutula, S. Sinaie, and S. P. Bordas (Elsevier, 2020) pp. 1–183.

[2] T. L. Anderson, *Fracture Mechanics: Fundamentals and Applications*, 4th ed. (CRC Press, 2017).

[3] C. Kuhn and R. Müller, Engineering Fracture Mechanics **77**, 3625 (2010), computational Mechanics in Fracture and Damage: A Special Issue in Honor of Prof. Gross.

[4] H. Edelsbrunner, D. Letscher, and A. Zomorodian, Discrete & Computational Geometry **28**, 511 (2002).

[5] R. Sato, S. Kawakami, H. Ejima, T. Ujii, K. Sato, T. Ichiki, and Y. Shibuta, Journal of Chemical Theory and Computation **20**, 10798 (2024).

[6] T. Shirai and T. Nakamura, Journal of the Physical Society of Japan **88**, 074801 (2019).

[7] A. Suzuki, M. Miyazawa, J. M. Minto, *et al.*, Scientific Reports **11**, 17948 (2021).

[8] T. Ichinomiya, I. Obayashi, and Y. Hiraoka, Physical Review E **95**, 012504 (2017).

[9] C. Hofer, R. Kwitt, M. Niethammer, and A. Uhl, Deep learning with topological signatures (2018), arXiv:1707.04041 [cs.CV] .

[10] P. Bubenik, J. Mach. Learn. Res. **16**, 77–102 (2015).

[11] M. Flammer and K. Hüper, Spatiotemporal persistence landscapes (2024), arXiv:2412.11925 [math.AT] .

[12] O. Vipond, Multiparameter persistence landscapes (2018), arXiv:1812.09935 [math.AT] .

[13] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, Journal of Machine Learning Research **18**, 1 (2017).

[14] M. Carrière and A. Blumberg, in *Advances in Neural Information Processing Systems*, Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, Inc., 2020) pp. 22432–22444.

[15] G. Kusano, Y. Hiraoka, and K. Fukumizu, in *Proceedings of The 33rd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 48, edited by M. F. Balcan and K. Q. Weinberger (PMLR, New York, New York, USA, 2016) pp. 2004–2013.

[16] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt, A stable multi-scale kernel for topological machine learning (2014), arXiv:1412.6821 [stat.ML] .

[17] C. Biscio and J. Møller, The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications (2017), arXiv:1611.00630 [math.ST] .

[18] Y.-C. Chen, D. Wang, A. Rinaldo, and L. Wasserman, Statistical analysis of persistence intensity functions (2015), arXiv:1510.02502 [stat.ME] .

[19] W. Wu, J. Kim, and A. Rinaldo, On the estimation of persistence intensity functions and linear representations of persistence diagrams (2023), arXiv:2310.11982 [math.ST] .

[20] H. Edelsbrunner, A. Ivanov, and R. Karasev, Modelirovanie i Analiz Informatsionnykh Sistem **19**, 5 (2012).

[21] F. A. Khasawneh, E. Munch, D. Barnes, M. M. Chumley, I. Guzel, A. D. Myers, S. Tanweer, S. Tymochko, and M. Yesilli, Journal of Open Source Software **10**, 7243 (2025).

[22] L. Xian, H. Adams, C. M. Topaz, and L. Ziegelmeier, Foundations of Data Science **4**, 1 (2022).

[23] W. Kim and F. Mémoli, in *Canadian Conference on Computational Geometry* (2018).

[24] S. Zeng, F. Graf, M. Uray, S. Huber, and R. Kwitt, in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24 (Curran Associates Inc., Red Hook, NY, USA, 2024).

[25] E. Hamdi and E. Lejeune, Towards robust surrogate models: Benchmarking machine learning approaches to expediting phase field simulations of brittle fracture (2025), arXiv:2507.07237 [cs.LG] .

[26] S. Chen, S. Feng, Y. Huang, Z. Lei, X. Jia, Y. Lin, and E. Rougier, Computational Materials Science **236**, 112846 (2024).

[27] Y. Wang, D. Oyen, W. Guo, *et al.*, npj Materials Degradation **5**, 10.1038/s41529-021-00151-y (2021).

[28] A. Marcato, A. Pachalieva, R. G. Hill, K. Gao, X. Wang, E. Rougier, Z. Lei, V. Agrawal, J. Chua, Q. Kang, J. D. Hyman, A. Hunter, N. DeBardeleben, E. Lawrence, H. Viswanathan, D. O'Malley, and J. E. Santos, A foundation model for material fracture prediction (2025), arXiv:2507.23077 [cs.LG] .

[29] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2008).

[30] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd ed. (John Wiley & Sons, New York, 2002).

[31] I. Ando, Y. Mugita, K. Hirayama, S. Munetoh, M. Aramaki, F. Jiang, T. Tsuji, A. Takeuchi, M. Uesugi, and Y. Ozaki, Materials Science and Engineering: A **828**, 142112 (2021).

[32] D. R. Cox and D. Oakes, *Analysis of Survival Data* (Chapman and Hall, London, 1984).

[33] D. R. Cox, Journal of the Royal Statistical Society: Series B (Methodological) **34**, 187 (1972).

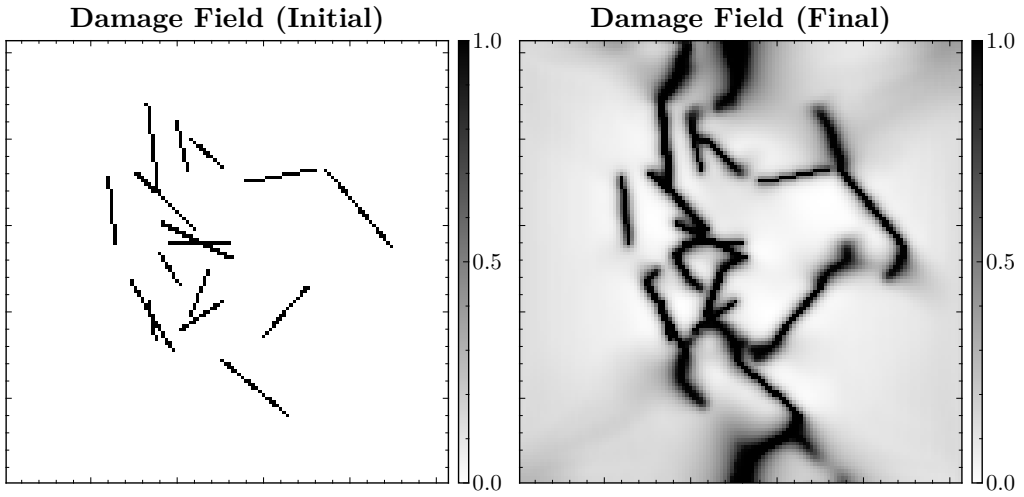
[34] D. Bates, A. Noack, S. Kornblith, M. Bouchet-Valat, M. K. Borregaard, A. Arslan, J. M. White, D. Kleinschmidt, P. Alday, G. Lynch, I. Dunning, P. K. Mogensen, S. Lendle, D. Aluthge, M. Dutta, pdeffebach, P. José Bayoán Santiago Calderón, A. Patnaik, B. Born, B. Setzler, C. DuBois, J. Quinn, O. Slámečka, P. Bastide, V. B. Shah, P. Anthony Blaom, and B. König, Juliastats/glm.jl: v1.9.0 (2023).

[35] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions (2017), arXiv:1710.05941 [cs.NE] .

[36] A. Pal, Lux: Explicit Parameterization of Deep Neural Networks in Julia (2023), if you use this software, please cite it as below.

[37] S. Danisch and J. Krumbiegel, Journal of Open Source Software **6**, 3349 (2021).

A Supplementary Figures



(a) A representative initial crack configuration.

(b) Crack configuration at the moment of failure.

Figure A1: Initial and final damage fields for Tungsten according to a phase-field simulation in which tension is applied horizontally.

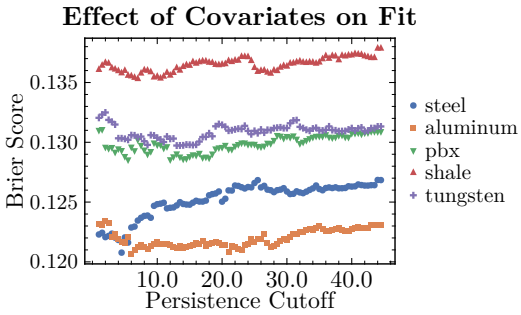
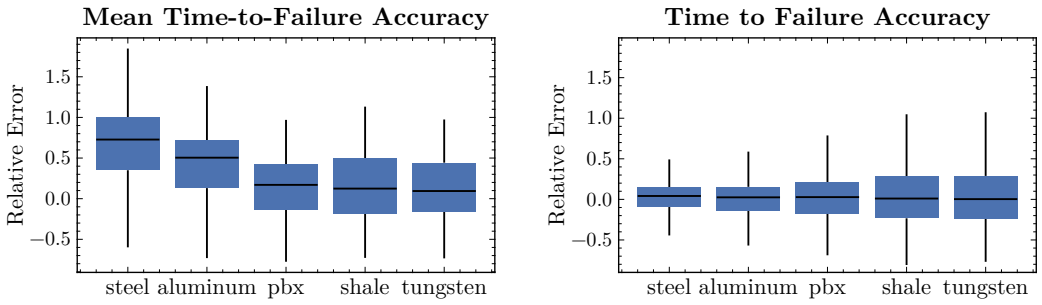


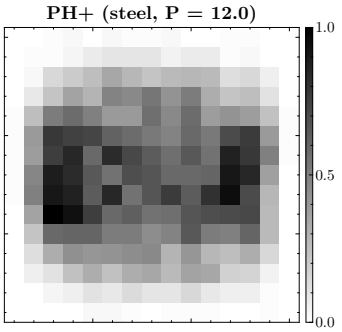
Figure A2: Visualization of survival model Brier scores as a function of selected covariates.



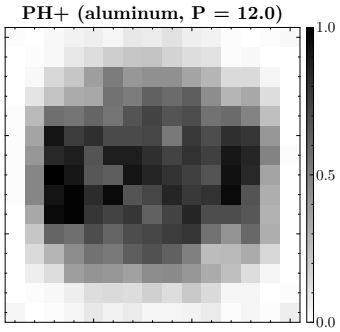
(a) Tukey plot of relative difference between GLM predicted mean-time-to-failure and empirical failure time.

(b) Tukey plot of relative difference between PH^+ predicted time-to-failure and empirical failure time.

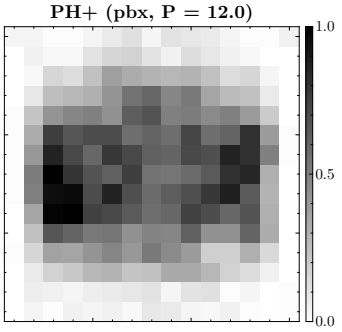
Figure A3: Comparison of time-to-failure prediction accuracy between GLM Figure A3a and PH^+ neural network Figure A3b.



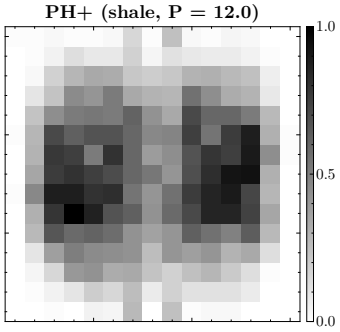
(a) Patched persistent homology of steel at failure for filtration distance $P = 12$ lattice units.



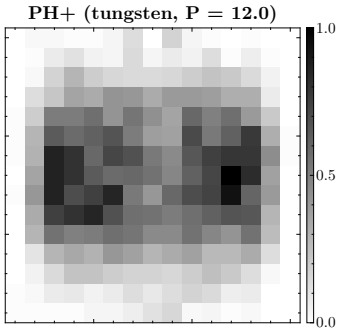
(b) Patched persistent homology of aluminum at failure for filtration distance $P = 12$ lattice units.



(c) Patched persistent homology of PBX at failure for filtration distance $P = 12$ lattice units.

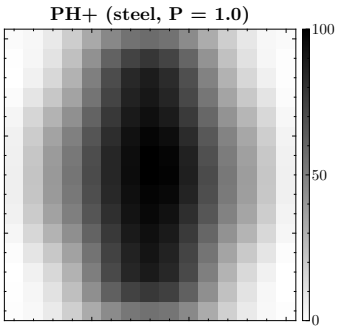


(d) Patched persistent homology of shale at failure for filtration distance $P = 12$ lattice units.

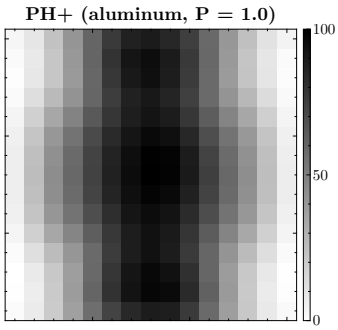


(e) Patched persistent homology of tungsten at failure for filtration distance $P = 12$ lattice units.

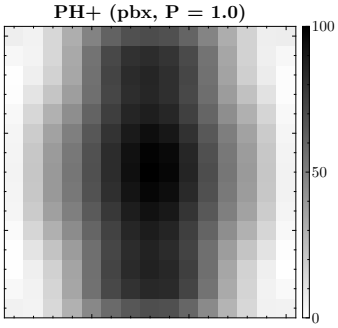
Figure A4: Averaged patched persistent homology across materials at failure.



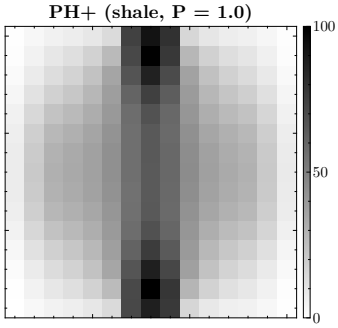
(a) Patched persistent homology of steel at failure for filtration distance $P = 12$ lattice units.



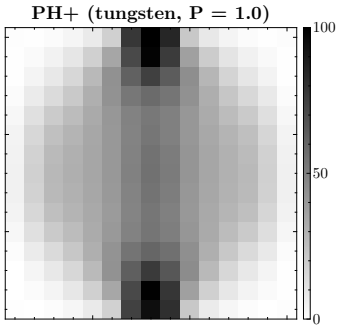
(b) Patched persistent homology of aluminum at failure for filtration distance $P = 12$ lattice units.



(c) Patched persistent homology of PBX at failure for filtration distance $P = 12$ lattice units.



(d) Patched persistent homology of shale at failure for filtration distance $P = 12$ lattice units.



(e) Patched persistent homology of tungsten at failure for filtration distance $P = 12$ lattice units.

Figure A5: Averaged patched persistent homology across materials at failure.