GEOMETRY-AWARE OOD GENERALIZATION FOR COMPOSITE MATERIALS

Abhiroop Bhattacharya
Department of Electrical Engineering
École de technologie supérieure
Montréal, Canada
abhiroop.bhattacharya.1@ens.etsmtl.ca

Hangwei Qian
CFAR and IHPC, Agency for Science
Technology and Research (A*STAR)
Singapore
qian_hangwei@a-star.edu.sg

Sylvain G. Cloutier
Department of Electrical Engineering
École de technologie supérieure
Montréal, Canada
SylvainG.Cloutier@etsmtl.ca

Ivor Tsang
CFAR and IHPC, Agency for Science
Technology and Research (A*STAR)
Singapore
ivor_tsang@a-star.edu.sg

ABSTRACT

Machine learning methods have recently shown promise in predicting composite material properties more efficiently than traditional finite element simulations. However, in real-world applications of composite materials, variations in filler characteristics or processing conditions can cause a significant mismatch between the training and test distributions, leading to degradation in model performance. Existing approaches to domain adaptation typically aim to learn feature representations that are invariant across distribution shifts or minimize worst-case risk across predefined environments, which are not suitable for composite materials where the distribution shifts are feature-driven and localized. In this work, we argue that such shifts are more naturally characterized by the geometry of the data manifold rather than Euclidean distances in the ambient feature space. We reinterpret OOD generalization under covariate shift as a problem of preserving geodesic relationships on the data manifold. Building on this perspective, we propose **Geometry Support Anchoring (GSA)**, a geometry-aware learning framework that anchors predictions using geodesic distances to data-dependent reference representations. For our approach, we have developed a multimodal conditional Joint Embedding Predictive Architecture (JEPA) that learns invariant, material-relevant representations by predicting latent targets using context from complementary modalities. By enforcing consistency in intrinsic geometry rather than raw features, our approach preserves physically meaningful variation while improving robustness to regime shifts. Experiments on both public and in-house composite material datasets with realistic distribution shifts demonstrate that our method consistently improves robustness and OOD generalization compared to state-of-the-art tabular learning and domain adaptation methods.

1 INTRODUCTION

Composite materials are widely used in construction, aerospace, and automotive industries, where precise control over mechanical and thermal properties is essential (Berry et al., 2026). Their performance derives from the deliberate introduction of heterogeneity, achieved by combining constituent materials with distinct properties (Chawla, 2012). By controlling the selection of filler constituents, processing conditions, and spatial organization, composite designs can realize substantial performance enhancements compared to monolithic materials (Ramesh et al., 2022). Traditionally, the development and optimization of such composite systems have relied on computational approaches such as finite element modeling (FEM), which are compute-intensive and difficult to scale (Nurhaniza et al., 2010). Over the past decade, machine learning models have achieved remarkable success

in analyzing composite systems (Mulenga et al., 2025). Despite their promise, the deployment of machine learning models for composite materials remains challenging due to the mismatch in distribution between the data on which the model was trained (in-domain data) and the deployment data (OOD data) (Karuppusamy et al., 2025).

A widely used approach to mitigate performance degradation caused by distribution shift is *transfer learning*. This technique leverages representations learned from a source task or domain to improve performance on a related target task, under the assumption that the underlying structure or mechanisms are shared between the two (Zhuang et al., 2020). Recently, researchers have used this technique to transfer knowledge across material systems, microstructures, and loading conditions (Nan et al., 2024; Zhou et al., 2025). However, in composite systems, the in-domain (ID) and OOD data often do not share the same underlying structure, leading to degradation in model performance.

More broadly, out-of-distribution (OOD) generalization has been widely studied, with approaches such as adaptive risk minimization (ARM) (Zhang et al., 2021) and unsupervised domain adaptation (Singhal et al., 2023), which aim to optimize worst-case risk or enforce feature invariance across domains. Domain generalization (DG) (Zhou et al., 2022) considers a related setting, where models are trained on multiple source domains and evaluated on unseen targets without access to target data. However, these methods rely on the assumption that covariate shift can be addressed through globally invariant representations. This assumption often fails in composite, feature-based systems. In such settings, OOD samples typically lie in low-density regions of the covariate space. Enforcing invariance can then be counterproductive, as it suppresses physically meaningful variations needed for accurate prediction.

To address these challenges, researchers have developed reweighting techniques to mitigate covariate shift (Sugiyama & Kawanabe, 2012; Sugiyama et al., 2007; Tsuboi et al., 2009). In contrast to classical covariate-shift reweighting methods that estimate explicit density ratios between training and test distributions, our approach **Geometry Support Anchoring (GSA)** is geometry-driven and operates directly in the learned representation space. Based on two classical assumptions in representation learning, i.e., *Cluster* and *Manifold* assumptions (Chapelle & Zien, 2005; Rigollet, 2007), we argue that OOD generalization is influenced by reachability within the learned representation, rather than by global invariance or pointwise density ratios. GSA exploits the intrinsic geometry of learned ID representations to relate OOD samples to supported regions of the data manifold. In raw input space, however, graph-based geodesics are often fragmented and dominated by irrelevant variation. We therefore leverage representations produced by a conditional joint embedding architecture (JEPA) to define distances along a smooth, semantically meaningful representation space. The key contributions of our work are

- We introduce **Geometry Support Anchoring (GSA)**, a geometry-aware anchoring framework that leverages intrinsic manifold structure to reason about OOD generalization in composite materials.
- We develop a conditional multimodal JEPA-based model which can be used to learn invariant representations for composite materials.
- We propose a theoretical upper bound for the ID-OOD risk gap.
- Our experiments demonstrate that the proposed geodesic anchoring improves robustness under different distribution shifts.

2 Related work

Early machine learning approaches for composite materials have focused on supervised prediction using tabular descriptors (Liang et al., 2025; Liu et al., 2020; Hashemi et al., 2021), with more recent work extending to multimodal models that incorporate microstructure images or textual metadata (Wu et al., 2025; Qian et al., 2025; Song et al., 2023; Saha et al., 2025). These methods primarily evaluate performance under in-domain settings and their performance deteriorates under covariate shift. Out-of-distribution (OOD) generalization under covariate shift has been extensively studied in machine learning (Zhou & Levine, 2021; Tripuraneni et al., 2021; Kimura & Hino, 2024). Several recent studies have examined the role of geometric structure in learned representations, using concepts such as manifold connectivity, geodesic distance, and curvature to analyze generalization behavior (Sáez de Ocáriz Borde et al., 2023; Bronstein et al., 2017; De Bortoli et al., 2022). However, these

geometric quantities are most often used for post-hoc analysis or diagnostics, rather than being directly incorporated into training objectives or sample weighting strategies. This work differs from prior approaches by using intrinsic, graph-based geodesic distances in a learned representation space to directly relate OOD samples to supported in-distribution regions and to guide geometry-aware reweighting under covariate shift, without assuming global invariance or explicit density estimation.

3 Proposed Method

3.1 Problem Formulation

We consider a dataset of composite materials, where the primary inputs $X \in \mathbb{X}$ denote tabular composite input features and the targets Y correspond to mechanical, thermal, or electrical properties. In addition to X, each sample may optionally include auxiliary modalities Z, such as text descriptions or microstructure images, yielding a multimodal input (X, Z). Let $(X, Y) \sim \mathbb{P}_{ID}$ denote the training distribution supported on a high-density subset $S_{ID} \subset \mathbb{X}$. The OOD split is defined with respect to the composite feature space X; auxiliary modalities Z are not used to construct the split.

At test time, we consider out-of-distribution samples drawn from $\mathbb{P}_{OOD}$, supported on a low-density subset $S_{OOD} \subset \mathbb{X}$. We assume that most OOD samples have a substantial mass in S_{OOD}, while allowing for partial overlap with the support of $\mathbb{P}_{ID}$. This study focuses on instances where the distribution shift is primarily a covariate shift, characterized by

$$P_{OOD}(x) \neq P_{ID}(x), \qquad P_{OOD}(y \mid x) = P_{ID}(y \mid x). \tag{1}$$

Under this assumption, OOD samples may fall outside the high-density support of the training data while still corresponding to valid semantic configurations governed by the same underlying relationship between the input and the output. While our analysis focuses on this setting, we additionally evaluate robustness under high-shift scenarios in which the conditional invariance in Eq. (1) does not hold.

3.2 Geometry Support Anchoring (GSA)

Euclidean distance in embedding space does not reflect physical similarity. Composite materials that are close in Euclidean space can belong to different physical regimes, making transfer unreliable. Conversely, materials that are far apart may still be physically related through smooth changes, such as ply reorientation or stiffness variation. This is shown in Fig. 1. This mismatch motivates learning representations that capture semantic similarity. Composite descriptors are heterogeneous and often violate smoothness assumptions, making raw feature space unreliable. We therefore adopt a representation learning approach that promotes invariance to noise and enforces local smoothness. To model this manifold, we use a multimodal conditional Joint Embedding Predictive Architecture (JEPA), described in Section 3.3. Further, we assume that the learned data manifold $\mathcal{M} \in \mathbb{R}^D$ is a compact, smooth manifold of known intrinsic dimension d. GSA exploits the insight that, for the distribution shift described in Eq. (1), the geodesic manifold effectively bridges the gap between $\mathbb{P}_{OOD}$ and $\mathbb{P}_{ID}$ by revealing that they share common support in the latent space. We assume that every two points in $\mathcal{M}$ are connected by a geodesic curve. Let $z = \phi(x) \in \mathbb{R}^D$ denote the embedding generated by the JEPA model. The in-distribution embeddings $\{z_i\}_{i=1}^N$ with $z_i = \phi(x_i), (x_i, y_i) \in P^{ID}$, are assumed to sample the manifold $\mathcal{M}$ sufficiently densely to approximate its intrinsic geometry. We construct a K nearest neighbor graph $G = (V, E)$ over in-domain embeddings z_i. As shown in Tenenbaum et al. (2000), shortest path distances on the constructed graph provide a consistent approximation to manifold geodesic distance when sampling is sufficiently dense. We denote the resulting graph based geodesic distance, computed using

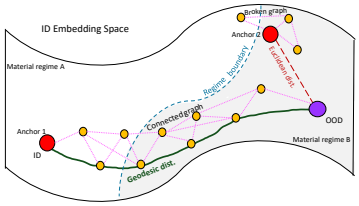


Figure 1: The illustration shows composite regimes linked by data-supported transitions; Euclidean proximity ignores this structure and selects the wrong anchor, while geodesic distance respects connectivity.

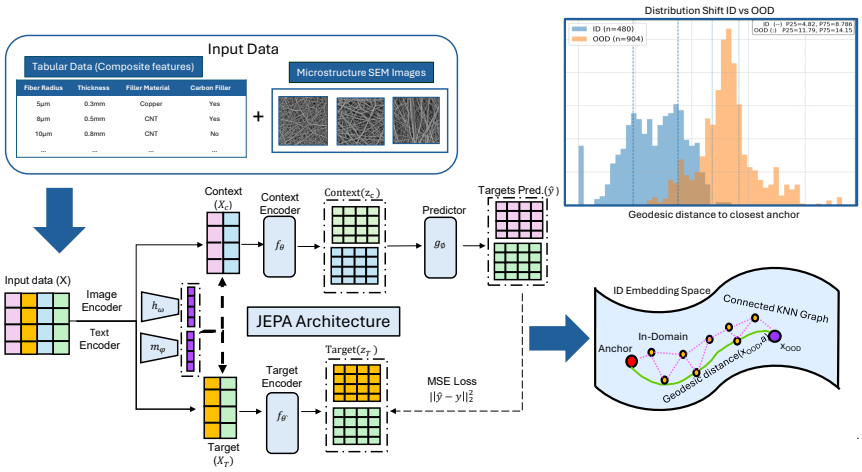


Figure 2: Overview of the proposed geometry-aware OOD generalization framework. Multimodal inputs (tabular descriptors and microstructure images) are embedded using a JEPA architecture to learn a shared representation space. Distribution shift between in-domain (ID) and out-of-domain (OOD) samples is reflected in geodesic distances to ID anchors on the learned manifold, enabling OOD samples to be anchored to supported regions and used for stable, geometry-driven weighting of ID data. The microstructure images have been taken from the $MM1$ dataset (Wu et al., 2025).

Dijkstra's algorithm (Dijkstra, 2022), by $d_M(z_i, z_j)$. Given an OOD sample $x^{OOD} \in P^{OOD}$ with the JEPA embedding $z^{OOD} = \phi(x^{OOD})$, we associate it with an **anchor point** by using

$$a(z^{OOD}) = \arg\min_{z_i \in \mathbf{M}} d_{\mathbf{M}}(z_{OOD}, z_i). \tag{2}$$

Unlike prior work (Kimura & Hino, 2024), we define **OOD distance with respect to the ID manifold and not among OOD points**. Fig. 2 shows a schematic of the GSA framework. The anchor identifies the in-distribution composite regime whose internal geometry best explains the OOD sample. In other words, it answers the question: *Which known material family is this new composite most physically compatible with?* To ensure reliability, we retain only *trusted* OOD samples whose geodesic distance is finite, whose assigned anchor is valid, and whose distance lies below the 90th percentile of the OOD distance distribution. Using these trusted samples, we aggregate anchor-level scores by taking the mean of negative geodesic distances per anchor, yielding a robustness weight for each anchor. Inspired by recent work by (Denize et al., 2023), we use **soft anchor geodesics** instead of a single nearest anchor, leading to robustness to anchor sparsity.

$$\bar{z} \;=\; \sum_{k=1}^{m} \frac{\exp(-\,d_k/T)}{\sum_{j=1}^{m} \exp(-\,d_j/T)}\, z_{a_k}, \tag{3}$$

The Eq. (3) illustrates the approach, assuming d_k denotes geodesic distance to anchor k, z_{a_k} is the embedding of anchor k and the softness is controlled by the hyperparameter temperature T. Further details about the hyperparameters for pre-training and downstream fine-tuning are described in Appendix section 10.

Finally, we propagate these anchor weights back to ID samples by assigning each ID point the weight of its closest anchor in the same geodesic framework, producing a per-sample weighting over the ID training set. Instead of normalizing out-of-distribution (OOD) weights independently within each mini-batch using a softmax function, we scale weights using robust statistics of the geodesic distance distribution. Specifically, distances are normalized by the **interquartile range (IQR)** (Whaley III, 2005), which reduces sensitivity to extreme values while preserving relative scale across batches.

3.3 PROPOSED JEPA MODEL ARCHITECTURE

The central modality of the proposed conditional JEPA architecture for composite materials consists of tabular features capturing key material properties. These features are encoded using a feature-wise Transformer, where each feature is treated as an independent token. Unlike a vanilla JEPA model, we condition the encoders with auxiliary multimodal input embeddings using vision and text encoders. These conditioned Featurewise transformers are used as the context and target encoders. The context encoder takes the normalized input dataset and produces a context embedding. We then mask around 30% of the input features and keep the rest visible ,as shown in Fig. 2. The visible context embeddings are passed as input to a multilayer perceptron (MLP) which predicts the masked features in latent space. The target encoder generates an embedding for the masked features in the same latent space. The error between the predicted masked features and the encoded masked features is measured with L2 loss. Further details about the model are provided in the Appendix 8.5.

3.4 THEORETICAL BOUNDS

In this section, we propose Proposition 1, an upper bound for the gap between estimated risk for ID and OOD as a function of the geodesic distance on the learned manifold. We assume **Lipschitz smoothness**, which means that moving a small distance along the data manifold in latent space does not significantly change the expected loss. We adopt this as a standard smoothness assumption in representation learning, consistent with prior work on predictive latent representations (Littwin et al., 2024).

Proposition 1. *Let $f_{\text{enc}} : \mathcal{X} \to \mathcal{Z}$ be the encoder of the JEPA model mapping inputs to a latent space. Denote $z = f_{\text{enc}}(x)$. Let P_{ID} and P_{OOD} be ID and OOD distributions on $\mathcal{X} \times \mathcal{Y}$ satisfying* covariate shift *assumption: $P_{ID}(y \mid x) = P_{OOD}(y \mid x)$ for all $x \in \mathcal{X}$. Let $P^{\mathcal{Z}}_{ID}$ and $P^{\mathcal{Z}}_{OOD}$ be the distribution of the latent representations z induced by ID and OOD input. Let d_G denote the graph geodesic distance approximated by a weighted kNN graph on $\mathcal{Z}$*

For a predictor $h : \mathcal{Z} \to \mathcal{A}$, and $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, define the conditional risk as a function of the latent representation:

$$g_h(z) \;:=\; \mathbb{E}[\ell(h(z), y) \mid z]\,. \tag{4}$$

*Assume that g_h is L-**Lipschitz with respect to** d_G, which implies it varies smoothly along the latent manifold.*

$$\big|g_h(z) - g_h(z')\big| \;\le\; L\, d_G(z, z') \qquad \forall z, z' \in \mathcal{Z}, \tag{5}$$

Then the difference between OOD and ID risks is bounded by

$$\big|R_{OOD}(h) - R_{ID}(h)\big| \;\le\; L\, W_1^{(G)}\big(P^{\mathcal{Z}}_{ID}, P^{\mathcal{Z}}_{OOD}\big)\,, \tag{6}$$

where

$$W_1^{(G)}\big(P^{\mathcal{Z}}_{ID}, P^{\mathcal{Z}}_{OOD}\big) \;:=\; \inf_{\gamma \in \Pi(P^{\mathcal{Z}}_{ID}, P^{\mathcal{Z}}_{OOD})} \mathbb{E}_{(z,z')\sim\gamma}[d_G(z, z')]\,, \tag{7}$$

$W_1^{(G)}$ denotes the Wasserstein-1 distance computed using the metric d_G.

Please refer to section 8.4 in the Appendix for a sketch of the proof for the proposition.

3.5 GEOMETRY-AWARE MANIFOLD METRICS

The graph-based geodesic distance captures intrinsic connectivity and curvature that Euclidean distance misses. To quantify this, we summarize the learned manifold using a small set of metrics derived from geodesic distances (Azarhooshang et al., 2020; Shvydun, 2025; Tenenbaum et al., 2000). Average Geodesic Stretch (AGS) measures the gap between Euclidean and intrinsic distances, reflecting curvature. Geodesic Reachability Ratio (GRR) captures global connectivity, where low values indicate fragmented support between ID and OOD regions. We also report local metrics, including median intrinsic dimension and bottleneck centrality, which capture changes in representation complexity and narrow transition regions. Together, these metrics help diagnose when OOD generalization is limited by geometric support rather than statistical mismatch. Formal definitions are provided in Appendix 8.4.1.

4 EXPERIMENTS

4.1 DATASETS

We study composite property prediction under distribution shifts using both unimodal and multimodal datasets. The tabular data captures process and material characteristics, while multimodal inputs additionally include microstructure images and textual descriptions. We simulate properties and microstructures using the RVE method (Bargmann et al., 2018) in Ansys Material Designer. **Nanofibers (UM1)** has 3010 samples of carbon composites. It includes composition ratios of carbon fiber, epoxy, and fillers (CNTs, graphene, copper, and nickel). The targets are yield strength and thermal conductivity. Carbon fiber proportion ranges from 0.50 to 0.67. By varying filler ratios, this dataset captures their effect on mechanical and thermal properties. **CF-Grade (UM2)** has 17,778 samples. It studies carbon fiber types (T300, T700, T800, T1100, and Pitch). Features include volume-weighted and mass-weighted fractions. We use volume-weighted features and focus on T300 and T800. This dataset reflects industry-relevant fiber types. **CF-T700** varies fiber alignment. Features include fiber volume fraction and rotation angle. Targets are yield strength and elongation. With 436 samples, this dataset supports fine-grained multimodal analysis linking images to properties. We also benchmark on *Nanofibers* ($MM1$), which has 471 samples. It includes process parameters, SEM images, and tensile properties under different loading directions (Wu et al., 2025). More details on inputs, targets, and dataset construction are in Appendix 9.

4.2 EXPERIMENTAL SETUP

We employ feature and dataset specific percentile thresholds and splitting strategies, as different material characteristics influence the target properties through distinct physical mechanisms. All splits are defined solely using input features and are fixed across methods to ensure a fair and unbiased evaluation. We consider the ID and OOD distributions to be comparable when the difference between their medians (m) does not exceed half of their interquartile range.

$$|m_{\text{OOD}} - m_{\text{I}}| \;\leq\; \frac{1}{2}\left(Q_{75}(\text{ID}) - Q_{25}(\text{ID})\right). \tag{8}$$

Unimodal datasets For $UM1$, we define four splits based on filler proportions. In $UM1S1$, we vary graphene content. Graphene strongly affects stiffness because of its high Young's modulus. The top and bottom 10% are OOD, and the 30–70% range is ID. This gives near-perfect GRR (1.0), with slightly higher BC_{max} and ORC_{median} due to tail effects. In $UM1S2$, we use CNT proportion. CNTs change stiffness through network formation. Again, the top and bottom 10% are OOD, and the 40–60% range is ID. GRR is 0.9999, with moderate boundary and residual effects. In $UM1S3$, we use copper proportion. Copper mainly affects thermal conductivity. The top and bottom 15% are OOD, and the 40–60% range is ID. GRR drops to about 0.48. This shows a strong shift, even though AGS improves and BC reduces. In $UM1S4$, we focus on nickel content. Nickel improves strength and thermal conductivity. Using the same 15% OOD split, GRR stays low (0.48). BC_{max} and ORC_{median} are minimal. This suggests weak ID coverage. For $UM2$, splits are based on carbon fiber type. In $UM2S1$, we isolate T300. In $UM2S2$, we isolate T800. OOD samples do not contain the selected fiber. The difference comes from higher stiffness of T800 due to its higher Young's modulus. **Multimodal Composite Data** For $MM1$, we focus on fiber radius. Radius affects mechanical performance, damage tolerance, and processing. We treat 1200 and 1600 as OOD, and the rest as ID. For $MM2$, we focus on the number of fibers. More fibers increase stiffness and load capacity along the fiber direction. The top and bottom 15% are OOD. The 40–60% range is ID. Metrics for all datasets are in Appendix table 6.

4.3 EVALUATION PROTOCOL

For $UM1$, we predict Young's Modulus (YM), Bulk Modulus (BM), Thermal Conductivity (TC), and Tensile Strength (TS). For $UM2$, we report YM, Electrical Conductivity (EC), and TS. For multimodal datasets, we augment tabular features with microstructure images and text to capture morphology and interfaces. Images are encoded using a ViT, and text using MatSciBERT, both are used to condition the JEPA model. For $MM1$, we predict Fracture Strength (FS), Yield Strength (YS), Elastic Modulus (EM), Elongation (EL), and Tangent Modulus (TM). For $MM2$, we report

Table 1: Test-time out-of-distribution (OOD) performance across datasets and distribution shifts. Higher R^2 indicates better performance.

			Coefficient of Determination (R^2)							
Dataset	Split	Target	XGBoost	CatBoost	NODE	TabNet	TabTransformer	CORAL	DANN	**GSA**
UM1	S1	YM	0.4358	0.3951	0.5291	-12.350	0.5166	0.5405	0.5099	**0.5816**
		TS	0.4568	0.3935	0.5477	-7.4876	0.5185	0.5454	0.5117	**0.5837**
	S3	YM	0.6741	0.6106	0.7263	-5.9591	0.4658	0.7152	0.5398	**0.7599**
		TS	0.6754	0.5951	0.7276	-10.212	0.4682	0.7115	0.5401	**0.7329**
UM2	S1	YM	0.6486	0.5382	0.5919	0.7696	0.7293	0.7934	0.8288	**0.9263**
		TS	0.3723	0.3448	0.7472	-0.0216	0.8265	0.8111	0.8260	**0.8535**
	S2	YM	0.5276	0.5382	0.8776	0.7847	0.7738	0.8781	0.8518	**0.9413**
		TS	-0.2371	0.1843	0.0759	0.7441	0.8402	0.8857	0.8949	**0.9098**

EL and YS. We fine-tune only the prediction heads while keeping the pretrained backbone frozen. For benchmarking, we compare against tabular SoTA methods. These include tree-based models (XGBoost (Chen, 2016), CatBoost (Prokhorenkova et al., 2018)) and deep models (NODE (Popov et al., 2019), TabNet (Arik & Pfister, 2021), TabTransformer (Huang et al., 2020)). All deep models use PyTorch-Tabular with default settings (Joseph, 2021). This ensures coverage across model families. We also compare with UDA methods, CORAL (Sun & Saenko, 2016) and DANN (Tzeng et al., 2017).

5 RESULTS

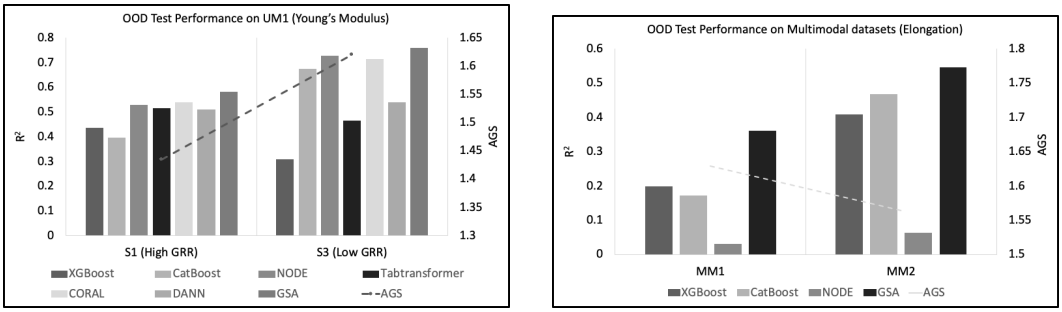
In this section, we report the coefficient of determination (R^2) on the ID/OOD distribution shift for each dataset. We report test OOD performance throughout and summarize results using compact visualizations that replace detailed tables, with full numerical results provided in the Appendix 9.1. Table 1 reports test-time out-of-distribution (OOD) performance for unimodal (UM1, UM2) composite datasets under representative distribution shifts. Baseline methods exhibit substantial variability across datasets and shifts, with OOD R^2 values ranging from strongly negative to moderately high, depending on model class and data modality. Tree-based methods such as XGBoost and CatBoost achieve OOD R^2 values between 0.40–0.67 in unimodal settings for predicting YM. Deep tabular models show mixed performance. NODE and TabTransformer reach up to 0.73 OOD R^2 on UM1 and 0.88 on UM2. However, their performance drops sharply under multimodal OOD shifts, with several cases producing near-zero or negative R^2. In contrast, GSA outperforms all baselines in the majority of settings. Further, we compare OOD performance for GSA against SoTA UDA methods on unimodal composite datasets. Across both datasets, GSA consistently outperforms CORAL and DANN under both distribution shifts. On the challenging split S3, DANN shows significant deterioration while, GSA performance is robust with a gain of 0.22 R^2. For multimodal datasets, as reported in Table 2, we observe that baseline performance remains below 0.40 and frequently near zero. GSA achieves OOD R^2 values of 0.29 on MM1 and 0.36 on MM2 for Elongation.

Table 2: Test-time out-of-distribution (OOD) performance across datasets and distribution shifts. Higher R^2 indicates better performance.

			Coefficient of Determination (R^2)					
Dataset	Split	Target	XGBoost	CatBoost	NODE	TabNet	TabTransformer	**GSA**
MM1	S1	EL	0.1989	0.1721	0.0311	-0.1047	-2.2554	**0.2885**
		YS	0.1015	-0.0051	-0.0413	-0.1285	-1.2764	**0.1931**
MM2	S1	EL	0.2499	0.3993	-0.0098	-8.0431	-6.092	**0.3613**
		YS	0.4078	0.4671	0.0629	-8.2869	-9.7161	**0.5452**

Fig. 3 illustrates test-time OOD performance in relation to manifold curvature quantified by AGS. In UM1 (Fig. 3a), the low-GRR split (S3) exhibits higher AGS, indicating increased geodesic distortion between in-distribution and OOD samples. Under this higher-curvature regime, baseline methods show larger variability and reduced OOD performance, whereas GSA maintains stronger performance, suggesting reduced sensitivity to global manifold distortion. In the multimodal setting 3(b), MM1

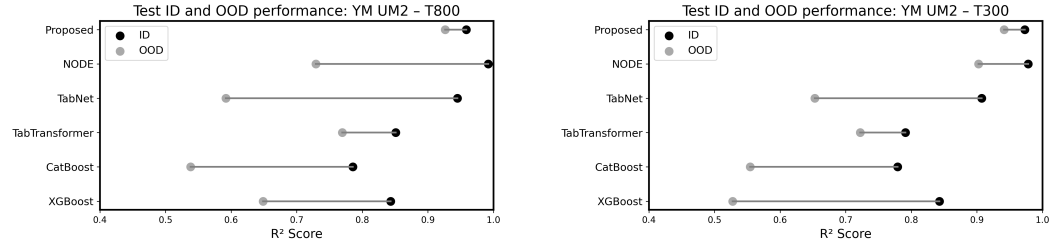
also exhibits elevated AGS, reflecting pronounced curvature induced by structural differences in microstructure. Despite this increased distortion, GSA consistently achieves higher OOD performance than competing methods.



(a) UM1 (b) MM1

Figure 3: The bar plots present a comparative picture of AGS on the OOD performance.

Fig. 4 compares in-distribution (ID) and out-of-distribution (OOD) test performance for Young's Modulus on the UM2 dataset under shifts induced by changes in carbon fiber type (T300 and T800). Across baseline methods, a clear degradation from ID to OOD performance is observed. Tree-based and transformer-based models exhibit particularly large ID–OOD gaps, indicating sensitivity to changes in fiber grade despite strong in-distribution accuracy. In contrast, GSA maintains high OOD performance while exhibiting a smaller separation between ID and OOD scores across both fiber types.



(a) S1 (b) S2

Figure 4: The dumbell plots show that GSA has the smallest gap for prediction of YM between Test ID and Test OOD data for UM2.

6 Discussion

In this section, we investigate the underlying mechanisms by ablating different components of the model to obtain insights. Further investigations regarding domain generalization (DG) algorithms, inductive bias and raw input space are described in the Appendix 8.1.

6.1 Residuals

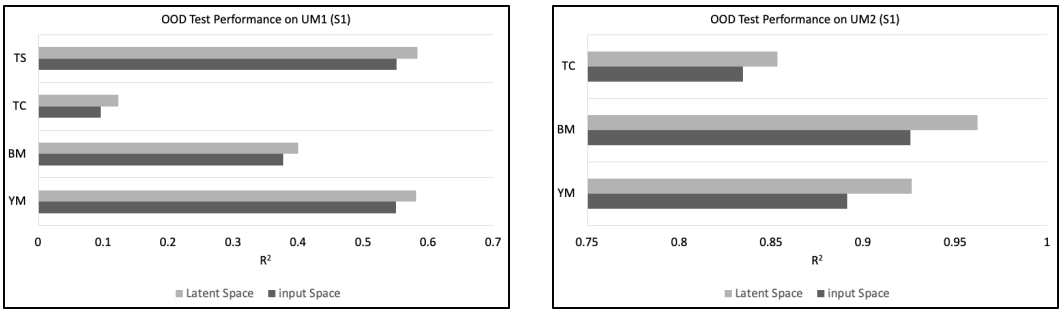
We ablate several formulations of the geodesic distance. Prior work uses residuals to capture distribution shift (Anirudh & Thiagarajan, 2023). As full residuals are noisy, we instead use the residual norm. However, Table 3 shows this degrades performance. We next consider the shortest geodesic distance from anchors to OOD points as a shift measure. Concatenating this distance to the features leads to a slight drop in OOD performance. Combining it with residuals yields similar degradation. Finally, we evaluate a dual loss formulation, where anchor weights emphasize ID regions near OOD points. This also reduces performance, suggesting that modifying the learned representation is detrimental.

Table 3: Ablation results for Split 1 for $UM1$ dataset

		Coefficient of Determination (R^2)				
Split	Model	YM	BM	TC1	TC2	TS
$S1$	Residuals(R)	0.5276	0.3221	-0.0321	0.0224	0.5873
	Distance(D)	0.5674	0.3942	0.0332	0.0915	0.5730
	R+D	0.5613	0.3477	-0.0292	0.0408	0.5532
	+Dual loss	0.5605	0.3476	-0.0301	0.0401	0.5531
	GSA	0.5816	0.3996	0.0664	0.1232	0.5837

6.2 Input Space

The JEPA architecture is robust to noise and material artifacts (Littwin et al., 2024). In this section, we compare it against reconstruction in raw input space for tabular data. This distinction is particularly important for composite materials since they have heterogeneous features. Fig. 5(a) shows that GSA outperforms input space reconstruction for both Graphene and CNT filler splits for $UM1$. A similar behavior is observed for T300 and T800 carbon fiber splits for $UM2$ as shown in fig 5(b). We believe input space reconstruction is brittle for composite materials which have multi scale features (Wu et al., 2025). The detailed numerical results are presented in Appendix 9.3.



(a) UM1 (b) UM2

Figure 5: Comparison of input-space and latent-space ablations under OOD evaluation.

7 Conclusion

In this work, we investigated out-of-distribution (OOD) generalization for fiber-reinforced polymer composites through the lens of learned manifold geometry. We first developed a conditional JEPA model to obtain representations that capture physically informed structure in the composite design space. Building on this representation, we introduced a geodesic anchoring framework, **Geometry Support Anchoring (GSA)**, that leverages intrinsic, graph-based distances to associate OOD samples with supported regions of the in-distribution manifold. By formulating anchoring as a geometric projection problem rather than a fixed or random reference selection, our approach uses connectivity information for local support and reachability under covariate shift. The proposed geometry-aware metrics provide quantitative diagnostics for distinguishing interpolation from extrapolation, revealing when and where learned models operate outside empirically supported regimes. Empirically, we show that GSA is most effective under structured covariate shifts characterized by local support mismatch.

Our results also show that improved OOD behavior does not always correspond to large accuracy gains. Instead, GSA provides insight into the geometric structure underlying generalization. This offers a principled, geometry-driven framework for scientific machine learning in composite systems. Future work includes extending geodesic reasoning to adaptive settings, incorporating uncertainty, and applying geometry-aware anchoring to broader scientific models.

Data and Code Availability The Nanofiber dataset is publicly available at https://github.com/wuyuhui-zju/MatMCL. Suitable requests for our in-house composite datasets should be addressed to Qian_Hangwei@a-star.edu.sg. The code used to train the models and reproduce the experiments will be made available upon reasonable request.

Author contributions AB developed and implemented the proposed method, performed the experiments, and led the manuscript preparation. HQ contributed to the conceptual development of the approach, supported manuscript writing, and coordinated the collection and organization of the composite datasets. SGC provided project supervision and contributed to manuscript revisions. IT led the overall project, contributed to method conceptualization, and oversaw the construction of the composite datasets. All authors reviewed and approved the final manuscript.

Acknowledgement This research is supported by the National Research Foundation, Singapore and DSO National Laborataries under the AI Singapore Programme(AISG Award No: AISG2-GC-2023-010), "Design Beyond What You Know": Material Informed Differential Generative AI (MIDGAI) for Light weight High-Entropy Alloys and Multi-functional composites(Stage 1b). This research is supported by A*STAR Career Development Fund $\langle Project\ No.C243512010\rangle$.

REFERENCES

Rushil Anirudh and Jayaraman J Thiagarajan. Out of distribution detection via neural network anchoring. In *Asian Conference on Machine Learning*, pp. 32–47. PMLR, 2023.

Sercan Ö Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 6679–6687, 2021.

Nazanin Azarhooshang, Prithviraj Sengupta, and Bhaskar DasGupta. A review of and some results for ollivier–ricci network curvature. *Mathematics*, 8(9):1416, 2020.

Swantje Bargmann, Benjamin Klusemann, Jürgen Markmann, Jan Eike Schnabel, Konrad Schneider, Celal Soyarslan, and Jana Wilmers. Generation of 3d representative volume elements for heterogeneous materials: A review. *Progress in materials science*, 96:322–384, 2018.

Nimit Berry, Muzammil Arshad, and Muazzam Arshad. Overview of polymer nanocomposites in aerospace and automotive applications. In *Polymer and Biopolymer Nanocomposites*, pp. 463–495. Elsevier, 2026.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Olivier Chapelle and Alexander Zien. Semi-supervised classification by low density separation. In *International workshop on artificial intelligence and statistics*, pp. 57–64. PMLR, 2005.

Krishan K Chawla. *Composite materials: science and engineering*. Springer Science & Business Media, 2012.

Tianqi Chen. Xgboost: A scalable tree boosting system. *Cornell University*, 2016.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

Julien Denize, Jaonary Rabarisoa, Astrid Orcesi, Romain Herault, and Stephane Canu. Similarity contrastive estimation for self-supervised soft contrastive learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2706–2716, 2023.

Edsger W Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: his life, work, and legacy*, pp. 287–290. 2022.

Mohammad Saber Hashemi, Masoud Safdari, and Azadeh Sheidaei. A supervised machine learning approach for accelerating the design of particulate composites: Application to thermal conductivity. *Computational Materials Science*, 197:110664, 2021.

Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. Tabpfn: A transformer that solves small tabular classification problems in a second. *arXiv preprint arXiv:2207.01848*, 2022.

Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2020.

Manu Joseph. Pytorch tabular: A framework for deep learning with tabular data. *arXiv preprint arXiv:2104.13638*, 2021.

Leonid Vasilevich Kantorovich and SG Rubinshtein. On a space of totally additive functions. *Vestnik of the St. Petersburg University: Mathematics*, 13(7):52–59, 1958.

Manickaraj Karuppusamy, Ramakrishnan Thirumalaisamy, Sivasubramanian Palanisamy, Sudha Nagamalai, Ehab El Sayed Massoud, and Nadir Ayrilmis. A review of machine learning applications in polymer composites: advancements, challenges, and future prospects. *Journal of Materials Chemistry A*, 2025.

Masanari Kimura and Hideitsu Hino. A short survey on importance weighting for machine learning. *arXiv preprint arXiv:2403.10175*, 2024.

Yi Liang, Xinyue Wei, Yongyue Peng, Xiaohan Wang, and Xiaoting Niu. A review on recent applications of machine learning in mechanical properties of composites. *Polymer Composites*, 46 (3):1939–1960, 2025.

Etai Littwin, Omid Saremi, Madhu Advani, Vimal Thilak, Preetum Nakkiran, Chen Huang, and Joshua Susskind. How jepa avoids noisy features: The implicit bias of deep linear self distillation networks. *Advances in Neural Information Processing Systems*, 37:91300–91336, 2024.

Fangyu Liu, Wenqi Ding, Yafei Qiao, and Linbing Wang. An artificial neural network model on tensile behavior of hybrid steel-pva fiber reinforced concrete containing fly ash and slag power. *Frontiers of Structural and Civil Engineering*, 14(6):1299–1315, 2020.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Timothy K Mulenga, Sanjay Mavinkere Rangappa, and Suchart Siengchin. Natural fiber composites: a comprehensive review on machine learning methods. *Archives of Computational Methods in Engineering*, pp. 1–27, 2025.

Chenyu Nan, Hongshi Ruan, Xiaozhe Ju, Junhan Hu, Lihua Liang, and Yangjian Xu. Transfer-learning-based strategy for enhancing prediction accuracy and computational efficiency of nonlinear mechanical properties in composite materials. *Composites Science and Technology*, 246:110388, 2024.

M Nurhaniza, MKA Ariffin, Aidy Ali, F Mustapha, and AW Noraini. Finite element analysis of composites materials for aerospace applications. In *IOP conference series: materials science and engineering*, volume 11, pp. 012010. IOP Publishing, 2010.

Sergei Popov, Stanislav Morozov, and Artem Babenko. Neural oblivious decision ensembles for deep learning on tabular data. *arXiv preprint arXiv:1909.06312*, 2019.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

Hangwei Qian, Yang He, Bingjin Chen, Mohit Sharma, and Ivor Tsang. Physics-constrained diffusion for lightweight composite material design. In *AI for Accelerated Materials Design-NeurIPS 2025*, 2025.

Jingang Qu, David Holzmüller, Gaël Varoquaux, and Marine Le Morvan. Tabicl: A tabular foundation model for in-context learning on large data. *arXiv preprint arXiv:2502.05564*, 2025.

Manickam Ramesh, Lakshmi Narasimhan Rajeshkumar, Nagarajan Srinivasan, Damodaran Vasanth Kumar, and Devarajan Balaji. Influence of filler material on properties of fiber-reinforced polymer composites: a review. *e-Polymers*, 22(1):898–916, 2022.

Philippe Rigollet. Generalization error bounds in semi-supervised classification under the cluster assumption. *Journal of Machine Learning Research*, 8(7), 2007.

Haitz Sáez de Ocáriz Borde, Alvaro Arroyo, Ismael Morales, Ingmar Posner, and Xiaowen Dong. Neural latent geometry search: Product manifold inference via gromov-hausdorff-informed bayesian optimization. *Advances in Neural Information Processing Systems*, 36:38370–38403, 2023.

Dhrubo Saha, Li Sun, and Chang Quan Lai. A science directed progressive neural network for multimodal prediction of elastoplastic behavior in composite materials. *Composites Part A: Applied Science and Manufacturing*, pp. 109179, 2025.

Sergey Shvydun. Zoo of centralities: Encyclopedia of node metrics in complex networks, 2025. URL `https://arxiv.org/abs/2511.05122`.

Peeyush Singhal, Rahee Walambe, Sheela Ramanna, and Ketan Kotecha. Domain adaptation: challenges, methods, datasets, and applications. *IEEE access*, 11:6973–7020, 2023.

Lei Song, Donglei Wang, Xuwang Liu, Aijun Yin, and Zhendong Long. Prediction of mechanical properties of composite materials using multimodal fusion learning. *Sensors and Actuators A: Physical*, 358:114433, 2023.

Masashi Sugiyama and Motoaki Kawanabe. *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT press, 2012.

Masashi Sugiyama, Matthias Krauledat, and Klaus-Robert Müller. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(5), 2007.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pp. 443–450. Springer, 2016.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

Nilesh Tripuraneni, Ben Adlam, and Jeffrey Pennington. Overparameterization improves robustness to covariate shift in high dimensions. *Advances in Neural Information Processing Systems*, 34: 13883–13897, 2021.

Yuta Tsuboi, Hisashi Kashima, Shohei Hido, Steffen Bickel, and Masashi Sugiyama. Direct density ratio estimation for large-scale covariate shift adaptation. *Journal of Information Processing*, 17: 138–155, 2009.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

Dewey Lonzo Whaley III. The interquartile range: Theory and estimation. Master's thesis, East Tennessee State University, 2005.

Yuhui Wu, Minmin Ding, Haonan He, Qijun Wu, Shaohua Jiang, Peng Zhang, and Jian Ji. A versatile multimodal learning framework bridging multiscale knowledge for material design. *npj Computational Materials*, 11(1):276, 2025.

Marvin Zhang, Henrik Marklund, Nikita Dhawan, Abhishek Gupta, Sergey Levine, and Chelsea Finn. Adaptive risk minimization: Learning to adapt to domain shift. *Advances in Neural Information Processing Systems*, 34:23664–23678, 2021.

Aurick Zhou and Sergey Levine. Bayesian adaptation for covariate shift. *Advances in neural information processing systems*, 34:914–927, 2021.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4396–4415, 2022.

Xun Zhou, Peng Zhang, Yanteng Wang, Zhenduo Zhang, and Keke Tang. A dual-transfer learning framework for predicting mechanical properties of non-circular fiber-reinforced composites with void defects. *Composite Structures*, pp. 119932, 2025.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

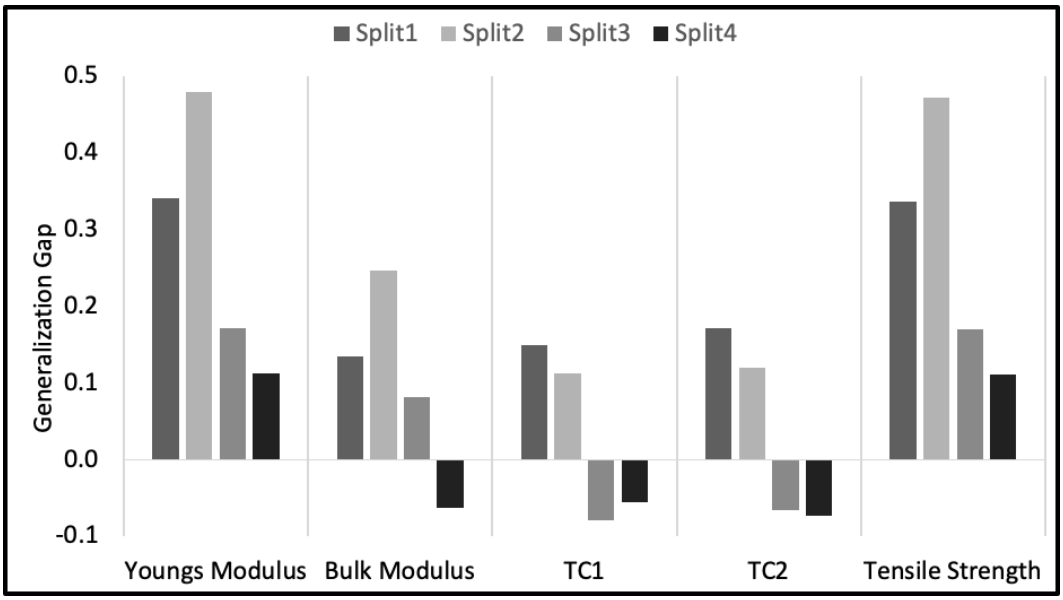


Figure 6: The bar plot shows the difference in generalization between the ID and OOD Splits ($R^2_{ID} - R^2_{OOD}$) across all properties for the proposed model.

8 Appendix

8.1 Additional Ablation studies

8.2 Generalization Gap

Fig. 6 captures the difference in generalization across ID and OOD splits for each property. The first insight is that a similar splitting rule results in different OOD severity depending on filler physics. The generalization gap is highest for CNT while it is lowest for Nickel. Further, the Youngs Modulus shows the largest generalization gap. The negative gap in certain cases indicates that the OOD split is more structured enabling the model to learn better patterns.

8.3 Reweighting Methods

We compare our GSA with two SoTA reweighting methods - Density weighting (Sugiyama et al., 2007) and Kernel reweighting (Sugiyama & Kawanabe, 2012). To ensure a fair comparison, we evaluate the model performance across 5 runs. The embeddings generated by the JEPA model were used as input for all the methods to ensure a fair comparison. Fig. 7 shows that the proposed model outperforms other reweighting methods. Density-based weighting methods assume that local Euclidean neighborhoods are meaningful and that data support can be approximated by local density in the ambient space. When these assumptions fail, geodesic distances provide a more correct notion of similarity by penalizing paths that traverse poorly supported or weakly connected regions of the learned representation. For a detailed comparison across all properties and unimodal datasets please refer to the Appendix 8.1.

8.3.1 Inductive bias

In recent years, in-context learning based models such as TabPFN (Hollmann et al., 2022) and TabICL (Qu et al., 2025) are pretrained on large collections of synthetic datasets and consequently encode a strong inductive bias toward a particular family of data-generating processes. This bias often enables strong out-of-distribution (OOD) performance, as the models implicitly capture structural regularities. However, this same inductive bias can lead to catastrophic failure in smaller composite datasets where complex patterns violate the implicit assumptions learned during pretraining as shown in Fig. 9.

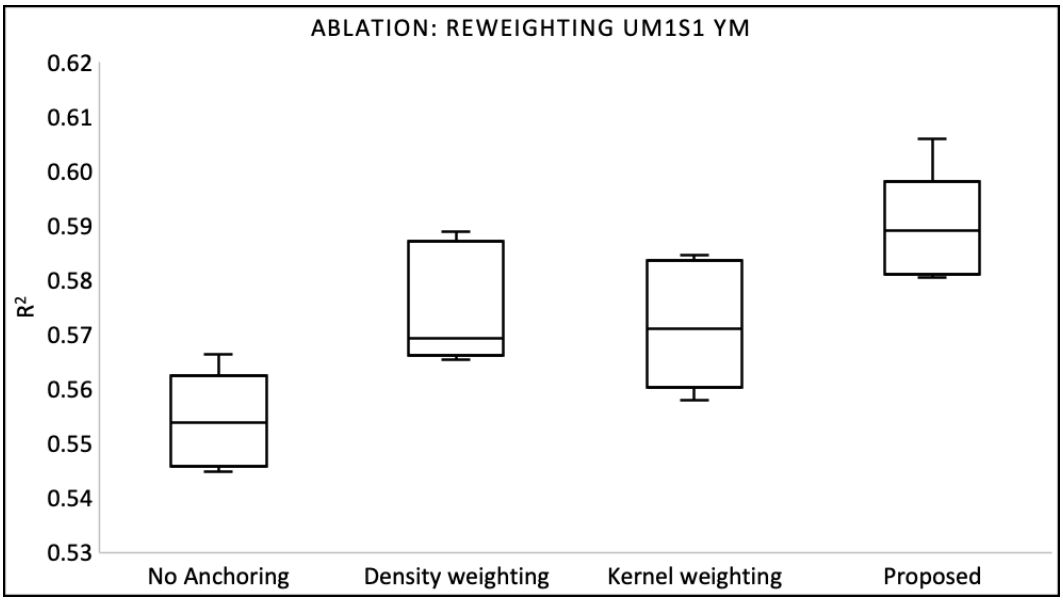


Figure 7: The figure shows that GSA outperforms other reweighting methods

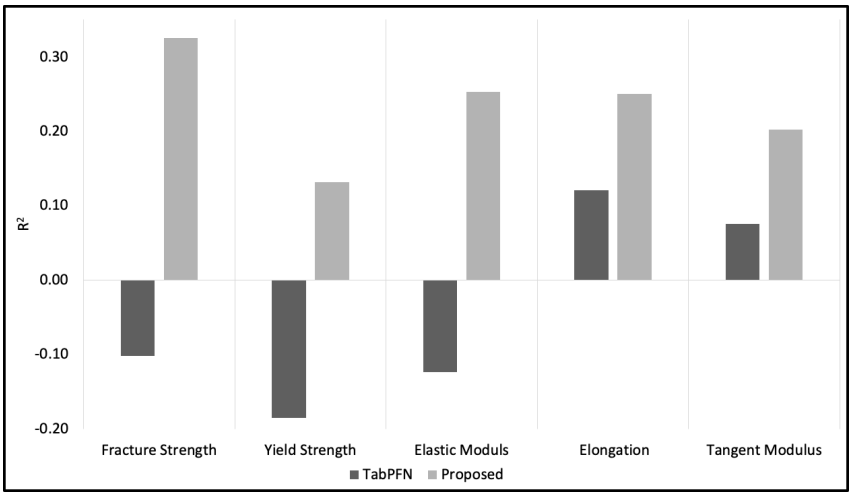


Figure 8: Multimodal composite data

Figure 9: The figure shows that TabPFN fails catastrophically resulting in negative R^2 when the inductive bias is misaligned.

8.3.2 REWEIGHTING

Table 4 compares the performance of the proposed algorithm with two SoTA reweighting methods - Density weighting (Sugiyama et al., 2007) and Kernel reweighting (Sugiyama & Kawanabe, 2012) for $UM1$. A similar comparison is reported for $UM2$ in the table 5.

Table 4: Reweighting methods for the $UM1$ dataset

		Coefficient of Determination (R^2)				
Split	Model	YM	BM	TC1	TC2	TS
$S1$	No Anchor	0.5468	0.3872	0.0344	0.0957	0.5509
	Density	0.5668	0.3881	0.0602	0.1099	0.5683
	Kernel	0.5711	0.3854	0.0618	0.1118	0.5727
	GSA	0.5816	0.3996	0.0664	0.1232	0.5837
$S2$	No Anchor	0.4291	0.3261	0.0918	0.1163	0.406
	Density	0.4372	0.2414	0.0762	0.0969	0.4371
	Kernel	0.3641	0.3288	0.1541	0.1835	0.3621
	GSA	0.4427	0.3336	0.0946	0.1249	0.4462
$S3$	No Anchor	0.7184	0.4607	0.2431	0.3008	0.7206
	Density	0.6609	0.4565	0.2447	0.3124	0.663
	Kernel	0.6415	0.4255	0.1906	0.2661	0.6551
	GSA	0.7299	0.4682	0.2577	0.3116	0.7329
$S4$	No Anchor	0.7484	0.5095	0.2695	0.3232	0.7403
	Density	0.7372	0.4942	0.2509	0.3131	0.7386
	Kernel	0.7406	0.4254	0.2094	0.3041	0.7424
	GSA	0.7552	0.5256	0.2888	0.3536	0.7569

Table 5: Reweighting methods for the $UM2$ dataset

		Coefficient of Determination (R^2)		
Split	Model	YM	EC	TS
$S1$	No Anchor	0.9146	0.9378	0.8397
	Density	0.8678	0.9472	0.7585
	Kernel	0.9017	0.9133	0.8523
	GSA	0.9263	0.9623	0.8535
$S2$	No Anchor	0.9292	0.9123	0.8884
	Density	0.8969	0.9368	0.8925
	Kernel	0.9359	0.8989	0.8570
	GSA	0.9413	0.9425	0.9098

8.4 PROOF SKETCH

In this section we provide the proof sketch for Proposition 1

Proof sketch. In composite materials, often the distribution shift is covariate in nature. Under the covariate shift assumption, the conditional distribution of labels given the input is the same in the ID and OOD distributions. As a result, the population risk $R(h)$ depends on the data distribution only through the marginal distribution of the latent representation z.

$$R_{ID}(h) = \mathbb{E}_{(x,y)\sim P_{ID}}\big[\ell(h(f_{\text{enc}}(x)), y)\big] = \mathbb{E}_{z\sim P^{Z}_{ID}}\big[g_h(z)\big] \tag{9}$$

Similarly, for the OOD distribution

$$R_{OOD}(h) = \mathbb{E}_{z\sim P^{Z}_{OOD}}\big[g_h(z)\big] \tag{10}$$

∴, the ID-OOD risk difference can be written as

$$R_{OOD}(h) - R_{ID}(h) = \mathbb{E}_{z\sim P^Z_{OOD}}\big[g_h(z)\big] - \mathbb{E}_{z\sim P^Z_{ID}}\big[g_h(z)\big]. \tag{11}$$

Assuming that g_h is Lipschitz smooth along the latent manifold with respect to geodesic distance (5). Define the rescaled function $\varphi(z) := \frac{1}{L} g_h(z)$ Then, φ is 1-Lipschitz w.r.t. d_G.

By the Kantorovich-Rubinstein duality theorem (Kantorovich & Rubinshtein, 1958) on the metric space $(\mathcal{Z}, d_G)$, the Wasserstein-1 distance between P^Z_{ID} and P^Z_{OOD} admits the dual representation

$$W_1^{(G)}(P^Z_{ID}, P^Z_{OOD}) = \sup_{\|\psi\|_{\mathrm{Lip}}\le 1} \left(\mathbb{E}_{z\sim P^Z_{OOD}}[\psi(z)] - \mathbb{E}_{z\sim P^Z_{ID}}[\psi(z)]\right). \tag{12}$$

Since φ is a valid 1-Lipschitz test function, plugging in $\psi = \varphi$ yields

$$\mathbb{E}_{P^Z_{OOD}}[g_h(z)] - \mathbb{E}_{P^Z_{ID}}[g_h(z)] = L\left(\mathbb{E}_{P^Z_{OOD}}[\varphi(z)] - \mathbb{E}_{P^Z_{ID}}[\varphi(z)]\right) \tag{13}$$

$$\le L\, W_1^{(G)}(P^Z_{ID}, P^Z_{OOD}) \tag{14}$$

□

Thus, assuming expected loss varies smoothly along the latent manifold, the degradation in performance under covariate shift is controlled by the distance moved by the probability mass from ID to OOD

8.4.1 Geometry-aware Manifold Metrics

Let $\{z_i\}_{i=1}^N \subset \mathbb{R}^D$ denote the in-distribution (ID) embeddings obtained from the conditional JEPA model, assumed to sample a compact, smooth manifold $\mathcal{M}$ of intrinsic dimension $d \ll D$. We approximate the intrinsic geometry of $\mathcal{M}$ by constructing a k-nearest-neighbor graph $G = (V, E)$ over the ID embeddings, with edge weights given by Euclidean distances in the ambient space. Shortest-path distances on this graph, computed using Dijkstra's algorithm (Dijkstra, 2022), define a graph-based approximation of the geodesic distance on $\mathcal{M}$, denoted by $d_{\mathcal{M}}(\cdot,\cdot)$. Using $d_{\mathcal{M}}$, we define the following geometry-aware metrics to characterize support, connectivity, and local structure of the learned manifold.

Average Geodesic Stretch (AGS) For a pair of points (z_i, z_j), the geodesic stretch is defined as the ratio between intrinsic and ambient distances. The Average Geodesic Stretch is given by

$$\mathrm{AGS} = \mathbb{E}_{i,j}\left[\frac{d_{\mathcal{M}}(z_i, z_j)}{\|z_i - z_j\|_2 + \epsilon}\right], \tag{15}$$

where ϵ is a small constant for numerical stability. Large values of AGS indicate tortuous or highly curved manifolds, while values close to 1 suggest near-Euclidean geometry.

Geodesic Reachability Ratio (GRR) GRR measures the fraction of point pairs that are mutually reachable via finite geodesic paths on the graph:

$$\mathrm{GRR} = \frac{1}{|\mathcal{P}|}\sum_{(i,j)\in\mathcal{P}} \mathbb{I}\big[d_{\mathcal{M}}(z_i, z_j) < \infty\big], \tag{16}$$

where $\mathcal{P}$ denotes a set of sampled point pairs. GRR close to 1 indicates a connected manifold, whereas lower values suggest disconnected or weakly connected components.

Median Intrinsic Dimension ($\mathrm{IDim}_{\mathbf{median}}$) The local intrinsic dimension at point z_i is estimated from the scaling behavior of geodesic distances to its neighbors. Let $\{r_{i,k}\}$ denote the sorted geodesic distances from z_i to its k nearest neighbors. A local dimension estimate $\hat{d}_i$ is obtained via standard maximum-likelihood estimators. We report the median intrinsic dimension across points:

$$\mathrm{IDim}_{\text{median}} = \mathrm{median}_i(\hat{d}_i). \tag{17}$$

This metric reflects the effective dimensionality of the learned manifold.

Bottleneck Centrality (BC) To quantify the presence of narrow passages or hubs on the manifold, we compute node betweenness centrality using geodesic shortest paths. For node v, the betweenness centrality is

$$\mathrm{BC}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{18}$$

where σ_{st} is the number of shortest geodesic paths between s and t, and $\sigma_{st}(v)$ is the number of such paths passing through v. We report

$$\mathrm{BC}_{\max} = \max_{v} \mathrm{BC}(v), \tag{19}$$

$$\mathrm{BC}_{\text{median}} = \mathrm{median}_{v} \mathrm{BC}(v). \tag{20}$$

High $\mathrm{BC}_{\max}$ indicates strong bottlenecks that dominate manifold connectivity.

Ollivier–Ricci Curvature (ORC) Local curvature of the graph is quantified using Ollivier–Ricci curvature. For an edge (i, j), the curvature is defined as

$$\kappa(i, j) = 1 - \frac{W_1(\mu_i, \mu_j)}{d_{\mathcal{M}}(z_i, z_j)}, \tag{21}$$

where W_1 is the Wasserstein-1 distance between local neighborhood measures μ_i and μ_j. We summarize curvature statistics via

$$\mathrm{ORC}_{\text{median}} = \mathrm{median}_{(i,j)\in E} \kappa(i, j). \tag{22}$$

Positive curvature indicates locally clustered, well-supported regions, while negative curvature reflects saddle-like or sparse regions. Together, AGS, GRR, $\mathrm{IDim}_{\text{median}}$, BC, and ORC provide complementary characterizations of the intrinsic geometry induced by $d_{\mathcal{M}}$, enabling principled analysis of support, reachability, and structural complexity of the learned manifold under distribution shift. Table 6 captures the metrics for the unimodal and multimodal datasets.

Table 6: Geometry Driven Metrics

		Geometry Aware Manifold Metrics				
Dataset	Split	AGS	GRR	IDim(Median)	BC(Max)	ORC(Median)
UM1	S1	1.4351	0.9999	3.3759	0.1736	0.3867
UM1	S2	1.4541	0.9999	3.6010	0.1368	0.3636
UM1	S3	1.6198	0.4781	3.5368	0.0789	0.3501
UM1	S4	1.5107	0.4781	3.6017	0.0526	0.3478
UM2	S1	1.514	0.9999	1.9996	0.1052	0.4000
UM2	S2	1.512	0.9999	1.9981	0.1100	0.3986
MM1	S1	1.629	0.9999	4.1747	0.1421	0.3529
MM2	S1	1.561	0.9999	4.1891	0.1525	0.3129

8.5 Proposed Model Architecture

We propose a self-supervised learning framework inspired by the Joint-Embedding Predictive Architecture (JEPA) paradigm. The central modality in the model are the tabular features which measure the key features of composite materials. We use a feature wise transformer to encode the tabular features. Each feature is encoded as a token and the transformer is able to capture the interaction between the different features. Unlike other modalities like text, the tabular features are independent and are not sequential in nature. To condition the feature wise transformer, we encode the microstructure images and the textual descriptions using a ViT encoder and MatSciBERT encoder respectively. The MatSciBERT encoder helps us leverage the pretrained scientific knowledge encoded in the MatSciBERT model. The text and image embeddings are concatenated and downscaled and then used as input to condition the Featurewise transformer. These conditioned Featurewise transformers are used as the context and target encoders. The context encoder takes the normalized input dataset

and produces a context embedding. We then mask around 30% of the input features and keep the rest visible. The visible context embeddings are passed as input to a MLP which predicts the masked features in latent space. The target encoder generates an embedding for the masked features in the same latent space. The objective of the proposed framework is to learn a rich representation. For each training instance. Let $t_{context}$ be the input context tabular features and t_{target} be the target tabular features. The text embedding is denoted by z_{text} and microstructure embedding is denoted by z_{image}.

$$z_{concat} = fusion(z_{text}, z_{image}) \tag{23}$$

$$\hat{t_{target}} = h([t_{context}|z_{concat}]) \tag{24}$$

Mean squared error (MSE) between the predicted and actual latent representation is used as the prediction loss

$$L_{pred} = ||t_{target} - \hat{t_{target}}||_2^2 \tag{25}$$

The model is trained using AdamW optimizer with weight decay (Loshchilov & Hutter, 2017). A cosine annealing scheduler is used for stable convergence.

8.6 ID/OOD Split

Fig. 10 captures the distribution shift of the dataset for in-domain and out-of-domain splits. For the first split, ID and OOD Fig. 10(a) show strong overlap with only a mild rightward shift for OOD, indicating limited distribution shift. ID and OOD distributions remain substantially overlapping, as shown in Fig 10(b), though OOD is slightly more dispersed than in Split 1. Fig. 10(c) shows a clear separation between ID and OOD, with OOD concentrated at much larger distances and minimal overlap. Fig. 10(d) shows the distribution shift along the geodesic distance. ID and OOD distributions are strongly shifted and weakly overlapping, similar to Cu but slightly smoother.

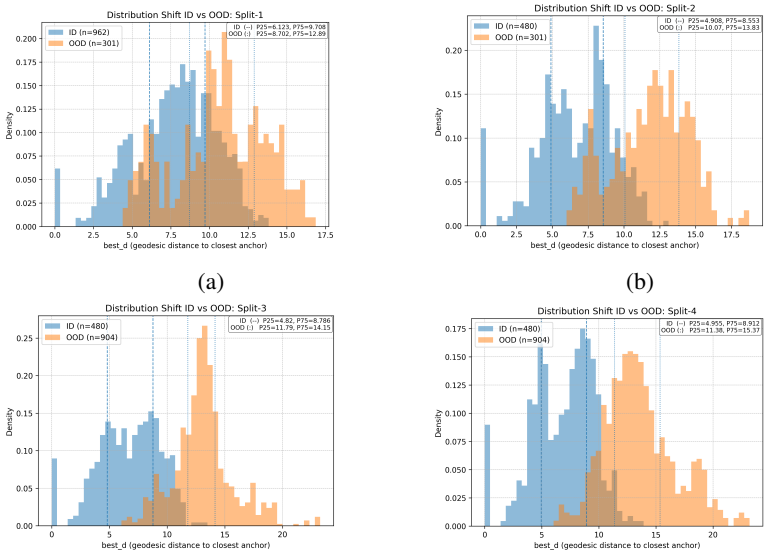


Figure 10: The histogram captures the distribution shift between the ID and OOD split. The data has been split based on the proportion of the filler.

9 Dataset Construction

UM1 : The dataset comprises 3,400 samples of carbon-based composite materials, each characterized by detailed compositional descriptors. These include the volume fractions of carbon fiber and epoxy, along with the presence and proportions of up to four filler materials: carbon nanotubes,

Table 7: Input Descriptor Ranges for in-house dataset.

Dataset	Input features	Range
Nanohybrid (UM1)	v_{cf}	0.51–0.67
	$v_{\text{cnt}}, v_{\text{gra}}, v_{\text{cu}}, v_{\text{ni}}$	0.00–0.10
	v_{resin}	0.09–0.44
CF-Grade (UM2)	$v_{\text{T300}}, v_{\text{T700}}, v_{\text{T800}}, v_{\text{T1100}}, v_{\text{Pitch}}$	0.00–0.73
	v_{epoxy}	0.27–0.65
CF-T700 (MM2)	Fiber volume fraction V_f	0.31–0.65

graphene, copper, and nickel. Variations in composition give rise to distinct mechanical and thermal behaviors, as different fillers contribute through different physical mechanisms. High-modulus carbon-based fillers, such as carbon nanotubes and graphene, are typically associated with increased mechanical strength, whereas metallic fillers primarily affect thermal transport properties. As a result, the observed material properties emerge from the combined influence of multiple interacting components rather than any single constituent. The dataset reflects this structured relationship between composition and performance, with target variables including yield strength and thermal conductivity.

UM2 : This dataset contains 17,700 samples of fiber-reinforced polymer composites composed of an epoxy matrix with carbon fiber reinforcement. Each sample is described by the type of carbon fiber used—T300, T700, T800, or T1100—along with the corresponding proportions of the constituent elements and resin. Differences among fiber grades introduce systematic variations in mechanical and electrical behavior, as higher-grade fibers typically exhibit increased stiffness and strength, while also influencing conductive pathways within the composite. The resulting material properties therefore depend jointly on fiber type and composition, reflecting both intrinsic fiber characteristics and matrix–fiber interactions. Target properties in this dataset include Young's modulus, electrical conductivity, and tensile strength. All material responses are obtained through physics-based simulations using Ansys.

MM1 : This dataset constructed by Wu et al. (2025) is based on electrospun nanofiber films. The dataset was generated by varying combinations of processing parameters, including flow rate, solution concentration, applied voltage, rotation speed, and ambient temperature and humidity. Microstructural features were characterized using scanning electron microscopy (SEM). Mechanical properties were measured through tensile testing in both longitudinal and transverse directions, yielding fracture strength, yield strength, elastic modulus, tangent modulus, and fracture elongation. A binary indicator is included in the processing descriptors to denote the tensile loading direction.

MM2 : The multimodal dataset $MM2$ consists of fiber-reinforced polymer composites composed of an epoxy matrix with carbon fiber reinforcement. We focus on composites based on CF-T700 fibers with randomly oriented unidirectional reinforcement. The dataset varies the fiber volume fraction (Vf) and the mean misalignment angle (MMA), which serve as the primary input descriptors. Using the Representative Volume Element (RVE) approach implemented in Ansys Material Designer, we simulate the corresponding microstructures and target properties, including tensile yield strength and elongation. In total, the dataset contains 436 samples spanning Vf values from 0.32 to 0.65 and MMA values from 0 to 5 degrees, with each sample represented by paired tabular descriptors and microstructure images. The range of input descriptors and target properties is given in table7.

9.1 Unimodal Composite data

Table 8 shows that the generalization is dependent on the filler material and how representative the ID region is of the extremes. For Graphene (S1) and CNT (S2), where we have a high GRR (around 0.99), the tree-based methods achieve constantly positive R^2 across all properties, while TabNet and MLP struggles. GSA performs best overall in both splits, achieving highest R^2 for Youngs modulus, bulk modulus and tensile strength. The Copper (S3) and Nickel (S4) splits which have a lower GRR (around 0.50), exhibit a higher variability in performance across properties. GSA is the most robust, delivering highest R^2 in S3 and competitive in S4, especially for bulk modulus and thermal conductivity.

Table 8: Benchmarking OOD generalization performance for UM1 against SoTA tabular models. The higher the better the model performance.

		Coefficient of Determination (R^2)				
Split	Properties	YM	BM	TC1	TC2	TS
S1	MLP	-0.3964	-0.0688	-0.0016	0.1229	-1.825
	XGBoost	0.4358	0.2536	-0.1031	-0.0895	0.4568
	CatBoost	0.3951	0.2125	-0.1058	-0.1056	0.3935
	NODE	0.5291	0.3985	0.0287	0.0707	0.5477
	TabNet	-12.350	-0.1886	-1.2711	-1.9525	-7.4876
	Tabformer	0.5166	0.4362	0.0871	0.1705	0.5185
	CORAL	0.5405	0.3296	0.0707	0.1063	0.5454
	DANN	0.5099	0.3489	0.0032	0.00733	0.5117
	GSA	0.5816	0.3996	0.0664	0.1232	0.5837
S2	MLP	-0.8810	-5.074	-0.6783	-0.6965	0.2335
	XGBoost	0.3796	0.2136	-0.0888	0.0022	0.4011
	CatBoost	0.3486	0.2132	-0.0551	-0.0781	0.3571
	NODE	0.3851	0.2313	0.0063	0.0284	0.3246
	TabNet	-16.0611	-3.1101	-1.0621	-1.0078	-10.034
	Tabformer	-0.3748	0.1499	0.1428	0.1657	-0.3139
	CORAL	0.3661	0.1815	-0.0481	0.0061	0.3657
	DANN	0.5099	0.3489	0.0032	0.0733	0.5117
	GSA	0.4427	0.3361	0.0946	0.1249	0.4462
S3	MLP	0.3086	0.1904	0.2602	0.3661	0.0911
	XGBoost	0.6741	0.3919	0.1703	0.1861	0.6754
	CatBoost	0.6106	0.3613	0.1458	0.1675	0.5951
	NODE	0.7263	0.4373	0.1946	0.27161	0.7276
	TabNet	-5.9591	-2.3803	-2.4153	-2.5124	-10.212
	Tabformer	0.4658	-0.0582	-0.0443	-0.0722	0.4682
	CORAL	0.7152	0.2699	0.0441	0.1296	0.7115
	DANN	0.5398	0.2979	0.0983	0.1895	0.5401
	GSA	0.7599	0.4682	0.2577	0.3116	0.7329
S4	MLP	-0.3121	-0.5466	-0.9711	-0.0933	0.1259
	XGBoost	0.6451	0.3641	0.1031	0.1826	0.6766
	CatBoost	0.6121	0.3822	0.1998	0.2339	0.6044
	NODE	0.7545	0.4775	0.2586	0.3393	0.7315
	TabNet	-6.5547	-11.7137	-5.9541	11.2251	-1.7062
	Tabformer	0.8079	0.4095	0.1052	0.1314	0.8089
	CORAL	0.6783	0.3051	0.0934	0.1639	0.6648
	DANN	0.6448	0.3238	0.0671	0.1334	0.6381
	GSA	0.7552	0.5256	0.2888	0.3536	0.7569

Table 9 presents OOD generalization results on the $UM2$ dataset, where the test domain excludes a specific carbon fiber grade (T300 in S1 and T800 in S2). YM generalizes well across fiber grades for most models, as it is mainly controlled by stiffness contributions from the fiber and composite architecture, which transfer reliably between grades. In contrast, EC and TS are more sensitive to changes in fiber grade, since they depend on contact networks, defects, and failure initiation, which vary with material processing and microstructure. As a result, several baseline models show large performance drops under these shifts, including negative R^2 values in the more challenging S2 split. While domain generalization methods partially mitigate this degradation, GSA remains consistently robust.

Table 9: Benchmarking OOD generalization performance for UM2 against SoTA tabular algorithms. The higher the better the model performance.

		Coefficient of Determination (R^2)		
Split	Properties	YM	EC	TS
S1	XGBoost	0.6486	0.2089	0.3723
	CatBoost	0.5382	0.2832	0.3448
	TabTransformer	0.7696	0.5274	-0.0216
	TabNet	0.5919	0.6488	0.7472
	NODE	0.7293	0.8259	0.8265
	CORAL	0.7934	0.8688	0.8111
	DANN	0.8288	0.8512	0.8260
	GSA	0.9263	0.9623	0.8535
S2	XGBoost	0.5276	-0.4795	-0.2371
	CatBoost	0.5382	0.0655	0.1843
	TabTransformer	0.7738	0.4083	0.0759
	TabNet	0.7847	0.3005	0.7441
	NODE	0.8776	0.8172	0.8402
	CORAL	0.8781	0.9061	0.8857
	DANN	0.8518	0.9138	0.8949
	GSA	0.9413	0.9425	0.9098

9.2 Multimodal Composite Data

Table 10 shows that the proposed model outperforms other SoTA models when we split the first multimodal dataset by radius. Most baseline models exhibit poor or negative R^2 values under OOD settings, while the proposed model maintains **stable and positive generalization** across all targets. Tree-based models (XGBoost, CatBoost) show modest positive performance on a few properties but lack consistency, while deep tabular models (TabNet, TabTransformer) degrade severely under OOD shift.

Table 10: Benchmarking OOD generalization performance for MM1 against SoTA Tabular models. The higher the better the model performance.

		Coefficient of Determination (R^2)				
Split	Properties	FS	YS	EM	EL	TM
MM1	XGBoost	0.0832	0.1015	0.1608	0.1989	0.0901
	CatBoost	0.1295	-0.0051	0.0804	0.1721	0.0701
	NODE	0.0354	-0.0413	-0.112	0.0311	-0.0528
	TabNet	-0.3417	-0.1285	-0.3513	-0.1047	-0.3953
	Tabformer	-1.5734	-1.2764	-2.4551	-2.2554	-1.5759
	GSA	0.3728	0.1931	0.2586	0.2885	0.1742

Table 11 reports the OOD generalization performance on $MM2$. Consistent with prior observations, GSA achieves superior predictive accuracy for both elongation and yield strength compared to

existing tabular baselines. Across all models, yield strength exhibits stronger OOD generalization than elongation. This behavior is expected, as yield strength is primarily governed by relatively monotonic and transferable relationships with characteristics, such as the number of fibers contributing to load-bearing capacity. In contrast, elongation is more sensitive to changes in damage evolution and failure mechanisms, which are less stable under distribution shifts. Owing to the limited number of samples in both training and test splits, transformer-based architectures exhibit pronounced overfitting, resulting in degraded OOD performance.

Table 11: Benchmarking OOD generalization performance for MM2 against SoTA Tabular models. The higher the better the model performance.

		Coefficient of Determination (R^2)	
Split	Properties	Elongation	Yield Strength
MM2	XGBoost	0.2499	0.4078
	CatBoost	0.3993	0.4671
	NODE	-0.0098	0.0629
	TabNet	-8.0431	-8.2869
	TabTransformer	-6.092	-9.7161
	GSA	0.3613	0.5452

9.3 Ablation: Latent space

Table 12: Comparison between raw input space and latent space for $UM1$ dataset

		Coefficient of Determination (R^2)				
Split	Representation	YM	BM	TC1	TC2	TS
$S1$	Input Space	0.5504	0.3772	0.0271	0.096	0.5510
	GSA	0.5816	0.3996	0.0664	0.1232	0.5837
$S2$	Input Space	0.4952	0.2851	0.105	0.1307	0.4971
	GSA	0.4427	0.3361	0.0946	0.1249	0.4462

Table 13: Comparison between raw input space and latent space for $UM2$ dataset

		Coefficient of Determination (R^2)		
Split	Representation	YM	EC	TS
$S1$	Input Space	0.8913	0.9257	0.8348
	GSA	0.9263	0.9623	0.8535
$S2$	Input Space	0.9176	0.9389	0.8925
	GSA	0.9413	0.9425	0.9098

9.4 Ablation: Domain generalization

Domain Generalization (DG) typically assumes access to multiple environments and leverages invariance across them using methods such as IRM, GroupDRO, and VREx. In contrast, our setting focuses on covariate shift, where there is overlap in support between the in-domain and out-of-domain distributions. To enable comparison with DG methods, we construct two pseudo-environments from the in-domain training data via random permutation, while reserving the OOD dataset strictly for evaluation. As shown in Table 14, DG methods exhibit high variability in performance across datasets and target tasks. This suggests that, while effective for multi-environment generalization, they may be less suitable for covariate shift settings in composite materials.

Table 14: Reweighting methods for the $UM1$ dataset

		Coefficient of Determination (R^2)				
Split	Model	YM	BM	TC1	TC2	TS
$S1$	IRM	0.4467	0.4703	0.2036	0.2928	0.4836
	VREx	0.5801	0.3974	0.1042	0.2384	0.5291
	GroupDRO	0.5421	0.4371	0.2410	0.3004	0.5396
	SWAD	0.5790	0.4012	0.2682	0.1996	0.5789
	GSA	0.5816	0.3996	0.0664	0.1232	0.5837
$S2$	IRM	-0.1307	0.4425	-0.0759	0.2406	0.5473
	VREx	0.5563	0.3049	0.0563	0.0644	0.4308
	GroupDRO	-0.4882	0.4507	0.0231	0.2683	0.4748
	SWAD	0.2716	0.4229	0.2134	0.0966	-0.1561
	GSA	0.4427	0.3336	0.0946	0.1249	0.4462
$S3$	IRM	0.7016	0.3723	-0.1108	0.1731	0.7038
	VREx	0.7228	0.4425	0.3119	0.2168	0.7005
	GroupDRO	0.6728	0.3061	0.0301	0.2799	0.7425
	SWAD	0.3294	-0.5185	-1.2875	-2.6012	-0.1403
	GSA	0.7299	0.4682	0.2577	0.3116	0.7329
$S4$	IRM	0.7102	-0.2588	-1.2501	-2.709	0.5469
	VREx	0.6925	0.3801	0.2913	0.1217	0.3006
	GroupDRO	0.7341	0.1724	-0.315	-1.3923	0.4515
	SWAD	-1.643	-1.3091	-3.7142	-6.2034	-0.7448
	GSA	0.7552	0.5256	0.2888	0.3536	0.7569

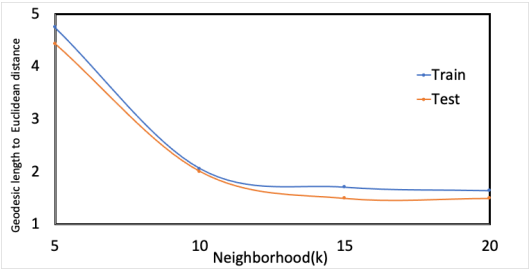


Figure 11: The line graph shows that as K increases the Geodesic to Euclidean distance decreases and stabilizes.

10 REPRODUCIBILITY

All experiments were run with fixed random seeds and deterministic dataloader ordering. Hyperparameters were selected via grid search and tuned across 5 independent runs. Unsupervised pretraining used a batch size of 128 for 25 epochs with AdamW with lr $1x10^{-4}$ and weight decay $1x10^{-4}$. The scheduler has a 5% linear warmup followed by a constant learning rate and gradient clipping at 1.0. Supervised fine tuning trained only the property predictor for 100 epochs using AdamW with the same configuration. The model is trained using MSE loss with a stepLR scheduler with step size of 10, gradient clipping at 1.0 and early stopping with patience 15. Geodesic-weighted fine tuning used 100 epochs, with a stronger regularization with weight decay $1x10^{-}2$, early stopping with patience 10. For the computation of geodesic weights, we used KNN graphs with k=15,20, farthest anchors, mutual neighbors and a 90th percentile trust threshold on finite distances. All hyperparameters were tuned using the validation split, final metrics correspond to the best performing hyperparameter configuration. The value of K is selected based on graph properties (connectivity and stability) computed on in-distribution data only. Fig. 11 shows that around k=15 the ratio of geodesic distance to Euclidean distance stabilizes.