RELIABLE SURROGATE MODELING OF PDES VIA PHYSICS-INFORMED CONFORMAL PREDICTION ON GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Deploying Deep Learning (DL) surrogate models in computational physics requires adherence to physical laws, generalization across diverse topologies of real-world applications, and the ability to quantify uncertainty. The high computational cost of traditional numerical solvers motivated the shift toward Artificial Intelligence (AI). Although Physics-Informed Neural Networks (PINNs) have enabled notable acceleration, the reliance on continuous automatic differentiation with respect to absolute coordinates removes spatial invariance and their ability to generalize to novel geometries. Graph Neural Network (GNN) and Geometric Deep Learning (GDL) surrogate models have addressed this by processing relative topological and graph-based features. Yet, the performance of standard message-passing architectures degrades under major distribution shifts in real-world deployment. Also, neither PINNs nor GNNs inherently quantify uncertainty, making them unreliable for safety-critical applications. To tackle these limitations, we propose the Physics-Informed Conformal Stacked Residual Network (PI-CSRN), based on a multi-resolution architecture with topology-preserving clustering and anisotropic, attention-driven message passing. PDEs are integrated into the training objectives of the surrogate models via discrete differential operators to preserve spatial invariance. Uncertainty estimation is provided via Conformal Prediction, with a residual learning mechanism to improve performance on novel manifolds. A custom benchmark of thousands of complex and diverse auto-generated 2D and 3D manifolds ensures the proper evaluation of surrogate models, using electrostatics and elasticity as the foundational physics. The proposed architecture achieves high accuracy, robustness to distribution shifts, and uncertainty quantification, thus offering a reliable blueprint for deployable AI physics surrogates.

1 INTRODUCTION

A major change in computational physics is the shift from traditional numerical solvers to PINNs and GDL models (Chenaud et al., 2024; Karniadakis et al., 2021). Conventional approaches used for the numerical solution of PDEs in computational physics, like the Finite Element Method (FEM) (Zienkiewicz et al., 2013), offer high mathematical accuracy. However, their computational cost and the need for high-quality discretization make them impractical for real-time deployment and fast prototyping purposes (Hughes et al., 2005; Benner et al., 2015). Therefore, the scientific community is increasingly turning to DL to construct computationally efficient surrogate models for PDEs (Hesthaven & Ubbiali, 2018; Boussif et al., 2022; Li et al., 2021).

To address these limitations of numerical solvers, PINNs were introduced, which embed governing differential equations into their training objectives (Raissi et al., 2019; Mishra & Molinaro, 2023). However, their reliance on automatic differentiation with respect to absolute spatial coordinates reduces the spatial invariance needed to extrapolate across novel manifolds (Wang et al., 2022; Franco et al., 2023), thus limiting their generalization power. In contrast, GNNs and other GDL architectures can achieve spatial invariance by operating on relative topological features (Gladstone et al., 2024). However, standard message-passing GNNs lack the global receptive field needed to model complex PDEs (Li et al., 2018), often leading to representation collapse and over-smoothing, a limitation that multi-resolution architectures have addressed (Pfaff et al., 2021; Tan et al., 2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

More critically, these models lack the uncertainty quantification required for reliable predictions in safety-critical engineering applications (Bülte et al., 2025).

The conventional methods of uncertainty quantification, such as Bayesian Neural Networks (Blundell et al., 2015; Gal & Ghahramani, 2016), suffer from additional computational costs or lack theoretical guarantees on coverage. Conformal Prediction (CP) addresses this problem by wrapping the prediction within distribution-free and model-agnostic confidence intervals (Huang et al., 2023; Gopakumar et al., 2025; Cini et al., 2025), providing valuable insights into the predictive uncertainty. The recent research in surrogate modeling of PDEs has allowed applying CP to complex physical surrogate models, based on either neural or graph neural networks, emphasizing the importance of this technique in safety-critical applications (Gopakumar et al., 2026; Ma et al., 2024). However, the primary theoretical assumption of CP of data exchangeability, required for marginal coverage, is not satisfied when applied to structurally different and novel geometries (Barber et al., 2023; Davis et al., 2025), while the existing studies typically consider models only in limited or in-distribution geometries.

To bridge these gaps, we propose the Physics-Informed Conformal Stacked Residual Network (PI-CSRN), a unified, shift-robust graph architecture trained and evaluated on a procedurally generated benchmark dataset of thousands of diverse manifolds, specifically designed to enforce the learning of topological invariants. At the foundational level, our architecture introduces a Multi-Resolution Graph U-Net (Gao & Ji, 2019) that combines topology-preserving geometric clustering with anisotropic attention-driven message passing to capture multi-resolution physical dynamics. Governing physical laws are integrated via the discrete differential operators (Sharp & Crane, 2020) that enforce the underlying PDE constraints directly in the training objectives, while preserving spatial invariance. Finally, to mitigate the breakdown of standard conformal guarantees under geometric covariate shift and the collapse of data exchangeability, we augment Conformalized Quantile Regression (CQR) (Romano et al., 2019) with a Stacked Residual Corrector. By learning to predict localized residual deltas, especially in regions of high epistemic uncertainty in novel manifolds, this Stacked Residual network increases robustness under covariate shift and provides a reliable solution for deployable AI physics surrogates in real-world applications.

2 Multi-Resolution Graph U-Net Architecture

The proposed Graph U-Net introduces a multi-resolution, topology-preserving hierarchical graph architecture with anisotropic attention-driven message passing that mimics the approach of standard multi-grid solvers. This architecture utilizes local features occurring at the fine-grained mesh resolution while simultaneously aggregating global influences at coarsened topological levels. By passing only relative spatial and graph-based features to the surrogate model, it maintains the required spatial invariance. This design overcomes the fundamental limitations of standard Graph Neural Networks, which face severe constraints when applied to complex irregular topologies. Specifically, as network depth increases to capture the global receptive fields needed to model PDEs, these architectures suffer from over-smoothing and representation collapse, as node embeddings converge to indistinguishable local averages.

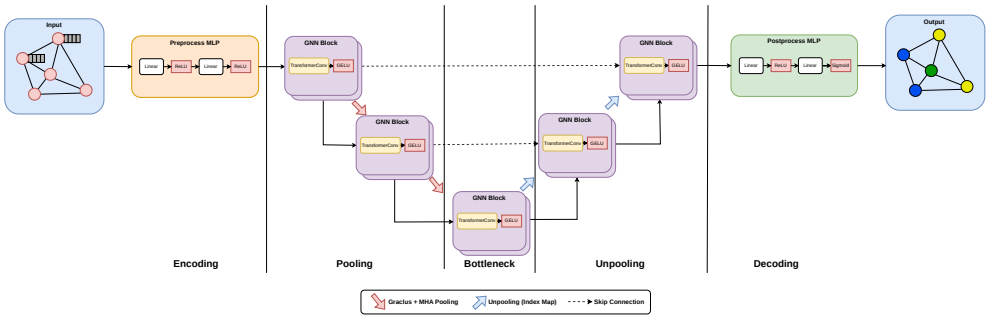


Figure 1: Multi-resolution and topology-preserving Graph U-Net architecture.

2.1 Geodesic-Preserving Coarsening and Anisotropic Message Passing

Preserving the manifold's topology is essential during hierarchical graph pooling. To ensure manifold integrity, the architecture employs the Graclus algorithm (Dhillon et al., 2007). The clustering edge weights w_{ij} are formulated as the inverse of the normalized ℓ_2 distance between the interacting nodes, ensuring that coarsened macro-elements retain the native geodesic connectivity of the input mesh. This strategy avoids the aggressive node removal typical of conventional pooling strategies that do not account for topology, which destroys the geodesic connectivity required for continuous PDE solutions.

$$w_{ij} = \frac{1}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 + \epsilon}. \tag{1}$$

Once the discrete topology is preserved, the next crucial step is to prevent representation collapse. To achieve this, static pooling operations are substituted with a learnable Multi-Head Attention Pooling mechanism. It utilizes a global query parameter, the seed vector $\boldsymbol{s} \in \mathbb{R}^{1\times H}$, to aggregate node features within the Graclus macro-clusters via a cluster-wise softmax operation. This attention-driven formulation selectively promotes salient topological features while explicitly bypassing feature homogenization, thereby preventing the loss of local variance inherent to conventional pooling strategies.

To further counteract over-smoothing during the intra-level localized message-passing phase, standard isotropic graph convolutions are replaced with anisotropic Transformer convolution layers (Veličković et al., 2018; Brandstetter et al., 2022; Karkaria et al., 2025). Standard convolutions treat features uniformly, inherently failing to capture the non-uniform physical properties present in complex PDEs. Anisotropic mechanisms dynamically promote features by weighting the influence of neighboring nodes based on learned similarities among hidden features. For a given node i, the updated feature representation $\boldsymbol{h}_i'$ is computed as the concatenation of K independent attention heads (Lee et al., 2019).

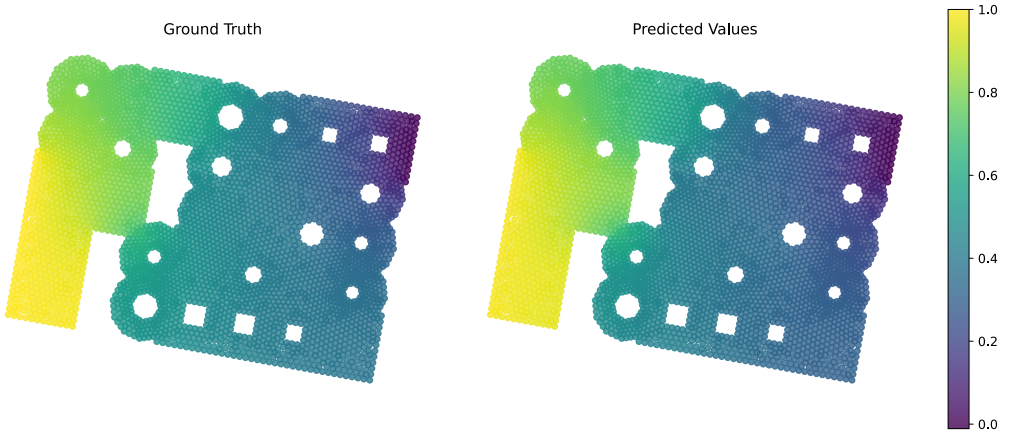


Figure 2: (Left) Target potential from the ElmerFEM electrostatic solver. (Right) Predicted potential from the proposed architecture.

This architecture defines a fully learnable, geometry-aware message-passing mechanism. By separating physical processes such as strong gradients near complex boundaries within individual attention heads, the architecture automatically prevents the homogenization of node embeddings. As a result, the non-linear physical dynamics can be preserved and propagated through the deep hierarchy of the network.

3 PHYSICS-INFORMED AND SHIFT-ROBUST UNCERTAINTY QUANTIFICATION

The proposed architecture directly integrates Conformal Prediction, a distribution-free and model-agnostic framework, to provide the predictive uncertainty quantification required for adopting and trusting deep learning surrogates in computational physics. This formulation overcomes the limitations of conventional probabilistic methodologies, such as Bayesian Neural Networks, which either incur high computational overhead or lack coverage guarantees.

3.1 GRAPH U-NET WITH CONFORMALIZED QUANTILE REGRESSION

The architecture utilizes Conformalized Quantile Regression (CQR) (Mollaali et al., 2026) to provide quantile regression with the strict marginal coverage guarantees of CP. For a specified target level of miscoverage $\alpha \in (0, 1)$ that corresponds to a confidence level of $1 - \alpha$, the network uses a custom multi-head output layer. This layer produces three distinct predictions for each node of the graph, a lower conditional quantile $\hat{q}_{\text{low}}$ targeting the $\alpha/2$ quantile, an upper conditional quantile estimate $\hat{q}_{\text{high}}$ targeting the $1 - \alpha/2$ quantile, and a central prediction $\hat{y}$.

During training, these quantile estimators are optimized using an asymmetric pinball loss. For any target quantile level $\tau \in (0, 1)$, given a ground truth physical state y and predicted quantile $\hat{q}$, the pinball loss is defined as:

$$L_\tau(y, \hat{q}) = \max\{\tau(y - \hat{q}),\ (\tau - 1)(y - \hat{q})\}\,. \tag{2}$$

Data loss in the N graph nodes aggregates the lower and upper quantile penalties with a Mean Squared Error (MSE) regularization term, designed to stabilize the central point prediction.

$$\mathcal{L}_{\text{data}} = \frac{1}{N}\sum_{i=1}^{N}\left[L_{\alpha/2}(y_i, \hat{q}_{\text{low},i}) + L_{1-\alpha/2}(y_i, \hat{q}_{\text{high},i}) + \lambda_{\text{mse}}(y_i - \hat{y}_i)^2\right]. \tag{3}$$

For theoretical marginal coverage of CP, a supplementary calibration set is needed. Based on the calibration set $\mathcal{D}_{\text{calib}}$ evaluation, the nonconformity score $s_i = \max(\hat{q}_{\text{low},i} - y_i,\ y_i - \hat{q}_{\text{high},i})$ is computed for each individual node of the graph in order to evaluate the discrepancy between the actual value and the predicted interval borders. Then, the calibration shift scalar, $\hat{\eta}$, is calculated using a predefined upper empirical quantile of the calibration scores. Symmetrically applying $\hat{\eta}$ to the predicted boundaries guarantees that the ground truth lies within the adjusted interval with a probability greater than or equal to $1 - \alpha$ (Fontana et al., 2023; Zargarbashi et al., 2023).

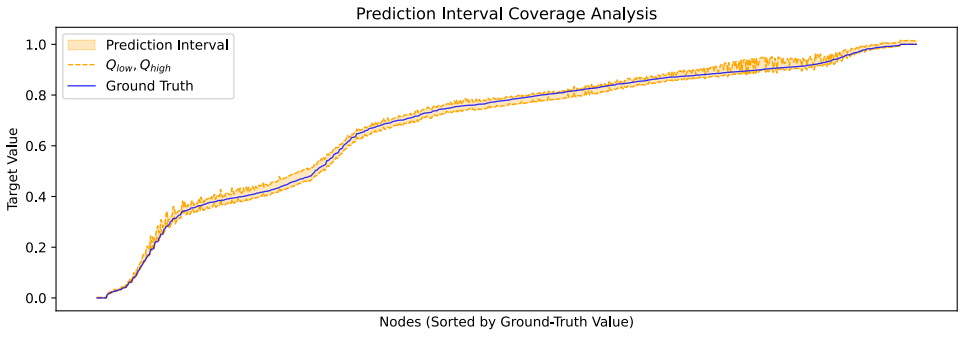


Figure 3: Coverage and prediction intervals of Conformalized Quantile Regression.

3.2 DISCRETE PHYSICS-INFORMED REGULARIZATION

To further restrict the predictions of the model to follow the governing physics, a multi-objective loss formulation is used. Two extra loss terms are added to inject physics information into the training objectives.

$$\mathcal{L}_{\text{total}} = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{pde}}\mathcal{L}_{\text{pde}} + \lambda_{\text{bc}}\mathcal{L}_{\text{bc}}. \tag{4}$$

The boundary condition penalty $\mathcal{L}_{\text{bc}}$ enforces the model to follow the known values on input boundary nodes through a boolean mask. However, formulating the loss term for the PDE, $\mathcal{L}_{\text{pde}}$, poses a considerable challenge. Traditional PINNs map PDE constraints using continuous automatic differentiation with respect to absolute spatial coordinates. Because the proposed framework operates exclusively on relative features to maintain spatial invariance, continuous automatic differentiation is incompatible with it. This is resolved by embedding discrete differential operators directly into the graph topology (Liao et al., 2025) and integrating the governing PDE directly in the loss term, acting as a physics-based regularization term. For a node i that is not configured as a boundary node, the discrete first-order gradient $(\nabla\hat{\phi})_i$ and the discrete Laplacian $(\Delta\hat{\phi})_i$ for 2D manifolds are approximated using the following formulations (Sharp & Crane, 2020):

$$(\nabla\hat{\phi})_i = \frac{1}{A_i}\sum_{f\in\mathcal{F}(i)}\frac{A_f}{3}(\nabla\hat{\phi})_f, \quad (\nabla\hat{\phi})_f = \frac{1}{2A_f}\sum_{j\in\mathcal{V}(f)}\hat{\phi}_j(\boldsymbol{n}_f\times\boldsymbol{e}_j), \tag{5}$$

$$(\Delta\hat{\phi})_i = \frac{1}{2A_i}\sum_{j\in\mathcal{N}(i)}(\cot\alpha_{ij}+\cot\beta_{ij})(\hat{\phi}_j-\hat{\phi}_i). \tag{6}$$

In this case, A_i corresponds to the barycentric dual area of node i, A_f to the area of the incident face f, while $\boldsymbol{n}_f$ represents the unit normal vector of face f. The vector $\boldsymbol{e}_j$ denotes the edge of face f opposite vertex j. Terms α_{ij} and β_{ij} correspond to the interior angles opposite the edge of nodes (i,j) in the incident triangles, where cot denotes the standard cotangent function. While the equations above define the operators for 2D manifolds, for 3D volumetric domains these operators are generalized using standard tetrahedral formulations (Crane, 2020). Therefore, the network is forced to produce predictions that adhere to the governing physics. Crucially, this physics-based regularization term makes use of purely relative geometric invariants, respecting the rigid-body spatial invariance requirement of the architecture.

3.3 Resolving Covariate Shift via Stacked Residual Generalization

To achieve the requested coverage in CP, the data exchangeability criterion must hold. When neural PDE solvers are deployed onto out-of-distribution topological manifolds that differ drastically from the training and calibration sets, this assumption is violated. This manifests as a severe degradation in theoretical marginal coverage in the test set (Tibshirani et al., 2019; Zhao et al., 2024).

To mitigate this coverage degradation and increase robustness under covariate shift, an architecture based on Stacked Residual Generalization is introduced that effectively captures localized predictive discrepancies to counteract the effects of non-exchangeability (Breiman, 1996; Marques, 2025; Song et al., 2024). Specifically, a dedicated secondary hierarchical network is trained to function as a self-calibrating post-processor by leveraging an augmented feature representation formed by concatenating the input graph features with the predictions and uncertainty intervals of the frozen ensemble of Graph U-Nets.

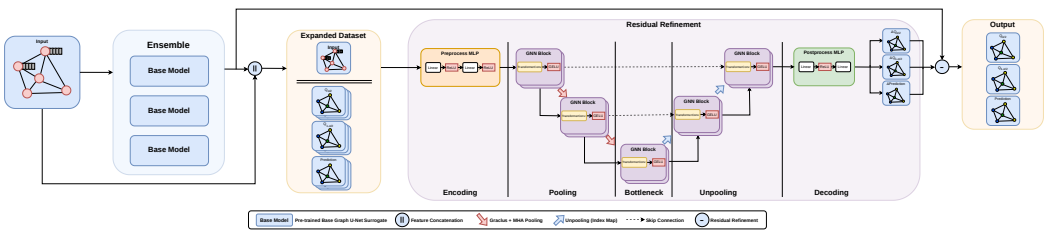


Figure 4: Physics-Informed Conformal Stacked Residual Network Architecture (PI-CSRN).

Rather than estimating absolute quantities, the secondary architecture generates residual corrections (Δq_{low}, Δq_{high}) to adjust the ensemble's baseline prediction intervals. The result is an in-

crease in uncertainty intervals in local regions with high complexity and epistemic uncertainty, and a decrease in uncertainty intervals across locally homogeneous and previously seen regions. Consequently, uncertainty quantification provides a twofold advantage. First and most critically, it enables a risk-aware deployment, where high-uncertainty topologies can be selectively routed to high-fidelity numerical solvers while safely relying on surrogate inference elsewhere. Second, the ensemble's predictive uncertainty is an informative feature for the stacked architecture, allowing the residual corrector to dynamically tune its localized refinements.

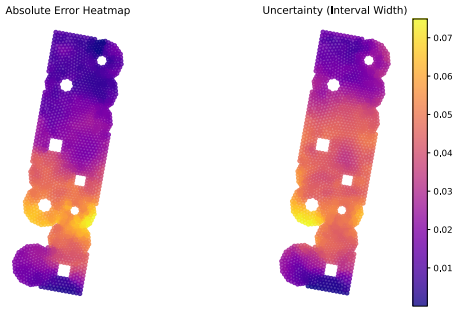


Figure 5: (Left) Absolute prediction errors on a novel manifold. (Right) Uncertainty interval widths.

4 A Large-Scale Graph Benchmark for Geometry Generalization

The robust evaluation of DL surrogates for PDEs is strictly bound by the geometric diversity of their training and testing distributions and their capacity for generalization. As noted, while PINNs demonstrate high interpolation capabilities within fixed spatial domains, they lack spatial invariance. On the other hand, GNNs can preserve spatial invariance, yet often require novel strategies for out-of-distribution (OOD) generalization (Nguyen et al., 2026; Fan et al., 2024; Li et al., 2025). Thus, assessing the generalization capabilities of these surrogate models requires evaluation beyond limited, in-distribution topologies. To address this, a novel large-scale graph dataset has been created to benchmark OOD generalization for continuous PDEs on complex, arbitrary manifolds. This benchmark, which includes over ten thousand unique simulations, deliberately omits absolute spatial features, forcing surrogate models to capture underlying physical laws exclusively from relative topological and graph-based features (Horie & Mitsume, 2022).

4.1 Procedural Synthesis and Finite Element Discretization

A fully automated generation pipeline is implemented to produce a diverse and unbiased dataset for evaluating surrogate models. Complex two and three dimensional continuous manifolds are synthesized by distributing and uniting geometric primitives across a grid space. Also, stochastic scaling factors are used to resize primitives, and dropout probabilities are used to remove arbitrary grid cells. Boolean subtract operations enabled the pipeline to generate arbitrary continuous manifolds with diverse topological characteristics, sharp corners, and irregular concavities. The generated benchmark is performed on various physical phenomena and geometric dimensions to assess the suitability of architectures as PDE surrogate models. Two data sets are created for two-dimensional electrostatic and three-dimensional elastic problems to demonstrate invariance to both the underlying PDE and the spatial dimensionality.

In the electrostatic problem, the domain Ω represents the continuous space. The unknown quantity is the scalar electric potential ϕ within the spatial domain Ω. The two subdomains on the boundary, Γ_1 and Γ_2, serve as the input ports with fixed boundary conditions. The PDE formulation is:

$$\Delta\phi(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \Omega, \tag{7}$$
$$\phi(\boldsymbol{x}) = 1 \quad \forall \boldsymbol{x} \in \Gamma_1, \tag{8}$$
$$\phi(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \Gamma_2, \tag{9}$$
$$\frac{\partial \phi}{\partial \boldsymbol{n}}(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega \setminus (\Gamma_1 \cup \Gamma_2). \tag{10}$$

The unknown in the elastic problem is the displacement vector field $\boldsymbol{u}$. The PDE formulation of the elasticity case is presented below.

$$\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{u})(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \Omega, \tag{11}$$
$$\boldsymbol{\sigma}(\boldsymbol{u}) = \lambda \operatorname{tr}(\boldsymbol{\varepsilon})\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T), \tag{12}$$
$$\boldsymbol{u}(\boldsymbol{x}) = [-1, 0, 0]^\top \quad \forall \boldsymbol{x} \in \Gamma_1, \tag{13}$$
$$\boldsymbol{u}(\boldsymbol{x}) = [1, 0, 0]^\top \quad \forall \boldsymbol{x} \in \Gamma_2, \tag{14}$$
$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}}(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega \setminus (\Gamma_1 \cup \Gamma_2). \tag{15}$$

After the procedural synthesis, these continuous domains are subjected to discretization. The Gmsh library (Geuzaine & Remacle, 2009) performs Delaunay conformal triangulation, and each PDE formulation is solved using the ElmerFEM and PyElmer multiphysics suite (Malinen & Råback, 2013).

4.2 Feature Engineering for Invariant Representation

To learn spatially invariant representations with the GNN model surrogates, the feature extraction process selects only relative topological and graph features. For the node feature matrix $\boldsymbol{X} \in \mathbb{R}^{|\mathcal{V}| \times F_n}$, a full range of spatially invariant scalar attributes is extracted. The feature vectors include the normalized geodesic distance from the nodes to the boundary conditions, the local mesh density, and graph properties such as centrality and clustering coefficient. The edge feature matrix $\boldsymbol{E} \in \mathbb{R}^{|\mathcal{E}| \times F_e}$ also contains only relative spatial features, such as normalized Euclidean edge length. Also, binary topological attributes that distinguish assigned boundary conditions for nodes and edges are extracted. Exporting this diverse dataset to a standardized format (Fey & Lenssen, 2019) provides a ready-to-use benchmark for evaluating PDE surrogate models.

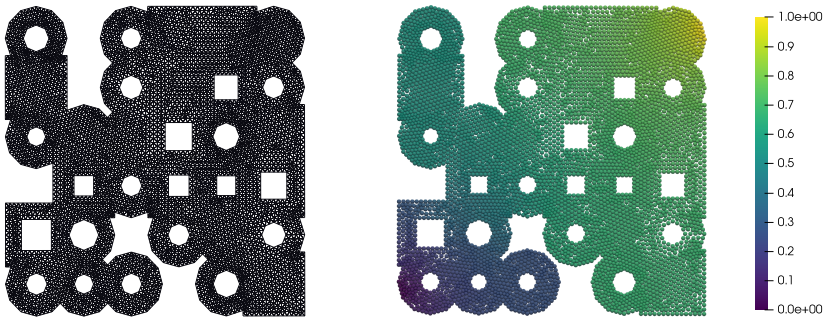


Figure 6: (Left) Discretized 2D mesh topology generated via Gmsh. (Right) Ground truth potential from the ElmerFEM electrostatic solver.

5 Experimental Results and Ablation Study

The comprehensive ablation study presented in Table 1 provides conclusive evidence for the validity of the proposed architecture compared to traditional message-passing methodologies on the

generated 2D electrostatic benchmark. Classical message-passing approaches such as GraphSAGE (Hamilton et al., 2017) achieve a Test Mean Absolute Error (MAE) of 0.0762 on unseen manifolds, which highlights their inability to model global dynamics due to the limited receptive field. The topology-preserving Graph U-Net gives a Test MAE of 0.018. This demonstrates the capability of the proposed geodesic-preserving coarsening and anisotropic message-passing to propagate long-range dynamics. The increase in Test MAE from 0.0180 to 0.0250 in the Conformal Anisotropic U-Net represents an expected multi-objective trade-off, as the additional Pinball loss term is introduced.

Moreover, the results suggest the susceptibility of conformal prediction under conditions of covariate shift, as the loss of data exchangeability in the test set causes the empirical coverage to drop to 0.74, significantly below the targeted 0.9 level. This problem is addressed via the Stacked Residual Generalization strategy. In particular, Stacked Residual GNN serves as a domain-adapted epistemic uncertainty calibrator that can perform residual correction on OOD topologies. Furthermore, the Physics-Informed multi-objective loss outperforms the standalone Pinball loss across the evaluation metrics of Test MAE, empirical coverage, and uncertainty interval width. The joint effect of residual correction and Physics-Informed regularization in PI-CSRN results in a Test MAE of 0.0159 and raises coverage to 0.901 while retaining interval widths.

Table 1: Ablation study evaluating architectural configurations, predictive accuracy, and uncertainty robustness across the diverse test manifolds. All models were optimized using Adam with an initial learning rate of 0.0005, decayed on plateau, for a maximum of 1000 epochs, governed by an early stopping criterion on validation loss, and were evaluated at the optimal validation checkpoint. **N/A** denotes inapplicable metrics. The final four rows represent the Stacked Conformal Prediction framework. † denotes experiments where the Stacked Residual GNN was explicitly fine-tuned on a small, disjoint dataset generated under a slightly perturbed configuration relative to the training set to enhance robustness under covariate shift. Optimal performance metrics for the test set are highlighted in **bold**.

Architecture	**Depth**	**Loss**	**MAE**			**Uncertainty**	
			Train	**Val**	**Test**	**Coverage**	**Width**
Simple NN	5	MSE	0.0840	0.0850	0.0870	N/A	N/A
SAGE GNN	5	MSE	0.0680	0.0700	0.0762	N/A	N/A
Residual SAGE GNN	5	MSE	0.0690	0.0700	0.0780	N/A	N/A
Residual Transformer GNN	5	MSE	0.0720	0.0730	0.0770	N/A	N/A
Isotropic U-Net GNN	10	MSE	0.0210	0.0290	0.0350	N/A	N/A
Anisotropic U-Net GNN	10	MSE	0.0070	0.0100	0.0180	N/A	N/A
Conformal Anisotropic U-Net GNN	10	Pinball	0.0105	0.0119	0.0250	0.740	0.066
Conformal Anisotropic U-Net GNN	10	Physics-Pinball	0.0120	0.0130	0.0220	0.800	0.064
Conformal Anisotropic Ensemble GNN	10	Physics-Pinball	N/A	N/A	0.0240	0.870	0.075
Conformal Stacked Residual GNN	10	Physics-Pinball	0.0190	0.0128	0.0215	0.790	**0.061**
Conformal Stacked Residual Ensemble GNN	10	Physics-Pinball	0.0118	0.0125	0.0208	0.840	0.066
Conformal Stacked Residual GNN†	10	Physics-Pinball	0.0109	0.0122	0.0197	0.890	0.063
Conformal Stacked Residual Ensemble GNN†	10	Physics-Pinball	0.0094	0.0107	**0.0159**	**0.901**	0.065

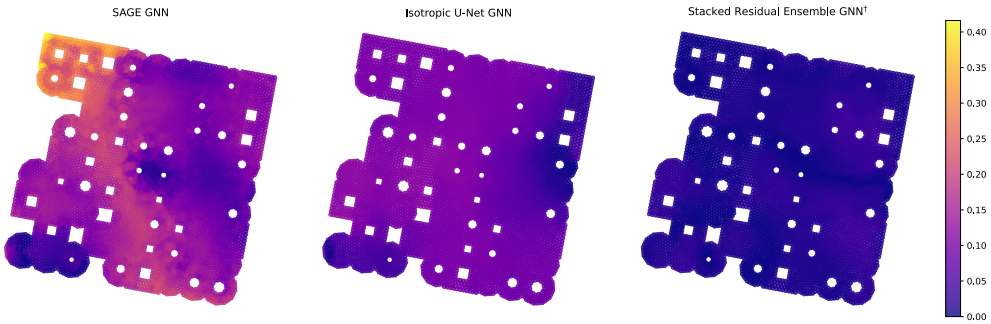


Figure 7: Absolute prediction error across evaluated architectures. (Left) SAGE GNN. (Middle) Isotropic U-Net GNN. (Right) Conformal Stacked Residual Ensemble GNN† (PI-CSRN).

To examine the proposed architecture's ability to generalize to a variety of physical problems and dimensions, PI-CSRN was also trained on a three-dimensional elastic problem. The progression from the electric potential to a multi-variable 3D displacement field explicitly tests the capability of the model to capture a variety of PDE dynamics within complex physical domains. The Conformal Stacked Residual Ensemble GNN† (PI-CSRN) showed high robustness when tested on the procedurally generated benchmark of 3D manifolds, achieving a Test MAE of 0.034 and an MSE of 0.0043.

The framework also achieved 90.5% empirical coverage under significant distributional shift, with a mean interval width of 0.13. These results prove that the PI-CSRN serves as a highly adaptable PDE surrogate capable of handling various physics and spatial dimensions.

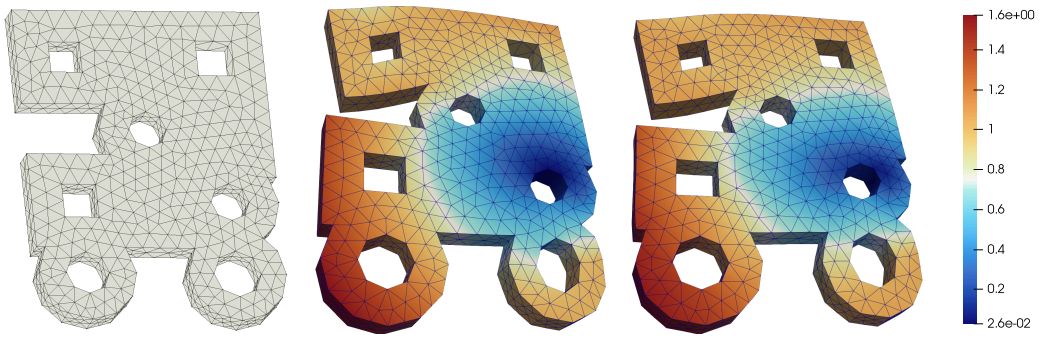


Figure 8: (Left) Discretized 3D mesh topology via Gmsh. (Middle) Target displacement from ElmerFEM. (Right) Predicted displacement via PI-CSRN.

6 Concluding Remarks

The introduction of the Physics-Informed Conformal Stacked Residual Network systematically addresses several critical vulnerabilities of modern neural PDE surrogates, including sensitivity to topological covariate shift, lack of spatial invariance, and the absence of uncertainty quantification.

Generalization capability was validated using a novel large-scale benchmark dataset, leading the graph models to learn relative invariants across thousands of diverse manifolds. Building on this foundation, the proposed Multi-Resolution Graph U-Net incorporates geodesic-preserving clustering and anisotropic message passing, which help eliminate the representation collapse problem faced by standard deep GNNs. Incorporating the discrete differential operators of the governing PDEs into its training objectives allows the method to satisfy the underlying physics while preserving spatial invariance. In addition, Conformalized Quantile Regression has been used to provide rigorous confidence intervals valuable for safety-critical applications, and the Stacked Residual Generalization of the final PI-CSRN architecture increases robustness to covariate shift.

The present study applies the electrostatic and elastic problem across various dimensions, validating the applicability of cross-physics generalization. Future development of the present methodology could include applications to highly heterogeneous and multi-material domains. Another important challenge is the transition from static elliptic PDEs to nonlinear, time-dependent continuous systems, such as the Navier-Stokes equations in fluid dynamics. Ultimately, this approach establishes a solid foundation for developing reliable surrogate models for PDEs applicable to real-world engineering environments.

AI use statement

The authors declare that generative AI tools were used exclusively for localized phrasing improvements to the manuscript and preliminary auxiliary code prototyping. All AI-assisted work was manually reviewed and verified by the authors. Generative AI was not used to formulate or prove mathematical claims, develop theoretical models or frameworks, propose hypotheses, design experiments, implement core methods, or generate synthetic datasets. All presented architectures,

methodologies, software implementations, and evaluation procedures were conceived, developed, and evaluated independently by the authors, who assume full responsibility for all content presented in this work.

Reproducibility statement

All theoretical foundations and mathematical formulations, including the Graclus graph pooling edge weight (Equation 1), the multi-objective loss formulations (Equations 2–4), the discrete differential operators (Equations 5–6), and the governing boundary value problems (Equations 7–15), are fully specified. The automated benchmark dataset generation pipeline, including the procedural synthesis, Gmsh and ElmerFEM solver utilization, is described in Section 4.1, while the invariant geometric node and edge feature engineering is detailed in Section 4.2. All key training hyperparameters, including the number of epochs, architectural depths, learning rates, early stopping criteria, optimizer settings, and selected loss formulations, are reported in Table 1. To ensure reproducibility, the source code implementing the PI-CSRN architecture, the procedural synthesis scripts along with input configuration files, and the complete benchmark dataset will be made publicly available upon publication.

References

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023.

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1613–1622, 2015.

Oussama Boussif, Dan Assouline, Loubna Benabbou, and Yoshua Bengio. MAgNet: Mesh agnostic neural PDE solver. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, volume 35. Curran Associates, Inc., 2022.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations*, 2022.

Leo Breiman. Stacked regressions. *Machine Learning*, 24(1):49–64, 1996.

Christopher Bülte, Philipp Scholl, and Gitta Kutyniok. Probabilistic neural operators for functional uncertainty quantification. *Transactions on Machine Learning Research*, 2025.

Marien Chenaud, Frédéric Magoulès, and José Alves. Physics-informed graph-mesh networks for PDEs: A hybrid approach for complex problems. *Advances in Engineering Software*, 197, 2024.

Andrea Cini, Alexander Jenkins, Danilo Mandic, Cesare Alippi, and Filippo Maria Bianchi. Relational conformal prediction for correlated time series. In *Forty-second International Conference on Machine Learning*, 2025.

Keenan Crane. *An Excursion Through Discrete Differential Geometry*, volume 76. American Mathematical Society, 2020.

Ed Davis, Ian Gallagher, Daniel Lawson, and Patrick Rubin-Delanchy. Valid conformal prediction for dynamic GNNs. In *The Thirteenth International Conference on Learning Representations*, 2025.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Weighted graph cuts without eigenvectors a multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11): 1944–1957, 2007.

Shaohua Fan, Xiao Wang, Chuan Shi, Peng Cui, and Bai Wang. Generalizing graph neural networks on out-of-distribution graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):322–337, 2024.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with PyTorch Geometric. In *Proceedings of the ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Nicola Rares Franco, Stefania Fresca, Filippo Tombari, and Andrea Manzoni. Deep learning-based surrogate models for parametrized PDEs: Handling geometric variability through graph neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(12):123119, 2023.

Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1050–1059. PMLR, 2016.

Hongyang Gao and Shuiwang Ji. Graph U-Nets. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 2083–2092. PMLR, 2019.

Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.

Rini Jasmine Gladstone, Helia Rahmani, Vishvas Suryakumar, Hadi Meidani, Marta D'Elia, and Ahmad Zareei. Mesh-based GNN surrogates for time-independent PDEs. *Scientific Reports*, 14, 2024.

Vignesh Gopakumar, Ander Gray, Lorenzo Zanisi, Timothy Nunn, Daniel Giles, Matt Kusner, Stanislas Pamela, and Marc Peter Deisenroth. Calibrated physics-informed uncertainty quantification. In *Proceedings of the 42nd International Conference on Machine Learning*. PMLR, 2025.

Vignesh Gopakumar, Ander Gray, Joel Oskarsson, Lorenzo Zanisi, Daniel Giles, Matt J Kusner, Stanislas Pamela, and Marc Peter Deisenroth. Uncertainty quantification of surrogate models using conformal prediction. *Machine Learning: Science and Technology*, 7(1):015025, 2026.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

Jan S. Hesthaven and Stefano Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018.

Masanobu Horie and Naoto Mitsume. Physics-embedded neural networks: Graph neural PDE solvers with mixed boundary conditions. In *36th Conference on Neural Information Processing Systems*, 2022.

Kexin Huang, Ying Jin, Emmanuel Candes, and Jure Leskovec. Uncertainty quantification over graph with conformalized graph neural networks. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Thomas J. R. Hughes, J. Austin Cottrell, and Yuri Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering*, 194(39):4135–4195, 2005.

Vispi Karkaria, Doksoo Lee, Yi-Ping Chen, Yue Yu, and Wei Chen. An attention-based spatio-temporal neural operator for evolving physics. *Machine Learning: Science and Technology*, 2025.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019.

Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Out-of-distribution generalization on graphs: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11): 10490–10512, 2025.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Yunfeng Liao, Yangxin Wu, and Xiucheng Li. Boundary-value PDEs meet higher-order differential topology-aware GNNs. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Ziqi Ma, David Pitt, Kamyar Azizzadenesheli, and Anima Anandkumar. Calibrated uncertainty quantification for operator learning via conformal prediction. *Transactions on Machine Learning Research*, 2024.

Mika Malinen and Peter Råback. *Elmer finite element solver for multiphysics and multiscale problems*, pp. 101–113. CSC-IT Center for Science, 2013.

Paulo Marques. Stacked conformal prediction. In *Proceedings of the Fourteenth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 266 of *Proceedings of Machine Learning Research*, pp. 305–316, 2025.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43(1):1–43, 2023.

Amirhossein Mollaali, Christian Bolivar Moya, Amanda A. Howard, Alexander Heinlein, Panos Stinis, and Guang Lin. Conformalized-KANs: Uncertainty quantification with coverage guarantees for kolmogorov-arnold networks in scientific machine learning. *Machine Learning: Science and Technology*, 2026.

Xuan Minh Vuong Nguyen, Nissrine Akkari, Fabien Casenave, Jonathan Viquerat, and Elie Hachem. Error-driven graph augmentation for mesh-based PDE surrogates. In *Forty-third International Conference on Machine Learning*, 2026.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021.

Maziar Raissi, Paris Perdikaris, and George Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.

Nicholas Sharp and Keenan Crane. A Laplacian for Nonmanifold triangle meshes. *Computer Graphics Forum*, 39(5):69–80, 2020.

Jianqing Song, Jianguo Huang, Wenyu Jiang, Baoming Zhang, Shuangjie Li, and Chongjun Wang. Similarity-navigated conformal prediction for graph neural networks. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2024.

Yingjie Tan, Quanming Yao, and Yaqing Wang. DGNet: Discrete green networks for data-efficient learning of spatiotemporal PDEs. In *The Fourteenth International Conference on Learning Representations*, 2026.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. *Conformal prediction under covariate shift*, volume 32, chapter 227. Curran Associates Inc., 2019.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.

Soroush H Zargarbashi, Simone Antonelli, and Aleksandar Bojchevski. Conformal prediction sets for graph neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 12292–12318, 2023.

Tianyi Zhao, Jian Kang, and Lu Cheng. Conformalized link prediction on graph neural networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024.

Olgierd Cecil Zienkiewicz, Robert Leroy Taylor, and Jian Z Zhu. *The Finite Element Method: Its Basis and Fundamentals*. Butterworth-Heinemann, 2013.