Bayesian Surrogates for Risk-Aware Pre-Assessment of Aging Bridge Portfolios

Sophia V. Kuhn[1,3,*] **Rafael Bischof**[2,3] **Marius Weber**[1] **Antoine Binggeli**[1]
Michael A. Kraus[4] **Walter Kaufmann**[1,3] **Fernando Pérez-Cruz**[4,5]

[1]Institute of Structural Engineering, ETH Zurich, Switzerland
[2]Computational Design Lab, ETH Zurich, Switzerland
[3]Design ++, ETH Zurich, Switzerland
[4]Institute of Structural Mechanics and Design, TU Darmstadt, Germany
[5]Department of Computer Science, ETH Zurich, Switzerland
[6]Bank for International Settlements, Switzerland
*Correspondence to `kuhnso@ethz.ch`

Abstract

Aging infrastructure portfolios pose a critical resource allocation challenge: deciding which structures require intervention and which can safely remain in service. Today's structural assessment approaches do not scale to portfolio level, as they require time-consuming manual digital modeling and computationally expensive simulations for each structure. We propose Bayesian neural network (BNN) surrogates for rapid structural pre-assessment of worldwide common bridge types, such as reinforced concrete frame bridges. Trained on a large-scale database of non-linear finite element analyses generated via a parametric pipeline and developed based on the Swiss Federal Railway's bridge portfolio, the models accurately and efficiently estimate high-fidelity structural analysis results by predicting code compliance factors with calibrated epistemic uncertainty. Our BNN surrogate enables fast, uncertainty-aware triage: flagging likely critical structures and providing guidance where refined analysis is pertinent. We demonstrate the framework's effectiveness in a real-world case study of a railway underpass, showing its potential to significantly reduce costs and emissions by avoiding unnecessary analyses and physical interventions across entire infrastructure portfolios.

1 Introduction

Bridges are critical components of transportation infrastructure, facilitating connectivity and economic activity. Yet, many of these structures around the world are aging, with a large share now operating close to or beyond their intended service life [3]. This presents asset managers with a difficult and resource intensive challenge: deciding which structures require immediate intervention, and how extensive those interventions should be. Accurate structural analysis is central to this process to determine whether a bridge meets safety requirements or must be strengthened or replaced.

Current assessment practices follow a tiered, so called levels-of-approximation approach [5]: Starting with simplified, conservative analyses that contain large safety margins to compensate for modeling abstractions. If these methods do not demonstrate compliance suggesting that an intervention is necessary, asset managers can commission more refined simulations such as non-linear finite element analysis (NLFEA). While NLFEA is expensive, time-consuming, and requires detailed inputs and expertise, it can uncover structural reserve capacity, avoiding costly strengthening measures and

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

enabling structures to remain in service beyond their original design life. However, the benefit of such detailed analyses is unknown a priori. In some cases, NLFEA merely confirms the findings of conservative methods at a significant extra cost. Furthermore, the typically high manual modeling and computational effort involved makes current assessment workflows unscalable to portfolio level.

In this work [10], we develop data-driven surrogate models that efficiently estimate refined structural analysis results and quantify the uncertainty associated with each prediction. These uncertainty-aware outputs enable risk-informed decisions that help prioritize allocation of scarce economic, material, and labor resources where they are most needed. Our main contributions are (i) a parametric simulation pipeline that generates a large-scale, high-fidelity dataset of reinforced concrete frame bridges covering common structures of existing infrastructure portfolios; (ii) Bayesian Neural Network (BNN) [6] surrogates trained to predict code compliance factors expected from detailed analyses, with calibrated epistemic uncertainty; and (iii) a triage policy that leverages this predictive distribution to inform decision-making at the portfolio level.

Data-driven surrogates have been widely applied to approximate computationally expensive simulations in engineering, including structural health monitoring and capacity estimation [2, 9, 11, 17]. However, for structural engineering most approaches remain deterministic and lack uncertainty quantification. Probabilistic methods such as Gaussian Processes, deep ensembles, and BNNs provide predictive distributions instead of point estimates [6], with recent studies demonstrating their effectiveness for pavement response prediction [1], bridge integrity assessment [14], and materials modeling [15].

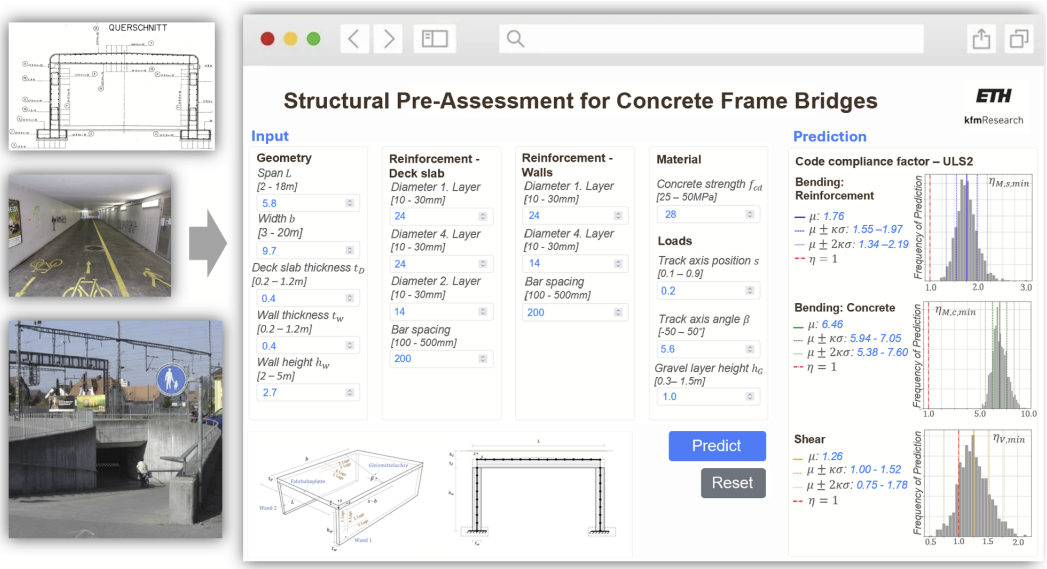

Figure 1: Graphical user interface of the developed ML-based structural pre-assessment tool for reinforced concrete frame bridges. Users input structural parameters (left) and receive rapid code compliance predictions with uncertainty estimates (right).

Table 1: Decision policy for classifying structures based on surrogate predictions.

Class	Condition	Interpretation
• Red	$\mu < 1$	Critical. Prioritize this structure for immediate detailed analysis. Even refined methods may not demonstrate sufficient load-bearing capacity.
• Orange	$\mu > 1$ and $\mu - 2\kappa\sigma < 1$	Likely safe, but uncertainty is high. A refined analysis is recommended, especially if the predicted compliance factor is only slightly above 1.
• Green	$\mu > 1$ and $\mu - 2\kappa\sigma \geq 1$	High confidence in compliance. Not to prioritize. Simplified analysis methods are predicted to be sufficient.

2 Methods

Let $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^k$ denote a set of bridge parameters (geometry, materials, loading, etc.). We define a parametric high-fidelity mechanical analysis model $f(\mathbf{x})$ that returns code-compliance factors $\boldsymbol{\eta}(\mathbf{x})$, defined as resistance-to-demand ratios evaluated for structural safety code verifications. Since evaluating f via materially non-linear finite element analysis (NLFEA) is computationally intensive, we approximate it using a Bayesian Neural Networks (BNNs) trained on a limited number of simulations. The BNN surrogate yields a predictive distribution $p(\hat{y} \mid \mathbf{x}, \mathcal{D})$; its mean μ serves as the best estimate of η, and its standard deviation σ quantifies epistemic uncertainty from limited data / model capacity. Decisions follow the predictive distribution, mapping to the triage policy in Table 1.

Data generation with parametric NLFEA. To obtain training data for supervised learning, we built an automated pipeline that generates models of reinforced concrete frame bridges from a parameter vector, evaluates them with NLFEA, and extracts code-compliance labels. The pipeline links the programs Rhino/Grasshopper (for geometry generation) with Ansys Mechanical APDL (for NLFEA), where the Cracked Membrane Model [7, 16] is implemented as a layered shell element user material. This orchestration is automated via the StrucEngLib plug-in [8]. For each simulation, the pipeline validates structural safety according to Swiss design codes and outputs three global compliance factors $\boldsymbol{\eta}(\mathbf{x}) = (\eta_{M,c,\min}, \eta_{M,s,\min}, \eta_{V,\min})$. These quantify the structural safety margin in bending (for concrete and reinforcing steel) and shear, computed as the minimum resistance-to-demand ratios across all elements, with $\eta \geq 1$ indicating code compliance. Since each NLFEA run is expensive (≈ 25 minutes per bridge, ≈ 2.5 minutes with 10-way parallelism)[1], we adopt a sampling strategy that balances coverage and efficiency. To ensure realistic variability, parameter ranges were derived from a database analysis of the Swiss Federal Railways (SBB) bridge inventory, with expert feedback. Latin Hypercube Sampling (LHS) is used to broadly cover the input space, followed by adaptive resampling based on kernel density estimation to focus simulations near the safety-critical region around $\eta \approx 1$ (cf. Table 3 in [10]). The resulting dataset of ≈ 11k samples spans common structures in the SBB portfolio while providing high resolution near the decision boundary.

Bayesian surrogate. We train three BNNs $\mathcal{B}_\theta$, one per compliance factor, adopting stochastic variational inference, approximating the true posterior as a product of independent Gaussian distributions for each model parameter θ following the mean-field approximation [13] and using Gaussian priors $p(\theta) = \mathcal{N}(0, 0.1I)$. The objective function, derived by maximizing the evidence lower bound (ELBO) [4], combines a KL divergence regularization term with an enhanced, domain-adapted weighted MSLE:

$$L(\theta, D) = \lambda \cdot \mathrm{KL}(q_\phi(\theta) \| p(\theta)) + \mathrm{wMSLE}(y, \mu(\hat{y})) \tag{1}$$

where D is a set of training samples, y is the true value, and $\mu_i(\hat{y})$ is the mean predicted value. The wMSLE heavily penalizes η-values ≈ 1, focusing learning near the decision boundary. The predictive mean μ and uncertainty σ are estimated using 20 forward passes during training and 1000 at inference. In addition to regularization during training, to further improve alignment between predicted and observed uncertainties, we apply post-hoc rescaling of the predictive standard deviations by a constant factor κ. Finally, to enable deployment when detailed structural information is unavailable, we use Kernel SHAP [12] to rank input features by importance and perform reduced-input inference by sampling missing features with LHS and propagating them through the trained BNNs.

3 Results

Model accuracy is evaluated on held-out data using RMSE and MAPE globally and stratified across η ranges to evaluate performance in regions with different importance for decision making. Calibration of the uncertainty quantification is assessed with calibration curves and summary metrics (Total Calibration Error (TCE) and Calibration Bias (CB) [10]).

Decision-focused accuracy. Prediction accuracy and uncertainty estimation are strongest in the safety-critical range $0.5 \leq \eta \leq 1.5$ (Fig. 2). Predictive intervals narrow near $\eta \approx 1$ and widen in uncritical regions, due to our targeted resampling and dedicated loss. Among the targets, $\eta_{M,s,\min}$ is predicted most reliably (RMSE 0.10, MAPE 4.8%), $\eta_{M,c,\min}$ shows moderate error, and $\eta_{V,\min}$ remains most challenging due to the fact that shear is governed by abrupt local effects (cf. Table 2).

[1] Note that in practice, the high cost of NLFEA is caused less by the simulation time itself and more by the manual effort required to build a validated digital model of a specific bridge.

Target	RMSE	MAPE	CB
$\eta_{M,s,\min}$	0.10	4.8%	-0.8
$\eta_{M,c,\min}$	0.40	37.5%	-0.9
$\eta_{V,\min}$	0.60	46.7%	-0.6

Table 2: Predictive performance in the safety-critical region $\eta \in [0.5, 1.5]$. CB reported after post-hoc scaling.

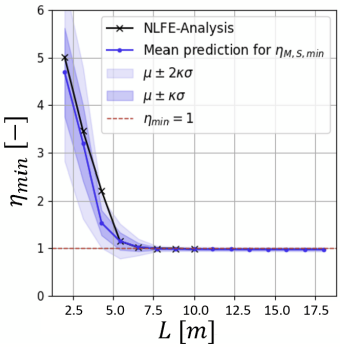

Figure 2: Representative 1D parameter study example comparing BNN predictions with NLFEA.

Calibration of predictive uncertainty. After training, the predictive intervals computed as $\mu \pm z(p)\sigma$, where $z(p)$ denotes the standard normal quantile for a given confidence level p, indicate that the BNN outputs are mildly overconfident in their uncertainty estimates. These intervals capture fewer true values than expected, causing calibration curves to fall below the diagonal and CB to be positive. We therefore apply a post-hoc scale factor per head (chosen on the validation set: $\kappa = 1.3$ for steel, 1.7 for concrete, 2.8 for shear) to align nominal and empirical coverage, reduce TCE, and obtain mildly under-confident intervals (CB < 0, Table 2) on the held-out data, which is desirable in this safety-critical setting. The Gaussian predictive distribution assumption holds well overall, but less so for shear, due to its brittle and locally governed behavior.

Reduced-input deployment. Using only the top five most influential and readily available input features (i.e., span, plate thicknesses, height and width), the reduced-input variant retains most of the predictive accuracy for steel and shear (predominantly governed by geometry) while accuracy for concrete degrades more noticeably, consistent with the SHAP analysis. Uncertainty calibration remains well-aligned, supporting the use of the reduced-input model for early screening, with optional input refinement when higher prediction resolution is required.

Case study: railway underpass. We demonstrate the surrogate on a Swiss railway underpass using the deployed graphical interface (Fig. 1). The model can be queried at multiple levels of input detail: from basic geometric parameters to detailed reinforcement and loading. SHAP-based feature attribution helps prioritize which inputs to acquire next to most effectively reduce uncertainty, enabling progressive refinement. At all input levels, the prediction intervals $\mu \pm 2\kappa\sigma$ cover the NLFEA ground truth, confirming robust calibration. Already with only high-level inputs, the model flags shear as potentially critical, though with high uncertainty. As detailed input is provided, intervals narrow. At full detail, the BNN confidently predicts compliance in bending and identifies shear as the decisive verification (shown in Fig. 1). The triage policy (Table 1) classifies this example structure as orange, and recommends its refined analysis. Subsequent NLFE analysis confirmed compliance and revealed reserve capacity, declaring interventions therefore as unnecessary. In contrast, simplified analysis would have incorrectly triggered strengthening of the structure. For this single structure alone, avoiding such unnecessary intervention saved approximately \$ 370,000 (estimate by SBB AG), as well as substantial material use, CO_2 emissions, and traffic disruptions on and under the overpass, which demonstrates our tool's economic and environmental value for portfolio-level decision-making.

4 Conclusion

We present a scalable, uncertainty-aware surrogate for structural pre-assessment of common bridge types. Trained on high-fidelity NLFEA data, our BNN model efficiently predicts code compliance based on a few structural input parameters with sufficient accuracy and calibrated uncertainty, supporting crucial triage decisions and whether refined analysis is warranted. Demonstrated on a real-world case study, the approach enables portfolio-scale screening even from limited input detail. It helps reduce unnecessary assessments and focuses scarce resources where they are most impactful, a critical need given today's aging infrastructure portfolios.

Acknowledgments and Disclosure of Funding

We acknowledge support and data access from Swiss Federal Railways (SBB). See journal publication [10] for extended acknowledgments.

References

[1] S. Arangio and J. L. Beck. Bayesian neural networks for bridge integrity assessment. *Structural Control and Health Monitoring*, 19(1):3–21, 2012. doi: https://doi.org/10.1002/stc.420.

[2] V. Balmer, S.V. Kuhn, R. Bischof, L. Salamanca, W. Kaufmann, F. Perez-Cruz, and M. A. Kraus. Design space exploration and explanation via conditional variational autoencoders in meta-model-based conceptual design of pedestrian bridges. *Automation in Construction*, 163: 105411, 2024. ISSN 0926-5805. doi: 10.1016/j.autcon.2024.105411.

[3] K. Bergmeister, F. Fingerloos, and J.-D. Wörner, editors. *Beton-Kalender 2010: Brücken, Betonbau im Wasser*. Ernst & Sohn, Berlin, 2009. ISBN 978-3-433-02931-2.

[4] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural networks, 2015. URL `https://arxiv.org/abs/1505.05424`.

[5] fib. *fib Model Code for Concrete Structures (2020)*. Lausanne, Switzerland, 2024. ISBN 978-2-88394-175-5.

[6] J. M. Hernandez-Lobato and R. Adams. Probabilistic backpropagation for scalable learning of bayesian neural networks. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1861–1869, Lille, France, 07–09 Jul 2015. PMLR. URL `https://proceedings.mlr.press/v37/hernandez-lobatoc15.html`.

[7] W. Kaufmann and P. Marti. Structural concrete: Cracked membrane model. *Journal of Structural Engineering*, 124(12):1467–1475, 1998. doi: 10.1061/(ASCE)0733-9445(1998)124:12(1467).

[8] kfmResearch. Strucenglibplugin 0.0.22, 2024. URL `https://strucenglib.ethz.ch/`.

[9] M.A. Kraus, R. Bischof, H. Riedel, et al. Strength lab ai: a mixture-of-experts deep learning approach for limit state analysis and design of monolithic and laminate structures made of glass. *Glass Structures & Engineering*, 9:607–655, 2024. doi: 10.1007/s40940-024-00277-w.

[10] S. V. Kuhn, M. Weber, A. Binggeli, M. A. Kraus, F. Pérez-Cruz, and W.Kaufmann. Predictive structural assessment with bayesian deep learning. *Advanced Engineering Informatics*, 68: 103721, 2025. ISSN 1474-0346. doi: 10.1016/j.aei.2025.103721.

[11] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf`.

[12] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. URL `http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf`.

[13] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press, Cambridge, MA, 2012. ISBN 978-0-262-01802-9.

[14] E. Okte and I. L. Al-Qadi. Prediction of flexible pavement 3-d finite element responses using bayesian neural networks. *International Journal of Pavement Engineering*, 23(14):5066–5076, 2022. doi: 10.1080/10298436.2021.1993219.

[15] A. Olivier, M. D. Shields, and L. Graham-Brady. Bayesian neural networks for uncertainty quantification in data-driven materials modeling. *Computer Methods in Applied Mechanics and Engineering*, 386:114079, 2021. ISSN 0045-7825. doi: 10.1016/j.cma.2021.114079.

[16] K. Thoma, P. Roos, and M. Weber. Finite-elemente-analyse von stahlbetonbauteilen im ebenen spannungszustand. Vol. 109:275–283, 2014. doi: 10.1002/best.201300087.

[17] C. Williams and C. Rasmussen. Gaussian processes for regression. In D. Touretzky, M.C. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8. MIT Press, 1995. URL `https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf`.