Academic Journal of Architecture and Civil Engineering
Print ISSN: 2960-0235
Online ISSN: 2960-0243
DOI: https://doi.org/10.61784/ajace3019

AN INTERPRETABLE, UNCERTAINTY-AWARE FRAMEWORK FOR BRIDGE DETERIORATION PREDICTION AND RISK-BASED MAINTENANCE PRIORITIZATION

ShuXin Zhang[1*], Lei Qiu[2]
[1]*University of California, Berkeley, Berkeley, CA 94720, USA.*
[2]*Ningbo University of Technology, Ningbo 315048, Zhejiang, China.*
[*]*Corresponding Author: ShuXin Zhang*

Abstract: Aging bridge stocks and constrained maintenance budgets make it essential to forecast structural deterioration and to prioritize interventions where they avert the most harm. Most machine-learning pipelines for bridge condition, however, emit point predictions and are coupled to prioritization rules that rank bridges solely by current condition or age, ignoring both predictive uncertainty and the consequences of failure. This paper proposes an integrated, interpretable and uncertainty-aware framework that (i) forecasts future component condition and the probability of entering a Poor state with gradient-boosted trees, (ii) equips the forecasts with distribution-free prediction intervals via split conformal prediction and with calibrated failure probabilities via isotonic regression, (iii) explains the learned deterioration drivers with SHAP, and (iv) ranks bridges with a Bridge Maintenance Priority Index (BMPI) that multiplies a risk-averse, uncertainty-inflated deterioration risk by a consequence score combining traffic exposure, network criticality, detour penalty and asset value. On a 9,800-bridge longitudinal benchmark that reproduces the schema and deterioration behaviour of the U.S. National Bridge Inventory, the classifier attains 0.84 ROC-AUC, conformal intervals achieve near-nominal 90% coverage, and isotonic calibration more than halves the expected calibration error. At a 15% maintenance budget the BMPI averts 51.6% of the consequence-weighted deterioration risk, versus 34.5% for condition-first and 21.6% for age-first ranking. The pipeline is reproducible and transfers directly to operational inventory data.
Keywords: Bridge management systems; Deterioration prediction; Machine learning; Gradient boosting; Conformal prediction; Uncertainty quantification; Probability calibration; SHAP; Maintenance prioritization; Risk-based decision making

1 INTRODUCTION

The United States maintains more than 617,000 highway bridges. The 2021 ASCE Report Card graded the stock a "C," noting that 42% of bridges are at least 50 years old, that about 7.5% (roughly 46,000 bridges) are in Poor condition, and that the bridge-repair backlog exceeds $125 billion [1]. With the average bridge now about 44 years old and rehabilitation proceeding slowly, agencies cannot inspect, repair or replace every deficient structure and must instead allocate scarce funds to the bridges where intervention yields the greatest benefit [1]. Effective allocation rests on two capabilities: predicting how each bridge will deteriorate, and prioritizing among the predicted risks in a way that reflects real-world consequences.

Since the 1970s, biennial inspections have populated the National Bridge Inventory (NBI), in which trained inspectors assign 0–9 condition ratings to a bridge's deck, superstructure and substructure [2]. These ratings underpin national performance reporting and resource allocation. Classical decision support has relied on Markov-chain and hazard-based deterioration models [3-9], while a growing body of work applies machine learning (ML) to NBI data [10-14]. Yet operational prioritization is still dominated by simple heuristics—repair the worst-rated or the oldest bridges first—that ignore how confident a prediction is and how costly a given failure would be.

We identify three gaps. First, most ML condition models output single-point predictions and provide no calibrated measure of predictive uncertainty, even though risk-based asset management is inherently probabilistic. Second, prediction and prioritization are usually treated separately: a model predicts condition, but the ranking that actually drives spending rarely couples predicted risk with consequence factors such as traffic exposure and network criticality. Third, black-box models impede the engineering review and accountability that public agencies require.

This paper addresses these gaps with an integrated framework (Fig. 1) and makes three contributions. (1) A two-stage pipeline that unifies probabilistic deterioration forecasting with distribution-free uncertainty quantification (UQ) and interpretability, using gradient-boosted trees, split conformal prediction and SHAP [15-18]. (2) A risk-based Bridge Maintenance Priority Index (BMPI) that multiplies a calibrated, uncertainty-inflated deterioration risk by a consequence score, so that high-risk, high-consequence and high-uncertainty bridges are elevated together—closing the loop between prediction and decision. (3) A controlled evaluation on a large NBI-schema longitudinal benchmark showing that the BMPI averts substantially more consequence-weighted deterioration than condition-first, age-first or risk-only ranking at equal budget, together with SHAP evidence that the model recovers established deterioration mechanisms.

Because agency-curated multi-year NBI extracts are frequently restricted from redistribution, and to make the study fully reproducible, our experiments use a longitudinal benchmark whose feature schema follows the FHWA coding guide and whose deterioration behaviour is calibrated to published statistics and mechanisms [1,2,4,8,11]. All models are trained and evaluated on this benchmark, and the entire pipeline transfers unchanged to operational inventory data.

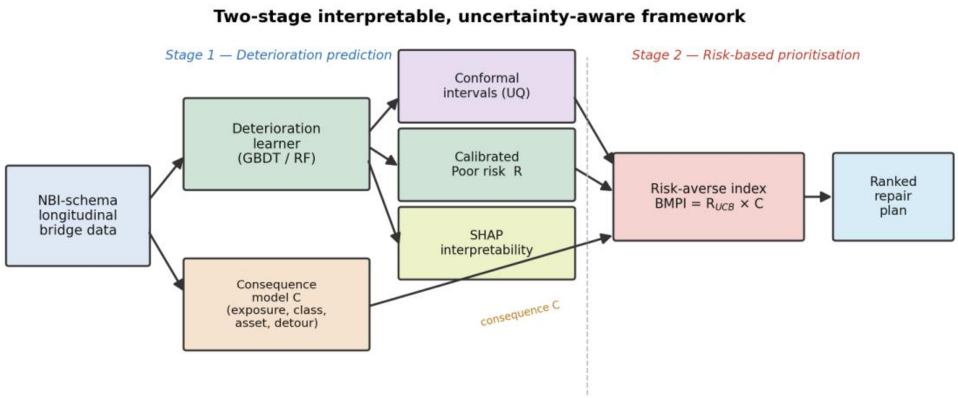

Figure 1 Overview of the Proposed Two-Stage Framework

Stage 1 forecasts deterioration and equips it with conformal intervals, calibrated probabilities and SHAP explanations; Stage 2 combines calibrated risk, uncertainty and consequence into the BMPI ranking.

2 RELATED WORK

2.1 Deterioration Modeling

Markov-chain models dominate classical bridge management, representing condition as discrete states with transition probabilities estimated from historical ratings [3,5,6]. Morcous examined the state-independence and stationarity assumptions of such models for bridge decks [6], and later work extended them to timber and railway elements and to corrosion-initiation and non-linear propagation phases [7-9]. Hazard- and duration-based models relax the discrete-state view: Mauch and Madanat fitted semiparametric hazard-rate models of reinforced-concrete deck deterioration [4]. These models are interpretable and well suited to network-level simulation, but they typically use few covariates, assume stationary transitions, and are not designed to exploit the many structural, traffic and environmental attributes recorded in the NBI.

2.2 Machine Learning for Bridge Condition

ML methods learn flexible, high-dimensional relationships between bridge attributes and condition. Lee et al. used artificial neural networks for backward prediction in a bridge management system [10]; Kim and Yoon identified environmental drivers of deterioration in cold regions [11]; and Bektas et al. applied classification and regression trees to predict NBI ratings [12]. Gradient-boosted trees have proven especially effective on tabular inventory data: Lim and Chi applied XGBoost with SHAP to estimate bridge damage [13], and Assaad and El-adaway compared several learners for deck-deterioration prediction [14]. These studies establish the predictive value of ML on NBI data, but they largely report point predictions, seldom quantify or calibrate predictive uncertainty, and rarely integrate the resulting predictions into an explicit, consequence-aware prioritization rule.

2.3 Uncertainty, Calibration and Interpretability

Conformal prediction produces prediction sets with finite-sample, distribution-free coverage guarantees under exchangeability [15,16]. Inductive (split) conformal prediction and its regression analysis [15,17], along with normalized and quantile-based variants [19], make the approach practical for modern learners. Separately, probability calibration corrects systematically over- or under-confident classifier scores; Platt scaling [20], isotonic regression and their evaluation are standard [21,22]. For interpretability, SHAP attributes a model's output to its inputs with game-theoretic consistency [18]. Despite their maturity, these tools are seldom combined within bridge-deterioration pipelines, and almost never used to inform maintenance prioritization.

2.4 Positioning

Our work differs by integrating all four capabilities—prediction, distribution-free UQ, calibration and interpretability—and, crucially, by feeding their outputs into a single risk-based priority index that also encodes failure consequence. This bridges the persistent gap between accurate condition prediction and defensible, uncertainty-aware maintenance decisions.

3 DATA AND PROBLEM FORMULATION

3.1 The National Bridge Inventory

The NBI records, for every public highway bridge, structural attributes and 0–9 condition ratings for the deck (Item 58), superstructure (Item 59) and substructure (Item 60), where 9 denotes excellent and ratings of 4 or below denote Poor condition [2]. FHWA classifies a bridge as Poor if any of the three component ratings is at most 4. Bridges are re-inspected on a nominal 24-month cycle, yielding, in principle, a longitudinal record of condition evolution.

3.2 A Reproducible NBI-Schema Longitudinal Benchmark

Multi-year, agency-curated NBI extracts are often unavailable for redistribution. We therefore construct a longitudinal benchmark of 10,000 bridges whose schema mirrors the FHWA coding guide and whose deterioration dynamics reproduce documented behaviour [2]. Static attributes (Table 1) are drawn from realistic distributions and include year built, material and design (concrete, steel, prestressed and timber families), number of spans, structure length, deck width, skew, functional class, average daily traffic (ADT), truck percentage (ADTT) and deck protection. Each bridge is aged from construction with a component-specific accumulated-damage process,

$$C_k(a) = 9 - r_k \cdot a^{p_k} \cdot \exp\left(\Sigma_j\, \beta_{kj} \cdot z_j \right) \tag{1}$$

where C_k is the latent condition of component k at age a, the parameters r_k and $p_k > 1$ set the base rate and the (convex) acceleration of deterioration, and the covariates z_j modulate the rate through an exponential factor. The sign and magnitude of each coefficient follow established engineering knowledge: decks deteriorate fastest; truck loading and freeze–thaw together with de-icing-salt or marine chloride exposure accelerate deterioration; prestressed concrete is durable while steel corrodes and timber decays; and deck protection slows deck deterioration [4,8,11]. Stochastic rehabilitation events—more likely as condition worsens—reset the effective age of the affected components, producing the characteristic saw-tooth trajectories seen in real inspection histories. Latent conditions are perturbed by inspection noise and rounded to integers on the 0–9 scale, and observations are sampled on a biennial cycle. The generator is fully seeded; its purpose is a realistic, shareable testbed for the framework, not a claim about any specific bridge population.

Table 1 NBI-Schema Feature Groups in the Benchmark

Group	Representative features (NBI item)
Structural	material/design (43), spans (45), length (49), deck width (52), skew
Traffic	ADT (29), truck % / ADTT (109), functional class (26)
Environment	climate/exposure region, freeze–thaw, chloride exposure
Maintenance	deck protection (108), reconstruction flag/age (106)
Condition	deck/super/sub ratings at t0 and t0−2 (58/59/60), trend
Consequence	ADT exposure, criticality, detour length, deck area

3.3 Realism

The benchmark reproduces salient features of the real inventory. The mean bridge age at the analysis year is 44.4 years, matching the roughly 44-year national average [1]; current deck ratings concentrate at 6–7 (Satisfactory–Good) as in the NBI; and 14.8% of bridges are Poor, consistent with the elevated deficiency rates of aging cohorts (the national figure is about 7.5% across all ages [1]). Fig. 2 shows that mean condition declines monotonically with age and that the decline is steepest for bridges in freeze–thaw-plus-salt environments and shallowest in temperate climates—the ordering expected from chloride-induced corrosion.

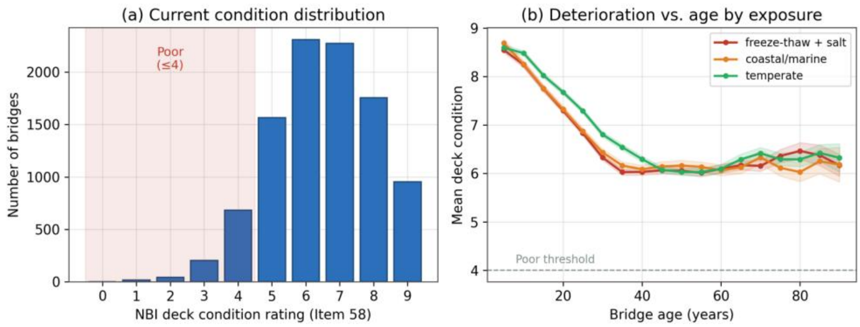

Figure 2 Benchmark Realism: (a) current deck-rating distribution concentrates at 6–7 with a Poor tail; (b) mean condition declines with age, fastest under freeze–thaw-plus-salt exposure

3.4 Prediction Tasks and Protocol

From each bridge's record at the analysis year t0 we predict its condition H = 10 years later. We consider (a) a regression task—the future deck condition rating—and (b) a binary classification task—whether the bridge will be Poor at t0 + H. Predictors comprise all attributes available at t0 (structural, environmental, traffic, current and previous component ratings, and short-term condition trend); no future information is used, precluding leakage. After removing bridges with incomplete inspection histories, 9,800 bridges remain, split 60/20/20 into training, calibration and test sets, stratified on the Poor label. The calibration split is reserved for conformal prediction and probability calibration and is never used for model fitting.

4 METHODOLOGY

4.1 Overview

Fig. 1 summarizes the framework. Stage 1 learns deterioration models, equips them with conformal prediction intervals and calibrated probabilities, and explains them with SHAP. Stage 2 combines the calibrated risk, its uncertainty and a consequence model into the BMPI, which yields a ranked repair plan.

4.2 Deterioration Learners

We benchmark a linear baseline (ridge/logistic regression), a random forest and two gradient-boosting implementations—histogram-based gradient boosting and XGBoost [23-25]—which are strong performers on tabular data [13,26]. Numeric features are standardized and categorical features one-hot encoded. Because Poor bridges are the minority (about 20%), the classifiers use class-balanced weighting; class-imbalance remedies such as re-weighting and synthetic oversampling are well studied [27].

4.3 Distribution-Free Uncertainty

For the regression task we use split conformal prediction [15,17]. Given a model f fit on the training set, we compute nonconformity scores $s_i = |y_i - f(x_i)|$ on the held-out calibration set of size n and set q to the empirical quantile of the scores at level $\lceil (n+1)(1-\alpha) \rceil / n$. The prediction interval for a new input x is

$$C(x) = [\ f(x) - q\ ,\ \ f(x) + q\ ] \tag{2}$$

which, under exchangeability, contains the true value with probability at least $1-\alpha$. We also evaluate a locally-adaptive normalized variant that scales residuals by an estimated difficulty $\mu(x)$, giving intervals $f(x) \pm q'\, \mu(x)$ whose width varies with input-dependent uncertainty [19]. We target 90% coverage ($\alpha = 0.1$).

4.4 Interpretability

We compute SHAP values for the gradient-boosted classifier using TreeSHAP [18], and rank features by mean absolute SHAP to identify the dominant deterioration drivers and check them against engineering expectations.

4.5 Probability Calibration

Tree ensembles can produce distorted class probabilities [21]. We fit isotonic regression on the calibration split to map raw scores to calibrated probabilities, and assess calibration with the Brier score and the expected calibration error (ECE) [21,22].

4.6 Risk-Based Prioritization: the BMPI

For each bridge i we define a deterioration risk R_i as the isotonic-calibrated probability of being Poor at $t0 + H$. We summarize predictive uncertainty by the (min-max normalized) half-width u_i of the conformal interval on the future condition. To be risk-averse—escalating bridges whose risk is high or uncertain—we form an uncertainty-inflated risk

$$R_i^{UCB} = \min\{\ 1\ ,\ \ R_i + \lambda \cdot u_i \cdot (1 - R_i)\ \} \tag{3}$$

with $\lambda \geq 0$ ($\lambda = 0.5$ here). The consequence of a bridge failing is captured by a normalized score

$$C_i = 0.40\ e_i + 0.30\ n_i + 0.15\ d_i + 0.15\ a_i \tag{4}$$

combining traffic exposure e_i (log ADT), network criticality n_i (functional class), detour penalty d_i and asset value a_i (log deck area); the weights are illustrative and can be set by agency policy. The Bridge Maintenance Priority Index is the expected consequence-weighted risk

$$BMPI_i = R_i^{UCB} \cdot C_i \tag{5}$$

and bridges are ranked in decreasing BMPI. We compare against four baselines: condition-first (worst current rating), age-first (oldest), risk-only (R_i), and risk×consequence without the uncertainty term ($R_i C_i$, an ablation). Because the benchmark provides the true future state of every bridge, we can measure prioritization quality directly. For a maintenance budget that treats the top-k bridges, we report the consequence-weighted risk averted—the share of the total consequence carried by truly-Poor bridges that is captured by the selected set—and the recall and precision of Poor-bridge detection, swept over budgets from 5% to 30%.

5 EXPERIMENTS AND RESULTS

Models are implemented with scikit-learn and XGBoost [25,28]; all randomness is seeded for reproducibility. Conformal prediction and isotonic calibration use only the calibration split.

5.1 Deterioration Prediction

Tables 2 and 3 report test-set performance. For regression (Table 2), all ensembles clearly outperform the linear baseline; the random forest attains the lowest error (MAE 0.849, RMSE 1.102, R^2 0.481), with histogram gradient boosting and XGBoost close behind. Predicting a rating a decade ahead is intrinsically uncertain—rehabilitation and inspection variability impose an irreducible component—so a mean absolute error below one rating point on the 0–9 scale is practically useful. For classification (Table 3), the tree ensembles reach 0.836–0.839 ROC-AUC and 0.48–0.50 PR-AUC against a 20% Poor prevalence, a substantial lift over the base rate; histogram gradient boosting gives the best AUC and the lowest Brier score among the strong models. Fig. 3 visualizes the comparison, tree ensembles outperform the linear baselines.

Table 2 Regression: Future Deck Condition Rating (Test Set)

Model	MAE	RMSE	R^2
Ridge (linear)	0.951	1.186	0.399
Random Forest	0.849	1.102	0.481
Hist. GBDT	0.859	1.128	0.456
XGBoost	0.856	1.119	0.465

Table 3 Classification: Future Poor State (Test Set)

Model	AUC	PR-AUC	F1	Brier
Logistic	0.829	0.472	0.545	0.174
Random Forest	0.836	0.500	0.525	0.131
Hist. GBDT	0.839	0.479	0.532	0.138
XGBoost	0.839	0.477	0.555	0.139

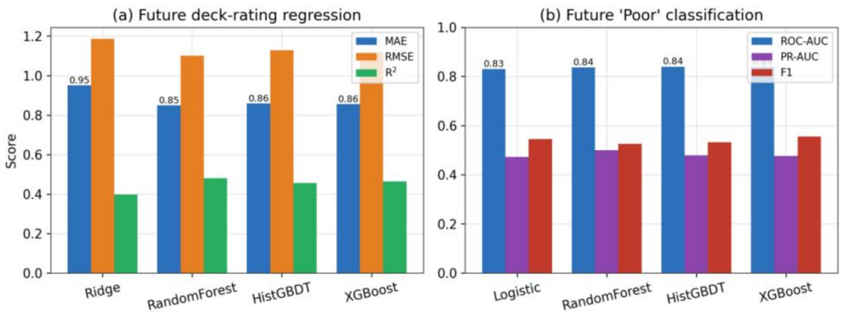

Figure 3 Model Comparison on the Test Set: (a) regression error and R^2; (b) classification ROC-AUC, PR-AUC and F1

5.2 Uncertainty Quantification

Fig. 4 shows the conformal intervals and their coverage. Standard split conformal achieves 88.7% empirical coverage against the 90% target with a mean interval width of 3.65 rating points; the normalized variant achieves 88.9% coverage with a slightly smaller mean width (3.52) by adapting to input-dependent difficulty. The small shortfall from the nominal level is within the finite-sample sampling variation of a single calibration/test draw, and confirms that the intervals are neither over- nor grossly under-confident—precisely the behaviour required for risk-based use.

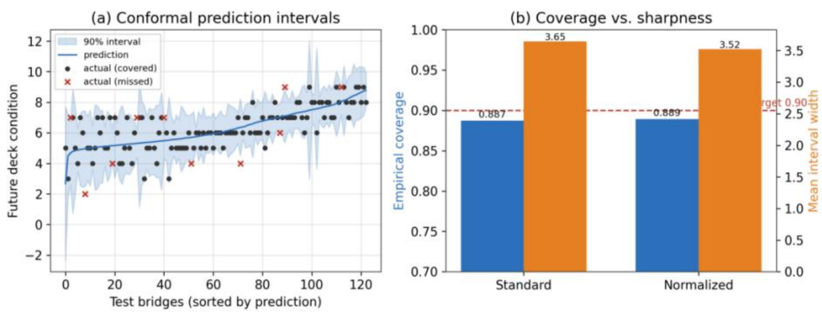

Figure 4 Split Conformal Prediction: (a) 90% intervals with covered (dots) and missed (crosses) actual values; (b) empirical coverage near the 0.90 target with the normalized variant slightly sharper

5.3 Calibration

Raw gradient-boosting probabilities are mildly miscalibrated (Brier 0.138, ECE 0.068). Isotonic regression fitted on the calibration split reduces the Brier score to 0.125 and the ECE to 0.028—more than halving calibration error (Fig. 5)—so that a predicted 30% Poor probability corresponds far more closely to a 30% empirical frequency. Well-calibrated probabilities are essential because the BMPI multiplies risk by consequence.

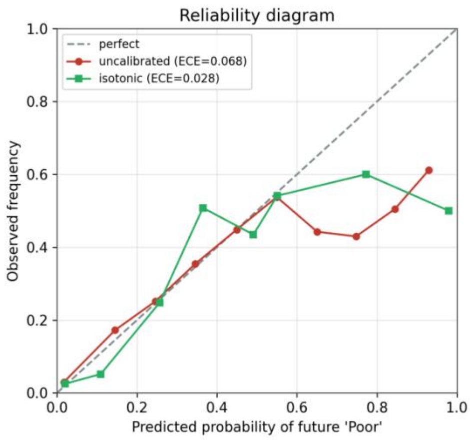

Figure 5 Reliability Diagram before and after Isotonic Calibration; the Calibrated Curve Tracks the Diagonal and the ECE Falls from 0.068 to 0.028

5.4 Interpretability

Fig. 6 shows the SHAP summary. The dominant drivers of future Poor status are the current minimum component rating and the current and previous deck ratings, followed by reconstruction status and years since reconstruction, bridge age, superstructure and substructure ratings, and traffic loading (ADT and truck percentage); material (prestressed) and deck protection also contribute. This ordering matches established bridge-engineering knowledge—present condition and age dominate, maintenance history and traffic loading modulate, and durable materials and protective systems reduce risk—providing evidence that the model has learned physically meaningful relationships rather than spurious correlations, which supports trust and transferability.

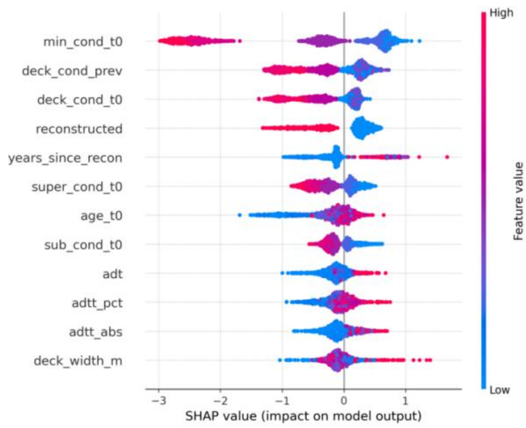

Figure 6 SHAP summary for the Future-Poor Classifier

Present condition and age dominate, with maintenance history, traffic loading, material and deck protection as secondary drivers—consistent with bridge-engineering knowledge.

5.5 Maintenance Prioritization

Fig. 7 report the central result. At a 15% budget, the BMPI averts 51.6% of the consequence-weighted deterioration risk, compared with 34.5% for condition-first and only 21.6% for age-first ranking—the two heuristics closest to current practice. The advantage holds across the entire 5–30% budget range (Fig. 7a). Notably, risk-only ranking attains almost the same Poor-bridge recall as the BMPI (35.2% versus 35.4% at 15%) yet averts far less consequence-weighted risk (37.8% versus 51.6%): detecting a similar number of Poor bridges is not the same as protecting the network, because the BMPI concentrates the budget on the Poor bridges that carry the greatest exposure and criticality. Removing the uncertainty term (risk×consequence) lowers the averted risk at most budgets (49.5% versus 51.6% at 15%), confirming that the risk-averse inflation contributes a small but consistent gain. Together these results show that coupling calibrated risk, predictive uncertainty and failure consequence yields materially better maintenance targeting than condition- or age-based rules.

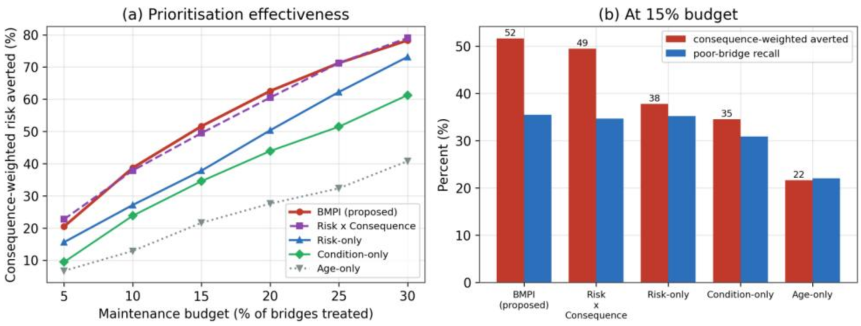

Figure 7 Prioritization Effectiveness: (a) consequence-weighted risk averted vs. budget; (b) averted risk and Poor-bridge recall at a 15% budget

The BMPI dominates condition- and age-based rules and, unlike risk-only ranking, converts similar recall into far greater averted consequence.

6 DISCUSSION

The framework offers transportation agencies a transparent, uncertainty-aware alternative to worst-first inspection triage. Rankings are driven by an explicit, auditable index whose factors—predicted risk, prediction confidence and failure consequence—map onto the concerns agencies already weigh informally, while SHAP explanations expose the engineering rationale of individual predictions for expert review. The prioritization gains are largest precisely where budgets are tightest, which is the regime of practical interest.

Several limitations qualify these findings. The experiments use a synthetic, if carefully grounded, benchmark; although its schema and deterioration behaviour match the NBI and its aggregate statistics align with national figures [1], absolute performance on a specific agency's data will differ, and the models should be re-fit and re-calibrated on operational inventories—an operation the pipeline supports directly. The consequence weights in (4) are illustrative; in deployment they should encode agency policy and could be elicited with multi-criteria methods. We predict a single 10-year horizon and treat traffic as static; multi-horizon and time-varying formulations, survival analysis for remaining-useful-life, and an explicit budget-constrained optimization (rather than top-k selection) are natural extensions. Finally, conformal coverage guarantees rest on exchangeability, which distribution shift over long horizons can violate; adaptive conformal methods are a promising remedy.

7 CONCLUSION

We presented an interpretable, uncertainty-aware framework that unifies bridge deterioration prediction with risk-based maintenance prioritization. Gradient-boosted models predict future condition and Poor-state probability with useful accuracy (0.84 ROC-AUC); split conformal prediction supplies near-nominal 90% intervals; isotonic calibration more than halves calibration error; and SHAP confirms that the learned drivers are physically sensible. The proposed Bridge Maintenance Priority Index, which multiplies an uncertainty-inflated, calibrated risk by a consequence score, averts 51.6% of consequence-weighted deterioration risk at a 15% budget—far exceeding condition-first (34.5%) and age-first (21.6%) practice—and outperforms risk-only ranking by concentrating resources on high-consequence bridges. The results argue for treating prediction and prioritization as a single, probabilistic, consequence-aware decision problem, and the reproducible pipeline transfers directly to operational National Bridge Inventory data.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] American Society of Civil Engineers. 2021 Report Card for America's Infrastructure: Bridges. ASCE, Reston, VA, USA, 2021.

[2] Federal Highway Administration. Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges. Rep. FHWA-PD-96-001, U.S. Dept. Transportation, Washington, DC, USA, 1995.

[3] Madanat S, Mishalani R, Ibrahim W H W. Estimation of infrastructure transition probabilities from condition rating data. J. Infrastruct. Syst., 1995, 1(2): 120-125.

[4] Mauch M, Madanat S. Semiparametric hazard rate models of reinforced concrete bridge deck deterioration. J. Infrastruct. Syst., 2001, 7(2): 49-57.

[5] Thompson P D, Johnson M B. Markovian bridge deterioration: developing models from historical data. Struct. Infrastruct. Eng., 2005, 1(1): 85-91.
[6] Morcous G. Performance Prediction of Bridge Deck Systems Using Markov Chains. Journal of Performance of Constructed Facilities, 2006, 20(2): 146-155.
[7] Ranjith S, Setunge S, Gravina R, et al. Deterioration Prediction of Timber Bridge Elements Using the Markov Chain. Journal of Performance of Constructed Facilities, 2013, 27(3): 319-325.
[8] O'Connor AJ, Sheils E, Breysse D, et al. Markovian Bridge Maintenance Planning Incorporating Corrosion Initiation and Nonlinear Deterioration. Journal of Bridge Engineering, 2013, 18(3): 189-199.
[9] Wellalage NK, Zhang T, Dwight R. Calibrating Markov Chain-Based Deterioration Models for Predicting Future Conditions of Railway Bridge Elements. Journal of Bridge Engineering, 2015, 20(3): 04014060.
[10] Lee J, Sanmugarasa K, Blumenstein M, et al. Improving the Reliability of a Bridge Management System (BMS) Using an ANN-Based Backward Prediction Model (BPM). Automation in Construction, 2008, 17(6): 758-772.
[11] Kim YJ, Yoon DK. Identifying Critical Sources of Bridge Deterioration in Cold Regions Through the Constructed Bridges in North Dakota. Journal of Bridge Engineering, 2010, 15(5): 542-552.
[12] Bektas BA, Carriquiry A, Smadi O. Using Classification Trees for Predicting National Bridge Inventory Condition Ratings. Journal of Infrastructure Systems, 2013, 19(4): 425-433.
[13] Lim S, Chi S. XGBoost Application on Bridge Management Systems for Proactive Damage Estimation. Advanced Engineering Informatics, 2019, 41: 100922.
[14] Assaad R, El-Adaway IH. Bridge Infrastructure Asset Management System: Comparative Computational Machine Learning Approach for Evaluating and Predicting Deck Deterioration Conditions. Journal of Infrastructure Systems, 2020, 26(3): 04020032.
[15] Papadopoulos H, Proedrou K, Vovk V, et al. Inductive Confidence Machines for Regression. In Proceedings of the 13th European Conference on Machine Learning (ECML). 2002: 345-356.
[16] Vovk V, Gammerman A, Shafer G. Algorithmic Learning in a Random World. New York, NY, USA: Springer, 2005.
[17] Lei J, G'Sell M, Rinaldo A, et al. Distribution-Free Predictive Inference for Regression. Journal of the American Statistical Association, 2018, 113(523): 1094-1111.
[18] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 4765-4774.
[19] Romano Y, Patterson E, Candès E. Conformalized Quantile Regression. Advances in Neural Information Processing Systems (NeurIPS), 2019, 32: 3543-3553.
[20] Platt J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In Advances in Large Margin Classifiers. Cambridge, MA, USA: MIT Press, 1999: 61-74.
[21] Niculescu-Mizil A, Caruana R. Predicting Good Probabilities with Supervised Learning. Proceedings of the 22nd International Conference on Machine Learning (ICML), 2005: 625-632.
[22] Guo C, Pleiss G, Sun Y, et al. On Calibration of Modern Neural Networks. Proceedings of the 34th International Conference on Machine Learning (ICML), 2017: 1321-1330.
[23] Breiman L. Random Forests. Machine Learning, 2001, 45(1): 5-32.
[24] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 2001, 29(5): 1189-1232.
[25] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016: 785-794.
[26] Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 3146-3154.
[27] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
[28] Pedregosa F, Varoquaux Gaël, Gramfort Alexandre, et al. Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research, 2011, 12: 2825-2830.