Improving Off-System Bridge Monitoring: An AI-Driven Methodology for Enhanced Condition Prediction

---

MAHMOUD BAYAT and SUBHAM KHAREL

ABSTRACT

Off-system bridges, especially in rural and underserved areas, play a vital role in connecting communities to essential services but often suffer from neglect and underinvestment. These structures are frequently excluded from comprehensive inspection programs, resulting in limited data and a high risk of undetected deterioration. This study addresses the critical challenge of class imbalance in bridge condition datasets by introducing a machine learning framework that integrates Boruta feature selection, Ridge regularization, TomekLinks sampling, and a Generative Adversarial Network trained with focal loss (GAN-FL). The model aims to enhance the accuracy and interpretability of bridge deck condition predictions. Despite improvements in performance for majority classes, minority class predictions remained a challenge, with certain conditions (e.g., 4, 5, and 8) showing low F1-scores. However, the use of GAN-FL led to measurable gains in balancing recall and precision across classes. The proposed approach demonstrates the potential to support risk-informed maintenance decisions, prioritize inspections, and improve safety outcomes for off-system bridges. Future work will focus on improving performance for underrepresented classes and increasing model transparency to support adoption by transportation agencies.

INTRODUCTION

Approximately 587,000 highway bridges in the United States that exceed 6.1 meters (20 feet) in length, nearly half are designated as off-system, meaning they fall outside the purview of the Federal Aid System. These bridges, commonly situated on rural and local roads, often endure significant deterioration due to insufficient funding and limited maintenance oversight [1]. Alarmingly, around 30% of these structures are either structurally deficient or no longer meet functional standards. Despite their shortcomings, they play a critical role in connecting isolated communities to essential services such as employment, healthcare, and education [2]. Their closure would disproportionately disrupt access for underserved populations, further deepening existing social inequalities [3]. To address this, the Bipartisan Infrastructure Law mandates that states dedicate no less than 15% of their Bridge Formula Program funds to support off-system bridges, under the administration of the FHWA [4].

Despite their importance, off-system bridges remain underrepresented in infrastructure research and condition assessments [5]. Given the high number of deficient bridges, it is imperative that engineers and inspectors prioritize these structures for evaluation and repair [6]. A recent analysis by Quadri (2024) [7] revealed a significant data gap, with Truck Average Daily Traffic (ADT) data absent or poorly estimated for about 30% of off-system bridges in the National Bridge Inventory (NBI), highlighting inconsistencies in data collection across jurisdictional levels.

Mahmoud Bayat, School of Architecture, College of Architecture, Planning and Public Affairs, The University of Texas at Arlington, Arlington, TX, USA, corresponding author email: mahmoud.bayat@uta.edu

Subham Kharel, Senior Data and Analytics Planner, Lehigh Valley Planning Commission, email: SKharel@lvpc.org

Managing off-system bridges involves navigating numerous challenges related to their aging, environmental exposure, and potential closure impacts. Nevertheless, maintaining these assets is essential to achieving both structural reliability and transportation equity. Traditional inspection techniques such as visual assessment and hammer sounding are still widely used, yet predictive degradation models offer more robust solutions for forecasting future conditions and planning maintenance interventions [8].

The NBI, established by the FHWA in 1972, remains the cornerstone data repository for U.S. bridge condition monitoring. It contains vital information on bridge ownership and structural ratings of key components, which are used in developing and validating degradation models [9-11].

Artificial intelligence techniques are increasingly gaining attention for their ability to capture complex patterns in large datasets and produce accurate condition forecasts. For example, Chencho et al. (2021) [6] demonstrated the use of a random forest regressor to identify relationships between damage severity and reduced stiffness parameters. Similarly, Fernandez-Navamuel et al. (2022) [12] employed Deep Learning Enhanced Principal Component Analysis to detect structural anomalies more effectively by integrating residual connections. Rajkumar et al. (2023) [8] developed a hybrid model combining autoencoders with random forests to forecast condition ratings using Florida’s NBI data.

Nevertheless, one of the primary challenges in bridge condition modeling is the presence of class imbalance, particularly for severely deteriorated bridges. Without addressing these imbalances, prediction models risk misclassifying critical cases, which may lead to misallocated resources and heightened infrastructure risks. To overcome this, the present study utilizes TxDOT bridge data to construct interpretable machine learning models that accurately predict bridge deck condition ratings while managing class imbalance issues [13].

Table 1. Condition rating codes as described in NBI [9]

Condition	Code	Description
Excellent	9	Excellent condition
Very Good	8	No problems noted
Good	7	Some minor problems
Satisfactory	6	Structural elements show some minor deterioration
Fair	5	All primary structural elements are sound but may have minor section loss, cracking, spalling, or scour
Poor	4	Advanced section loss, deterioration, spalling, or scour
Serious	3	Loss of section, deterioration, spalling, or scour have seriously affected primary structural components. Local failures are possible. Fatigue cracks in steel or shear cracks in concrete may be present
Critical	2	Advanced deterioration of primary structural elements. Fatigue cracks in steel or shear cracks in concrete may be present or scour may have removed substructure support. Unless closely monitored, it may be necessary to close the bridge until corrective action is taken
Imminent failure	1	Major deterioration or section loss present in critical structural components or obvious vertical or horizontal movement affecting structure stability. Bridge is closed to traffic
Failed	0	Out of service—beyond corrective action

METHODOLOGY

After data cleaning and normalization using a Standard Scaler, the dataset was partitioned into training, validation, and testing sets with a 60-20-20 split. To identify the most influential predictors of bridge defects, the Boruta algorithm—an all-relevant feature selection method based on random forests—was employed. This method evaluates the importance of real variables by comparing them against randomized shadow attributes [14]. The key variables retained for modeling included construction year, improvement cost, condition ratings, deck type, span length, deck width, structure length, and traffic volume metrics (see Figure 1). To mitigate overfitting in early modeling stages, Ridge regression (L2 regularization) was applied, and Tomek Links were used to perform preliminary class balancing by removing borderline inter-class samples that are likely to confuse the classifier [15].

Building on this foundation, a comprehensive strategy was implemented to further address the significant issue of class imbalance, which frequently undermines the accurate detection of minority classes, such as rare or severe bridge defects. The proposed method integrates a Generative Adversarial Network (GAN) trained with focal loss a loss function that reduces the impact of easily classified examples and forces the model to focus on harder, misclassified samples. This allows the GAN to generate high-quality synthetic instances of underrepresented classes, effectively enriching the dataset and improving balance across class labels [16]. The synthetic samples produced by the GAN are then merged with the original dataset to create a more equitable class distribution. To capitalize on this improved training set, a Random Forest classifier is trained on the augmented data. This hybrid framework termed GAN-FL-RF (Generative Adversarial Network with Focal Loss and Random Forest) leverages the complementary strengths of generative augmentation, focal loss-based learning, and ensemble classification to enhance predictive performance. The model's effectiveness is rigorously assessed using precision, recall, support, and F1-score, which are particularly informative in evaluating classifier performance on imbalanced data.

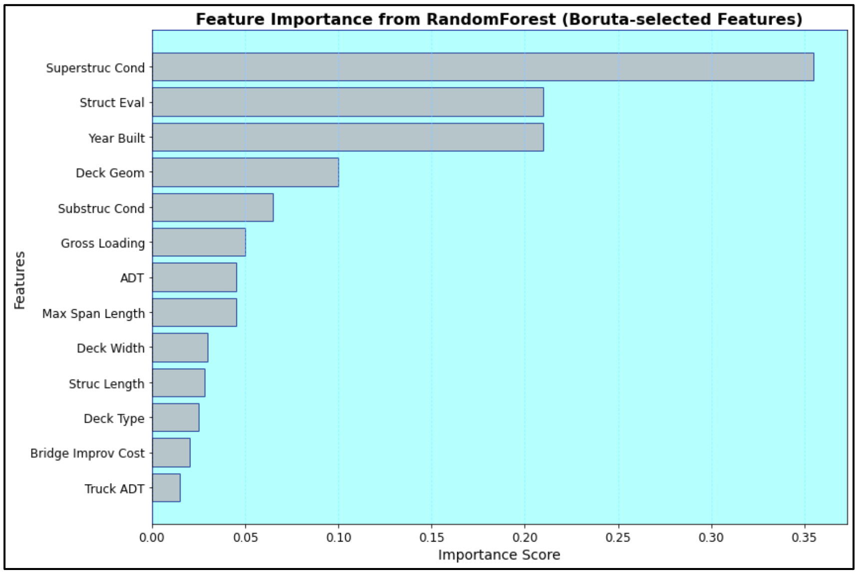

Figure 1. Feature Importance Values

RESULTS AND DISCUSSIONS

This section presents the results and analysis of the study, focusing on the effectiveness of the proposed methods in addressing class imbalance while improving model accuracy and interpretability. Figure 2 illustrates the frequency distribution of all condition ratings, highlighting the pronounced class imbalance within the dataset. To establish a baseline, a Random Forest classifier was trained using features selected via the Boruta algorithm, in conjunction with Ridge regularization and the TomekLinks technique.

Then, the baseline model was trained using GAN-FL and a more comprehensive GAN-FL-RF model was developed.

The outcomes of the comparison between Pre- and Post- GAN-FL-RF models are detailed in Figure 3, which reports key performance metrics including precision, recall, F1-score, and support. The comparison between the classification metrics before and after applying the GAN-FL-RF method reveals notable improvements, particularly in underrepresented classes.

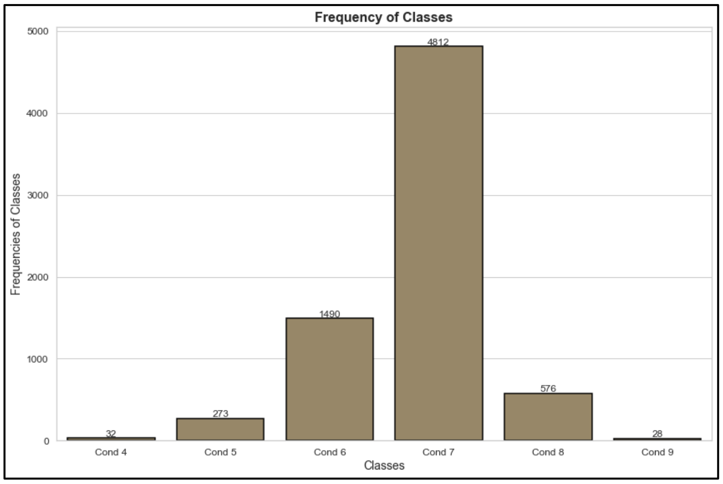

Figure 2. Distribution of Frequency of Occurrence for all conditions

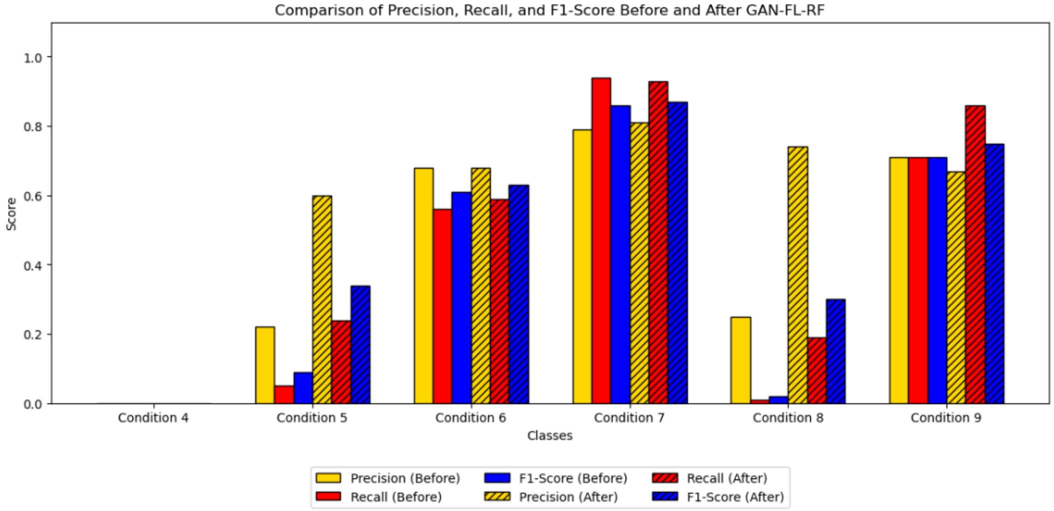

Figure 3. Comparison of Pre- and Post- GAN-FL-RF Results

While precision, recall, and F1-score were already high for Condition 7 the majority class with 977 samples and remained relatively stable, substantial gains were observed in minority classes with much smaller support. For instance, Condition 5, which had only 38 samples, saw its precision increase from 0.22 to 0.60 and its F1-score from 0.09 to 0.34, suggesting a marked improvement in correctly identifying this underrepresented group. Similarly, Condition 8, with 105 samples, experienced a dramatic jump in F1-score from 0.02 to 0.30, indicating improved sensitivity and generalization. Although Condition 4, with just 8 samples, remained unclassified, and Condition 9 (7 samples) saw modest changes, the overall increase in macro-averaged F1-score from 0.38 to 0.48 and weighted average F1 from 0.72 to 0.76 indicates that GAN-FL-RF helped the model better balance performance across classes. These results highlight the method's ability to enhance detection of minority classes without sacrificing accuracy on the dominant class. In contrast, the model performed better for more frequently occurring classes. Condition 6 achieved a F1-score of 0.61 with a precision of 0.68 and recall of 0.56 across 308 samples. The highest performance was observed for Condition 7, with strong results across 977 samples precision of 0.79, recall of 0.94, and an F1-score of 0.86 indicating the classifier’s reliability when sufficient training data is available.

CONCLUSION

Off-system bridges, particularly in rural and local areas, are essential yet often neglected in inspections and research. This study addressed that gap by developing a machine learning framework that improves defect prediction using Boruta feature selection, Ridge regularization, TomekLinks, and a GAN with focal loss to manage class imbalance. The model performed well on the majority class, Condition 7, with an F1-score of 0.86, while showing notable improvements for minority classes. For example, Condition 5’s F1-score increased from 0.09 to 0.34, and Condition 8’s from 0.02 to 0.30. However, some minority classes, such as Condition 4, remained unclassified. Despite these challenges, the macro-averaged F1-score improved from 0.38 to 0.48, and the weighted average F1-score rose from 0.72 to 0.76, reflecting better performance across both majority and minority classes. The proposed framework shows promise for supporting more efficient and proactive maintenance of off-system bridges. It can complement traditional inspections by enabling risk-based prioritization. Future work should focus on improving minority class predictions, incorporating additional data, and enhancing model interpretability to support broader adoption by infrastructure agencies.

ACKNOWLEDGEMENTS

The authors would like to thank the Federal Highway Administration for providing the inventory of data used in this study.

REFERENCES

1. Wipf, T. J., Russo, F. M., & Klaiber, F. W. (2002). Cost-effective structures for off-system bridges (NCHRP Synthesis of Highway Practice Project 20-5, Topic 32-08). *Draft report. Transportation Research Board*, National Research Council.
2. Stommes, E. S. (2003). Rural Roads and Bridges, 1994–2000: How Did the South Fare?. *Journal of Agricultural and Applied Economics,* 35(2), 263-278.
3. Sharkey, J. (2022). *Infrastructure Investment and Jobs Act (Iija)/Bipartisan Infrastructure Law (Bil)*.
4. Alipour, M., Harris, D. K., Barnes, L. E., Ozbulut, O. E., & Carroll, J. (2017). Load-capacity rating of bridge populations through machine learning: Application of decision trees and random forests. *Journal of Bridge Engineering*, 22(10), 04017076.
5. Russo, F. M., Wipf, T. J., & Klaiber, F. W. (2003). Cost-effective Structures for off-system bridges. *Transportation research record*, 1819(1), 397-404.
6. Chencho, Li, J., Hao, H., Wang, R., & Li, L. (2021). Development and application of random forest technique for element level structural damage quantification. *Structural Control and Health Monitoring*, *28*(3), e2678.
7. Quadri, M.W. (2024). Statistical analysis on the effects of environmental factors on bridge condition rating in Texas (Master's thesis). *Civil Engineering Theses*, 315.
8. Rajkumar, M., Nagarajan, S., & Arockiasamy, M. (2023). Bridge infrastructure management system: Autoencoder approach for predicting bridge condition ratings. *Journal of Infrastructure Systems*, *29*(1), 04022042.
9. FHWA (Federal Highway Administration). (1995). Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges. In *Report No. FHWA-PD-96-001*. US Department of Transportation Federal Highway.
10. Chase, S. B., Small, E. P., & Nutakor, C. (1999). An In-Depth Analysis of the National Bridge Inventory Database Utilizing Data Mining, GIS and Advanced Statistical Methods TRB Transportation Research Circular 498. PC-6/1~ C-6/17.
11. Frangopol, D. M., Kallen, M. J., & Noortwijk, J. M. V. (2004). Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Progress in structural engineering and Materials*, 6(4), 197-212.
12. Fernandez-Navamuel, A., Magalhaes, F., Zamora-Sánchez, D., Omella, A. J., Garcia-Sanchez, D., & Pardo, D. (2022). Deep learning enhanced principal component analysis for structural health monitoring. *Structural Health Monitoring*, 21(4), 1710–1722.
13. Xu, Z., Shen, D., Nie, T., & Kou, Y. (2020). A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data. *Journal of Biomedical Informatics*, 107, 103465.
14. Kumar, S. S., & Shaikh, T. (2017). Empirical evaluation of the performance of feature selection approaches on random forest. In 2017 *international conference on computer and applications (ICCA)* (pp. 227-231). IEEE.
15. Zeng, M., Zou, B., Wei, F., Liu, X., & Wang, L. (2016). Effective prediction of three common diseases by combining SMOTE with Tomek links technique for

imbalanced medical data. *In 2016 IEEE International Conference of Online Analysis and Computing Science* (ICOACS) (pp. 225-228). IEEE.

16. Li, W., Liu, D., Li, Y., Hou, M., Liu, J., Zhao, Z., ... & Deng, W. (2024). Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data. *Structural Health Monitoring*, 14759217241254121.