Calculating Impact of Ongoing Deterioration to National Bridge Inventory

Christopher Schneider
Department of Engineering, Computing, and Mathematical Sciences
Lewis University
Romeoville, IL, USA
christopherjschnei@lewisu.edu

Mahmood Al-khassaweneh
Department of Engineering, Computing, and Mathematical Sciences
Lewis University
Romeoville, IL, USA
malkhassaweneh@lewisu.edu

***Abstract*—The Federal Highway Administration (FHWA) performs inspections and maintains data regarding the condition of bridges within the United States roadway network. Each bridge is given a classification of Good, Fair, or Poor. These inspection records are used for critical funding decisions, yet the most recent information available is often significantly outdated.**

By narrowing the scope of the NBI to relevant records and carefully selecting or engineering features, we will seek to train a model that can accurately predict an overall structural evaluation based on available data. Then, by modeling the deterioration of individual bridges, we will re-evaluate to see what proportion of bridges are likely to slipped into a lower category of condition, and attempt to predict which are at greatest risk.

***Index Terms*—transportation, USDOT, Federal Highway Administration, FHWA, National Bridge Inventory, NBI, infrastructure projects, BIL, machine learning.**

I. Introduction

ACCOMMODATING nearly 49 billion vehicle crossings per year, roadway bridges are a vitally important link in the United State transportation network. The condition of that infrastructure is thus of utmost importance to the economy as well as the freedom of individuals traveling from one place to another. It is well-understood by policymakers that many of these bridges are aged and in need of restoration, as evidenced by numerous legislative efforts over many years to pass funding to make infrastructural repairs.

To gain a sense for the scope of this need, we turn to data from the Federal Highway Administration (FHWA). This body maintains a resource called the National Bridge Inventory (NBI), a database of over 621,000 bridges within the United States, with each record containing statistics about roadway/traffic characteristics, bridge type, condition, and inspection details [1]. This impressive resource has long been the gold-standard benchmark for assessing infrastructure priorities, and funding decisions are routinely made using the data found within the NBI resource [4], including the signature "Bipartisan Infrastructure Law" passed by Congress and signed into law in 2021 [5].

However, in reviewing this data, we begin to understand that inspecting 621,000 bridges with the number of engineers available is not a rapid process. Even restricting our purview to 113,000 bridges serving designated Interstate or U.S. Highways, the age of these records varies widely. As shown in Figure 1 below, the data is bimodal with peaks around 20 and 35 months. Mean is 26.14 months with a standard deviation of 8.13.

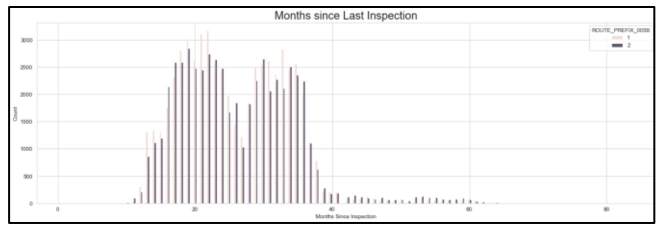

Fig. 1. Time from last inspection date to present (March 2024). Major ticks on the x-axis are in periods of 20 months. There is little observed difference in distribution between Interstate (pink) and U.S. Highway (gray) route maintenance.

With this in mind, we perceive the opportunity to use machine learning models to simulate the continued deterioration that will have taken place in the ensuing months. This will permit policy makers to better understand the situation as it exists today and base their funding and repair project decisions on a more representative set of data.

The task at hand can be performed in three phases: (1) designing a classification model that can accurately assign an arbitrary bridge record into a rating class of Good, Fair, or Poor; (2) calculating a deterioration factor to apply to each bridge based on observable parameters; and (3) using these two tools to simulate the current condition of every bridge, then re-classify them based on those new characteristics.

II. Defining and Evaluating Models

The NBI is a formidable data source. For each of the 621,000 records contained herein, there are over 100 categories and sub-categories providing identifying information, descriptive details of the bridge itself, the type and volume of traffic accommodated, what feature(s) the bridge spans, information about reconstruction projects already completed, and a host of statistical data points from the most recent inspection completed. The FHWA publishes both a recording/coding guide [2] and a specification guide [3], which together serve as a data dictionary and reference guide for engineering measurements into numerical categories; the documents together are over 400 pages long. Clearly, the inspection of just one bridge is no simple task, and one begins to appreciate the herculean effort that goes into maintaining over half a million of them.

Data preparation and cleaning is conducted with pandas 2.2.1 [12] in Python 3.11 in Anaconda Spyder [6]. In reviewing the data at hand, we find that many fields do not contain useful predictive value. Some are descriptive, identifying latitude/longitude, state, county, city, route numbers, mile markers, road and rail crossings, bodies of water, and other landmarks to describe the location in whatever manner may be needed for a given purpose. Physical characteristics are provided ranging from subtended height to structural material to total square meters of deck area.

At the same time, we have the definitional data: the classification of each bridge as given at last inspection, along with the numeric scores that directly define the class category. These data points make it a matter of rote arithmetic to successfully evaluate a bridge—but they consist of whole-number stepped grades, and thus copying the grade calculation exactly will not yield the level of detail we need in the next phase, as the decay model is logarithmic and thus continuous (to be discussed in part III).

Thus, we find ourselves seeking the middle ground: including data dimensions that provide predictive value without trivializing the model. We perform a series of numerous steps combining, binning, and reshaping variables, as follows:

- Three fields for last date of inspection (fracture, underwater, and special) are converted to binary. The existence of the inspection is predictive, while the relative recency may be based on non-predictive factors.
- The dates for year built and most recent reconstruction are merged when both are present.
- We impute mean value for % of truck traffic where this statistic does not exist.
- We impute mean values for evaluation scores where the evaluation is marked "N" for not relevant (i.e. waterway evaluation for a bridge that does not span water).
- We use date of inspection, an alphanumeric string, to compute number of months since last inspection, a linear variable.
- We combine lanes of traffic above and below the bridge into one statistic "total lanes" for determining magnitude of traffic impact.
- We review inter-quartile outliers for the dataset. Due to the inherently normalizing process of the engineering grading process, these don't exist for many variables. In other cases, the existence of the variable is itself an outlier, so we convert to a binary variable (e.g. special inspection dates above).

At the conclusion of these steps, we land upon 16 predictive features to use for our supervised learning process. Six of these are highly correlated or overlap in definition, and these are reduced to three in order to simplify our model. Figure 2 shows a correlation matrix for the remaining 13 features.

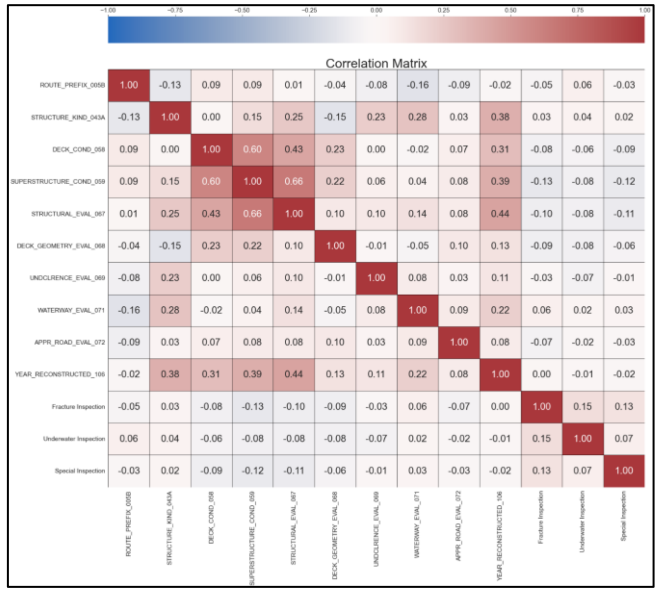

Fig. 2. Correlation of data features. The highest are between Deck Condition, Structural Eval, and Superstructure Condition; these are all high-information features and there is no fundamental overlap in definitions. All other pairs are below 0.50 correlation.

With features established, we now shape our dataset for model training. Given the high variability in local road standards and conditions, we limit our scope to Interstate and U.S. Highway routes, as mentioned above. Within this subset, the smallest rating class by far is 'Poor' with 2,905 records listed. We use this value as our target size and select a random permutation for each of 'Good' and 'Fair' records to balance the dataset. From this corpus, we take a train/test split of 80/20.

Reviewing potential models for a multi-class analysis, we determine that 4 are best suited to the task: linear discriminant analysis (LDA) [15], support vector machines (SVM) [10], gradient boosted trees (GBC) [14], all implementations and deep neural network (DNN) as implemented with Tensorflow and Keras [8, 9]. K-nearest neighbors is a valid option, but given our intent to apply the model back to a full-scope dataset, the time complexity will be prohibitive, so we opt not to use this method.

Each of our four models was run with a range of hyperparameters using Scikit-Learn's GridSearchCV [16] to select the best-performing model to our training data. We evaluated these using a loss function to confirm their optimality. For GBC and SVM, the log-loss function was used as it represents a multiclass model well. For LDA, all choices of solver performed identically. The graph below (Fig. 3) is an example of how different parameters performed, and how we selected the optimal models to use for classifying our test data. (For log-loss, maximizing the inverted loss function corresponded to minimizing the true loss.)

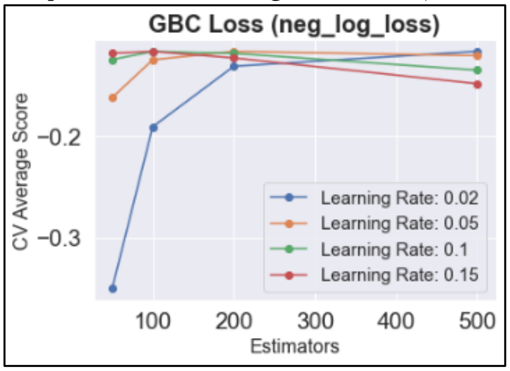

Fig. 3. A graph of calculated loss functions for various hyperparameters. This shows that the best_estimator_ selected by GridSearchCV was indeed the optimal value with learning rate 0.05 and 200 estimators/boosting stages.

These optimized models are then applied to the test dataset. Confusion matrices are shown in figures below (class labels [0,1,2] → [Good, Fair, Poor]), with key performance statistics given in Table I.

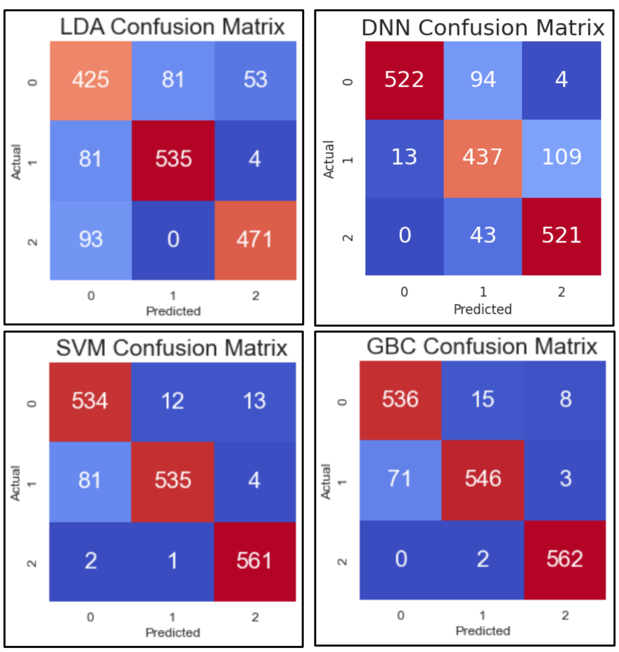

Fig. 4-7 (top row first, left-right).

LDA with singular value decomposition.

DNN with 4 hidden layers of 16 units each (2 ReLU, 2 sigmoid) to 3 softmax outputs.

SVM with a 5th-degree polynomial kernel.

GBC with 200 boosting stages and a learning rate of 0.05.

TABLE I

CLASSIFIER MODEL PERFORMANCE

Model	Accuracy	Precision	Recall
GBC	94.32%	94.45%	94.53%
SVM	93.52%	93.74%	93.76%
DNN	83.59%	83.65%	83.47%
LDA	82.10%	82.34%	81.94%

Table I. Gradient boosted trees classifier (GBC) had the highest accuracy as well as best F1-score metrics of all multi-class models tested.

III. MODELING DECAY FUNCTION

With our predictive classifier model in hand, we now seek to find a mathematical basis by which to "age-forward" our records. The objective will be to apply a decay factor onto each of the affected features from the classifier, with that factor defined in part by additional data which we held back during the classifier training. We will seek to account for two general factors: how much wear & tear the bridge experiences, and how much time has passed.

A feature representing time between inspections is already available to us (see Figure 1). Using the dates from most inspections recorded in the two snapshots, we can calculate the number of months elapsed between them, which allows us to fully capture the element of passing time (t).

Evaluating wear and tear is more nuanced. The given values for daily vehicles carried, even for our highway subset, ranges from under 100 to nearly 800,000. Furthermore, all traffic is not created equal; we have a measurement for percentage of traffic consisting of trucks, which ranges in whole number increments from 0% to 99% (mean 15.6%, mode 8%). And certainly we can imagine that an 8-lane reinforced steel suspension bridge will deteriorate less, on a per-vehicle basis, than a covered wooden bridge built for horse carts. We do have coded data corresponding to building materials and construction methods, but these don't correspond to measurements where interpolation or arithmetic factors make sense. Instead, we will use the lane capacity of the bridge as a representative factor to which we can perform regression analysis. Thus, we will estimate the wear-and-tear factor on a given bridge as

$$\sqrt{\frac{Traffic_{All} \times (1 + Traffic_{Truck\%})}{Total\ Lanes}} = F. \qquad (1)$$

We expect the deterioration of the bridge to be very minor in any given month, but it should build up over time. We make the assumption that this accumulation will be geometric in nature, and thus we model the condition of the bridge as an exponential decay function:

$$C(t,F) = C_0 + \left(e^{-\beta_i F}\right)^t \tag{2}$$

where β_i is a parameter, specific to each of the $i \in [1..7]$ features, to be determined by best fit.

To conduct that optimized fitting, we need relevant data to define our curve. For this, we can refer to a past snapshot of the NBI to look for instances where a bridge was inspected between assessments: the inspection time stamp will change, and we'll be able to calculate both the number of elapsed months between inspections (thus providing a value for *t* by defining the earlier inspection to be at month zero in all cases) and the degradation of each measurable factor, if any, by direct comparison. These data points are then fit to an exponential decay function as described above, with best-fit parameters identified using the curve_fit function from scipy.optimize [7]. This process is repeated for each variable to obtain seven individual models.

Prior to conducting the curve fitting, we inspect and sanitize for outliers. Notably, of records relevant to our analysis, approximately 15% (12,108 out of 81,345) indicate an improvement in rating for one or more of the condition ratings without an associated reconstruction record between 2019 and 2023. Improper recordkeeping, perhaps due to overlapping maintenance jurisdictions between local/state/federal agencies without consistent practices, may be likely explanations. As structural conditions should decline monotonically in the absence of such repair work, we therefore remove the contraindicating records from our regression model. In addition to this abridgement, values above three standard deviations are truncated to the 99.5% percentile.

An example of the exponential decay curve fit graph discovered for each feature are shown in Figure 8, and all best-fit functions are provided in Table II.

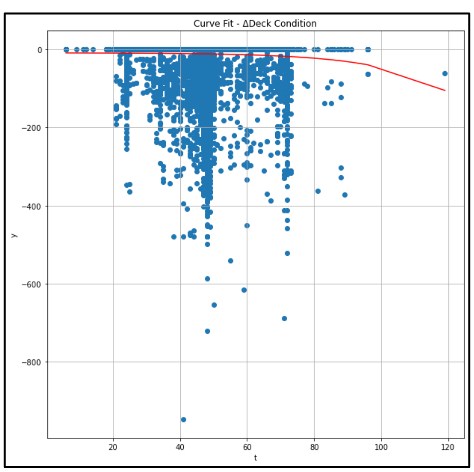

Fig. 8. The "Deck Condition" feature shown with the exponential decay curve of best fit. The horizontal axis *t* shows that most records are 36 to 72 months old, indicating the aging needed to bring records up to date.

TABLE II
REPARAMETRIZING FEATURES: CURVES OF BEST FIT

Feature	**Exponential Decay Curve**
Deck	$-0.26e^{0.05t} + -9.26$
Superstructure	$-0.10e^{0.06t} + -8.58$
Structure	$-1.17e^{0.04t} + -10.62$
Geometry	$-1.63e^{0.02t} + 0.88$

Table II. The equations above are used to simulate the aging of each bridge record, with *t* expressed in months.

IV. SIMULATION OF AGING; RECLASSIFICATION

With the variable modeling complete, it is a simple matter to apply them to the data. Using the value of *t* specific to each entry, a new simulated data record can be created with the outputs of each variable model standing in place of the original values. These are normalized using the same scale parameters, with any outlying feature values above the highest or below the lowest defined values per the FHWA recording/coding guide [2] are abridged to these highest and lowest values respectively. Once remapped in this way, the values are directly comparable to the original integer-based condition value and can be used in our classifier model from part II above.

Figures 9 through 12, below, illustrate the impact of this decay calculation on the feature values themselves, expressed by percentile to reflect the breadth of data within each range.

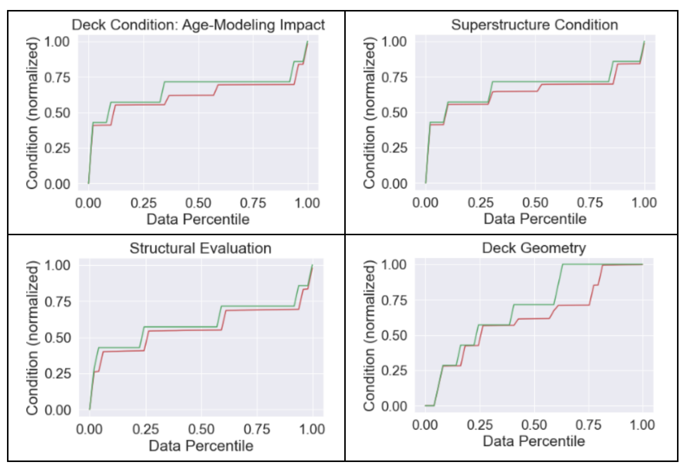

Fig. 9-12. Pre- and post- modeling of key structural features. Green lines reflect the composition of data as of publication of the 2023 NBI; red lines reflect those same records subject to the exponential decay factors in Table II.

Once the age-simulated dataset is constructed, we use our gradient-boosted trees ensemble classifier to identify the condition status of that bridge with aging factors applied. Checking each record against its original classification, we can calculate our estimate for the individual data points, as well as the proportion in aggregate, of the NBI which have likely changed class since their most recent inspection.

V. Conclusions

Figure 13 shows the results of applying the exponential decay regression model, recasting these transformed features into composite records, and using our gradient boosted trees model to predict the resulting classification.

Fig. 13. The matrix above shows what proportion of each original class was reclassified into which of the three categories. The largest change by a substantial margin was 31.4% of "Good" records veering into "Fair" territory.

As shown above, nearly one-third of records previously rated "Good" have been reclassified as "Fair" after post-, with a smaller portion declining to "Poor." Taking the relative class sizes into account, in aggregate we see 0.2% of records (154 out of 70,998) improving in status but nearly 15% (10,319 out of 70,998) declining.

From this, we draw the conclusion that ongoing aging of bridges is a significant factor in evaluating their condition, and we urge that further study be conducted to ensure that the records used for key governance decisions are, if not perpetually kept up-to-date, at least used with consideration of the latent bias they are apt to contain.

VI. Related Work

The National Bridge Inventory is employed for the purpose of statistical data mining in a number of published works. In particular, prior literature demonstrates several efforts to predict bridge conditions using component ratings. A team from Iowa State University in 2013 employed a methodology using classification and regression trees [17], which are a precursor to the gradient-boosted trees algorithm identified in part II as the optimal model selection.

The work most parallel to this was conducted by a team of ASCE civil engineers collaborating between the Illinois Institute of Technology and Atuturk University in Turkey in 2004 [13]. This group assessed a set of 2,601 bridges within the state of Illinois and did a longitudinal analysis from 1976 to 1998, focusing similarly on the NBI component ratings. The approach in that work was to develop a regression model using third-degree polynomial curves, with the time variable under study being the overall age of the bridge. The major improvements in our present work include (1) the incorporation of supervised machine learning techniques to improve the translation of component rating data into an overall bridge evaluation; (2) the utilization of a more sophisticated curve-modeling process to apply aging factors to each individual record; and (3) simplification of modeling to a single iterated process rather than a branching process where different formulas are used depending on the age and primary structural material of each bridge. This streamlining allowed us to consider a larger pool of NBI entries in a tractable way.

VII. Further Opportunities

We wish to acknowledge some opportunities for further refinement and potential improvement of these results.

As stated above, we reduced the feature set to ensure an acceptably low level of collinearity between features. A more thorough analysis utilizing knowledge of civil/structural engineering principles may yield a different result that hews more closely to the real world. Linear discriminant analysis (LDA) and principal component analysis (PCA) both indicate that dimensionality can be reduced further with minimal loss to data fidelity.

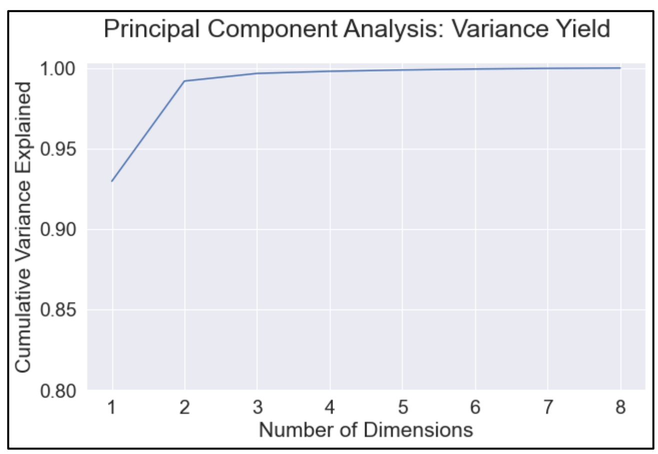

Fig. 14. PCA shows dimensional reduction is apt to be highly effective on the NBI dataset. Over 99.8% of variance is explained with 3 orthogonal features.

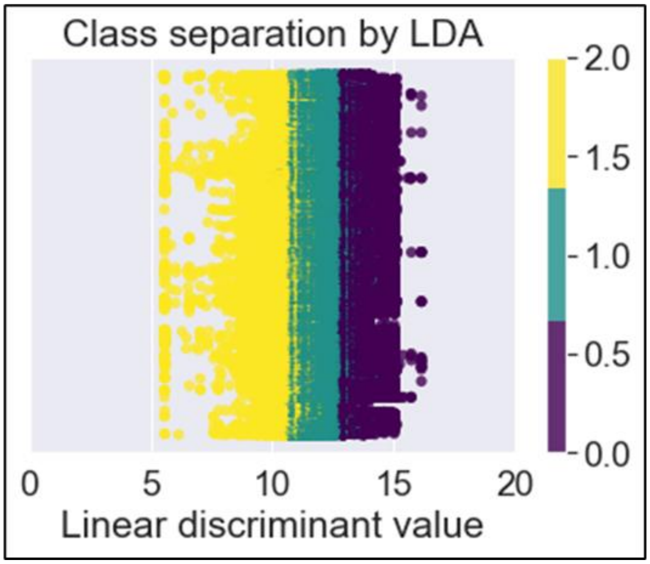

Fig. 15. LDA shows a clean separation of data between the three classes (0 = 'Good', 1 = 'Fair', 2 = 'Poor') using eigenvector decomposition to optimally map points onto a single axis. (The vertical axis is not data-valued and is stretched to showcase relative density.)

We also note that thousands of records had to be thrown out because untracked improvements led to structural and/or status changes that were otherwise inexplicable with only the data at hand. The best outcome would be identifying alternative (perhaps state and/or municipal) record sources to cross-reference and ensure that maintenance dates are accurate reflections of the bridges' history. In the absence of this, a more thorough analysis of the potential bias propagated by large-scale exclusion of records may also have potential to improve model accuracy.

References

[1] Federal Highway Administration; Bridges & Structures – National Bridge Inventory. Website; https://www.fhwa.dot.gov/bridge/nbi.cfm, accessed March 8, 2024.

[2] "Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges." Publication No. FHWA-PD-96-001. Accessed via https://www.fhwa.dot.gov/bridge/mtguide.pdf on March 8, 2024.

[3] "Specification for the National Bridge Inventory." Publication No. FHWA-HIF-22-017. Accessed via https://www.fhwa.dot.gov/bridge/snbi.cfm on March 8, 2024.

[4] Federal Highway Administration; Bridge Programs. Website; https://www.fhwa.dot.gov/bridge/bripro.cfm, accessed March 8, 2024.

[5] 117th United States Congress, H.R. 3684, "Infrastructure Investment and Jobs Act." Accessed via https://www.congress.gov/117/bills/hr3684/BILLS-117hr3684enr.pdf on March 9, 2024.

[6] Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Website; https://anaconda.com.

[7] Virtanen, Gommers, et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272.

[8] Abadi, Agarwal, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[9] Chollet, François and others. Computer software. Keras 3, 2015-2024. Website; https://keras.io.

[10] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[11] Google Colaboratory. Website; https://colab.research.google.com/ accessed March 28, 2024.

[12] McKinney, W. Data structures for statistical computing in Python. Proc. 9th Python in Sci. Conf., 2010. pandas computer software via website; https://pandas.pydata.org accessed March 28, 2024.

[13] Bolukbasi, M., Mohammadi, J., & Arditi, D. "Estimating the Future Condition of Highway Bridge Components Using National Bridge Inventory Data." Pract. Period. Struct. Des. Constr., pp. 16-25, vol. 9, no. 1, Feb. 1, 2004 (ASCE).

[14] Chen, T. & Guestrin, C. "XGBoost: A Scalable Tree Boosting System." KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 2016, pp. 785–794.

[15] Fisher, E.M. (1936) "Linear Discriminant Analysis." Statistics & Discrete Methods of Data Sciences, 392, 1-5.

[16] Pedregosa et al. "Scikit-learn: Machine Learning in Python." JMLR 12, pp. 2825-2830, 2011.

[17] Bektas, B., Carriquiry, A., & Smadi, O. "Using Classification Trees for Predicting National Bridge Inventory Condition Ratings." Journal of Infrastructure Systems, vol. 19, issue 4, pp. 425-433, December 2013.