Real-Time Crack Detection in Bridges Using Monitoring and Machine Learning—Verified with an Actual Damage Case

IMANE BAYANE[1], JACOB NYMAN[2], JENS HÄGGSTRÖM[3] and JOHN LEANDER[4]

[1]Structural Engineering and Bridges, KTH-Royal Institute of Technology, Brinellvägen 23, 10044, Stockholm, Sweden < bayane@kth.se>
[2]IoTBridge AB, Stockholm Sweden
[3]Trafikverket, Stockholm Sweden
[4]Structural Engineering and Bridges, KTH-Royal Institute of Technology, Brinellvägen 23, 10044, Stockholm, Sweden

ABSTRACT

Detecting damage in bridges that present signs of deterioration or have exceeded the expected lifespan is critical for ensuring safety in service. This paper suggests an approach for real-time damage detection for such bridges through monitoring and machine learning algorithms, which serve as timely alarms for decision-making and subsequent damage identification.

The approach involves five steps: monitoring, data collection, data separation, feature extraction, and anomaly detection in real-time. Monitoring is ensured by strain gauges, accelerometers, and a temperature sensor. Data collection is ensured at high frequency continuously to capture the dynamic effects of loading. Data separation is provided to classify monitoring data according to loading events, which is in the case of the study characterized by the bridge opening, the bridge closing, and train passages. Feature extraction is provided to characterize monitoring data for each loading event. Anomaly detection is performed by the Isolation Forest and the One-Class Support Vector Machine algorithms. The algorithms are implemented in real-time for each new event.

The approach is illustrated in a full-scale post-damage case study of a steel-bascule-railway bridge, in service since 1916, with signs of corrosion and fatigue. The results demonstrate the ability of the approach to capture a cracking event in real-time. The Isolation Forest algorithm is found to be more robust for damage detection compared to the One-Class Support Vector Machine. It assigned high scores to the events occurring during and after the cracking, highlighting its ability to capture such incidents promptly. These findings have significant implications for bridge owners as they can identify damage in components in real time, enabling them to take timely measures such as traffic interruption and subsequent repairs.

Keywords: Bridge, Damage, Real-Time, Machine Learning, Isolation Forest (IF), One-Class Support Vector Machine (OCSVM)

INTRODUCTION

The replacement of bridges exceeding the expected lifespan or showing signs of damage is not always feasible with the increasing demand for infrastructure, limitations on resources for new construction, and requirements to maintain traffic flow. There is a growing need for innovative approaches to managing such bridges, as a significant number of bridges currently in service are near or have exceeded their expected lifespan.

Monitoring the bridge's critical parts and ensuring real-time damage detection can be good alternatives to avoid stopping traffic and ensure bridge safety in service. Bridge monitoring over the last two decades has seen significant improvements in sensor technology, data acquisition systems, and data processing techniques. However, the interpretation of monitoring data in real-time is still challenging due to the large volume of data and the difficulties related to converting it to valuable information.

Machine learning is a promising tool to address data interpretation challenges. One of the most used cases of machine learning to find and identify outliers in large data is anomaly detection algorithms. Previous studies have been conducted on damage detection using supervised machine learning algorithms and simulated damages in numerical models and laboratory experiments. However, the developed algorithms have often poor performance in practice where bridges are exposed to different uncertainties and uncontrollable conditions [1].

Real-time damage detection in bridges remains challenging, with limited success achieved under operational loadings. There are few related literatures and publications in the field of anomaly detection applied to bridge assessment.

Recently, techniques such as Support Vector Machine and Isolation Forest have emerged as promising approaches for detecting anomalies in large data sets, offering improved prediction accuracy and computational efficiency [2]–[4].

This paper presents an approach to interpret monitoring data for real-time damage detection by using the Isolation Forest (IF) algorithm and the One-Class Support Vector Machine (OCSVM) algorithm. The approach is illustrated with a case study of a currently in-service steel-bascule-railway bridge that experienced a damage occurrence event to demonstrate its applicability in practical bridge monitoring scenarios.

METHODOLOGY

The methodology for real-time damage detection is outlined in the following subsections. A description of the monitoring system and the sensors employed for data collection is presented. The features extracted from the collected data and used to construct the feature matrix for the anomaly detection algorithms are then detailed. Finally, the unsupervised machine learning algorithms applied in this study are described.

Monitoring

The local behavior of bridge components is captured at critical components using strain gauges. The global dynamic behavior of the bridge is captured using accelerometers at different components of the bridge. Ambient temperature is measured using a weather station to identify temperature effects on measurements.

Data Collection

Monitoring data is collected continuously at high frequency to capture the dynamic response of the bridge and detect possible damage initiation or propagation. Data has to be made available for analysis at the site to allow real-time damage detection. With IoT solutions for communication, all computations can be performed in a cloud solution for storage and result presentations.

Data Separation

In the data separation process, the monitoring data is organized into subsets based on the corresponding loading events. Analyses can be performed directly on subsets from the database. For the specific case study, inclinometer measurements are utilized to divide the data into three sets, corresponding to the events associated with the opening of the bridge, the closing of the bridge, and train passages.

Feature Extraction

Features are extracted from strain and acceleration measurements for each sensor and event in real time. For a given measurement $x(x_1, \dots, x_n)$ in the time domain [1, n], kurtois, root-mean squares (RMS), peak to RMS, root sum of squares, peak to peak, sinad and the mean frequency are extracted and computed using established functions as shown in Equations 1-5.

Kurtois K is a measure of how outlier-prone a distribution is.

$$K = \frac{\frac{1}{n}\sum_1^n (x_i - \mu_x)^4}{\left(\frac{1}{n}\sum_1^n (x_i - \mu_x)^2\right)^2} \tag{1}$$

where μ_x is the mean of the vector x.

Root-Mean Square (RMS) is the root mean square of the measurement vector x.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2} \tag{2}$$

Peak to RMS (PRMS) is the ratio of the largest absolute value in x to the root-mean-square (RMS) value of x.

$$PRMS = \frac{\max\left(|x_1|, \dots, |x_2|\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} |x_i|^2}} \tag{3}$$

Root-Sum of Squares (RSS) is the root-sum-of-squares of the measurement vector x.

$$RSS = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{4}$$

Peak to Peak is the difference between the maximum and minimum values in x.
Sinad is the signal to noise and distortion ratio of the measurement vector.
Mean Frequency estimates the mean normalized frequency of the power spectrum of the time-domain vector x.

To these features are added the maximum, the minimum, the mean and the standard deviation of measurements for each given event and sensor. Moreover, the correlation coefficient between the adjacent strain gauges or accelerometers are used as feature inputs to anomaly detection algorithms. Correlation coefficient $\rho(x, y)$ is the correlation coefficient between two measurements $x(x_1, \dots, x_n)$ and $y(y_1, \dots, y_n)$.

$$\rho(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \mu_x}{\sigma_x}\right)\left(\frac{y_i - \mu_y}{\sigma_y}\right) \tag{5}$$

where $\mu_x, \sigma_x, \mu_y, \sigma_y$ are respectively the mean and the standard deviation of the vectors x and y.

To eliminate the direct effect of temperature, measurements are reset to zero for each event before generating the feature matrix. This method is supported by the clear correlation observed between temperature variations and initial values of strain measurements for each event. Once the feature matrix is created, it is normalized before applying any anomaly detection algorithms.

Anomaly Detection by One-Class Support Vector Machines (OCSVM)

One-Class Support Vector Machines (OCSVM) is a machine learning algorithm used for anomaly detection in data where labels (normal/ anomalous) are not available. It has the ability to handle high-dimensional data and the potential to detect novel or previously unseen anomalies.

OCSVM maps the input data into a higher-dimensional space using a Kernel function and tries to separate the data points from the origin. By using the training data, OCSVM creates a decision function that captures the boundary between normal and anomalous data points. This boundary separates normal data points in a lowdimensional feature space while isolating anomalous data points that deviate from the normal data distribution.

Anomaly Detection by Isolation Forests (IF)

The Isolation Forest algorithm is a machine learning technique used for anomaly detection that is based on the concept of isolating anomalies from the majority of normal data points using an ensemble of isolation trees. It can handle high-dimensional data and detect global and local anomalies in large datasets.

The algorithm creates an ensemble of decision trees that recursively partition the data points by randomly selecting a feature and a split value. This process continues until each data point is isolated in its own tree or until a predefined maximum depth is reached [5]. Anomalies are more likely to be isolated in shorter paths in a separate leaf node closer to the root node since they are different from the majority of normal data points. The Isolation Forest algorithm calculates an anomaly score for each data point based on the average path length required to isolate the data point across all trees. Higher average path lengths indicate a higher likelihood of being an anomaly.

CASE STUDY

The approach is illustrated in a full-scale post-damage of a steel-bascule-railway bridge in service since 1916 (Figure 1), and it is evaluated using a dataset with an actual damage case provided by the company IoTBridge® [6]. An anomaly was detected in March 2023 using their technology and verified by visual inspection as a crack.

The bridge has a movable leaf made of a riveted truss and a concrete counterweight. It has an overall length of 68 m, a truss main span length of 42 m and a height equal to 7.6 m. The bridge's lower deck consists of steel members built up by riveted plates and L profiles, and the upper chords have box sections laterally braced by transversal trusses consisting of L profiles.

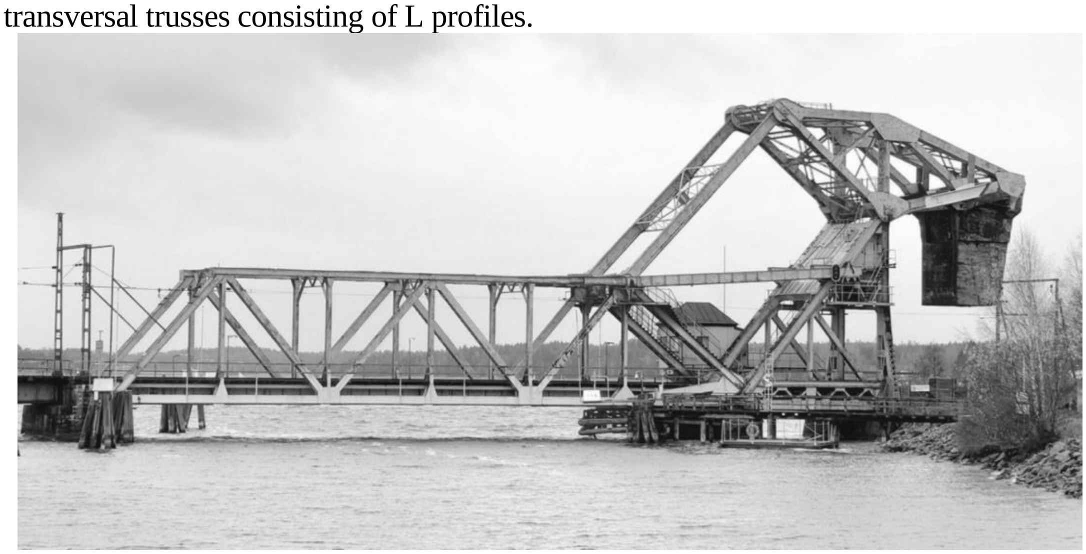

Figure 1. View of the case study of the steel-bascule-railway bridge

The bridge has exceeded the expected service life and has been presenting signs of corrosion and fatigue. It has been instrumented since December 2021. The monitoring system comprises 16 strain gauges *SG1-SG16*, 5 piezoelectric uniaxial accelerometers *A1-A5*, 1 inclinometer *Inc1*, and 1 weather station to measure ambient temperature, wind direction and speed. Sensor deployment is shown in Figure 2.

The monitoring system comprises an HBM quantum data-acquisition system that is connected to a 4G router to ensure real-time transmission of the collected monitoring data to the cloud server. Data was sampled continuously with 200 Hz and was stored every 10 minutes and accessed through the IoTBridge® cloud-based solution (See [6] for more details).

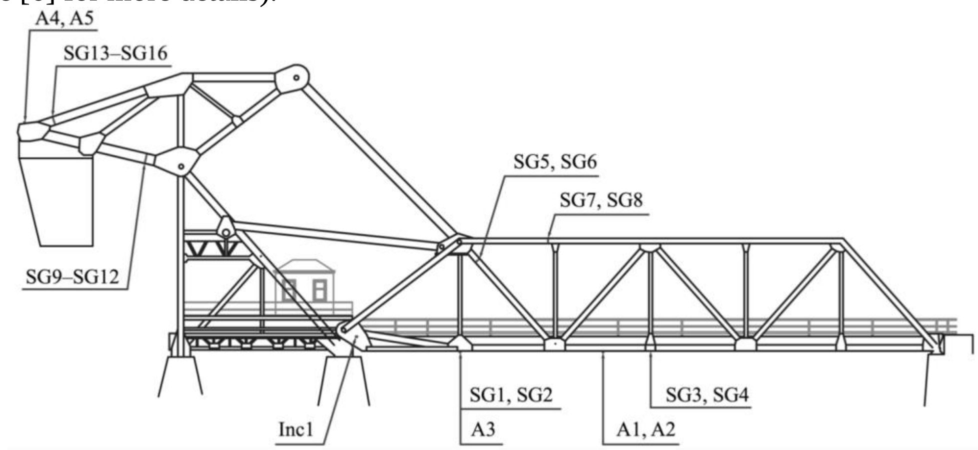

Figure 2. Sensor Deployment

Figure 3 illustrates a typical sensor output for train passages as well as the opening and closing of the bridge. The first value of each event measurement is subtracted, and all values are presented on the same time scale for clarity.

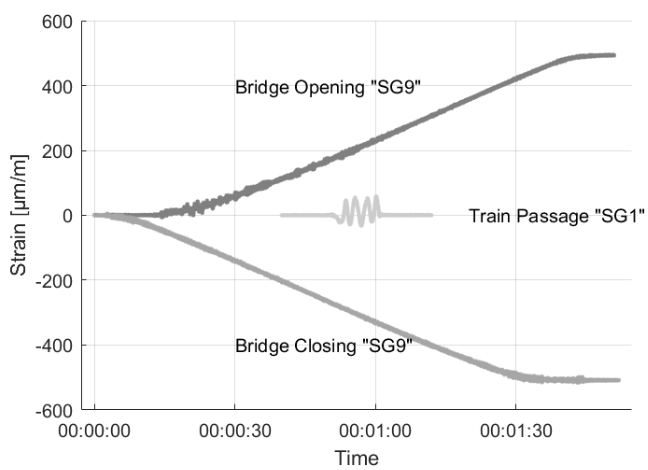

Figure 3. Typical sensor outputs for train passages, and the opening and the closing of the bridge

RESULTS

Unsupervised machine learning is used to analyze strain and accelerometer data sets obtained from the bridge during opening, closing, and train passages. The objective is to identify anomalies in measurements using raw data without labels and detect possible damage occurrence. This is achieved by using IF and OCSVM, two anomaly detection algorithms as described in the methodology part.

Algorithm inputs include the computed feature matrix for each new event with the feature matrix of a selected number of previous events along with the contamination rate. The contamination rate is the expected fraction of anomalies in the data set that is in the range of [0,1].

Algorithm outputs include two types of information: a binary classification label (0 or 1) indicating whether the event is normal or anomalous, and an assigned score for anomalous events. Scores close to 0 are assigned to normal observations, and scores close to 1 are assigned to anomalies. For example, when the contamination rate is set to 0.05, the 5% of events with the highest scores are detected as anomalous.

Figure 4 shows the scores of the detected anomalies in SG9 measurements during the opening of the bridge for the period January to March 2023 giving 500 events and a contamination rate of 5%.

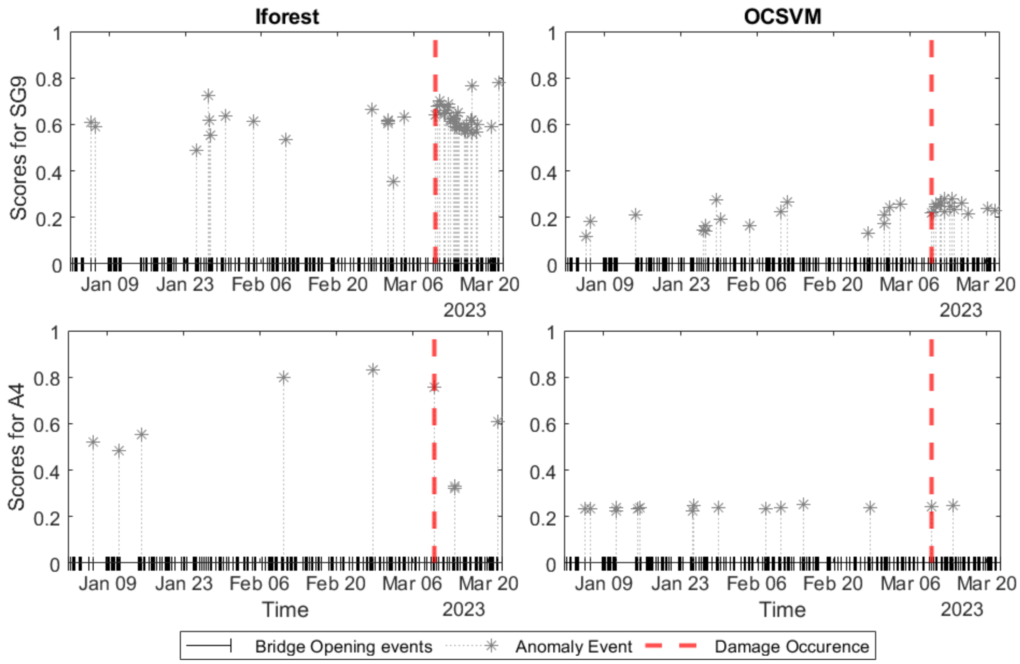

Figure 4. Scores of the detected anomalies in SG9 and A4 measurements during bridge opening; scores close to 0 are assigned to normal observations, and scores close to 1 are assigned to anomalies.

Around 15 anomalies are detected by both algorithms during the opening of the bridge for the period 1st of January to 8th of march. The detected anomalies are analyzed manually to determine the cause, which varies from signal shifting and loss to malfunctions during the opening of the bridge.

Anomalies have been continuously detected for SG9 since March 9th, with the IF algorithm showing particularly high scores. The raw data was further processed and revealed a prominent acceleration peak at sensors A2, A3, A4, and A5, along with a shift in strain measurements. Specifically, there was an increase in strains at SG9 while a decrease was observed at SG10, as demonstrated in Figure 5

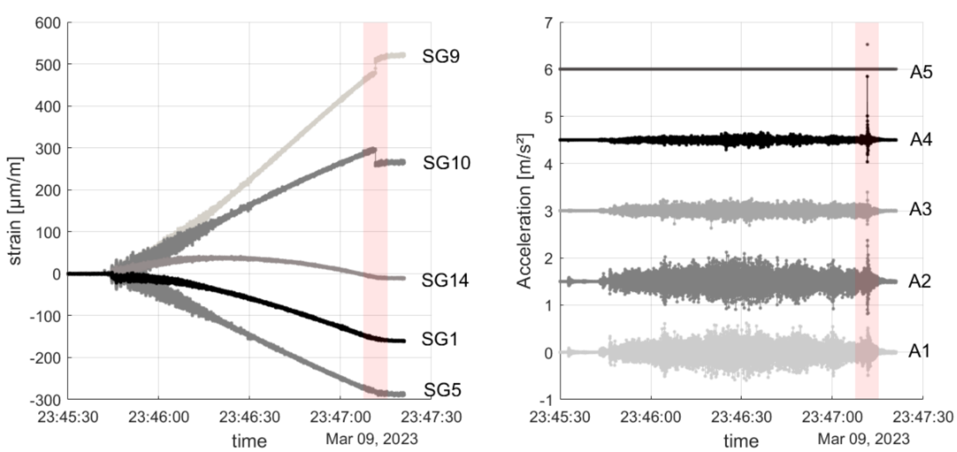

Figure 5. Sensor measurement during damage occurrence

These anomalies were due to a cracking event as it was confirmed by a visual inspection for which it was possible to identify a visible crack in the component where SG9 and SG10 are installed.

Both IF and OCSVM algorithms were efficient in detecting the occurrence of damage. Most accelerometers (A2-A5) detected the anomaly instantly. However, strain gauges (SG9 and SG10) near the crack detected the anomaly instantly and subsequently. This finding underscores the importance of having strain gauges near the crack for localizing the damage and detecting it during and after the occurrence.

Unsupervised anomaly detection requires the estimation of the contamination rate in data sets and the choice of the feature matrix. Sensitivity analysis on the anomaly detection algorithms is performed to explore better the effect of the contamination rate and the feature matrix.

Different inputs were considered, including changing the sensor types used to construct the feature matrix, combining feature matrixes of adjacent sensors like *S* (*SG9* and *SG10*) and *A* (*A4* and *A5*), altering the number of events used to construct the feature matrix from 100, 500 to 1000, and modifying the percentage of expected anomalies from 1%, 5% to 10% for the given number of events. This estimation is based here on the examination of the data points classified as anomalies. Damage occurrence in bridges is rare and so are the anomalies considered in the data set. The resulting scores for the cracking event are presented in Table 1.

Table 1. Resutling scores for the damage event for different contamination rates and events. Scores close to 0 are assigned to normal observations, and scores close to 1 are assigned to anomalies.

Inputs		**Outputs**											
N° events	**Contamination Rate**	**IF**						**OCSVM**					
		SG9	**SG10**	**S**	**A4**	**A5**	**A**	**SG9**	**SG10**	**S**	**A4**	**A5**	**A**
100	**10%**	0.73	0.68	0.72	0.78	0.91	-	0.25	0.21	0.29	0.39	0.39	-
	5%	0.73	0.68	0.72	0.78	0.91	-	0.24	0.22	0.26	0.29	0.31	-
	1%	0.73	0.68	0.72	-	0.91	-	0.13	0.15	0.16	-	-	-
500	**10%**	0.64	0.54	0.62	0.76	0.86	-	0.18	0.24	0.24	0.30	0.32	-
	5%	0.64	-	0.62	0.76	0.86	-	0.22	0.23	0.23	0.24	0.20	-
	1%	0.59	-	0.62	0.76	0.86	-	0.13	-	0.15	0.07	0.06	-
1000	**10%**	0.62	0.58	0.60	0.74	0.83	-	0.13	0.22	0.17	0.32	0.32	-
	5%	0.62	0.58	0.60	0.74	0.83	-	0.10	0.17	0.13	0.29	0.22	-
	1%	0.62	-	-	0.74	0.83	-	0.10	-	0.06	0.09	0.06	-

A contamination rate of 10% is found to detect the cracking event by all sensors for both algorithms with high scores for the IF algorithm. However, too many false-positive anomalies are detected, making it difficult to identify the cracking event manually. Using a contamination rate of 1% detects fewer anomalies, but the cracking event may only be detected by specific sensors. A contamination factor of 5% is found to detect the cracking event with high scores and less false positive anomalies for the different numbers of events.

The IF algorithm is more robust than OCSVM for detecting damage in real-time, as it consistently attributed high scores to the cracking event even with varying feature matrix size and estimated anomaly fraction.

Using combined strain measurements to construct the feature matrix is effective in detecting the cracking event, whereas combining accelerations is found to be ineffective. Choosing the feature matrix construction method and estimating the anomaly fraction are found to be more important than choosing the algorithm for detecting the cracking event.

CONCLUSIONS

The results and contributions of the investigation are concluded as follows.

- The efficiency of the unsupervised machine learning approach for detecting damage in real time has been verified through a real-case study of a bridge in service by implementing Isolation Forest and the One-Class Support Vector Machine algorithms on chosen features from strain and acceleration measurements. In particular, the approach was successful in detecting a cracking event.
- The results indicate that strain gauges located near the damage site are necessary to detect and localize the damage. However, accelerometers can still detect the damage even when placed at a distance from the damaged component. Therefore, accelerometers are effective for damage detection, while strain gauges are essential for damage localization and subsequent detection.
- Choosing the appropriate feature matrix construction method and accurately estimating the contamination rate are crucial for detecting the cracking event, with combined strain measurements being effective and combining accelerations being ineffective. Additionally, the Isolation Forest algorithm is more robust in detecting damage in real-time compared to the One-Class Support Vector Machine.

The results of this case study demonstrate the potential of unsupervised machine learning in promoting maintenance for bridges, enabling more targeted visual inspections and better allocation of maintenance budgets.

REFERENCES

[1] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, “Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection,” *J. Struct. Eng.*, vol. 146, no. 5, p. 04020073, May 2020, doi: 10.1061/(ASCE)ST.1943-541X.0002535.

[2] Y. Chen, Z. Zhao, H. Wu, X. Chen, Q. Xiao, and Y. Yu, “Fault anomaly detection of synchronous machine winding based on isolation forest and impulse frequency response analysis,” *Measurement*, vol. 188, p. 110531, Jan. 2022, doi: 10.1016/j.measurement.2021.110531.

[3] Z. Sun, J. Santos, and E. Caetano, “Vision and Support Vector Machine–Based Train Classification Using Weigh-in-Motion Data,” *J. Bridge Eng.*, vol. 27, no. 6, p. 06022001, Jun. 2022, doi: 10.1061/(ASCE)BE.1943-5592.0001878.

[4] Z. Sun, D. M. Siringoringo, S. Chen, and J. Lu, “Cumulative displacement-based detection of damper malfunction in bridges using data-driven isolation forest algorithm,” *Eng. Fail. Anal.*, vol. 143, p. 106849, Jan. 2023, doi: 10.1016/j.engfailanal.2022.106849.

[5] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[6] J. Nyman, P. Rosengren, P. Kool, R. Karoumi, J. Leander, and H. Petursson, “Smart condition monitoring of a steel bascule railway bridge,” no. IALCCE 2023.