lajss.org© Latin American Journal of Solids and Structures

www.lajss.org

Convolutional neural network for highway bridge indirect structural health monitoring

Pedro Gasparotti[a*], Thiago Fernandes[b], Leandro Fadel Miguel[c], Tiago Morkis Siqueira[d], Rafael Holdorf Lopez[e], Gabriel Padilha Alves[f]

[a] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email:pedrovgasparotti@gmail.com
[b] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email: morenof.thiago@gmail.com
[c] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email: leandroffm@gmail.com
[d] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email: tiago.morkis@ufsc.br
[e] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email: rafaelholdorf@gmail.com
[f] Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, 88040-900, Santa Catarina, Brazil. Email: gpadilhaalves@gmail.com

Abstract

The integrity of highway bridges is critical for infrastructure management, necessitating efficient structural health monitoring (SHM) strategies. This study utilizes indirect SHM - where vehicle-mounted sensors capture dynamic interaction responses - due to its cost-effectiveness and ability to monitor the entire bridge continuity without traffic interruption. This work addresses the challenge of detecting early-stage scour damage, a hidden, leading cause of roadway bridge failures, using indirect monitoring data. The main contribution is the systematic investigation of the Convolutional Neural Network (CNN) capability to detect low-intensity damage, representing the initial stages of foundation scour, under a comprehensive set of Environmental and Operational Variabilities (EOVs). The dataset is generated from a numerical finite element code (VBI-2D) simulating the vehicle-structure dynamic interaction. Damage is modeled as a reduction in foundation support stiffness at the midspan. The EOVs incorporated include variations in vehicle properties, speed fluctuations, and stochastic road surface irregularities (ISO Class 'A'). Additionally, the study evaluates the damage-classification performance for different sensor locations along the vehicle and quantifies the uncertainties arising from both data variability and the stochastic nature of the learning algorithm. The optimal CNN architecture, determined via Bayesian hyperparameter tuning, demonstrated high performance. The network achieved an overall accuracy the CNN achieved mean accuracies of ≈93.0% (DOF-6), ≈90.3% (DOF-1), and ≈82.2% (DOF-5), with the broadest spread at DOF-5; 95% confidence intervals for accuracy are 88.9-97.1% (DOF-6), 87.9-92.6% (DOF-1), and 77.8-86.5% (DOF-5), which consisted of samples unseen during the training and testing phases. This rigorous validation demonstrates the feasibility of using supervised CNNs with indirect monitoring data for early scour detection in highway bridges, serving as proof-of-concept and giving a step toward the practical implementation of robust SHM systems.

Keywords
Structural health monitoring, Machine learning, Concrete bridges, Artificial Intelligence, Finite element methods, Numerical modeling, Structural dynamics

1 INTRODUCTION

Among the various deterioration mechanisms threatening structures such as bridges and overpasses, scour, the erosion of soil around bridge foundations due to water flow, stands as one of the leading causes of bridge failures worldwide. According to Wang et al. (2017), scour is responsible for more than 60% of bridge collapses in the United States, with similar trends observed globally. This erosion process significantly undermines structural stability by reducing foundation support capacity, potentially leading to catastrophic failures with severe economic losses and humanitarian tragedies, as highlighted by Balageas (2010). The critical nature of scour damage lies in its hidden progression beneath water level, making it particularly challenging to detect through conventional visual inspection methods. Deng et al. (2016) emphasize that traditional inspection approaches often fail to identify scour-related damage in early stages, when preventive interventions would be most effective and economical.

Drive-by monitoring, also referred to as indirect Structural Health Monitoring (SHM), has emerged as a promising alternative to conventional direct monitoring techniques for bridge condition assessment. This methodology, pioneered by Yang et al. (2004) in their seminal work on extracting bridge frequencies from the dynamic response of a passing vehicle, leverages the principle that vehicles traversing a bridge act as moving sensors, capturing information about the structure's condition through their own dynamic responses. The vehicle-bridge interaction (VBI) generates acceleration signals that contain information sensitive to changes in structural properties, including stiffness reduction caused by damage. Unlike direct monitoring, which requires sensors to be permanently installed on the bridge structure, drive-by methods utilize sensors mounted on vehicles, offering significant advantages in terms of spatial coverage, implementation cost, and operational flexibility. This approach enables continuous monitoring of multiple bridges along a roadway without traffic interruption, addressing logistical challenges and high costs associated with instrumenting each structure individually, as highlighted by Spencer Jr. et al. (2017), Malekjafarian et al. (2015) and Carnevale et al. (2019).

However, the practical implementation of drive-by monitoring faces substantial challenges due to operational and environmental variabilities (EOVs) inherent to the vehicle-bridge system. These variabilities, according to Locke et al. (2020) include fluctuations in vehicle properties (mass, suspension characteristics, tire pressure), operational conditions (speed, load distribution), environmental factors (temperature, humidity), and road surface irregularities. As demonstrated by Corbally and Malekjafarian (2022), temperature variations alone can significantly affect the dynamic behavior of materials and consequently mask or amplify damage indicators, potentially leading to false positives or missed detections. Additionally, the stochastic nature of road surface roughness introduces noise that can obscure damage-related features in the acquired signals. Fallahian et al. (2018) further emphasize that without proper consideration of these EOVs, damage detection algorithms may misclassify normal operational variations as structural damage, severely compromising the reliability of monitoring systems. This sensitivity to multiple confounding factors represents one of the primary obstacles to transitioning drive-by methodologies from controlled experimental settings to real-world operational environments, an issue pointed out by Fernandes et al. (2025).

Recent advancements in machine learning and deep neural networks have opened new pathways for processing drive-by data and addressing the challenges posed by operational variabilities. Using unsupervised learning methodologies, several studies have focused on detecting reductions in flexural stiffness through advanced signal-processing techniques. For instance, Cantero et al. (2024) developed a comprehensive numerical benchmark comparing four data-driven methods for bridge damage detection, demonstrating the potential of machine learning approaches to manage the complexity of vehicle-bridge interaction signals. Demirlioglu and Erduran (2024) investigated the use of continuous wavelet transforms for drive-by damage detection. With respect to scour damage, Zhang et al. (2022) examined scour detection in highway bridges using a wavelet-based approach, with the feasibility of their method validated through both numerical simulations and laboratory-scale experiments.

On the other hand, supervised learning approaches using Convolutional Neural Networks (CNNs) have demonstrated remarkable capabilities in extracting damage-sensitive features from complex signals to classify the damage severity. Corbally and Malekjafarian (2024a) developed a data-driven algorithm for drive-by condition monitoring, to classify damage presence, while their subsequent work (Corbally and Malekjafarian, 2024b) extended this framework to identify damage type (girder crack and support condition damage), location, and severity using deep learning architectures. Studies using CNNs and drive-by measurements to classify scour damage may be found for railway bridges. Fernandes et al. (2025) employed a CNN-based framework to detect scour and other simultaneous damage mechanisms by processing vehicle acceleration data through a Bayesian-optimized network architecture, demonstrating strong robustness to EOVs and achieving high classification accuracy across different sensor locations. In a complementary study, Fernandes et al. (2026) showed that incorporating vehicle speed responses as additional CNN inputs further improves the model's accuracy in classifying scour-related damage.

Despite these significant advances, critical gaps remain in the current state of research, especially in highway bridges. Most existing studies focus on damage scenarios that consider crack girder or support condition damages. The literature lacks the investigation of early scour damage detection using indirect monitoring in highway bridges, especially in the presence of EOVs, and detecting low-intensity damage (less than 10% stiffness reduction). Hence, this work addresses these gaps by investigating the effectiveness of supervised Convolutional Neural Networks for identifying early scour-induced damage using indirect monitoring data. The methodology is validated through comprehensive numerical simulations that incorporate realistic operational variabilities including vehicle property variations, speed fluctuations, and road surface irregularities. Damage is modeled as foundation stiffness reduction at different severity levels (5% and 10%), and the CNN's capability to detect these damage states is evaluated through rigorous classification performance metrics, providing insights into the feasibility of early damage detection under challenging operational conditions. In addition to addressing early scour detection, this study advances the field by systematically exploring different sensor locations along the vehicle to assess their influence on classification performance. Furthermore, it quantifies the uncertainties associated with damage classification arising from data variability induced by EOVs and from the intrinsic stochasticity of the CNN training process. Together, these analyses provide a deeper understanding of the robustness and practical feasibility of indirect scour detection under challenging operating conditions.

2 NUMERICAL MODELING AND DATA ACQUISITION

This section presents the numerical framework adopted in this study, which integrates a detailed Vehicle-Bridge Interaction model with realistic representations of scour damage and key Environmental and Operational Variabilities. The simulations capture the dynamic interaction between a heavy vehicle and the bridge structure, incorporate foundation stiffness degradation to represent different scour levels, and include stochastic factors such as road roughness and vehicle speed fluctuations. In addition, different degrees-of-freedom along the vehicle were selected to capture its vertical acceleration response, ensuring a richer and more informative representation of the drive-by measurements. This integrated approach ensures that the synthetic dataset reflects the essential complexities of real-world drive-by monitoring, providing a robust and representative basis for evaluating the proposed damage-classification methodology.

2.1 VEHICLE-BRIDGE INTERACTION MODEL

The Vehicle-Bridge Interaction (VBI-2D) model is a numerical simulation tool for analyzing the dynamic response of vehicles traversing a bridge, developed by Cantero (2024). This MATLAB-based software enables the simulation of traffic scenarios, including the effects of road profile irregularities and the interaction between vehicles and bridge structures, employing the Finite Element Method (FEM) to represent structural behavior.

The vehicle model adopted represents a truck-trailer combination with a total average mass of 35×10^3 kg. The system comprises 14 degrees of freedom (DOFs): Body 1 (Tractor), with 3 DOFs (vertical displacement, pitch rotation, and center of mass displacement); Body 2 (Trailer): 3 DOFs (vertical displacement, pitch rotation, and center of mass displacement); Axle Groups: 8 DOFs (4 axle groups × 2 DOFs each: vertical displacement and unsprung mass displacement). Figure 1 provides an overview of the vehicle-bridge interaction model, from which acceleration responses are collected at vehicle DOFs 1, 5, and 6, corresponding to vertical acceleration measurements.

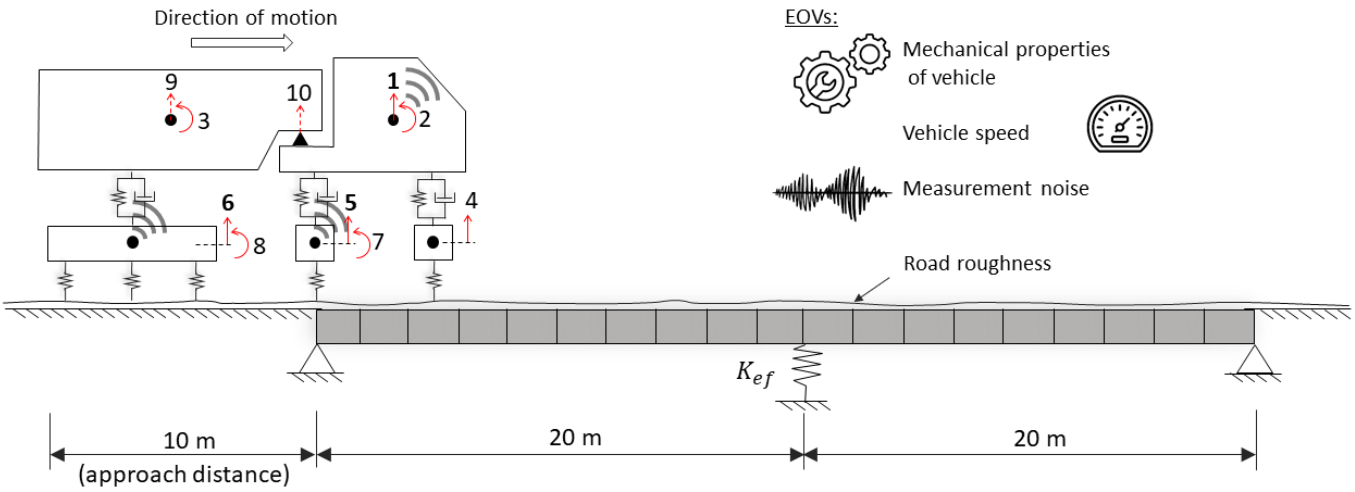

Figure 1 *Schematic representation of the vehicle-bridge interaction model and the considered Environmental and Operational Variabilities.*

Vertical displacements of sprung and unsprung masses, pitch rotations of tractor and trailer bodies, suspension dynamics (compression and extension) and tire deformations at contact points where considered as the vehicle motions. The axle masses, ranging from 700 to 2,250 kg, are incorporated into the model as lumped masses connected to the vehicle bodies through spring-damper systems representing the suspension. Each axle group is modeled as a concentrated mass element with two DOFs: the vertical displacement of the axle itself and the vertical displacement of the connection point to the vehicle body. This lumped mass approach simplifies the computational model while maintaining accuracy in capturing the dynamic interaction effects, as validated by Yang et al. (2004).

Multiple axle groups with masses distributed across tractor and trailer units feature the vehicle model. The system exhibits tire stiffness values between 1.6×10^6 to 3.9×10^6 N/m and tire viscous damping of approximately 1×10^4 N·s/m, with suspension stiffness ranging from 3.6×10^5 to 2.5×10^6 N/m and damping coefficients from 900 to 3,300 N·s/m.

The bridge is modeled as an Euler-Bernoulli beam using finite element discretization. Each finite element consists of two nodes, with each node having two degrees of freedom: vertical displacement and rotation. The adoption of Euler-Bernoulli beam theory is justified for this study given the span length of 25 meters and the slenderness ratio of the bridge cross-section, similar to the employed method by Sarwar and Cantero (2023). For the geometric and material properties considered, the contribution of shear deformations to the total deflection is less than 2%, well within acceptable limits for engineering applications. The Euler-Bernoulli assumption (plane sections remain plane and perpendicular to the neutral axis) is valid when the span-to-depth ratio exceeds 10, which is satisfied in this configuration.

The contact between the vehicle wheels and the bridge surface is modeled as an interaction force applied at the contact point, rather than as a displacement compatibility condition. At each time step, the contact force is calculated based on the relative displacement and velocity between the tire and the bridge surface. Dynamic interaction models are employed to compute the vehicle-bridge interaction (VBI) forces, where the coupled system is governed by a set of equations of motion:

$$Ms\,\ddot{q} = Cs\,\dot{q} + Ks\,q = Fs \tag{1}$$

in which Ms, Cs, and Ks denote the time-dependent mass, damping, and stiffness matrices of the vehicle-pavement-bridge system. The vector q contains the generalized displacements of the vehicle and bridge subsystems, while Fs represents the external and interaction forces acting on the coupled system.

The bridge degrees of freedom are treated through modal superposition to reduce computational cost while maintaining accuracy. For the 25-meter simply supported bridge, the first n=10 modes are considered in the analysis, capturing all significant dynamic effects within the frequency range of interest (0-50 Hz). This modal reduction technique significantly improves computational efficiency while maintaining numerical accuracy, as the contribution of higher modes to the total response becomes negligible beyond the 10th mode.

2.2 SCOUR DAMAGE MODELING

Scour refers to the erosion of soil around bridge foundations caused by water flow, which can significantly compromise structural stability. According to Wang et al. (2017), scour-related damage is particularly challenging to detect because it occurs below the waterline, making traditional visual inspection methods ineffective. In this study, scour is simulated by reducing the stiffness of the foundation support element at the bridge midspan, representing the loss of soil-structure interaction due to erosion around the footing. The healthy-condition stiffness is set to $Kf = 3.44 \times 10\hat{}8\ Pa$, a value consistent with the vertical stiffness of shallow pad foundations modeled in the literature—for example, Fitzgerald et al. (2019) report a comparable foundation stiffness for medium-dense sand supporting a similar shallow footing configuration. In this sense, the damage can be simulated by reducing the vertical stiffness Kf by:

$$Kef = \beta Kf \tag{2}$$

where β is a reduction factor $0 \leq \beta \leq 1$ representing the severity of the scour. A value of $\beta = 0$ implies no scour. In this study, two cases were evaluated to understand the impact of varying levels of structural integrity. Minor damage is represented with $\beta = 0.95$, corresponding to a 5% reduction in foundation support stiffness, while more significant damage is modeled with $\beta = 0.90$ reflecting a 10% reduction in vertical stiffness.

2.3 ENVIRONMENTAL AND OPERATIONAL VARIABILITIES

To ensure that the numerical simulations approximate real-world operational conditions, several sources of variability are incorporated into the VBI-2D model. These environmental and operational variabilities (EOVs) are critical for evaluating the robustness of the damage detection methodology under realistic conditions. The road profile significantly affects the vehicle-bridge interaction. This investigation adopts the ISO-8608 standard for generating random road profiles. Specifically, Class 'A' profiles are employed, representing well-maintained highway surfaces. The road profile is generated in the frequency domain using the Power Spectral Density (PSD) approach. The ISO-8608 standard defines the PSD function S(n) as:

$$S(n) = S(no)\left(\frac{n}{no}\right)^{-w} \tag{3}$$

where n is the spatial frequency [cycles/m], $no = 0.1$ cycles/m is the reference spatial frequency, $S(no)$ is the PSD value at the reference frequency and $w = 2$ is the waviness exponent. The frequency range considered for road profile generation spans from $n\ min$ = 0.01 cycles/m to $n\ max$ = 10 cycles/m, corresponding to wavelengths from 100 m down to 0.1 m. This range captures all relevant irregularities that influence vehicle-bridge dynamic interaction, from long undulations to short-wavelength roughness. The road profile is generated at each simulation with a new random realization, as well as 5% of random measure noise, ensuring that each vehicle passage encounters a unique (but statistically equivalent) roughness pattern. This stochastic approach prevents the model from learning specific road features rather than damage-related patterns.

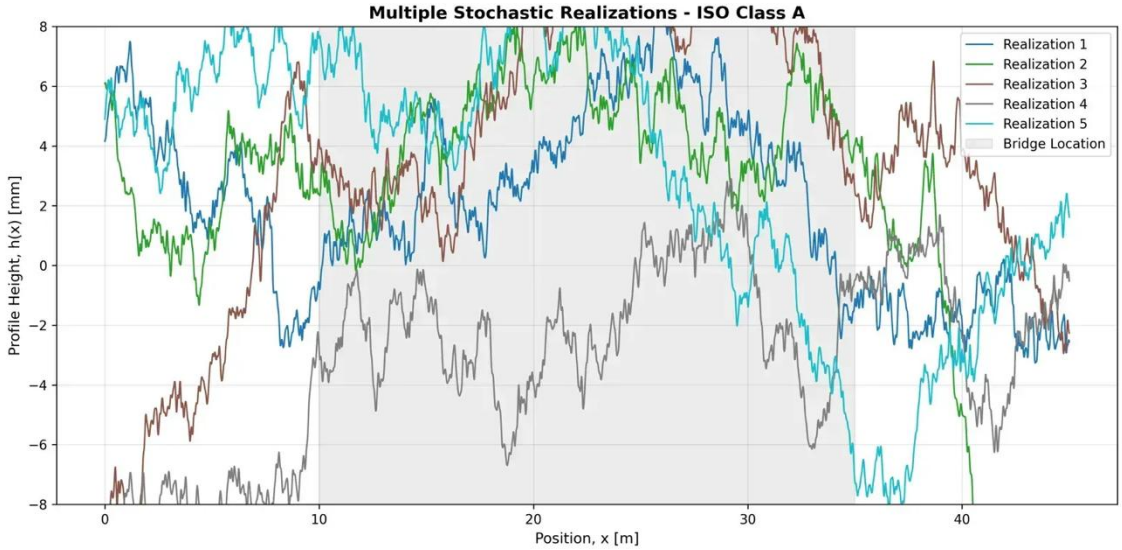

Figure 2 *Different vertical profiles generated for each algorithm iteration.*

To ensure realistic operational conditions, environmental and operational variabilities (EOVs) are incorporated through stochastic sampling of vehicle properties and road profiles. Vehicle parameters are sampled from normal gaussian distributions within the following ranges: total mass averaging 35×10^3 kg with axle masses from 700 to 2,250 kg, tire stiffness from 1.6×10^6 to 3.9×10^6 N/m (damping fixed at 1×10^4 N·s/m), suspension stiffness from 3.6×10^5 to 2.5×10^6 N/m (damping from 900 to 3,300 N·s/m), and velocity from 19 to 21 m/s.

Temperature variability is not included in this study, as Fernandes et al. (2025) showed that, in scour damage classification problems using drive-by measurements and CNNs, temperature had no significant influence on classification accuracy when compared with other EOVs.

2.4 DATA ACQUISITION

The data acquisition strategy simulates a realistic drive-by monitoring scenario where accelerometers mounted on the vehicle capture dynamic responses during the bridge crossing event. The numerical simulations employ a sampling frequency of fs=1000 Hz, corresponding to a time step of Δt=0.001s. This sampling rate is sufficiently high to capture the dynamic content of the vehicle-bridge interaction, which is dominated by frequencies below 50 Hz. According to the Nyquist-Shannon sampling theorem, a sampling rate of 1000 Hz allows accurate representation of frequency components up to 500 Hz, providing a substantial safety margin above the frequency range of interest.

The time integration of the coupled vehicle-bridge system is performed using the Newmark β method with parameters γ= 0.5 and β=0.25, corresponding to the average acceleration method. This integration scheme is unconditionally stable and provides second-order accuracy in time. The time step size of 0.001 s ensures numerical stability, with the stability criterion satisfied as:

$$\Delta t < \frac{Tmin}{10} \tag{4}$$

where $Tmin = 0.025\ s$ is the period of the highest mode considered (10^{th} mode at approximately 40 Hz). The vehicle's trajectory extends from an initial position 10 meters before the bridge entrance to a final position 10 meters beyond the bridge exit, totaling a monitored length of 45 meters for the 25-meter span bridge, and the acceleration data is recorded for three degrees of freedom (DOFs) of the vehicle: DOF 1, DOF 5 and DOF 6, as presented in Figure 1. While the simulations are performed in the time domain with fixed time steps, the vehicle position varies continuously due to its forward motion. To create consistent input dimensions for the CNN, the spatially-interpolated signals are resampled to

a fixed length of 1,936 points, corresponding to the passage over the 25-meter bridge span at 0.0129 m spacing. This resampling is performed using linear interpolation to avoid introducing artificial frequency components.

For each bridge condition (healthy, 5% damage, 10% damage), 600 vehicle passages are simulated for training the neural network, and an additional 100 passages are reserved for testing and validation. The acceleration data for each passage is stored as a vector of length 1,936, corresponding to the spatial discretization over the bridge span. The complete dataset is organized in a training set, containing 480 passages per condition (1,440 total), a validation set, containing 60 passages per condition (180 total) and a test set: 60 passages per condition (180 total).

3 DATA PROCESSING

The present investigation evaluates damage identification and classification for discrete outcomes in limited classes using 1D Convolutional Neural Network. A 1D Convolutional Neural Network is designed to extract meaningful temporal features from sequential data such as acceleration signals. At its core, the convolution operation applies learnable filters that slide across the input sequence to detect localized patterns. The discrete convolution used in 1D CNNs is expressed as:

$$y_i = (w * x)_i + b = \sum_{j=1}^{k} x_{i+j-1} \cdot w_j + b \quad (6)$$

where x is the input signal, w is the convolutional filter of length k, b is a bias term, and $y_i$denotes the feature-map output at position i. Each filter learns to detect specific characteristics of the signal, such as peaks, trends, or frequency-related signatures. Nonlinear activation functions, ReLU (Rectified Linear Unit), are then applied to introduce nonlinearity into the model (Wang et al. 2020).

To stabilize and accelerate training, Batch Normalization is applied after convolutional layers. It normalizes intermediate activations, reducing internal covariate shift and enabling the use of higher learning rates, thus improving convergence in deep models. Pooling layers, particularly max-pooling, follow the convolutional blocks to down sample the feature maps. Max-pooling retains only the highest value within each pooling window, preserving the most salient features while reducing dimensionality and computational cost. These compacted representations are then passed to fully connected layers, which aggregate extracted features to perform the final classification. Training is performed by minimizing a loss function, typically cross-entropy, via stochastic gradient-based methods (Robbins and Monro 1951), with adaptive optimizers such as Adam being widely used due to their efficiency. To improve generalization and mitigate overfitting, dropout regularization is also applied, especially in dense layers (Srivastava et al. 2014).

3.1 Training and validation dataset split

Three distinct datasets are loaded, each corresponding to a specific experimental condition: healthy baseline state (labeled yo), 5% damage state (labeled $y1$) and a 10% damage state (labeled $y2$) with one-hot-encoding. Each dataset is programmatically assigned its corresponding integer and categorical label with shape (2000x1). Subsequently, each signal is independently normalized with z-score transformation, so the dataset can be reshaped into the standard input shape for 1D CNN architectures (samples, timesteps, channels).

Dataset partitioning starts with the separation of a holdout set, reserved to attest the trained models efficiency in predicting the class of unseen data. Then, the dataset undergoes stratified subsampling, separating a balanced dataset with 100 samples from each class scenario, from which train-test split occurs with a rate of 80% for training, 10% for testing and 10% for validation. All resulting datasets, holdout, train and test are serialized to serve as input for the CNN training and validation processes.

3.2 Search for minimum sample size

Data availability may be a difficulty for real case scenarios, as pointed out by Fernandes and Lopez (2025) and Golpayegani and Moloney (2019). To systematically search for the minimum number of samples in a balanced dataset that reaches a satisfactory accuracy, this investigation used a process to generate a learning curve to evaluate the generalization capability of the classification model repeatedly building and training a generic CNN with different train set sizes (5, 10, 20, 40, 60, 80, 100, 150, 200) of each class five times to generate a confidence interval, as Figure 3 shows.

The initial convolutional architecture used to find the minimum sample size is described in Table 1, and its training strategy was dividing the dataset $Dset$ into training, validation, and testing subsets. Let ($Dtrain$), ($Dval$), and ($Dtest$) denote these subsets respectively. The division is done to ensure that the model is trained on one portion of the data ($Dtrain$), fine-tuned on another ($Dval$), and finally evaluated on unseen data ($Dtest$) to assess its generalization capability. Figure 3 displays the mean accuracy of the search for minimum training sample size, each run trains the model from scratch multiple times (5 trials) on balanced subsets of increasing size (5-200 samples per class), evaluates each on a fixed test set, and plots mean test accuracy (±1 std) vs. total training samples to show how performance scales with data volume.

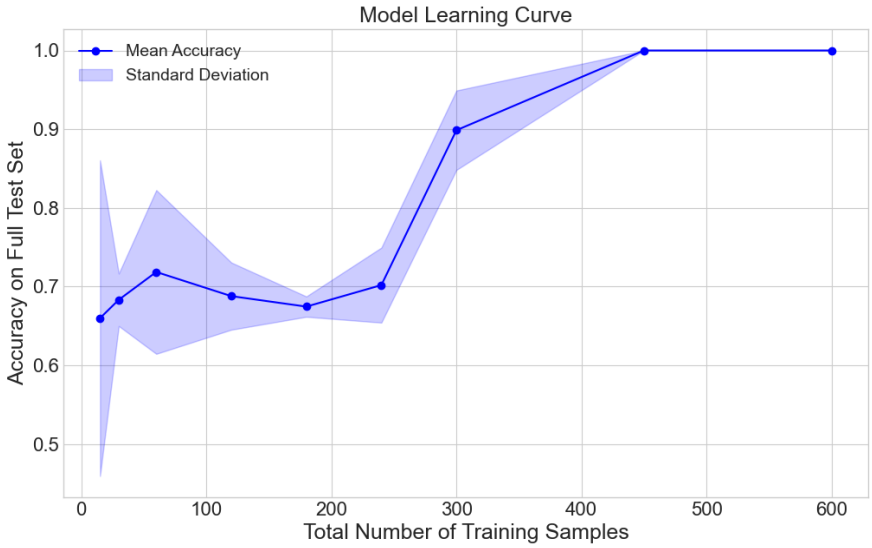

Figure 3 *Search for minimum training sample size.*

Layer	Output shape	Parameters
Input Layer	(1, 2000,1)	0
Conv1D	(1, 2000, 32)	192
Batch Normalization	(1, 2000, 32)	128
Activation	(1, 2000, 32)	0
MaxPooling1D	(1, 2000,32)	0
Conv1D (1)	(1, 1000, 64)	10 304
Batch Normalization (1)	(1, 1000, 64)	256
Activation (1)	(1, 1000, 64)	0
MaxPooling1D (1)	(1, 500, 64)	0
Dropout	(1, 500, 128)	0
Conv1D (2)	(1, 500,64)	41 086
BatchNormalization (2)	(1, 500, 128)	512
Activation (2)	(1, 500, 128)	0
Global Average Pooling	(1, 128)	0
Dense	(1, 128)	16 512
Dropout	(1, 128)	0
Dense (1)	(1, 3)	387

Table 1 *CNN architecture for minimum training sample size search.*

Further on, to find the best combination of hyperparameters, a search is performed based on the implementation of a Bayesian optimization provided by the library Keras tuner Python, with the objective function to minimize validation accuracy for 50 trials. The result architecture is described in Figure 4, and with this topology, the model is tested in the dataset to generate accuracy results presented in the following section.

Layer	Output shape	Parameters
Input Layer	(1, 2000,1)	0
Conv1D	(1, 2000, 48)	288
Batch Normalization	(1, 2000, 48)	192
Activation	(1, 2000, 48)	0
MaxPooling1D	(1, 1000, 48)	0
Dropout	(1, 1000, 48)	0
Conv1D (1)	(1, 1000, 16)	3 856
BatchNormalization (1)	(1, 1000, 16)	64
Activation (1)	(1, 1000, 16)	0
MaxPooling (1)	(1, 500, 16)	0
Dropout (1)	(1, 500, 16)	0
Global Average Pooling	(1, 16)	0
Dense	(1, 16)	8 704
Dropout (2)	(1, 16)	0
Dense (1)	(1, 3)	1 539

Table 2 *Optimal CNN architecture after Bayesian hyperparameter tuning.*

4 RESULTS

This section investigates the performance of the proposed Convolutional Neural Network (CNN) approach for scour damage classification in a highway bridge. The study utilizes data from three distinct sensor positions: the tractor box (DOF 1), tractor rear suspension (DOF 5), and trailer suspension (DOF 6), as illustrated in Figure 1. To account for inherent variability in both the data distribution and the CNN training process, each classification case was executed 20 times.

These runs were performed using different data splits and randomized seeds. Consequently, the performance metrics are presented in boxplot format to visualize the range and distribution of the results, further details can be found in Fernandes et al. (2025). All model training is conducted on a macOs equipped with an M2 Pro chip with 10 cores and 16 GB of RAM. Figure 4 presents the F1-Scores and Macro-Average Precision of 20 independent runs for the three sensor positions (DOF-1, DOF-5 and DOF-6), and the three classes (Healthy, 5% and 10%). These results show that sensor placement materially affects class-wise F1 and macro-precision: DOF-6 yields the highest macro F1 on average, followed by DOF-1, while DOF-5 lags and exhibits the largest run-to-run dispersion.

The Healthy class attains near-perfect F1 across all DOFs, whereas performance drops for the 5% and 10% classes - most noticeably at DOF-5 - reflecting residual confusion between adjacent severities, showcasing how vehicle sensor location (DOFs 1, 5, 6) influences supervised CNN performance under identical preprocessing, sampling, and training conditions across 20 independent runs per DOF (seeds 42-61). For each run, a balanced subsample of 100 signals per class is drawn after isolating a holdout set of 100/class; the subsample is then split 80/10/10 (train/test/val) with stratification.

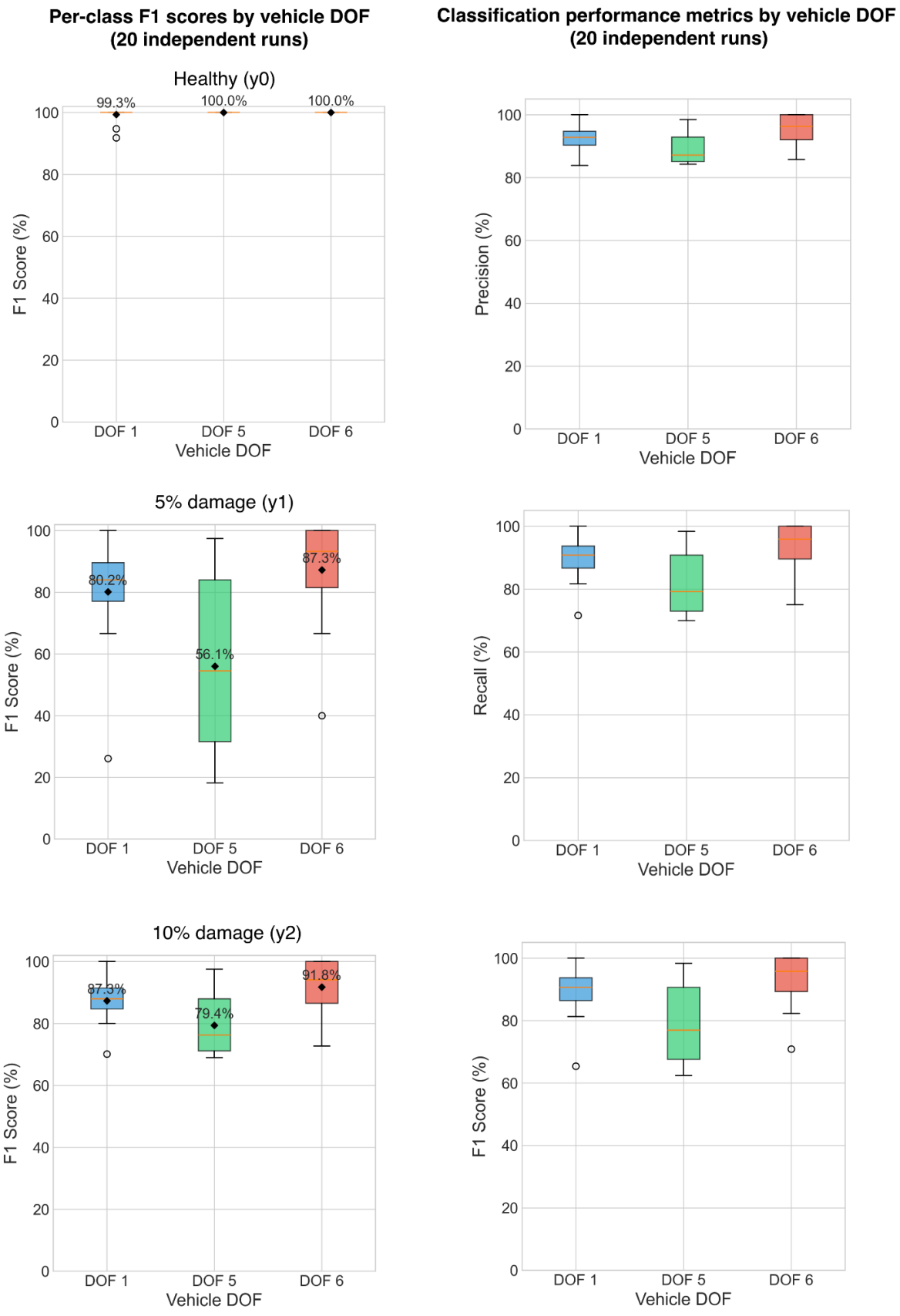

Figure 4 *F1-scores and performance metrics results by vehicle DOF*

Figure 5 shows the best confusion matrix across the 20 runs for each DOF, highlighting dominant diagonals (true positives) and any systematic off-diagonal confusions. When misclassifications occur, they are most likely to arise between the 5% and 10% damage classes (y1 vs y2), consistent with their proximity in severity, whereas the healthy class (y0) typically remains better separated. Comparing the three panels (one per DOF) reveals whether specific DOFs concentrate residual errors in particular class pairs.

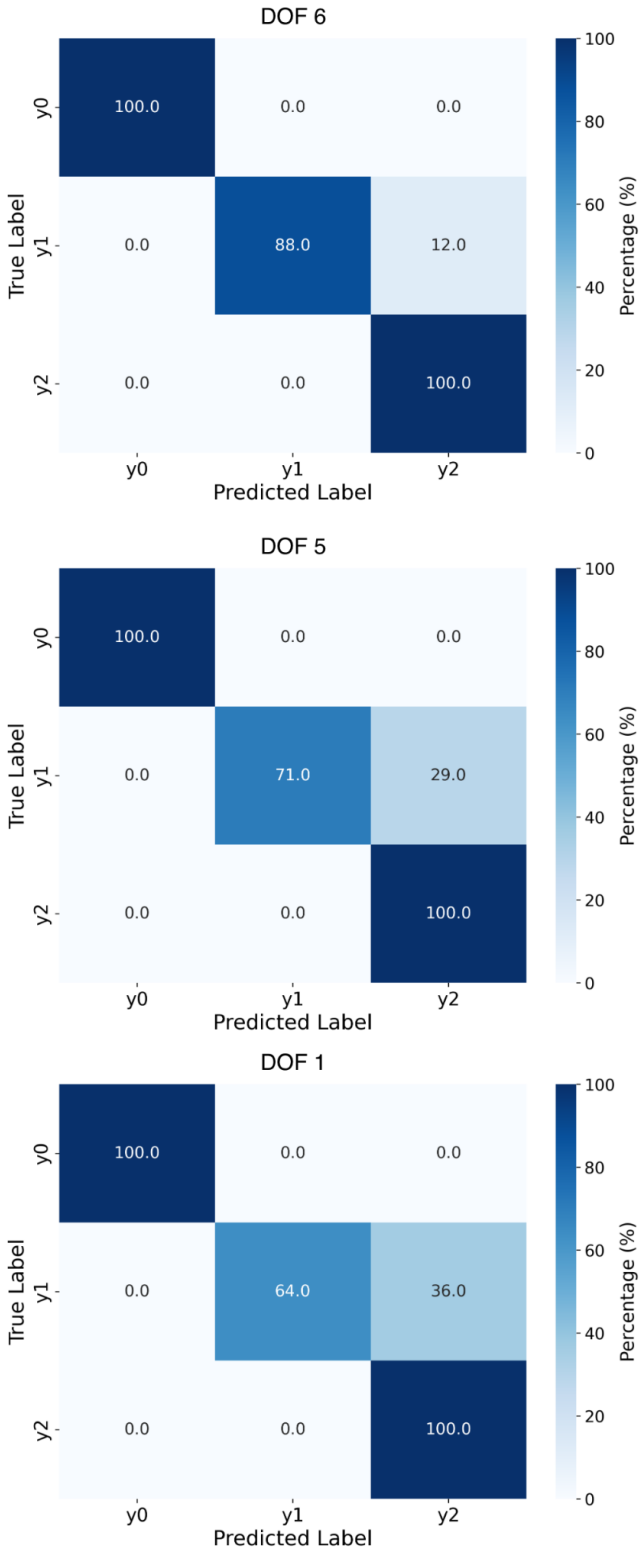

Figure 5 *- Average confusion matrixes across 20 runs for each DOF*

Figure 6 summarizes the confusion matrixes across the 20 runs for each DOF with boxplots presenting the standard deviation of the calculated accuracies. The distributions (median, IQR, whiskers) reflect the run-to-run variability driven by data resampling and training under fixed architecture and hyperparameters. Mean accuracy markers and mean ± std annotations allow a comparison of central tendency and dispersion among DOFs. This figure is the primary indicator of stability and potential DOF sensitivity under repeated trials. A one-way ANOVA confirmed that sensor placement significantly affects performance, with DOF 5 statistically worse than DOFs 1 and 6, and no significant difference between DOFs 1 and 6 for accuracy.

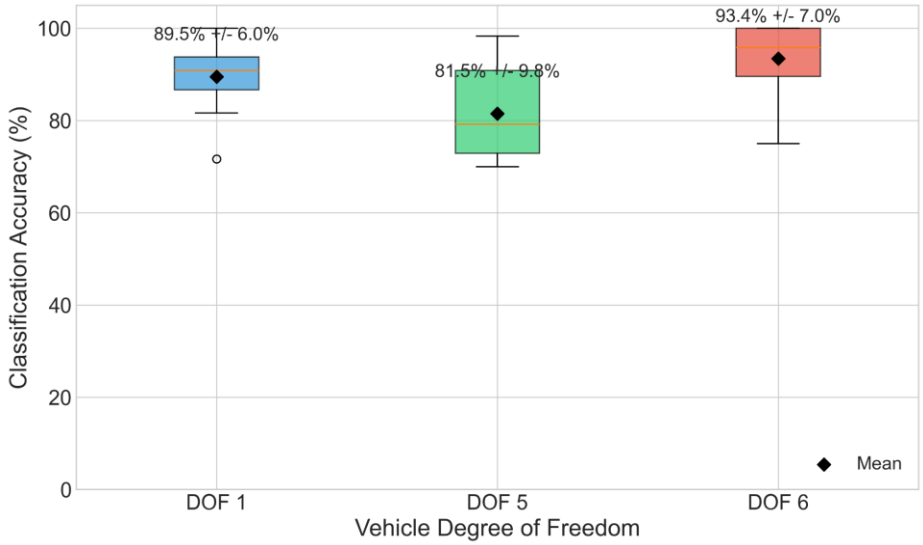

Figure 6 - *Classification accuracy standard deviation by vehicle DOF*

5 FURTHER COMMENTS: PRACTICAL FEASIBILITY AND IMPLEMENTATION CHALLENGES

Although simulation-based results demonstrate that damage-sensitive features can be reliably extracted from vehicle accelerations, translating this capability into practice involves nontrivial constraints related to data acquisition, environmental variability, and system robustness under real traffic conditions. In real-world applications, precise synchronization between vehicle location and sensor readings, as well as stable power and communication systems, is essential to avoid incomplete or misaligned datasets. Challenges such as intermittent connectivity, limited access to power in remote regions, and potential equipment vandalism further complicate long-term deployment and require robust strategies, including autonomous power supplies, redundant data storage, and secure sensor enclosures.

Additional implementation barriers arise from the vehicle itself. Depending on the bridge span, stiffness, and dynamic characteristics, reliably capturing the subtle vibration signatures associated with early scour may require a sensor-equipped vehicle with sufficient mass to excite measurable changes in the structural response. Light vehicles may fail to generate adequate dynamic interaction, reducing the method's sensitivity to incipient damage. Moreover, the approach depends on aggregating data from a large number of vehicle passages, making operational feasibility a decisive factor. In urban roads, this requirement can be met by equipping existing high-mass vehicles, such as public buses that repeatedly traverse the same route. In rural or hard-to-access regions, where installing fixed sensors is often impractical, drive-by monitoring could leverage heavy trucks involved in agricultural or mining logistics, which typically follow stable origin-destination patterns and provide both the necessary vehicle mass and the high passage frequency required for statistical learning.

These considerations highlight that practical implementation of road-based drive-by monitoring must balance technological reliability with realistic traffic and operational conditions. External disturbances such as variable speed, heterogeneous traffic flows, and uneven road profiles introduce significant noise, necessitating adaptive processing, advanced filtering, and long-term data aggregation strategies. Nonetheless, periodic multi-passage data collection

remains attractive because it enables the gradual accumulation of evidence regarding structural changes without the cost and complexity of permanent installations. When properly integrated into transportation networks, this strategy supports scalable monitoring programs capable of identifying early deterioration trends, such as progressive scour, while maintaining flexibility in inspection frequency and resource allocation across different bridge typologies and risk levels.

6 CONCLUSION

This paper explored the effectiveness of supervised Convolutional Neural Networks (CNNs) for early damage classification in highway bridges using indirect monitoring data. The methodology was validated through comprehensive numerical simulations of the Vehicle-Bridge Interaction (VBI) system, where damage was modeled as stiffness reduction at the foundation support element located at the bridge midspan, representing scour-induced damage. To ensure the robustness and practical relevance of the approach, the simulations incorporated a comprehensive set of Environmental and Operational Variabilities (EOVs), including variations in vehicle properties, speed fluctuations, and random road surface irregularities (ISO Class 'A' profiles). Specifically, the study systematically investigated CNN's ability to detect low-intensity damage scenarios of 5% and 10% stiffness reduction. The data processing involved normalization and stratified splitting to create robust training, validation, and holdout sets. A Bayesian optimization approach was used to determine the optimal CNN architecture.

The trained CNN demonstrated promising performance in classifying bridge signals based on their healthiness levels. Utilizing the optimal hyperparameter configuration, the CNN achieved mean accuracies of ≈93.0% (DOF-6), ≈90.3% (DOF-1), and ≈82.2% (DOF-5), with the broadest spread at DOF-5; 95% confidence intervals for accuracy are 88.9-97.1% (DOF-6), 87.9-92.6% (DOF-1), and 77.8-86.5% (DOF-5), which consisted of samples unseen during the training and testing phases. The confusion matrix visualized the best classification performance across all three classes: healthy, 5% damage, and 10% damage. This result showcases the potential of CNNs to accurately categorize low-intensity structural impairment under the influence of significant operational noise and variability, paving the way for broader adoption of machine learning in infrastructure monitoring.

Despite the promising numerical results, there are key limitations and areas for future research. The current work employs a 2D VBI model, and future studies should incorporate more realistic 3D models to capture complex spatial dynamics. More critically, while simulation-generated data has facilitated the initial development, the models must be validated and refined using data collected from laboratory experiments, and then, real-case scenarios and operational bridges. Hence, despite the promising results, this paper functions as a proof-of-concept, and gives a step toward the practical implementation of a drive-by scour damage detection system.

Author's Contribuitions: Conceptualization, Thiago Fernandes Rafael, Holdorf Lopez and Pedro Gasparotti; Investigation, Thiago Fernandes, Rafael, Holdorf Lopez, Pedro Gasparotti, Tiago Morkis, Gabriel Padilha Alves, Leandro Fadel Miguel - original draft, Pedro Gasparotti - review & editing Thiago Moreno Fernandes, Rafael Holdorf Lopez and Pedro Gasparotti; Supervision, Rafael Holdorf Lopez and Leandro Fadel Miguel.

Data Availability: To confirm data availability, inform one of the following information:

- Research data is available in a repository: https://github.com/pedrogasparotti/cnn_classifier

Editor: Eduardo Alberto Fancello and Paulo de Tarso Mendonça

References

Balageas, D.L., (2010). Structural health monitoring. John Wiley and Sons.

Cantero, D., (2024). VBI-2D - Road vehicle-bridge interaction simulation tool and verification framework for Matlab. SoftwareX 26:101725.

Cantero, Daniel, et al. "Numerical benchmark for road bridge damage detection from passing vehicles responses applied to four data-driven methods." Archives of Civil and Mechanical Engineering 24.3 (2024): 190.

Carnevale, M., Collina, A., Peirlinck, T., (2019). A feasibility study of the drive-by method for damage detection in railway bridges. Applied Sciences 9:160.

Corbally, R., Malekjafarian, A., (2022). A data-driven approach for drive-by damage detection in bridges considering the influence of temperature change. Engineering Structures 253:113783.

Corbally, Robert, and Abdollah Malekjafarian. "Experimental verification of a data-driven algorithm for drive-by bridge condition monitoring." Structure and Infrastructure Engineering 20.7-8 (2024a): 1174-1196.

Corbally, Robert, and Abdollah Malekjafarian. "A deep-learning framework for classifying the type, location, and severity of bridge damage using drive-by measurements." Computer-Aided Civil and Infrastructure Engineering 39.6 (2024b): 852-871.

Demirlioglu, Kultigin, and Emrah Erduran. "Drive-by bridge damage detection using continuous wavelet transform." Applied Sciences 14.7 (2024): 2969.

Deng, L., Wang, W., Yu, Y., (2016). State-of-the-art review on the causes and mechanisms of bridge collapse. Journal of Performance of Constructed Facilities 30:04015005.

Fallahian, M., Khoshnoudian, F., Meruane, V., (2018). Ensemble classification method for structural damage assessment under varying temperature. Structural Health Monitoring 17(4):747-762.

Fawcett, T., (2006). Introduction to ROC analysis. Pattern Recognition Letters 27:861-874.

Fernandes, T.M., R. H. Ribeiro, D. R. F., & Miguel, L. F. F. (2025). Early Multi-Damage Classification in Railway Bridges Using Drive-by Numerical Measurements with Piecewise Aggregate Approximation and Convolutional Neural Networks. International Journal of Structural Stability and Dynamics, 2650316.

Fernandes, T. et al. (2026) “Early scour damage detection using drive- by monitoring data through supervised learning,” Journal of Structural Design and Construction Practice. doi:10.1061/JSDCCC/SCENG- 1785 (accepted for publication).

Golpayegani, S., Moloney, C., (2019). A machine learning approach to bridge-damage detection using responses measured on a passing vehicle. Sensors 19(18):4035.

Jang, J., Jo, H., Cho, S., Mechitov, K., Rice, J.A., Sim, S.H., Jung, H.J., Yun, C.B., Spencer Jr., B.F., Agha, G., (2010). Smart Structures and Systems. Smart Structures and Systems 6(5-6):439-459.

Kamariotis, A., (2024). Monitoring-Supported Value Generation for Managing Structures and Infrastructure Systems.

Lin, C.W., Yang, Y.B., (2005). Use of a passing vehicle to scan the fundamental bridge frequencies: An experimental verification. Engineering Structures 27(13):1865-1878.

Locke, W., Sybrandt, J., Redmond, L., Safro, I., Atamturktur, S., (2020). Using drive-by health monitoring to detect bridge damage considering environmental and operational effects. Journal of Sound and Vibration 468:115088.

Malekjafarian, A., McGetrick, P.J., Obrien, E.J., (2015). A review of indirect bridge monitoring using passing vehicles. Shock and Vibration, article 286139.

Melo, F., (2013). Receiver Operating Characteristic (ROC) Curve. Encyclopedia of Systems Biology, pp. 1818-1823. Springer New York, New York, NY.

Robbins, H. and Monro, S. (1951). “A stochastic approximation method.” The Annals of Mathematical Statistics, 400-407.

Sarker, I.H., (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science 2(6):420.

Sarwar, M.Z., Cantero, D., (2023). Vehicle Assisted Bridge Damage Assessment Using Probabilistic Deep Learning. Measurement 206:112216.

Spencer Jr., B.F., Park, J.W., Mechitov, K., Jo, H., Agha, G., (2017). Next generation wireless smart sensors to sustainable civil infrastructure. Procedia Engineering 171:5-13.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). “Dropout: a simple way to prevent neural networks from overfitting.” The journal of machine learning research, 15(1), 1929-1958.

Wang, C., Yu, X., Liang, F., (2017). A review of bridge scour: mechanism, estimation, monitoring and countermeasures. Natural Hazards 87(3):1881-1906.

Wang, Y., Li, Y., Song, Y., and Rong, X. (2020). “The influence of the activation function in a convolution neural network model of facial expression recognition.” *Applied Sciences*, 10(5), 592 1897.

Yang, Y-B., C. W. Lin, and J. D. Yau. "Extracting bridge frequencies from the dynamic response of a passing vehicle." *Journal of Sound and Vibration* 272.3-5 (2004): 471-493.

Zhang, B., Zhao, H., Tan, C., OBrien, E. J., Fitzgerald, P. C., & Kim, C. W. (2022). Laboratory investigation on detecting bridge scour using the indirect measurement from a passing vehicle. *Remote Sensing*, *14*(13), 3106.

Zhang, T., Zhu, J., Xiong, Z., Zheng, K., & Wu, M. (2023). A new drive-by method for bridge damage inspection based on characteristic wavelet coefficient. *Buildings*, 13(2), 397.