Available online at www.eccomasproceedia.org
Eccomas Proceedia COMPDYN (2025) 1282-1291

ECCOMAS
Proceedia

COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-18 June 2025

# BRIDGING ANALYTICAL AND EXPERIMENTAL DOMAINS WITH EXPLAINABLE AI: GRAD-CAM++ FOR DAMAGE DETECTION IN LIGHTWEIGHT STRUCTURES UNDER MOVING MASSES

**Georgios I. Dadoulis[1], George D. Manolis[1] and Konstantinos Katakalos[1*]**

[1]Laboratory for Strength of Materials and Structures, Aristotle University of Thessaloniki,
Aristotle University Campus, 54124, Thessaloniki, Greece
{dadoulis, gdm, kkatakal}@civil.auth.gr

[*] Corresponding author

## Abstract

*Structural health monitoring (SHM) is crucial for the safety and reliability of engineering structures, especially for lightweight bridges subjected to moving mass loading. While convolutional neural networks (CNNs) have demonstrated strong performance in vibration-based damage detection, their limited interpretability remains a barrier to real-world deployment. Building on our previous work in classifying vibration response using CNNs, this study introduces Gradient-weighted-Class Activation Mapping++ (Grad-CAM++) to enhance CNN explainability. By applying Grad-CAM++ to spectrograms derived from experimental vibration data, we identify the structural regions that most influence the CNN's damage classification. This approach bridges the gap between black-box machine learning outputs and engineering insight, revealing frequency-dependent damage signatures embedded in the learned features. The proposed framework advances the transparency and trustworthiness of CNN-based SHM systems, offering a field-validated tool for operational diagnostics under dynamic conditions.*

**Keywords:** Structural Health Monitoring (SHM), Explainable AI, Grad-CAM++, Moving masses on bridges, Vibration-based damage detection.

ISSN:2623-3347 
Peer-review under responsibility of the organizing committee of COMPDYN 2025.
doi: 10.7712/120125.12488.27760

# 1 INTRODUCTION

In recent decades, structural health monitoring (SHM) has become an important issue regarding civil engineering infrastructure [1], aiming to ensure their stable operation and overall functionality. To this end, vibration testing is frequently employed for system identification and damage detection, a process classified under the umbrella of non-destructive testing evaluation of structures. For example, vibration testing is usually applied in the form of experimental modal analysis (typically, input-output modal identification) as well as operational modal analysis (typically, output-only modal identification), which are based on collecting acceleration response data under a controlled loading environment and under the realm of environmentally induced excitations, respectively [2]. Drawing from the field of SHM, in which vibration testing plays a central role, the common feature among vibration-based damage detection techniques is the extraction of damage-sensitive structural and/or mechanical features, which essentially serve as “damage indicators”. Decisions regarding the existence of damage are based therefore based on comparisons between damage indicators involving two structural states, one designated as “undamaged” and one as “current” [3].

Early laboratory-validated approaches using modal parameters (eigenfrequencies, mode shapes) as damage indicators were promising, but transferring them to real-life SHM has proven challenging because of the low sensitivity of dominant modal parameters to the onset of structural damage [4]. Furthermore, there is the implicit requirement that both external excitations and the structural parameters are relatively stationary, a requirement that rarely holds true. Finally, tracking eigenfrequencies and mode shapes may be particularly difficult in structures exhibiting complex dynamic behavior, in which modal parameters can hardly be inferred by engineering judgment [5]. One such case is lightweight bridges subjected to vehicular traffic. Specifically, if vehicular masses are comparable to the structural mass of a lightweight bridges, they behave as “traveling masses”, resulting in a strong coupling effect [6].

To address this problem, the damage detection approach presented here builds upon feature extraction for identifying structural damage during laboratory testing. Machine learning (ML) is introduced that uses the acceleration response of a bridge span under a moving mass as the extracted feature, followed by its classification into predefined damage scenarios. Specifically, a convolutional neural network (CNN) is created and trained with acceleration time histories obtained from numerical simulations of the dynamic response of a steel beam (representing a lightweight bridge deck) under different damage scenarios. Once trained, this CNN is fed with acceleration response data recorded during laboratory testing of the actual steel beam subjected to traveling masses. The damage scenarios examined focused on support settlement and on the presence of cracks across the web and flange of the beam [7], with the reference case being the intact beam [8,9].

To address the interpretability challenges inherent in CNN-based damage detection, we implement the Grad-CAM++ algorithm [10], an enhanced version of Gradient-weighted Class Activation Mapping (Grad-CAM) [11]. Unlike conventional approaches that focus solely on prediction accuracy, Grad-CAM++ generates detailed heatmaps by analyzing activation gradients in the final convolutional layer. These heatmaps visually identify the specific regions of input spectrograms that most significantly contribute to the model's classification decisions. Crucially, this approach enables us to: (1) verify whether the CNN focuses on physically meaningful features (e.g., eigenfrequency variations caused by structural damage), (2) detect potential overreliance on artifacts or noise, and (3) validate the model's alignment with engineering principles. By making the CNN's decision-making process transparent, Grad-CAM++ advances explainable AI (XAI) in structural health monitoring, fostering greater trust in deep learning applications for critical infrastructure assessment.

## 2 EXPERIMENTAL SETUP AND DATA GENERATION

Building upon our previous experimental work in structural health monitoring [8,9], we employ data from tests conducted on a simply supported HEB100 steel beam with a 5.83 m clear span and 1.0 m guide segments at each end (Figure 1). The experimental setup simulates bridge dynamics under traffic loading through a heavy mass sliding along the beam at controlled, adjustable speeds. This configuration enables the study of forced vibration during mass traversal and free vibration after mass departure, providing comprehensive data for damage detection under realistic dynamic conditions.

Figure 1: Experimental setup for moving mass tests.

More specifically, the aim of this research work is to identify damage in the beam based on acceleration recordings at various locations, as the mass slides along the bridge span (forced vibration regime), as well as when the mass has left the beam (free vibration regime). Two basic types of damage are considered, namely (a) support settlement and (b) cracking in the span. Given the fact that three wireless acceleration sensors were used [12], four convolution neural networks (CNNs) are developed, the first three trained by data from each individual sensor and the fourth by the combined data furnished by all three sensors. Thus, the first three CNNs were of the single-entry type receiving information from a single sensor, while the fourth one was of the triple-entry type receiving information from all sensors. More specifically, training of the CNNs was accomplished using numerically generated spectrograms from a highly accurate analytical model [8,9] that was capable of accounting for both the moving mass dynamics and the damage mechanics. Finally, validation of the CNNs was successfully carried out based on the processing of recorded acceleration time histories recovered from the aforementioned experiments.

The spectrograms developed here are based on the Fourier transform of the product of the acceleration time history times a fixed-width, time window that translates along the time axis [13]. This technique is useful in tracing changes over time of key dynamic system parameters. In our case, the heavy moving mass creates a non-autonomous dynamic system [6], where the eigenproblem of the original Bernoulli-Euler beam becomes time-dependent. As shown in Figure 2, the first and second eigenfrequencies of the intact beam exhibit variability with time during the forced vibration regime caused by the sliding mass. Furthermore, if the beam is damaged, this time dependence of the eigenfrequencies is different compared to that observed for the intact beam, see Figure 3 where the results of both numerical simulations and experimental measurements are shown and contrasted. Although these changes may not be quite detectable to the eye, they are certainly detectable by a trained CNN.

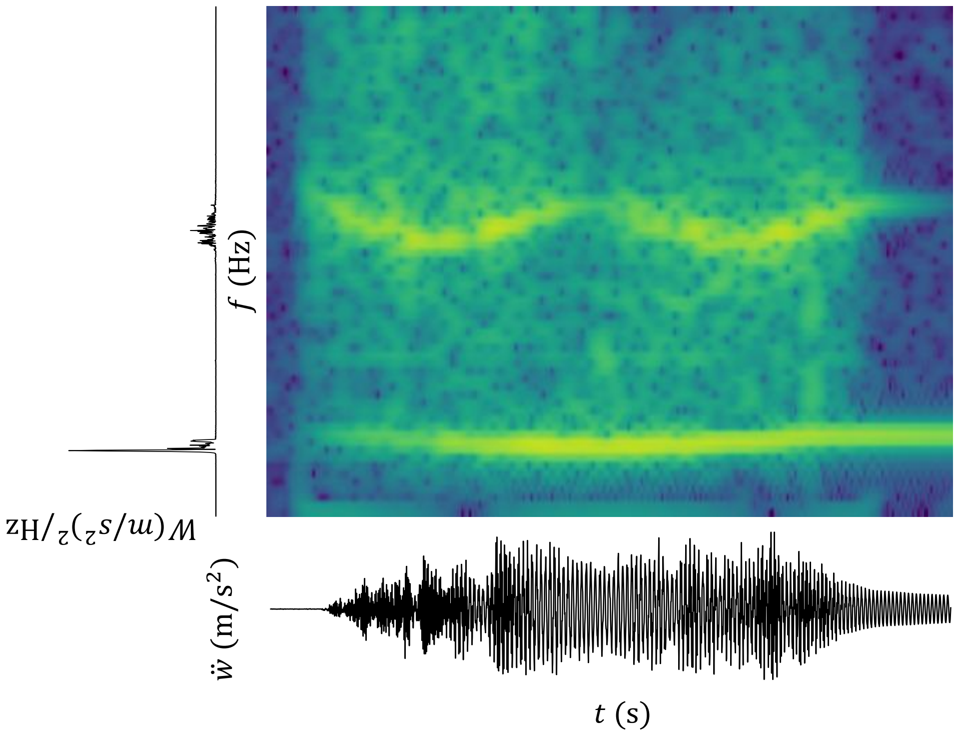



Figure 2: Spectrogram at beam location $x=L/4$ derived from the analytical solution. Axis *x-x* depicts the acceleration time history and the *y-y* axis plots the Power Spectra Density (PSD) function for a $m$=24 kg mass rolling at a speed of $v$=42 cm/s.

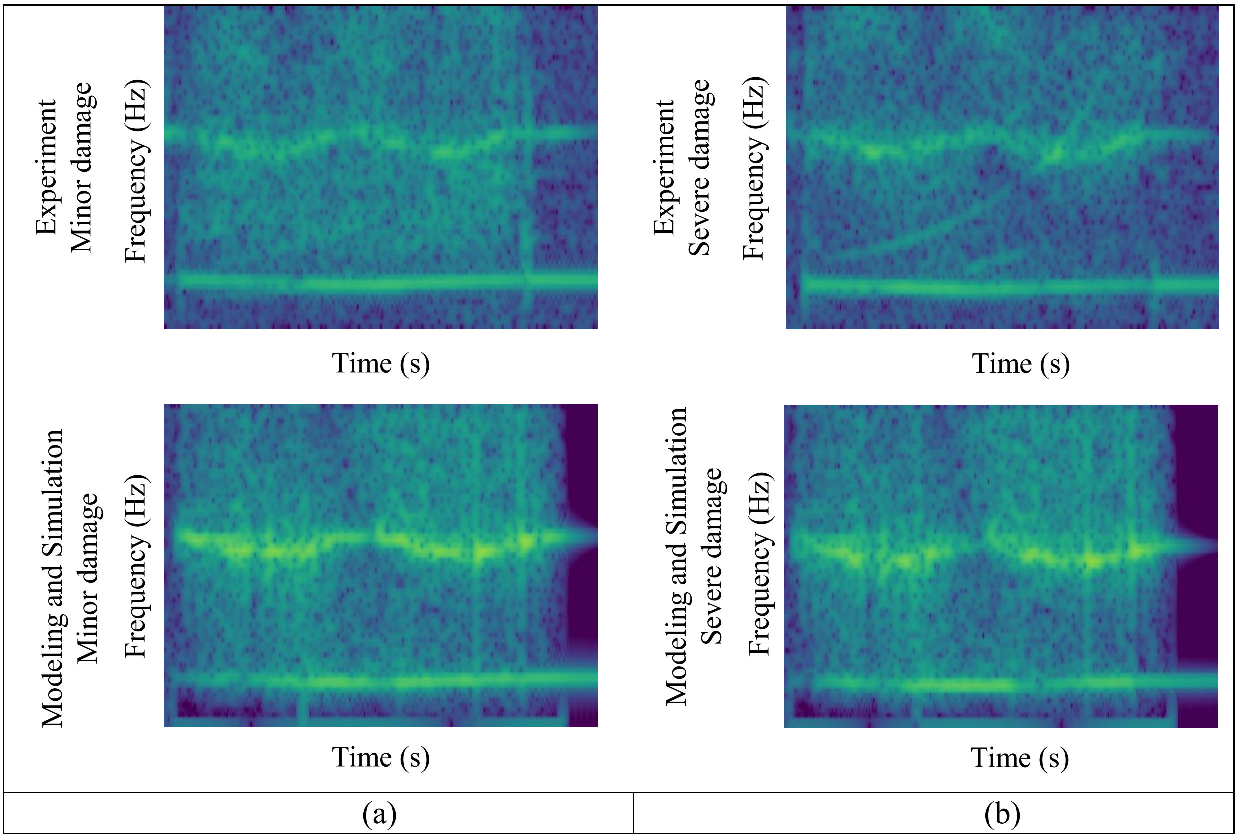



Figure 3: Spectograms for a crack at center span corresponding to minor damage (left column) and to severe damage (right column): (a) Experimental measurements and (b) numerical simulations. In all cases, the mass is $m$=13 kg moving with at a speed *of* $v$=25 cm/s.

## 3 DAMAGE SCENARIOS: CLASSIFICATION AND QUANTIFICATION

Damage in a bridge span can be attributed to a number of factors besides normal operating conditions, such as environmentally induced temperature fluctuations, ground shaking, humidity, etc. [3]. In here, we will consider CNN development for a simply-supported beam with loading in the form of a moving mass so that six damage states (or classes) can be identified. These are labelled as follows:

(1) L0 for the intact beam.
(2) L1 for a crack in the first one-third part of the span.
(3) L2 for a crack at center span.
(4) L3 for a crack in the last one-third part of the span.
(5) L4 for left support settlement.
(6) L5 for right support settlement.

From a numerical viewpoint, damage is represented by springs at the specific location where it is expected and/or observed. We distinguish two basic cases, namely (a) support settlement where support fixity is replaced by a translational spring $k_T$ and (b) vertical cracking of the span where the cross-section continuity is destroyed and replaced by a rotational spring $k_R$ [7]. This latter case is illustrated in Figure 4, where a vertical crack was artificially sliced from the bottom web and moving upwards at center span. Depending on the extent of the crack depth $d$, two sub-states can be identified, namely a shallow crack corresponding to minor damage and a deep crack corresponding to severe damage, as can be seen in Table 1.

(a) (b)

Figure 4: Scenario L2 for artificially induced vertical crack damage at the center of the HEB100 steel beam. (a) minor damage, (b) severe damage.

| Damage scenario L2 | Crack depth $d$ (mm) | Eigenfrequency $f_1$ (Hz) | Eigenfrequency $f_2$ (Hz) | Equivalent rotational spring $k_R$ (kN-m/rad) |
|---|---|---|---|---|
| Minor damage | 16.5 | 9.4 | 39.1 | 3558.0 |
| Severe damage | 35.0 | 8.1 | 39.0 | 655.0 |

Table 1: Quantification of damage demonstrated by the beam's first two eigenfrequencies for a center-span crack.

## 4 CNN ARCHITECTURE AND PERFORMANCE

Since damage in a bridge span can be attributed to both normal operating conditions and environmentally induced loads, it is obvious that CNN development is particular to the type of damage and/or performance degradation that is of interest to the end user. In our case, we have bridges supporting vehicular traffic, a topic that is the responsibility of state and/or federal transportation authorities. To this end, a typical architecture behind a single-entry CNN, as developed from spectrograms corresponding to acceleration time histories at fixed locations on the beam's span, is shown in Figure 5. The three-input CNN follows a nearly identical architecture.

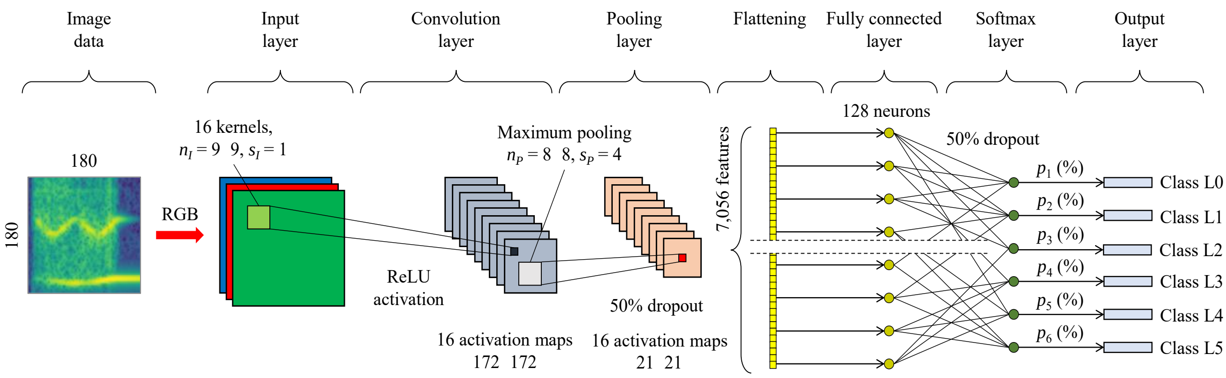



Figure 5. CNN architecture for a single-entry setup.

Of all four CNN that were developed and validated against numerical data, the three-entry CNN gave the best results for classification of all damage scenarios that was based on experimental measurements, see the 'confusion' matrices of Figure 6. The totality of correct predictions can be quantified using the ratio of correct predictions to the total number of requested predictions, for all six damage scenarios. This ratio is around 95% for scenarios that exhibited mild damage to around 97% for severe damage. By way of contrast, the single-entry CNN that was trained from data generated at sensor location $x=3L/4$ (a location that avoids the nodal point of the first eigenmode that is at the center), provided less accurate results, see Figure 7. In more detail, the best prediction accuracy was obtained for damage scenario L2, where out of 30 mild damage cases, 27 were correctly identified, while for severe damage, correct identification reached 29 cases. Note that mild versus severe damage for scenario L2 of Figure 4 was quantified in Table 1.

(a)

| | | Actual scenario | | | | | |
|---|---|---|---|---|---|---|---|
| | | **L0** | **L1** | **L2** | **L3** | **L4** | **L5** |
| Classification | **L0** | 30 | | | | | |
| | **L1** | 2 | 27 | 1 | | | |
| | **L2** | 2 | | 28 | | | |
| | **L3** | 3 | | | 25 | 2 | |
| | **L4** | | | | | 30 | |
| | **L5** | | | | | | 30 |

(b)

| | | Actual scenario | | | | | |
|---|---|---|---|---|---|---|---|
| | | **L0** | **L1** | **L2** | **L3** | **L4** | **L5** |
| Classification | **L0** | 30 | | | | | |
| | **L1** | | 25 | | 4 | 1 | |
| | **L2** | | | 30 | | | |
| | **L3** | | | | 29 | 1 | |
| | **L4** | | | | | 30 | |
| | **L5** | | | | | | 30 |

Figure 6: Confusion matrices for the triple-entry CNN architecture: (a) mild damage, (b) severe damage.

(a)

| | | Actual scenario | | | | | |
|---|---|---|---|---|---|---|---|
| | | **L0** | **L1** | **L2** | **L3** | **L4** | **L5** |
| Classification | **L0** | 30 | | | | | |
| | **L1** | | 14 | 2 | 14 | | |
| | **L2** | 3 | | 27 | | | |
| | **L3** | | 21 | 1 | 6 | 2 | |
| | **L4** | | | | | 29 | 1 |
| | **L5** | | | | | | 30 |

(b)

| | | Actual scenario | | | | | |
|---|---|---|---|---|---|---|---|
| | | **L0** | **L1** | **L2** | **L3** | **L4** | **L5** |
| Classification | **L0** | 30 | | | | | |
| | **L1** | | 16 | 1 | 13 | | |
| | **L2** | | 1 | 29 | | | |
| | **L3** | | 11 | 2 | 13 | 4 | |
| | **L4** | | | | | 29 | 1 |
| | **L5** | | | | | | 30 |

Figure 7: Confusion matrices for the single-entry CNN architecture based on sensor location at $x=3L/4$: (a) mild damage, (b) severe damage.

## 5 EXPLAINABLE AI WITH GRAD-CAM++

In developing CNNs, the question that arises is which areas in a spectrogram are responsible for guiding the network to reach a correct decision regarding damage classification. Furthermore, for a specific type of damage, how do these areas change as damage progresses from mild to severe. Thus, to help elucidate these questions, we introduce in our CNN development the Grad-CAM++ algorithm [10], which is an improved version of the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [11]. This improvement helps visualize and interpret the decisions formulated by the deep learning AI systems and more specifically by the CNNs, thus exposing the regions in a signal that are most important in the decision taxonomy. Applications for such improvements in formulating CNN systems include medical imaging, damage identification from robots and drones, and finally the development of Explainable Artificial Intelligence (XAI) capable of explaining to the user what parts of the entry data provided are important in the decision-making process.

More specifically, the Grad-CAM [11] technique is based on the computation of the derivatives at the exit point, for a specific class in terms of the characteristics of a convolution layer. Following that, it uses the mean value of the gradients (global average pooling) to quantify the importance of each feature map, thus resulting in a heatmap that illustrates the regions of interest. A further improvement is achieved by the Grad-CAM++ [10] algorithm that focuses on certain weaknesses of the Grad-CAM method, such as the inability to accurately identify minor or multiple objects at the entry point. Thus, the improvement achieved by the Grad-CAM++ algorithm comes from a further computation of the first, second and third derivatives at the exit point with respect to the network characteristics, while at the same time accounting for the importance of a feature map, plus the relative contribution of each of its pixels. We note that the tf.GradientTape [14] software was used for the computation of all these derivatives.

To give some details, let $y^c$ be the exit point of the network for class c , and let $A^k$ be the *k-th* feature map of the last convolution layer. The elements of $A^k$ are further identified as $A^k_{ij}$, where $(i, j)$ is a specific location on the characteristics table. The aim is to produce a 'heat map' $L^c$ that will identify the most important regions for class $c$, and is defined for each point $(i, j)$ as follows:

$$L^c_{ij} = \sum_k w^c_k A^k_{ij} \tag{1}$$

In the above, $w_k^c$ are the weights for each *k-th* feature map. In the Grad-CAM++ algorithm, these $w_k^c$ weights are not simply the mean values of the gradients, but are computed from a more complex expression involving their second and third derivatives as follows:

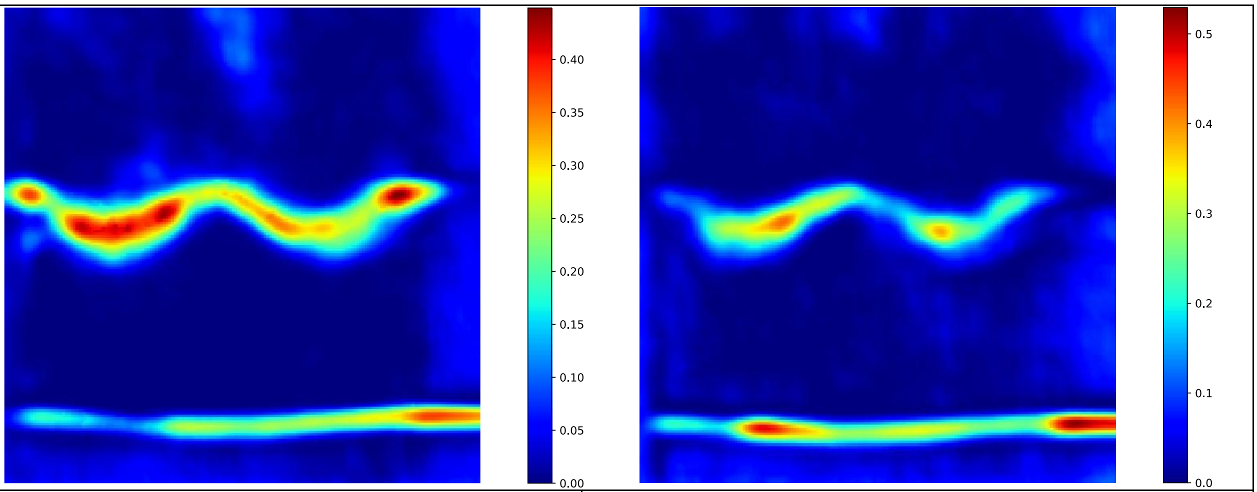
$$w_k^c = \sum_i \sum_j a_{ij}^{kc} \cdot relu\left(\frac{\partial y^c}{\partial A_{ij}^k}\right) \quad (2)$$

Finally, coefficients $a_{ij}^{kc}$ are given below as

$$a_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{\left(\partial A_{ij}^k\right)^2}}{2 \cdot \frac{\partial^2 y^c}{\left(\partial A_{ij}^k\right)^2} + \sum_{a,b} A_{ab}^k \cdot \frac{\partial^3 y^c}{\left(\partial A_{ab}^k\right)^3}} \quad (3)$$

## 6 DECISION VISUALIZATION

The single-input convolutional neural network (CNN) architecture (Figure 5) consists of $k$=16 feature maps (size 172×172) generated by the last convolutional layer. The Grad-CAM++ algorithm used 27 mild-damage and 29 severe-damage spectrograms—derived from acceleration time histories recorded at station $x=3L/4$ as input. The network correctly classified all spectrograms as belonging to the L2 class (center-span crack damage). For each damage subgroup, 27 and 29 heatmaps were generated, respectively, followed by mean-value computation per category. Figure 8 displays the resulting mean heatmaps for both mild and severe damage subclasses.




(a) (b)

Figure 8: Mean heatmaps for (a) minor damage and (b) severe damage, case L2.

It is clear that the single entry CNN focuses around in the first two eigenfrequency zones of the dynamic bridge-moving mass system. The areas with low activation (shown in deep blue color) do not contribute to the process of categorization and are practically inactive. On the other hand, the highly activated areas are the ones that correspond to the zones of variability in the first two eigenfrequencies of the dynamic system. Now looking at the case where damage is considered to be mild, most activity is observed around the zone of the second eigenfrequency, a fact that implies that the information furnished by this eigenproperty is sufficient for a correct damage classification. On the other hand, when damage is considered to be severe, this zone extends to encompass the first eigenfrequency. This fact implies that as damage increases, the CNN utilizes a wider spectrum of eigenmode information. These findings are consistent with the physics of the problem, meaning that at high levels of damage, the lower eigenfrequencies begin to be affected. On the other hand, for mild damage one has to look at the higher eigenfrequencies to be able to see their effect on the dynamic system.

## 7 SUMMARY AND DISCUSSION

Damage detection using vibration testing typically relies on damage-sensitive features, which serve as damage indicators. Decisions upon the existence of damage are based on comparing these damage indicators, as retrieved from at least two distinct structural states. However, the relatively low sensitivity of damage indicators to the onset of structural damage remains an open question, despite the considerable research efforts over the years. Low-sensitivity problems may be particularly exacerbated by the complex dynamic behavior of lightweight structures, such as metallic bridges subjected to vehicular traffic. In particular, vehicles traversing lightweight bridges essentially act as traveling masses comparable to the structural mass and result in a complex, coupled dynamic system that may obscure typical damage indicators recovered from vibration testing.

This work presents a machine learning (ML)-driven approach for damage detection in such systems, leveraging a convolutional neural network (CNN) to classify acceleration responses from vibration testing. Training data were generated from numerically simulated damage scenarios based on calibrated analytical models. The Grad-CAM++ algorithm was implemented to interpret the CNN's decision-making process during analysis of experimental vibration data. Results indicated the network's reliance on variations in the first two eigenfrequency bands for damage classification. Within the same damage category (L2 in this case), these frequency regions contained distinct sub-zones demonstrating differential sensitivity patterns. Specifically, mild damage cases predominantly activated higher eigenfrequency components, while severe damage cases exhibited significant activation across both higher and lower frequency ranges. This systematic behavior provides compelling evidence that the CNN's decision framework captures physically significant damage characteristics rather than superficial statistical correlations, validating its effectiveness for structural health monitoring applications.

## ACKNOWLEDGMENTS

The authors acknowledge support offered by the Hellenic Foundation for Research and Innovation (HFRI) under grant 6522. Results presented in this work have been produced using the Aristotle University of Thessaloniki High-Performance Computing Infrastructure. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the HFRI.

## REFERENCES

[1] Farrar, K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A*, **365**, 303–315, 2007.

[2] A. Cunha, E. Caetano E. Experimental modal analysis of civil engineering structures. *Sound and Vibration* **6(40)**, 12–20, 2006.

[3] P.C. Chang, A. Flatau, S.C. Liu. Health monitoring of civil infrastructure. *Journal of Structural Health Monitoring* 2003, 2(3): 257–267.

[4] M.F. Green, D. Cebon. Dynamic response of highway bridges to heavy vehicle loads: Theory and experimental validation, *Journal of Sound and Vibration* 1994, 70(1): 51-78.

[5] R. Brincker, C. Ventura. *Introduction to Operational Modal Analysis*. Wiley, Chichester, UK, 2015.

[6] C.I. Bajer, B. Dyniewicz. *Numerical Analysis of Vibrations in Structures under Moving Loads*, Springer, Heidelberg, 2012.

[7] A.D. Dimarogonas. Vibration of cracked structures: A state-of-the-art review, *Engineering Fracture Mechanics*, **55(5)**, 831-857. 1996.

[8] G.I. Dadoulis, G.D. Manolis. Dynamic response of a damaged bridge model traversed by a heavy point mass. *Journal of Sound and Vibration*, **551**, 117613, 2023. *https://doi.org/10.1016/j.jsv.2023.117613*

[9] G.I. Dadoulis, G.D. Manolis, K. Katakalos, K. Dragos, K. Smarsly (2025). Damage detection in lightweight bridges with traveling masses using machine learning. Engineering Structures, **322(B)**, 119216, 2025. *https://doi.org/10.1016/j.engstruct.2024.119216*

[10] A. Chattopadhyay, A. Sarkar, P. Howlader, V.N. Balasubramanian. Grad-CAM++: Improved visual explanations for deep convolutional networks. Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV 2018), pp. 839-847, 12-15 March 2018, Lake Tahoe, Nevada, USA. *https://doi.org/10.1109/WACV.2018.00097*

[11] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra. Grad-CAM: visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, **128(2)**, 336–359. 2016. https://doi.org/10.1007/s11263-019-01228-7

[12] Parker Hannifin. *Microstrain Sensing Product Datasheet: G-Link-200 Ruggedized Wireless Triaxial Accelerometer Node*. Williston, Vermont, USA, 2020.

[13] A. Brandt. *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures*. Wiley, Chichester, UK, 2011

[14] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. https://www.tensorflow.org/