



Assessment of Damage in a Reinforced Beam by K-Nearest Neighbor Algorithm

Amanda Aryda Silva Rodrigues de Sousa¹, Jefferson da Silva Coelho², Marcela Rodrigues Machado² and Maciej Dutkiewicz³

¹ Post-Graduate Program - Integrity of Engineering Materials, University of Brasília, Campus Universitário Darcy Ribeiro, Asa Norte, Brasília, 70910-900, DF, Brazil

² Department of Mechanical Engineering, University of Brasília, Campus Universitário Darcy Ribeiro, Asa Norte, Brasília, 70910-900, DF, Brazil

³ Faculty of Civil, Environmental Engineering and Architecture, University of Science and Technology, Bydgoszcz, 85-796, Cujávia-Pomerânia, Poland

Abstract: Structural damage is considered any change in a structure's local flexibility or mass that creates undesirable displacements and vibrations. Mass loss in the structure can increase sensitivity to local damage and potentiate the effects of discontinuities in the dynamic response, which can be used in damage identification and location. Therefore, this work aims to investigate the application of k-nearest neighbor (k-NN) machine learning for damage detection in a cantilevered beam with reinforcement. Under undamaged and damaged conditions, a damage index (DI) is assumed to build the dataset from the frequency response function. Structure reinforcement mass loss directly influences the system's vibration; hence, the DI is applied to detect damage and quantify its severity. Numerical results demonstrate that the Euclidean, Manhattan, and Braycurtis metrics are considered robust for both datasets with and without noise, and k values lower than 75, are considered fast and accurate in the detection problem and estimation of damage in beam structures.

Keywords: damage identification, machine learning, k-nearest neighbor, beam with additional auxiliary mass

INTRODUCTION

Structural systems can progressively deteriorate and accumulate damage over their lifetime due to overload, fatigue, and extreme weather events. Early damage identification and periodic structural integrity assessment are necessary for the system to operate correctly and for damage to be identified, monitored, and corrected. Damage can be considered any change in the local flexibility of a structure that creates undesirable displacements and vibrations (Léornad *et al.*, 2001). Structural Health Monitoring (SHM) provides practical means to assess and predict structural performance under operational conditions (Pathirage *et al.*, 2018). The monitoring procedure involves observing a structure or mechanical system over time and acquiring measured data. Damage identification methods using structural vibration characteristics are widely used because changes in structural physical parameters, such as stiffness and mass, also change the structural vibration characteristics, that is, the modal parameters. Therefore, some structural monitoring techniques are used to add an auxiliary transverse mass change in the structure to increase the effects of discontinuities in the dynamic response and facilitate the identification and location of the damage. (Zhang *et al.*, 2013) proposed uses the frequency shift curve with an auxiliary mass in a structure for damage detection. (Wang *et al.*, 2016) used a Frequency Shift (FRESH) method to explain the dynamic behaviour of merged structures with auxiliary mass. The method is considered rational and computationally operational in identifying the loss.

One of the challenges associated with damage identification vibration-based methods is modelling errors, noise in the measured vibration data, and environmental effects. Artificial intelligence techniques based on machine learning were trained to generate the most probable result and validate the model based on unseen datasets (Yuan *et al.*, 2020). Machine learning has been successfully applied in SHM systems. The concept of ML enters this paradigm of feature selection and statistical modelling for feature discrimination described in (Machado *et al.*, 2017; Farrar and Worden, 2013). According to (Sun *et al.*, 2022), the most used supervised learning algorithms in structural design construction and structure performance evaluation are Linear Regression, Kernel Regression, Tree-Based Algorithms, Logistic Regression, Support Vector Machine (SVM), K-Nearest-Neighbor (k-NN), and Neural Networks. (Vitola *et al.*, 2017), and (Tibaduiza *et al.*, 2018) used different types of kNN algorithms, applying a piezoelectric sensor network to obtain the database. In Vitola's work, the authors inspect and evaluate the damage on a rectangular aluminium profile, an aluminium plate, and a composite plate. The results show that fine kNN and weighted kNN better performed among the algorithm studied. In Tibaduza's paper, the authors analyzed a sandwich structure composed of carbon fiber reinforced polymer (CFRP) and a composite plate of CFRP, obtaining similar outcomes as Vitola according to the k-NN. (Tsuchimoto and Kitagawa, 2004) proposed a structural damage detection system that uses natural frequencies in a five-story structure, modelled with a five-mass shear system, damage locations were first detected globally using the artificial neural network (ANN) method. The failure was identified locally by determining changes in structure eccentricity between centres of stiffness and weight

due to damage. The ANN showed good accuracy in detecting damage.

This work investigates the application of k-nearest neighbor (k-NN) machine learning for damage detection in a reinforced cantilevered beam in the context of SHM. Damage is assumed as a mass loss on the reinforcement, and the DI is estimated from the system's frequency response function (FRF) under undamaged and damaged conditions, with and without noise. Therefore, the DI is applied as input to the ML algorithm, and the monitoring is performed via the k-NN algorithm. Results show the efficiency of the ML algorithm and compare the k-NN algorithm metrics.

DAMAGE DETECTION USING K-NEAREST-NEIGHBOR

Damage detection methods have been applied to locate and quantify structural damage through changes using the dynamic characteristics change. Therefore, those changes in dynamic characteristics can be used as indicators of damage when compared to the reference signal. Hence, the damage index based on the beams' FRF is used for damage detection. The system is a reinforced cantilever beam modelled by the spectral element method with additional auxiliary mass. The beam is excited with a unitary force applied on the free edge, and the response is obtained at the same point. The beam has an $L = 0.38m$ length, a width of 0.025 m, and a height of 0.004 m. Material properties are Young's modulus of 2.1 GPa and mass density of 7800 kg/m^3 . The auxiliary mass position is at $L1 = 0.05L$. The reinforced beam shown in figure 1 is considered a healthy state of the reinforced beam. The damage is when the reinforcement loses a mass of 2.96, 5.92 e 8.87% of the total weight of the auxiliary mass. The DI is calculated by comparing a reference signal, usually derived from the system considered undamaged or with a healthy signature, to the one provided by the system under the presence of discontinuing or damage Barreto *et al.* (2021). Therefore, the DI is presented in values between zero to unity, where the unit accuses no damage. A lower value up to zero indicates the presence of damage and its severity within the analysis scenario. This work uses the DI as structure information for the training and testing data in the multiclass ML algorithms.

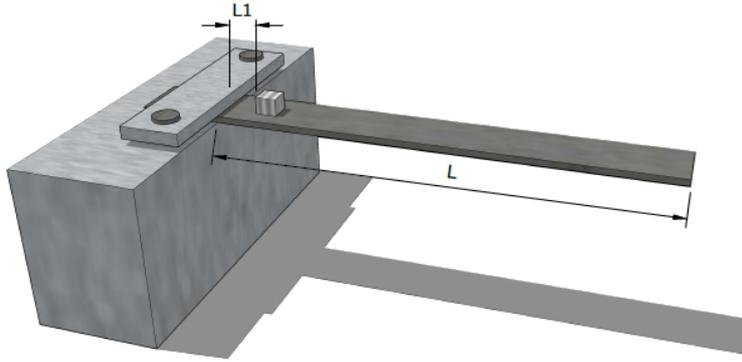


Figure 1: Reinforced cantilever beam.

Dataset based in damage index

The Frequency Response Assurance Criterion (FRAC) Heylen and Lammens (1996) is used in this work. FRAC is a damage index that correlates FRF signals, where a strong correlation is indicated by a unity representing no damage state. In contrast, the lowest correlation to zero means damage condition and severity. Equation (1) formulates the FRAC that compares the FRF signal of the cracked beam (H_{ij}^d) and for healthy beam indicated by (H_{ij}^u). Because a crack directly influences the system vibration, the DI can detect and quantify the damage.

$$FRAC_{ij}(\omega) = \frac{\|H_{ij}^d(\omega)(H_{ij}^u(\omega))^*\|^2}{[H_{ij}^u(\omega)(H_{ij}^u(\omega))^*][H_{ij}^d(\omega)(H_{ij}^d(\omega))^*]} \quad (1)$$

where * defines the complex conjugate operator. The excitation is applied at the j th coordinate, and the response function at the i th coordinate. The index compares the FRFs of the damaged and undamaged beam responses, thus the entire spectrum energy response information. The FRAC DIs were calculated using the beam's FRF under undamaged and damaged conditions. In this simulation, 3% e 5% of white noise was incorporated into the FRFs to investigate the robustness of the ML algorithm in damage detection. The dataset reached 100 samples for each class. Henceforth, following the DI values, the multiclass dataset was labelled as health, 2,96%, 5,92%, and 8,87% Damage.

K-Nearest-Neighbor Classifier

The machine learning process starts with dataset extraction, followed by splitting the data for training and testing, 75% and 25%, respectively, and applying the ML algorithm to classify and provide information on damage identification and quantification. All damage detection ML implementation uses scikit-learn Machine Learning in Python. In this work, the algorithm used will be the K-nearest neighbor(k-NN).

K-NN is one of the simplest supervised learner methods Cutler and Dickenson (2020); Malekloo *et al.* (2021) and is widely used for pattern recognition Kurian and Liyanapathirana (2020). k-NN can be used for classification and regression, where data with discrete labels usually uses classification and data with continuous labels regression. The k-NN algorithm, in its simplest version, only considers exactly one nearest neighbor, which is the closest training data point to the point we want to predict. The prediction is then simply the known output for this training point. Depending on the value of k, each sample is compared to find similarity or closeness with k surrounding samples. For example, when $k = 3$, the individual samples undergo comparison with the nearest three samples, and hence the unknown sample is classified accordingly Cutler and Dickenson (2020), according to the figure 2. The optimal choice of the value of k is highly data-dependent, in general, a larger suppresses the effects of noise but makes the classification boundaries less distinct.

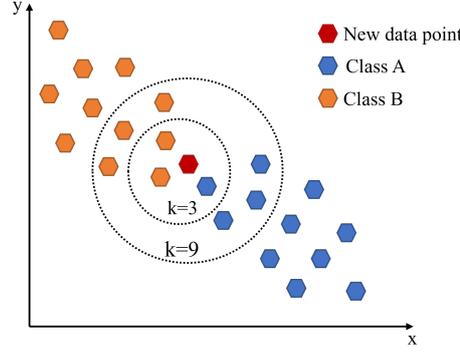


Figure 2: K-Nearest-Neighbor Classifier

. The most common metrics used in the literature are **Euclidean**, the distance between two vectors, **Braycurtis**, the distance between two 1-D arrays, **Manhattan**, the distance between two points is the sum of the absolute differences of their cartesian coordinates. **Cosine**, the distance metric is used mainly to calculate the similarity between two vectors. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in the same direction as

$$\text{Euclidean} - d(x_i, y_i) = \sqrt{\sum_{l=1}^d (x_l^i - y_l^i)^2} \quad (2)$$

$$\text{Braycurtis} - d(x_i, y_i) = \frac{\sum |x_i - y_i|}{\sum |x_i + y_i|} \quad (3)$$

$$\text{Manhattan} - d(x_i, y_i) = \sum_{i=1}^n |x_i - y_i| \quad (4)$$

$$\text{Cosine} - \cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (5)$$

where x_i and y_i are objects represented by vectors in \mathfrak{R}^d space, and x_l^i and y_l^i are elements of the vectors, which correspond to the values of the coordinate l(attributes). After identifying the k closest data classes, the largest number of classes is determined as the new data class. The class to be determined depends on the value of k, and the accuracy of the classification can change according to the value of k. In this work, 11 values of k were considered, with $k = 5, 15, 25, 35, 45, 55, 65, 75, 85, 95,$ and 105 . To analyze whether the K-NN algorithm can detect structure damage, 4 distance metrics, such as Euclidean, Manhattan, Braycurtis, and Cosine, were compared to find the best distance metric of the k-NN algorithm. For general evaluation and comparison of the damage detection capacity of the k-NN algorithm, precision and confusion matrix metrics were used. Accuracy close to 100% is considered a good performance.

DAMAGE QUANTIFICATION

Damage assessment using FRAC DI considered the undamaged and damaged beam with the severity of 2.96, 5.92, and 8.87%, computing four classes for the damage identification. In the figure, 3, the accuracy of the Euclidean, Manhattan, and Braycurtis metrics achieved good damage detection and quantification, with 100% when the data are affected by 3% and 5% of white noise, and k value between 5 to 65. However, when the value of K reaches 75, the model's accuracy decreases, reaching 66, 90, 86%, for the case considering noise-free, 3% noise, and 5% noise, respectively. The higher the value of k, the lower the accuracy. Therefore, in this case, it is recommended to use values of k less than 75. In

figure 3 (d), it shows the accuracy values for the cosine metric, which was considered to have the lowest accuracy of the 4 addressed metrics, reaching accuracy values between 14 and 38, during the analysis, both for data with and without noise.

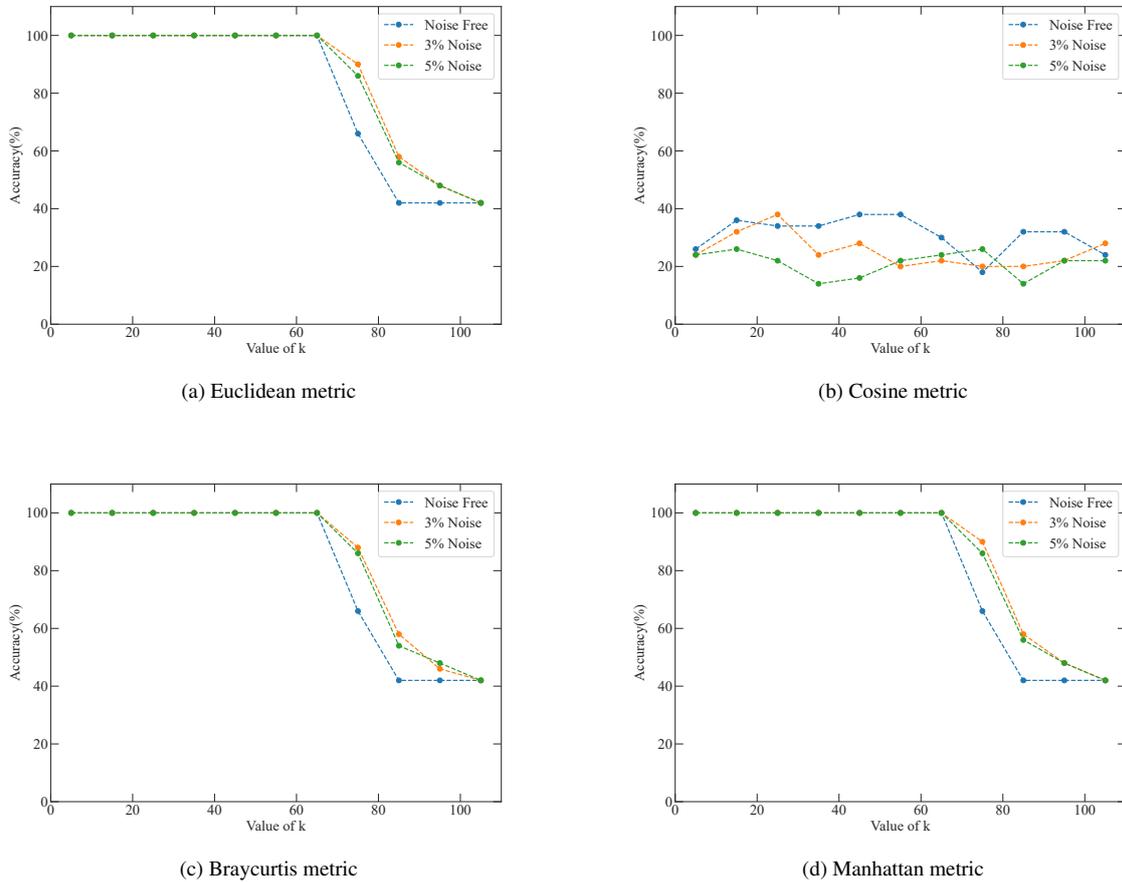


Figure 3: Accuracy of the k values studied

The algorithm's general performance is evaluated by accuracy, which is the proportion of health and damage conditions that were correctly classified. Therefore, only the accuracy does not provide enough information to diagnose possible errors made by the algorithms. Hence, the confusion matrix is also used to track the dataset classification. The confusion matrix has rows and columns representing the class prediction of the damage severity. This matrix allows us an understanding of the damage detection patterns and errors of the ML techniques classification. The matrix elements indicate the conditions of the data as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The diagonal of the confusion matrix represents the correct damage detection rate. Thus, the ideal model has high values on the diagonal and minimum values elsewhere.

Figure 4 shows the confusion matrices containing values and percentages predicted by ML techniques. Where Figs 4(a) are estimated with Euclidean, Manhattan and Braycurtis metrics, for values of k less than 75, Figs 4(b-c) metric for values of k= 75 and 105, Figs 4(d-f) cosine metrics for values of k=5, 75 and 105. As the accuracy reached 100%, the confusion matrix correctly classified the classes. When using k=75, the accuracy of the metrics reached 66%, due to misclassification in the sample for the 8.87 damage condition, with four classes assumed to be 5.92 damage and one sample classified as healthy. In the cosine metric, all samples were misclassified, as seen in the confusion matrix for three different k. Therefore, only Euclidean, Manhattan and Bray Curtis metrics are considered robust for noise and noiseless datasets, with k values less than 75.

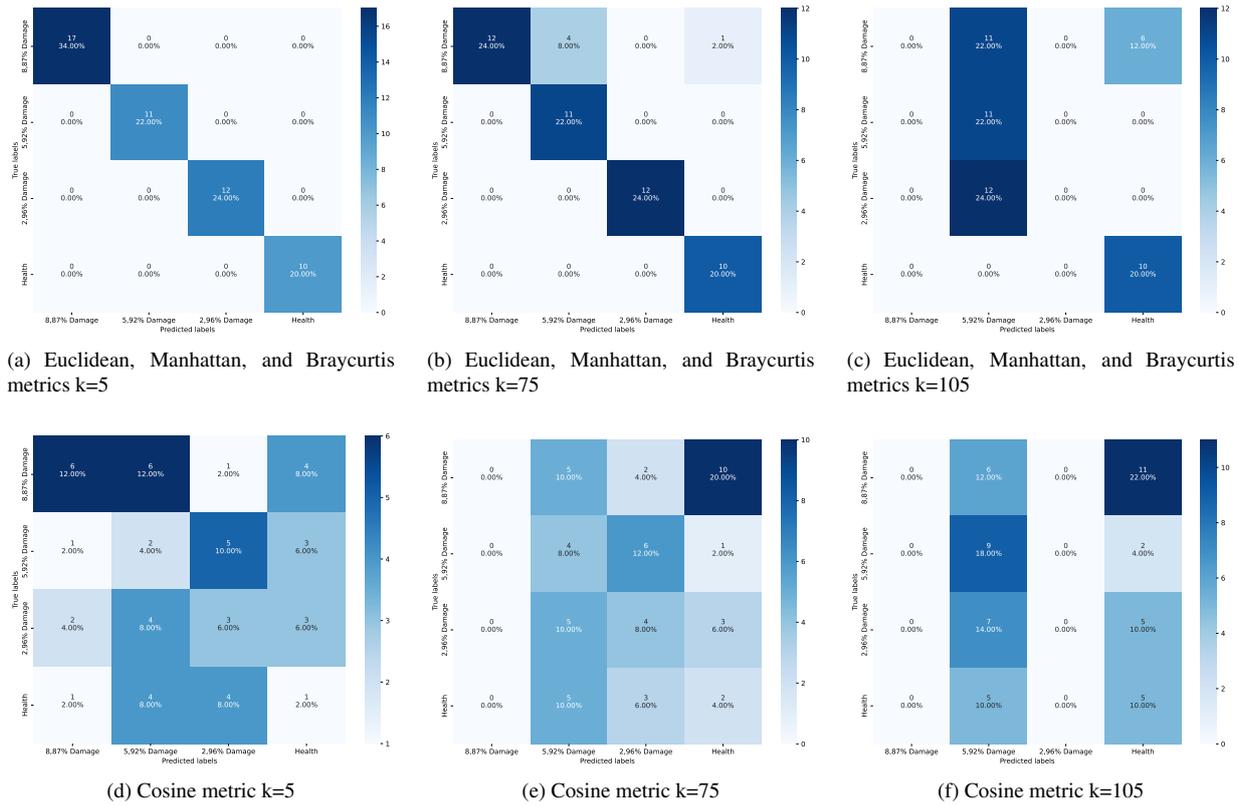


Figure 4: Confusion matrix of the multiclass classification damage classification from FRAC DI with and without noise

CONCLUSION

This paper investigated the k-NN technique for different metrics and k values to detect damage in a reinforced cantilever beam by adding a mass using FRF. The vibration method was calculated using the spectral element method, and FRAC was used as the damage index. The ML algorithm was trained and evaluated through the data set to determine the structural condition of the beam. ML algorithm analysis includes accuracy and confusion matrix in damage detection assessment. The numerical results demonstrate that the Euclidean, Manhattan, and Braycurtis metrics are considered robust for both datasets with and without noise, with k values lower than 75, considered fast and accurate in the detection and estimation problem of damage in beam structures.

REFERENCES

- Barreto, L.S., Machado, M.R., Santos, J.C., De Moura, B.B. and Khalij, L., 2021. “Damage indices evaluation for one-dimensional guided wave-based structural health monitoring”. *Latin American Journal of Solids and Structures*, Vol. 18, No. 2, pp. 1–17. ISSN 16797825.
- Cutler, J. and Dickenson, M., 2020. *Introduction to Machine Learning with Python*. ISBN 9781449369415.
- Farrar, C.R. and Worden, K., 2013. “SHM- ML Perspective - Operational Evaluation”. *Structural health monitoring: a machine learning perspective*, pp. 55–63. URL <http://linkinghub.elsevier.com/retrieve/pii/B9780121676322500101>.
- Heylen, W. and Lammens, S., 1996. “Frac: a consistent way of comparing frequency response functions”.
- Kurian, B. and Liyanapathirana, R., 2020. *Machine Learning Techniques for Structural Health Monitoring*. Springer Singapore. ISBN 9789811383304. URL http://dx.doi.org/10.1007/978-981-13-8331-1_1.
- Léornad, F., LANTEIGNE, J., LALONDE, S. and TURCOTTE, Y., 2001. “Free-vibration behaviour of a cracked cantilever beam and crack detection”. *Mechanical Systems and Signal Processing*, Vol. 15, No. 3, pp. 529–548.
- Machado, M., Adhikari, S. and Santos, J.D., 2017. “A spectral approach for damage quantification in stochastic dynamic systems”. *Mechanical Systems and Signal Processing*, Vol. 88, pp. 253–273.

- Malekloo, A., Ozer, E., AlHamaydeh, M. and Girolami, M., 2021. *Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights*, Vol. 0. ISBN 1475921721.
- Pathirage, C.S.N., Li, J., Li, L., Hao, H., Liu, W. and Ni, P., 2018. "Structural damage identification based on autoencoder neural networks and deep learning". *Engineering Structures*, Vol. 172, pp. 13–28.
- Sun, Z., Santos, J. and Caetano, E., 2022. "Vision and Support Vector Machine–Based Train Classification Using Weigh-in-Motion Data". *Journal of Bridge Engineering*, Vol. 27, No. 6, pp. 1–8. ISSN 1084-0702.
- Tibaduiza, D., Torres-Arredondo, M.Á., Vitola, J., Anaya, M. and Pozo, F., 2018. "A Damage Classification Approach for Structural Health Monitoring Using Machine Learning". *Complexity*, Vol. 2018. ISSN 10990526.
- Tsuchimoto, K., W.N. and Kitagawa, Y., 2004. "DIAGNOSTIC ASSESSMENT SYSTEM FOR STRUCTURAL SEISMIC SAFETY - DAMAGE IDENTIFICATION BASED ON NEURAL NETWORKS AND TORSION". , No. 1842.
- Vitola, J., Pozo, F., Tibaduiza, D.A. and Anaya, M., 2017. "A sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications". *Sensors (Switzerland)*, Vol. 17, No. 2. ISSN 14248220.
- Wang, L., Lie, S.T. and Zhang, Y., 2016. "Damage detection using frequency shift path". *Mechanical Systems and Signal Processing*, Vol. 66-67, pp. 298–313. ISSN 10961216. URL <http://dx.doi.org/10.1016/j.ymsp.2015.06.028>.
- Yuan, F.G., Zargar, S.A., Chen, Q. and Wang, S., 2020. "Machine learning for structural health monitoring: challenges and opportunities". Vol. 1137903, No. April 2020, p. 2. ISSN 1996756X.
- Zhang, Y., Lie, S.T. and Xiang, Z., 2013. "Damage detection method based on operating deflection shape curvature extracted from dynamic response of a passing vehicle". *Mechanical Systems and Signal Processing*, Vol. 35, No. 1-2, pp. 238–254. ISSN 08883270. URL <http://dx.doi.org/10.1016/j.ymsp.2012.10.002>.

RESPONSIBILITY NOTICE

The author(s) is (are) the only responsible for the printed material included in this paper.