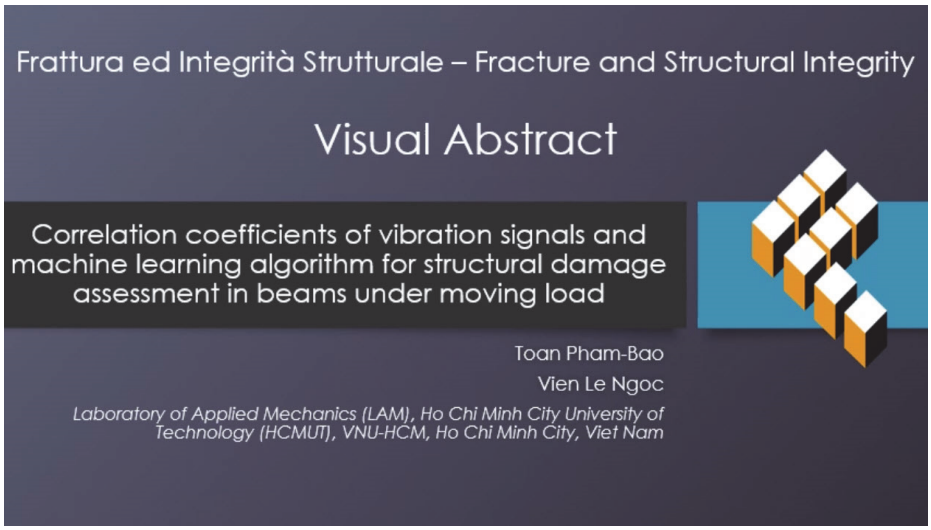
# Correlation coefficients of vibration signals and machine learning algorithm for structural damage assessment in beams under moving load

Toan Pham-Bao*, Vien Le-Ngoc
*Laboratory of Applied Mechanics (LAM), Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Viet Nam*
*baotoanbk@hcmut.edu.vn, https://orcid.org/0000-0002-2105-2403*
*lnvien.sdh19@hcmut.edu.vn, http://orcid.org/0000-0002-8154-1014*




**Citation:** Pham-Bao, T., Le-Ngoc, V., Correlation coefficients of vibration signals and machine learning algorithm for structural damage assessment in beams under moving load, Frattura ed Integrità Strutturale, 70 (2024) 55-70.

**Received:** 24.05.2024
**Accepted:** 09.07.2024
**Published:** 16.07.2024
**Issue:** 10.2024



**KEYWORDS.** Beam structures, Correlation coefficient, Machine learning, Artificial neural network, Structural health monitoring.

## INTRODUCTION

Bridges, buildings, aircraft wings, and other civil infrastructure applications rely on beam structures. It is clear that maintaining the structural integrity of beam structures to ensure safety, reliability, and longevity is essential. It is common for beam structures to degrade over time due to fatigue, corrosion, and external loading, all of which can negatively impact their mechanical characteristics and ultimately affect their structural integrity. It is common for structures to be subjected to dynamic loads from moving vehicles, which are the leading cause of gradual deterioration and damage over time. Excitation of heavy vehicles, in particular, can cause significant stresses and strains, which can lead to cracks,

deformations, and other forms of structural damage. In spite of their effectiveness, traditional inspection methods are time-consuming, labor-intensive, and costly. It is, therefore, becoming increasingly important to develop automated, non-destructive monitoring techniques for these structures. For example, Dipendra Gautam et al. studied safety evaluations of a prestressed concrete bridge by using vibration properties obtained from dynamic identification. The finite element model is updated based on vibration frequencies calculated using parametric and non-parametric system identification methods. Through periodic measurements and system identification, the calibrated model can be used for further analyses, including non-linear response and damage detection [1]. Furthermore, Xiang Zhu et al. provided a concise survey of damage identification in bridges by processing dynamic responses to moving loads. It includes four methods: three direct (Fourier Transform, Wavelet Transform, Hilbert-Huang Transform) and one indirect (Heuristic Interrogation of Damage) [2]. To improve damage detection capabilities, advanced techniques based on structural dynamics and data analytics have been increasingly popular. Vibration-based damage identification (VBDI) has emerged as one of the most promising techniques, as it is sensitive to subtle changes in structural behaviour caused by damage [3, 4]. The method is based on the fact that damage alters the structural dynamics, resulting in changes to vibration characteristics like frequencies, mode shapes, or modal parameters. Through the use of sensors for data collection and advanced algorithms, structural damage and degradation can be identified based on vibration signals. Additionally, VBDI involves analysing the dynamic response of a structure to moving loads. Vibration signatures change when the stiffness, mass distribution, and damping characteristics of a structure are altered. As a result, Early detection and localisation of damage is possible when these changes are monitored before major damage occurs. For instance, Weiwei Zhang introduced a non-model based approach for detecting damage in bridge structures under moving loads by analysing the phase trajectory changes in multi-type vibration measurements [5].

Vibration signals collected by sensors are considered big data, so data processing is essential to assess their structural nature. An integral part of data analysis is the preprocessing step, which has a significant impact on the quality and reliability of the results that follow. It is possible to improve the accuracy, reduce bias, and ensure the robustness of the model by preprocessing data correctly [6, 7]. In addition, preprocessing is essential for selecting and extracting features, reducing dimensionality, and normalising data, which are all essential to improving machine learning performance [8, 9]. The random decrement technique (RDT) stands as a prominent example of the significance of the preprocessing technique. It has been widely utilised in various fields due to its ability to isolate structural responses from noise, which makes it a popular choice among structural health monitors, vibration analysis, and system identification [10, 11]. As a result of statistical methods employed by RDT, random noise is effectively reduced, and underlying trends or patterns are extracted, thereby enhancing the clarity of the signal and the accuracy of subsequent analysis. Hadi Kordestani et al. proposed a two-stage, fully time-domain damage detection method to locate and quantify damage in a simply supported beam under a moving sprung mass. The RDT calculates Random Decrement Signatures (RDSs) from acceleration responses. Then, these RDSs are filtered using the Savitzky–Golay Filter (SGF) to produce SGF-RDSs. The energy-based damage index processes the filtered RDSs to locate and quantify potential damage. According to the results, the authors show that The method is robust to noise due to the use of RDT and SGF [12]. Moreover, Min Qin et al. proposed a novel method for parameter identification under non-white noise excitation, utilising a combination of transformer encoder and long short-term memory networks. RDT is used to remove noise from data. The results show that combining deep learning with traditional methods, such as RDT, yields appropriate accuracy for parameter identification [13]. Additionally, Xingxian Bao et al. proposed a new method for detecting structural damage in offshore systems using a combination of the RDT and long and short-term memory (LSTM) networks. A vibration analysis is carried out, regardless of noise level, using RDT, while a defect identification and evaluation is carried out using LSTM. It has been demonstrated that this method can detect a variety of levels of damage, both simulated and real-world, with high accuracy [11].

Besides preprocessing techniques, machine learning is an excellent support tool for structural damage assessment. A significant characteristic of this method is that it empowers computers to learn from data, identify patterns, and make predictions or decisions without the need to provide explicit information. Machine learning algorithms can be used to analyse data, test hypotheses, and model predictive outcomes, facilitating the exploration of phenomena, the validation of theories, and the development of informed decisions. Many studies have used machine learning either as a primary method or combined with other techniques to assess structural damage [3, 14-17]. Mohammad Abedin et al. investigated the long-term performance of an Accelerated Bridge Construction using machine-learning techniques. In order to accurately identify and assess joint damage, a new approach is proposed that combines machine-learning algorithms with data from load tests. The approach involves utilising bridge response data from a detailed finite element model, simulating various damage scenarios, and training supervised learning algorithms to predict joint damage locations and severity. The results show that this approach can effectively estimate and locate bridge joint damages with a high degree of confidence [18]. Furthermore, Mojtaba Salkhordeh et al. proposed a rapid machine-learning framework to detect damage in primary and secondary

components of reinforced concrete bridges following earthquakes. In this study, two bridge models were used for validation, and Support Vector Machines were determined to be the most effective ML algorithm for assessing bridge component damage [19]. The feedforward net is also one of the countless machine learning methods that experts favour because of its clear structure and ability to discover non-linear relationships between inputs and outputs. Standard optimisation techniques such as gradient descent and back-propagation can effectively train feedforward networks. This helps to expedite the process of testing and improving the model. Additionally, FNN is often combined with other methods to enhance pattern recognition. For instance, Long Viet Ho et al. proposed a novel approach for structural damage detection using a hybrid method combining feedforward neural networks (FNN) with the marine predator algorithm (MPA). The objective of this method is to use MPA's foraging strategy and memory to optimise FNN parameters for enhanced learning. By comparing MPAFNN to other optimisation algorithms, MPAFNN outperforms them in classification tasks. Besides, the method has been able to detect damages in various structural models effectively [20].
In conclusion, this scientific paper stands at the intersection of structural dynamics and artificial intelligence, exploring the symbiotic relationship between machine learning and vibration signal analysis for damage detection. By navigating through theoretical foundations, methodological considerations, and real-world applications, the paper aims to contribute to the ongoing discourse in dam-age detection, empowering researchers to leverage the full potential of machine learning in safeguarding the integrity and resilience of critical infrastructures. In this paper, we use the RDT method to extract the free response of signal and ANN with the input values as the correlation coefficient to detect the damage appearance and location. In total, this paper consists of five sections. Section one is an introduction to the impact of damage determination and related methods. Sections two, three, four and five are a theory, experimental model, data analysis and results, and conclusion, respectively.

## THEORY

### *Random decrement technique (RDT)*

When traditional methods are inadequate due to the non-stationary nature of the signals, RDT offers a robust approach for identifying these parameters from random vibration data. A response signal's decay over time can be captured by RDT's decrement function. A stochastic process auto-RD function represents its mean value under a given $T$ condition. The function is defined as [21]:

$$D_{XX}(\tau) = E\left[ X(t+\tau) \middle| T_{X(t)} \right] \tag{1}$$

where $X(t)$ is the response signal at time $t$; $\tau$ is a time delay and $T_{X(t)}$ is denoted triggering conditions.
It is crucial to assume that the random process is both stationary and ergodic in order to estimate the conditional mean value from a single observation. As a result, autocorrelation functions can be estimated as empirical conditional means derived from a single instance.

$$\hat{D}_{XX}(\tau) = \frac{1}{N}\sum_{i=1}^{N} x(t_i+\tau) \middle| T_{x(t_i)} \tag{2}$$

where $N$ represents the number of points in the process that meet the triggering condition, and $x(t)$ represents realisations of $X(t)$.
It is the number of trigger points, $N$, that makes up the absolute decisive variable in the estimation of RD functions. $N$ must be large enough to ensure that Eqn. (2) has converged sufficiently towards Eqn. (1). In formula (2), unbiased estimates of the RD functions are essential.

$$E\left[ \hat{D}_{XX}(\tau) \right] = \frac{1}{N}\sum_{i=1}^{N} E\left[ x(t_i+\tau) \middle| T_{x(t_i)} \right] = D_{XX}(\tau) \tag{3}$$

The function RD in Eqn. (1) assumes $X(t)$ has a continuous time index. An analysis of the response of a structure is based on the simultaneous sampling of equidistant time points at the sampling interval $\Delta T$ at equidistant points in time.

The obtained results from the triggered conditions are important. The correlation functions of stationary zero-mean Gaussian distributed processes can be derived directly from the RD functions of any triggering condition. The applied general triggering condition $T_{X(t)}^{G_A}$ of the stochastic process *X(t)* is defined as:

$$T_{X(t)}^{G_A} = \left\{ a_1 \le X(t) < a_2, b_1 \le \dot{X}(t) < b_2 \right\} \tag{4}$$

The term "$G_A$" is used to refer to the General Applied triggering condition. When the correlation function and its time derivative are used as the triggering conditions, the RD functions can be derived as a weighted sum.

$$D_{XX}(\tau) = \frac{R_{XX}(\tau)}{\sigma_X^2}.\tilde{a} - \frac{R_{XX}^{'}(\tau)}{\sigma_X^2}.\tilde{b} \tag{5}$$

where $\tilde{a}$ and $\tilde{b}$ are functions of the triggering bounds and density function are defined as:

$$\tilde{a} = \frac{\int_{a1}^{a2} x p_X(x) dx}{\int_{a1}^{a2} p_X(x) dx}; \; \tilde{b} = \frac{\int_{b1}^{b2} \dot{x} p_{\dot{X}}(\dot{x}) d\dot{x}}{\int_{b1}^{b2} p_{\dot{X}}(\dot{x}) d\dot{x}} \tag{6}$$

Eqn. (5) demonstrates the versatility of the RDT. It is possible to change the correlation functions and the time derivative of the correlation functions by adjusting the triggering bounds $a_1$; $a_2$ and/or $b_1$; $b_2$.

*Damage-sensitive feature*

In this study, the vibration response of the slender beam under the effect of moving load is analysed. The load is modelled as a time-varying force moving on a simple Euler-Bernoulli beam. The dynamic beam equation is written as:

$$\frac{\partial^2}{\partial x^2}\left[ EJ(x) \frac{\partial w^2(x,t)}{\partial x^2} \right] + m \frac{\partial w^2(x,t)}{\partial t^2} + c \frac{\partial w(x,t)}{\partial t} = \delta(x - vt) P(t) \tag{7}$$

which $w(x,t)$ – beam deflection at point $x$ and time $t$, $EJ(x)$ – flexural stiffness, $m$ – mass per unit length, $\delta(x)$ - the Dirac delta and $P(t)$ – moving force.

Displacement response $w(x,t)$ can be shown as the modal coordinates:

$$w(x,t) = \sum_{r=1}^{\infty} \phi_r(x) q_r(\mathrm{t}) \tag{8}$$

In which $\phi_r(x)$ and $q_r(t)$ are $r^{\text{th}}$ modal shape and $r^{\text{th}}$ corresponding modal coordinate, respectively.

The modal shapes $\phi_r(x)$ of the beam are the roots of the following expression:

$$m\omega_r^2 \phi_r(x) = \frac{\partial^2}{\partial x^2}\left[ EJ(x) \frac{\partial^2 \phi_r(x)}{\partial x^2} \right] \tag{9}$$

From Eqn. (7) and Eqn. (8), the differential Equation of the beam in each modal coordinate is shown as:

$$\ddot{q}_r(t) + 2\xi_r \omega_r \dot{q}_r(t) + \omega_r^2 q_r(t) = \phi_r(vt) P(t) / \mu_r = f_r(t) \tag{10}$$

with $\omega_m$, $\xi_r$, $\mu_r$ and $f_r(t)$ are natural frequency, damping ratio and corresponding force of $r^{\text{th}}$ modal shape, respectively.

The solution of Eqn. (10) can be determined through the impulse response function h(t) and the damping frequency $\omega_{dr} = \sqrt{1-\zeta_r^2}$ , as follows:

$$q_r(t) = \int_{-\infty}^{\infty} f_r(t-\tau) h_r(\tau) d\tau$$
$$h_r(\tau) = \frac{e^{-\xi_r \omega_r \tau}}{\omega_{dr}} \sin \omega_{dr} \tau \tag{11}$$

After substituting $q_r(t)$ in Eqn. (11) into formula (8), we obtain the vibration response of the beam under the moving load:

$$w(x,t) = \sum_{r=1}^{\infty} \frac{1}{\omega_{dr}} \phi_r(x) \int_{-\infty}^{\infty} f_r(t-\tau) e^{-\xi_r \omega_r(\tau)} \sin \omega_{dr}(\tau) d\tau \tag{12}$$

The average value ($m_w$) and variance ($\sigma_w$) of the response signal at position $x_i$ on the beam are determined by calculating the average and variance of Eqn. (13).

$$m_w(x_i,t) = E[w(x_i,t)] = \sum_{r=1}^{\infty} \frac{\phi(x_i)}{\mu_r \omega_{dr}} \int_{-\infty}^{\infty} m_p \phi(v(t-\tau)) e^{-\xi_r \omega_r(\tau)} \sin \omega_{dr}(\tau) d\tau$$
$$\sigma_w^2(x_i,t) = \sum_{r=1}^{\infty} \frac{1}{(\mu_r \omega_{dr})^2} \phi_r^2(x_i) \int_{-\infty}^{\infty} \sigma_P^2 \phi_r^2(v(t-\tau)) e^{-2\xi_r \omega_r \tau} \sin^2(\omega_{dr}\tau) d\tau \tag{13}$$

where $m_p$ and $\sigma_P$ are the mean value and standard deviation of the input force.
We observe that the mean and variance of the vibration response depend on the natural frequency (corresponding to stiffness) and the mode shape function. When damage occurs at position $x_i$, it alters the $m_w$ and $\sigma_w$ values. The study suggests using a correlation coefficient between the mean value and variance of the response signal for damage identification.

$$R\left(w(x_i,t), w(x_j,t)\right) = \frac{E\left[\left(w_i - m_{w_i}\right)\left(w_j - m_{w_j}\right)\right]}{\sigma_{w_i} \sigma_{w_j}} \tag{14}$$

*Artificial neural network (ANN)*

Artificial neural networks (ANNs) have gained prominence as powerful computational models to identify non-linear features. A feedforward neural network (FNN) is one of the most widely used machine learning architectures due to its simplicity and effectiveness in learning complex mappings between input and output. A FNN consists of an input layer, one or more hidden layers, and an output layer. The connections between neurons in a layer are determined by their associated weights, which determine how strong the connection is between neurons. Using a series of weighted summations and activation functions, an FNN can compute its output y based on an input vector x.
Activation functions introduce nonlinearity into the neural network, allowing it to understand complex patterns. Common activation functions include sigmoid, hyperbolic tangent, rectified linear unit, purelin, and softmax functions. Activation functions and weighted summations determine the output y of an according to the formula:

$$y_j = f_{ac} \sum w_{ji} x_i + b_j \tag{15}$$

where for each neuron, $x_i$ represents its inputs, $y_j$ represents its outputs, and $b_j$ represents its biases. $w_{ij}$ is the weight coefficient, the notation "ji" indicates that the input is referred to is *i*, and the neuron being referred to is j. $f_{ac}$ is an activation function.

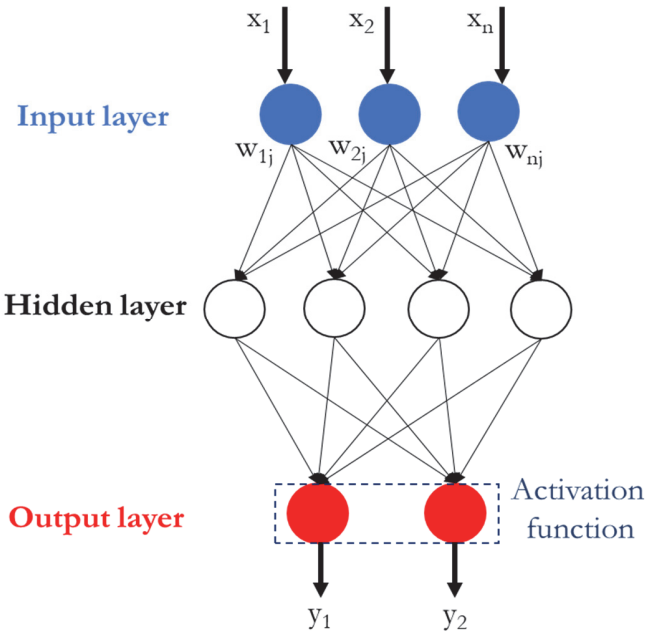



Figure 1: The simple architecture of three layers of FNN.

At the hidden layer, the value bj corresponds to selecting the largest derivative of the error function to initialise each neuron's value. This speeds up the process of minimising errors. Feedforward networks typically use the average squared error (MSE) as their error function. In terms of MSE, the output of the network is compared with the output of the target according to the following formula:

$$MSE=\sum_{j=1}^{N}\frac{(y_j-y_t)^2}{N} \tag{16}$$

where $y_j$ is the output of ANN, $y_t$ is the target output, and N is the number of samples.
The MSE is a function that depends on the weight and bias. The rate of error improvement is determined by the derivative of the activation function.

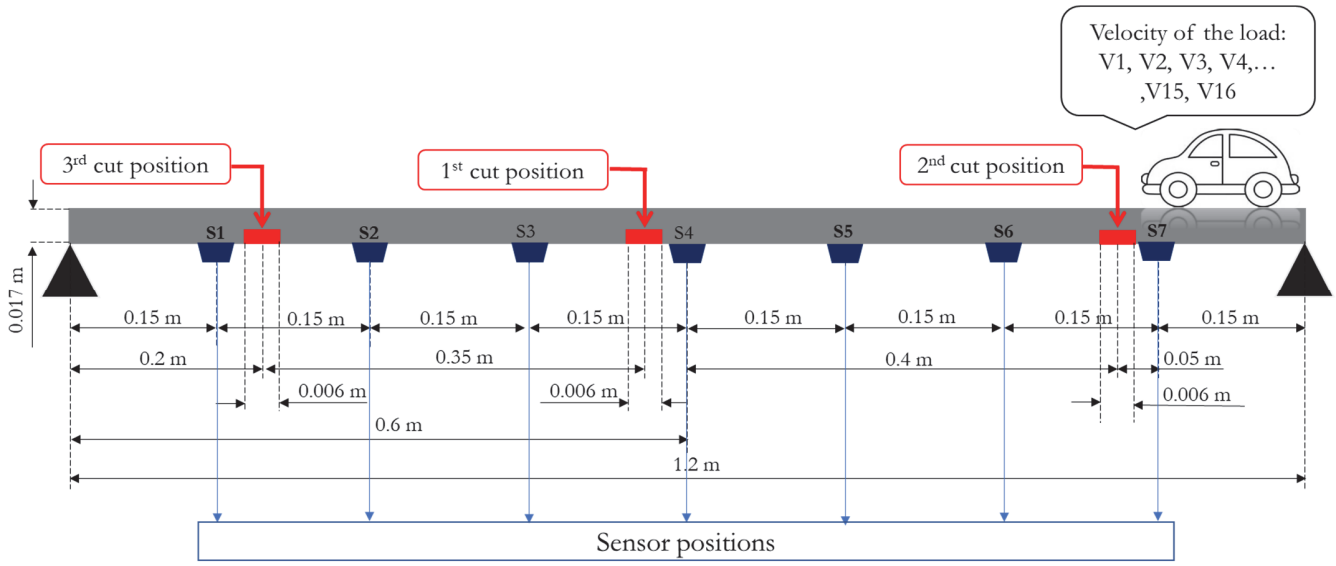



Figure 2: The general diagram of the experimental model.

## EXPERIMENTAL MODEL AND DATA COLLECTION

The investigation involved testing a simulated bridge girder using a wooden beam with dimensions of 1.2 m × 0.1 m × 0.017 m (length × width × thickness). The beam supported two ends and was subjected to a moving load ranging from 37.7 cm/s to 94.25 cm/s, controlled by an inverter with frequencies from 20Hz to 50Hz. The cuts were consistently 0.006 m wide, with depths of 0.003 m, 0.006 m, 0.009 m, and 0.011 m, extending across the entire beam width. Seven accelerometers (S1, S2, S3, S4, S5, S6, S7) were strategically placed along the beam's length at 1/8, 2/8, 3/8, 4/8, 5/8, 6/8, and 7/8 positions. The study explored twelve scenarios, including an intact beam and eleven damaged states with varying damage locations. The initial measurement captured vibrations of the undamaged beam, followed by making cuts near sensor positions (S1 and S7) with increasing depths.
Accelerometers recorded data continuously for 400 seconds under each of the 16-speed states (V1 to V16) during 10-second intervals at a sampling frequency of 2000. This process was repeated 12 times for different beam scenarios. Fig. 2 and Fig. 4 illustrated the experimental model and implementation, while Fig. 2 depicted the shape of the cuts, and Tab. 2 listed the 12 damage situations. Tab. 3 shows the configurations and specifications of the accelerometer.

| Load mass | Symbol velocity (m/s) | Load mass | Symbol velocity (m/s) |
|---|---|---|---|
| 3 kg | V1 = 0.377 | 3 kg | V9 = 0.6786 |
| | V2 = 0.415 | | V10 = 0.7163 |
| | V3 = 0.452 | | V11 = 0.7540 |
| | V4 = 0.490 | | V12 = 0.7917 |
| | V5 = 0.528 | | V13 = 0.8294 |
| | V6 = 0.566 | | V14 = 0.8671 |
| | V7 = 0.603 | | V15 = 0.9048 |
| | V8 = 0.641 | | V16 = 0.9425 |

Table 1: Symbol of the speed and mass of the load.

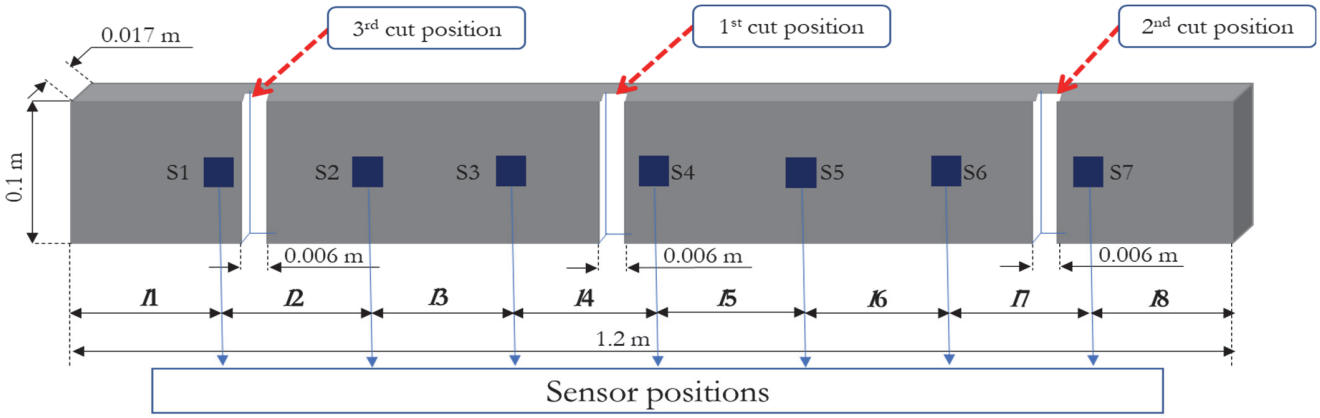



Figure 3: The shape cuts and symbol of distance between sensors.

| Damage state (symbol) | Quantity of damage | Location of damage 1st Cut (depth) | 2nd Cut (depth) | 3rd Cut (depth) |
|---|---|---|---|---|
| D00 (Intact Beam) | 0 | - | - | - |
| D01 | 1 | *l*4 (3mm) | - | - |
| D02 | 1 | *l*4 (6mm) | - | - |
| D03 | 1 | *l*4 (9mm) | - | - |
| D04 | 1 | *l*4 (11mm) | - | - |
| D05 | 2 | *l*4 (11mm) | *l*7 (3mm) | - |
| D06 | 2 | *l*4 (11mm) | *l*7 (6mm) | - |
| D07 | 2 | *l*4 (11mm) | *l*7 (9mm) | - |
| D08 | 2 | *l*4 (11mm) | *l*7 (11mm) | - |
| D09 | 3 | *l*4 (11mm) | *l*7 (11mm) | *l*2 (6mm) |
| D10 | 3 | *l*4 (11mm) | *l*7 (11mm) | *l*2 (9mm) |
| D11 | 3 | *l*4 (11mm) | *l*7 (11mm) | *l*2 (11mm) |

Table 2: The summary of damage scenarios.

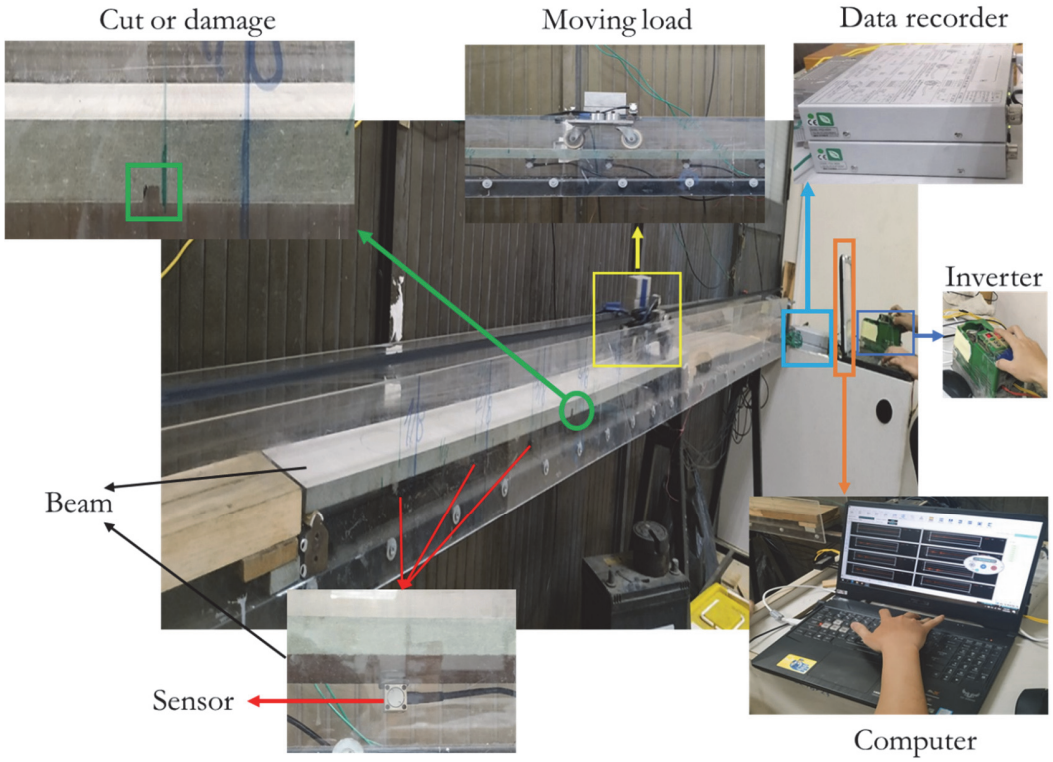



Figure 4: Design the model in the laboratory.

| Parameter | Magnitude |
|---|---|
| Rated capacity | ± 19.61 m/s² (± 2G) |
| Nonlinearity | Within ± 1% RO |
| Hysteresis | Within ± 1% RO |
| Rated output | 0.5 mV/V (1000×10$^{-6}$ strain) or more |
| Safe temperature | -15 to 65$^{0}$C |
| Safe excitation | 6V AC or DC |
| Recommended excitation | 1 to 3V AC or DC |
| Input resistance | 121Ω ± 1.7% |
| Output resistance | 121Ω ± 1.7% |
| Cable | 4-conductor (0.08 mm²) vinyl shielded cable, 3.2mm diameter by 5m long<br>Terminated with a connector plug PRC03-12A10-7M (Shield wire is connected to the case). |
| Safe overload | 300% |
| Frequency response | DC to 60Hz ± 5% |
| Transverse sensitivity | 4% RO or less |
| Weight | Approx. 25g (excluding cable) |

Table 3: Configurations and specifications of accelerometer.

After conducting the load to move through the beam at each speed, we obtain 640 samples in each damage case. However, in the case of cutting D09, we only collected 610 samples due to some objective reasons when collecting data. Therefore, the total number of samples we collected in this experiment is 7650 samples.

## DATA ANALYSIS AND RESULTS

### *Data analysis*

Fig. 5 summarises the data processing procedures in this paper. The raw data collection part has been mentioned in the above section. Next, As shown in Fig. 4, the raw signal is converted into an RDT signal by formula (1). The duration of the subsegments is set to 5 seconds, and other parameters depend on each signal considered. The results for the signal of damage case D00 at sensor location S1 are shown in Fig. 6.

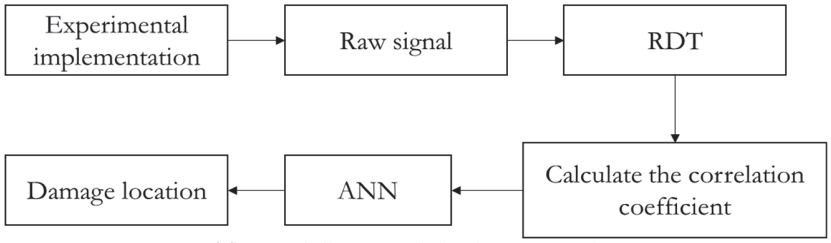



Figure 5: The total diagram of the data processing process.

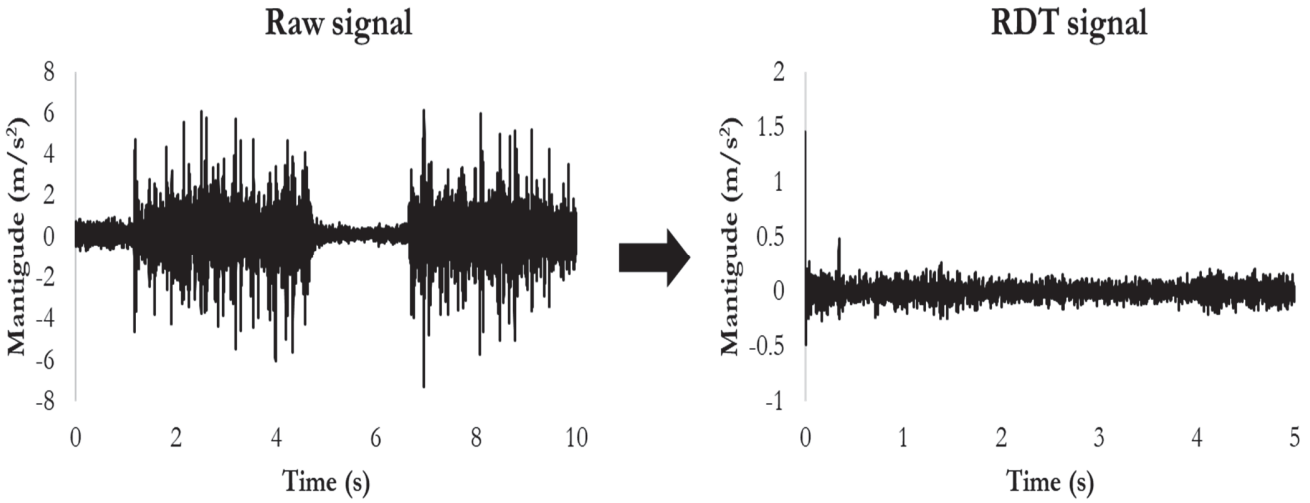



Figure 6: This figure illustrates the process of converting a raw signal into an RDT signal.

After converting raw data to random decrement data, we calculated the correlation between sensor locations according to the formula (17). We use correlation values as the features of this study.

$$r_{x_i x_j} = \frac{\sum_{k=1}^{n}(X_{x_i,k} - \bar{X}_{x_i})(Y_{x_j,k} - \bar{Y}_{x_j})}{\sqrt{\sum_{k=1}^{n}(X_{x_i,k} - \bar{X}_{x_i})^2 . \sum_{l=1}^{n}(Y_{x_j,l} - \bar{Y}_{x_j})^2}} \tag{17}$$

where
$x_i$ is the signal at i$^{th}$ location (i = 1, 2, …, 7)
$x_j$ is the signal at j$^{th}$ location (j = 1, 2, …, 7)
$X$ are the elements in the signal $x_i$
$\bar{X}$ is the mean value of the signal $x_i$
$Y$ are the elements in the signal $x_j$

$\bar{Y}$ is the mean value of the signal $x_i$

Because of seven sensors from S1 to S7, so there will be a total of 7 x 7 = 49 correlation coefficient values shown in Tab. 3.

| Sensors | S1 | S2 | S3 | S4 | S5 | S6 | S7 |
|---|---|---|---|---|---|---|---|
| S1 | 1 | 0.594728 | 0.372957 | 0.352154 | 0.339963 | 0.352570 | 0.326428 |
| S2 | 0.594728 | 1 | 0.538598 | 0.245095 | 0.499330 | 0.560030 | 0.352909 |
| S3 | 0.372957 | 0.538598 | 1 | 0.321962 | 0.347568 | 0.412174 | 0.261094 |
| S4 | 0.352154 | 0.245095 | 0.321962 | 1 | 0.329459 | 0.198569 | 0.349146 |
| S5 | 0.339963 | 0.499330 | 0.347568 | 0.329459 | 1 | 0.537905 | 0.341837 |
| S6 | 0.352570 | 0.560030 | 0.412174 | 0.198569 | 0.537905 | 1 | 0.546458 |
| S7 | 0.326428 | 0.352909 | 0.261094 | 0.349146 | 0.341837 | 0.546458 | 1 |

Table 4: The following table displays the correlation values of a sample in a scenario where there is no damage, denoted by D00.

The correlation coefficient results for other scenarios have been calculated in a similar way. Additionally, we remove duplicate values in Tab. 3, reducing the input values from 49 correlation coefficient values per damage case to 21 values (Tab. 4). Therefore, with a total of 7,650 samples, we have a grand total of 160,650 correlation coefficient values (7,650 x 21) for all damage scenarios.

| Sensors | S1 | S2 | S3 | S4 | S5 | S6 | S7 |
|---|---|---|---|---|---|---|---|
| S1 | // | 0.594728 | 0.372957 | 0.352154 | 0.339963 | 0.352570 | 0.326428 |
| S2 | // | // | 0.538598 | 0.245095 | 0.499330 | 0.560030 | 0.352909 |
| S3 | // | // | // | 0.321962 | 0.347568 | 0.412174 | 0.261094 |
| S4 | // | // | // | // | 0.329459 | 0.198569 | 0.349146 |
| S5 | // | // | // | // | // | 0.537905 | 0.341837 |
| S6 | // | // | // | // | // | // | 0.546458 |
| S7 | // | // | // | // | // | // | // |

Table 5: After reducing similar correlation values.

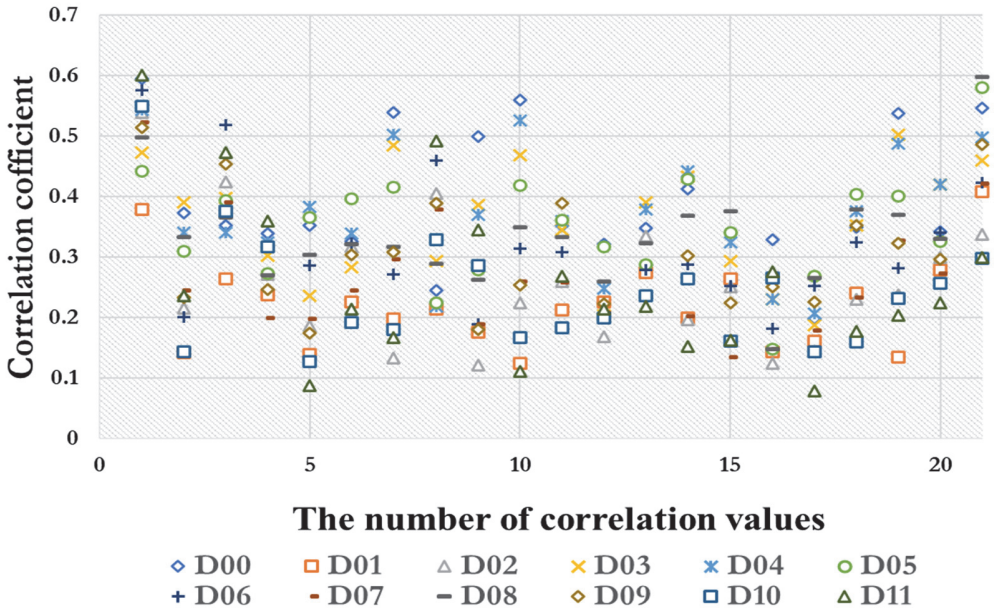



Figure 7: The figure illustrates 21 correlation values of twelve damage scenarios (D00 to D11) at a random sample.

According to the results in Fig. 7, the correlation value ranges from 0.1 to 0.6 in all damage cases. These values do not follow linear patterns that help identify damages. Therefore, machine learning methods should be applied to compute these non-linear features.

As damage cannot be identified through conventional methods through these values (Fig. 7), we employed ANN as a supportive tool to diagnose it. We have randomly segregated 7650 samples into two sets - one for training and the other for retesting purposes. In this case, 90% of the data (6885 samples) is used for training and 10% (765 samples) for retesting. Then, we designed an Artificial Neural Network (ANN) with a feedforward architecture. The network architecture and dimensions are defined with one input layer, one hidden layer, and one output layer (Fig. 6). The input layer consists of 21 neurons, each corresponding to a correlation coefficient value for a sample. The output layer has eight neurons, which correspond to the length of the beam (***l1***, ***l2***, ..., ***l8***), and the hidden layer consists of 25 neurons, as shown in Fig. 8.

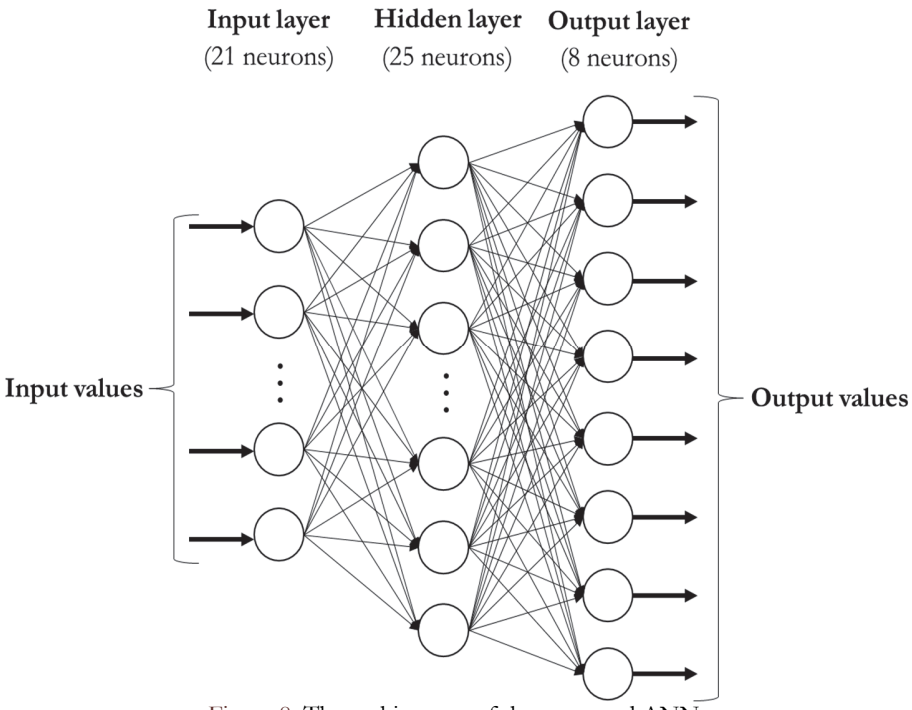



Figure 8: The architecture of the proposed ANN.

A network's weights and biases are initialised once its architecture has been established; weights and biases are randomly assigned in this study. Following this, selecting an appropriate activation function for each neuron is essential to introduce nonlinearity and enable the network to learn complex patterns. We select a log-sigmoid transfer function (logsig) for the hidden layer and hyperbolic tangent sigmoid transfer function (tansig) for the output layer according to formulas (18) and (19), respectively.

$$\text{logsig}(n) = \frac{1}{1+e^{-n}} \tag{18}$$

$$\text{tansig}(n) = \frac{2}{1+e^{-2n}} - 1 \tag{19}$$

The ANN results will be a number between 0 and 1, indicating the presence or absence of damage. When the output is close to 0, it indicates no damage, while an output approaching 1 is assumed to indicate damage. For example, In the event that the first neuron of the ANN outputs a value close to 1 and the remaining neurons values close to 0, it indicates that the location *l*1 is damaged. If there are multiple instances of damage, the amount will increase accordingly. With the proposed ANN model, we can determine whether and where damage is likely based on the output values of the eight neurons.

The proposed ANN architecture was trained using a databank of 6,885 samples, representing 90% of the total databank and covering eleven damage and integrity scenarios. This dataset is divided into three parts for training, validation, and

testing, with 80%, 10%, and 10% distribution, respectively. The Levenberg-Marquardt back-propagation algorithm is employed to update the weights and biases during the training process. The training is limited to a maximum of 50 epochs, and the learning rate is set to 0.001. If there are six consecutive validation failures, the training is terminated. As shown in Fig. 9, the optimal validation performance is achieved at the 15th epoch, and training concludes at the 21st epoch due to meeting the validation criterion.

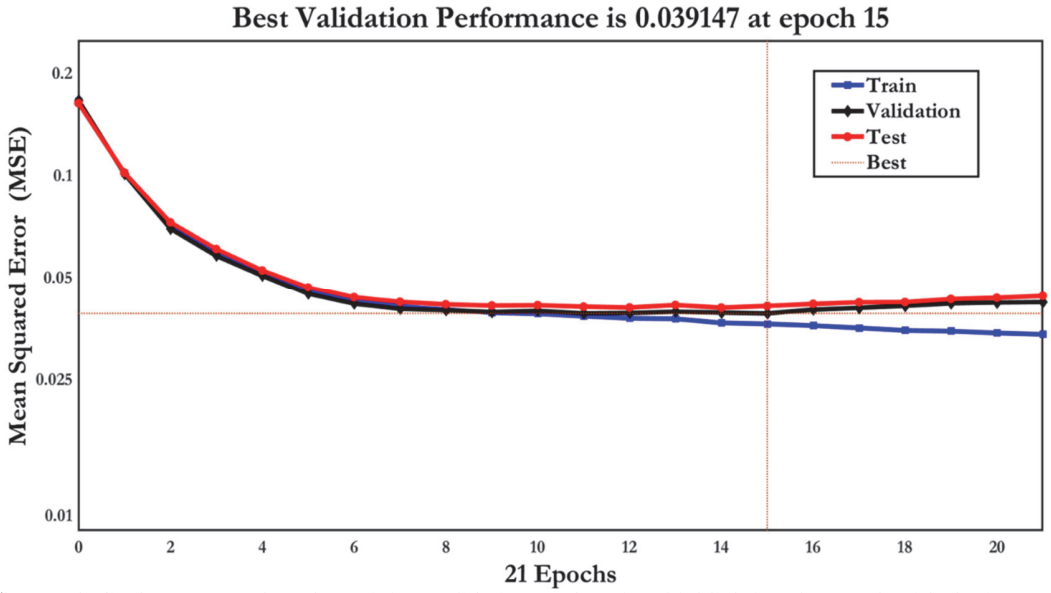



Figure 9: The graph displays MSE values for training, validation, and testing, highlighting the epoch with the best performance.

*Results*

After completing the training process, we test the remaining 765 samples in the common data bank to verify the accuracy of ANN. As shown in Fig. 10, Fig. 11, Fig. 12, and Fig. 13, we present the retest results with a total of 10 representative samples for each damage state due to the large number of samples used for testing.

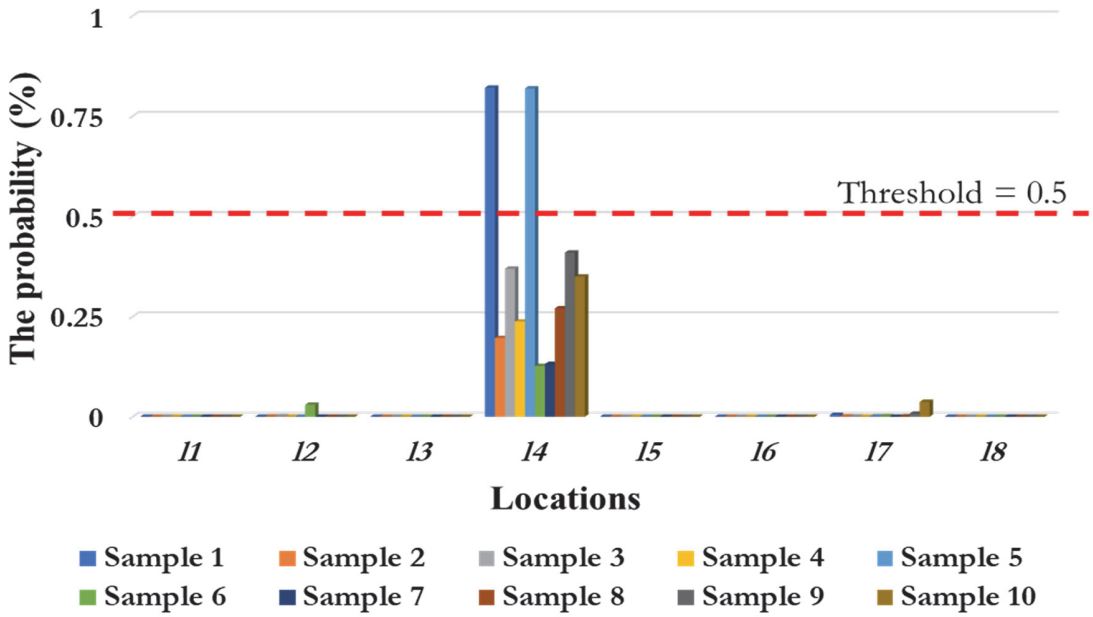



Figure 10: The chart illustrates the probability of damage occurring at eight locations (*l*1 to *l*8) in a scenario where there is no damage.

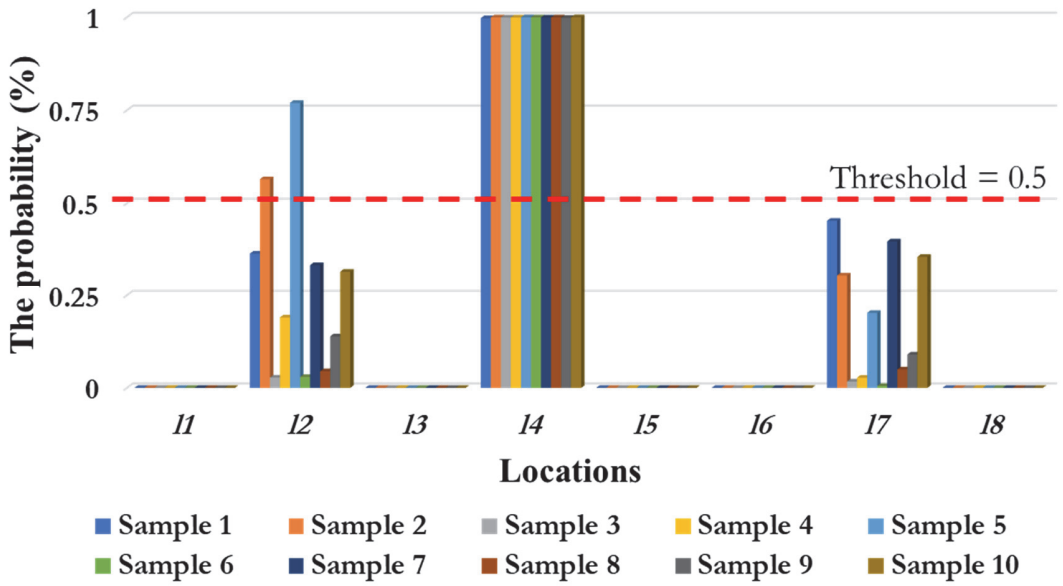



Figure 11: The chart demonstrates the likelihood of damage at eight locations (*l*1 to *l*8) in a scenario where there is one instance of damage at the *l*4 location with a severity rating of four.

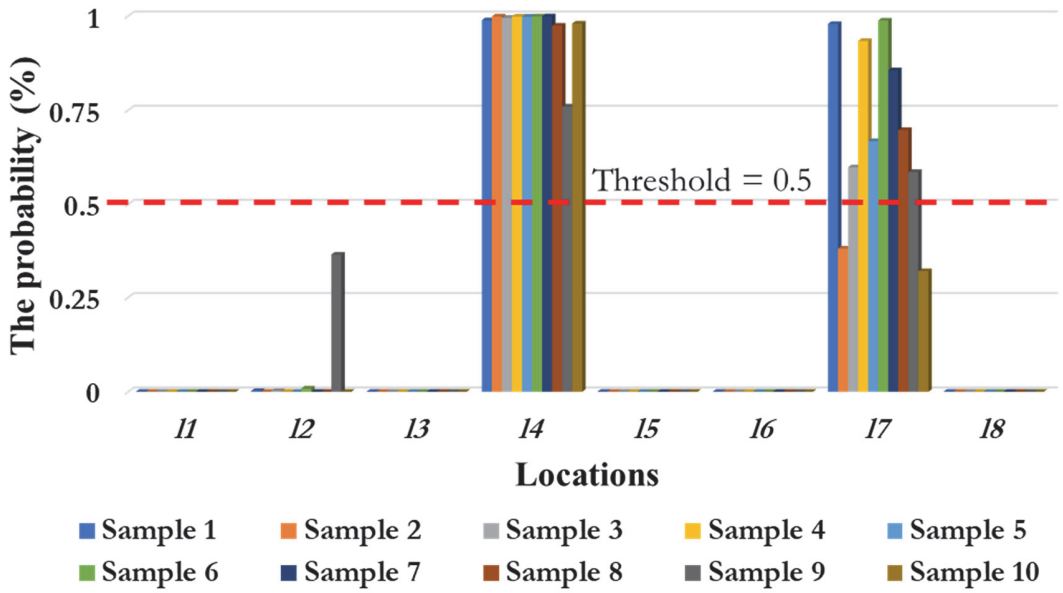



Figure 12: The chart demonstrates the likelihood of damage at eight locations (*l*1 to *l*8) in a scenario where there are two instances of damage at the *l*4 and *l*7 locations with a severity rating of four.

Based on the results shown in Figs. 10 through 12, it is evident that the method accurately detects damage appearance and location with a damage threshold of 0.5. Despite significant superiority in damage assessment methods, there are still cases where the results are inaccurate. This is a problem that persists across all types of damage. However, the proposed method is highly effective with 765 retest samples with an accuracy of up to 98.16% for each position is predicted. Furthermore, if there are more damages, the method will be more effective at identifying them.

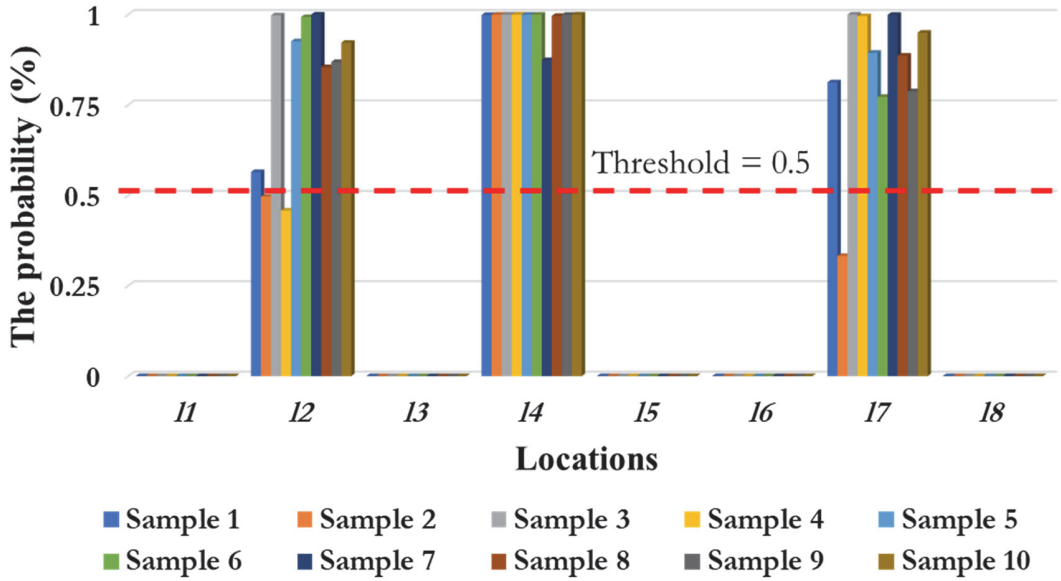



Figure 13: The chart illustrates the probability of damage occurring at eight different locations (*l*1 to *l*8). There were three instances of damage reported at *l*4, *l*7, and *l*2 locations. The severity rating of the damage was four for both *l*4 and *l*7 and three for *l*2.

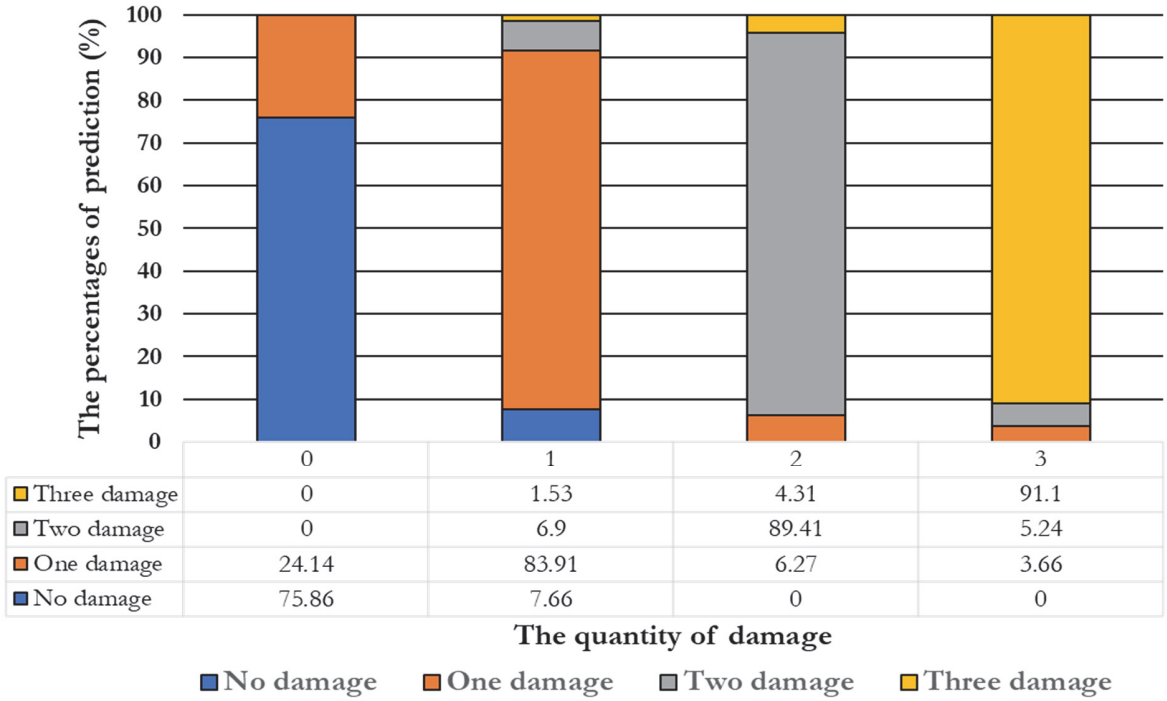



Figure 14: The chart shows the percentage predicting damage in the contexts of no damage, one damage, two damages and three damages.

In Fig. 14, we show the prediction ability of ANN with 765 samples, which are used for retesting. The results show that the ability to predict the beam without damage accurately is 75.86%, while 24.14 % of incorrect predictions are mainly the beam with one damage. The prediction accuracy for one, two, and three damage is 83.91%, 89.41%, and 91.1%, respectively. This proves that as more damages occur, this method will be more effective in identifying them. On the other hand, it only reaches 86.93% accuracy when considering each case of damage.

## CONCLUSION

This study demonstrates the efficacy of utilising experimental data through the RDT for free response extraction. RDT is also considered a signal preprocessing in this research. Thereafter, the features for the ANN input were derived from calculating the correlation coefficient of the RDT signal between measurement locations. By implementing a three-layer FNN, we have successfully identified and localised damage on beams with good predictability. The results indicate a high accuracy of 98.16% in identifying the specific location of damage within the structure. However, when considering individual damage cases, the accuracy slightly decreases to 86.93%. More specifically, the effectiveness of ANN for identifying the beam without damage is 75.86%. Then, this number slightly increases when the beam has damage with 83.91%, 89.41% and 91.1% for the beam with one damage, two damages and three damages, respectively. Additionally, The results show that the proposed method has good identification ability with high levels of structural damage. A combination of advanced signal processing techniques (RDT) and machine learning (FNN) could enhance structural health monitoring systems, as shown by our findings.

However, several shortcomings of the method were identified during the study. One fundamental limitation is the low accuracy when classifying between structures without damage and structures with damage. Raw signal correlations may not be enough to distinguish subtle variations in damage patterns in the ANN model. For the model to become more effective in distinguishing between different types of damage scenarios, more sophisticated methods of feature extraction or additional input data are needed. Additionally, We still cannot determine the severity of damage to the beam using the correlation values of the original signal in this article.

Overall, While raw signal correlations and ANN-based approaches show promise in identifying structural damage, further research and improvements are needed to address methodological limitations and make the model more accurate, robust, and applicable to real-world situations.

## ACKNOWLEDGEMENT:

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, for supporting this study.

## REFERENCES

[1] Gautam, D., Adhikari, R., Olafsson, S.and Rupakhety, R. (2023). Dynamic identification of an aging prestressed concrete bridge using heavy vehicle excitation. Soil Dynamics and Earthquake Engineering, 173, p. 108127.

[2] Zhu, X., Cao, M., Ostachowicz, W.and Xu, W. (2019). Damage identification in bridges by processing dynamic responses to moving loads: features and evaluation. Sensors, 19 (3), p. 463.

[3] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M.and Inman, D.J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical systems and signal processing, 147, p. 107077.

[4] Das, S., Saha, P.and Patro, S. (2016). Vibration-based damage detection techniques used for health monitoring of structures: a review. Journal of Civil Structural Health Monitoring, 6, pp. 477-507.

[5] Zhang, W., Li, J., Hao, H.and Ma, H. (2017). Damage detection in bridge structures under moving loads with phase trajectory change of multi-type vibration measurements. Mechanical Systems and Signal Processing, 87, pp. 410-425.

[6] Alasadi, S.A. and Bhaya, W.S. (2017). Review of data preprocessing techniques in data mining. Journal of Engineering Applied Sciences, 12 (16), pp. 4102-4107.

[7] Çetin, V.and YILDIZ, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 28 (2), pp. 299-312.

[8] Alam, S.and Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. Computational Mathematical Organization Theory, 25, pp. 319-335.

[9] Zelaya, C.V.G., (Year). Towards explaining the effects of data preprocessing on machine learning. 2019 IEEE 35th international conference on data engineering (ICDE), pp. 2086-2090.

[10] Bao, X., Liu, M., Fu, D., Shi, C., Cui, H., Sun, Z., Liu, Z.and Iglesias, G. (2023). Damage identification for jacket offshore platforms using Transformer neural networks and random decrement technique. Ocean Engineering, 288, p. 115973.
[11] Bao, X., Wang, Z.and Iglesias, G. (2021). Damage detection for offshore structures using long and short-term memory networks and random decrement technique. Ocean Engineering, 235, p. 109388.
[12] Kordestani, H., Zhang, C.and Shadabfar, M. (2019). Beam damage detection under a moving load using random decrement technique and Savitzky–Golay filter. Sensors, 20 (1), p. 243.
[13] Qin, M., Chen, H., Zheng, R.and Gao, T. (2021). An adaptive operational modal analysis method using encoder LSTM with random decrement technique. Journal of Sensors, 2021, pp. 1-11.
[14] Azimi, M., Eslamlou, A.D.and Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. Sensors, 20 (10), p. 2778.
[15] Caicedo, D., Lara-Valencia, L.and Valencia, Y. (2022). Machine learning techniques and population-based metaheuristics for damage detection and localisation through frequency and modal-based structural health monitoring: A review. Archives of Computational Methods in Engineering, 29 (6), pp. 3541-3565.
[16] Figueiredo, E., Park, G., Farrar, C.R., Worden, K.and Figueiras, J. (2011). Machine learning algorithms for damage detection under operational and environmental variability. Structural Health Monitoring, 10 (6), pp. 559-572.
[17] Santos, A., Figueiredo, E., Silva, M., Sales, C.and Costa, J. (2016). Machine learning algorithms for damage detection: Kernel-based approaches. Journal of Sound Vibration, 363, pp. 584-599.
[18] Abedin, M., Mokhtari, S.and Mehrabi, A.B., (Year). Bridge damage detection using machine learning algorithms. Health Monitoring of Structural and Biological Systems XV, 11593, pp. 532-539.
[19] Salkhordeh, M., Mirtaheri, M., Rabiee, N., Govahi, E.and Soroushian, S. (2023). A rapid machine learning-based damage detection technique for detecting local damages in reinforced concrete bridges. Journal of Earthquake Engineering, 27 (16), pp. 4705-4738.
[20] Ho, L.V., Nguyen, D.H., Mousavi, M., De Roeck, G., Bui-Tien, T., Gandomi, A.H.and Wahab, M.A. (2021). A hybrid computational intelligence approach for structural damage detection using marine predator algorithm and feedforward neural networks. Computers and Structures, 252, p. 106568.
[21] Asmussen, J.C. (1997). Modal analysis based on the random decrement technique. Department of Building Technology Structural Engineering University of Aalborg.