DAMAGE DETECTION IN STEEL BEAMS THROUGH NATURAL FREQUENCY USING A RANDOM FOREST MODEL

Van Tuan Vu[1,*], Anh Dung Dang[1], Hai Dang Lam[1], Duc Thinh Nguyen[1], Trung Duc Tran[1]

[1]*Institute of Techniques for Special Engineering, Le Quy Don Technical University*

Abstract

Recently, machine learning (ML) algorithms have proven to be highly effective tools for predicting structural damage. However, the data used in structural health monitoring often consists primarily of normal operational conditions or slight deviations from the original state, with a scarcity of data representing potentially dangerous conditions. This limitation makes it challenging to create realistic datasets for training ML models to detect structural damage. If such data were available, it would likely involve parameters like the stress intensity factor range and stress ratio, which are difficult to measure in real-world structures. In this study, a random forest (RF) model was developed to predict the locations, widths, and depths of saw-cuts in steel beams based on variations in natural frequencies. These natural frequencies under various damage scenarios were determined using the Finite Element Method (FEM). To ensure accuracy, the natural frequencies in the undamaged state were compared between the FEM and Frequency Domain Decomposition (FDD). After training, the RF model showed an R-squared value of 0.996 for location, 0.876 for width, and 0.880 for depth. The mean squared error (MSE) was found to be 0.0003 for location, 0.0313 for width, and 0.0420 for depth. The results indicate that combining the FEM and FDD with the RF model holds significant potential for applications in structural health monitoring.

Keywords: *Saw-cut prediction; random forest; natural frequency; Frequency Domain Decomposition; FEM dynamic analysis.*

1. Introduction

Beams have long been fundamental in engineering applications and are frequently used to model civil engineering challenges. Numerous models and techniques have been developed to detect damage in beams. For instance, Yang *et al.* [1] applied Galerkin's method and the energy approach to identify cracks in vibrating beams. In another study, Swamidas *et al.* [2] employed both Timoshenko and Euler formulations to detect cracks in beams. Research by G. R. Gillich *et al.* [3], Zhou *et al.* [4] and G. R. Gillich *et al.* [5] focused on detecting damage cracks through vibration measurements. Additionally, Zhou *et al.* [6] explored the forced vibration behavior of cracked beams. The findings from these studies have shown strong performance in structural damage detection.

[*] Corresponding author, email: vantuanvu@lqdtu.edu.vn

DOI: 10.10.56651/lqdtu.jst.v7.n02.878.sce

In recent years, machine learning algorithms have emerged as powerful tools for predicting structural damage. Samir *et al.* [7] addressed damage identification by employing a Genetic Algorithm (GA) approach that capitalizes on changes in natural frequencies. Ghadimi *et al.* [8] developed a crack detection technique using a modified extreme learning machine, which takes mode shapes and the first three frequencies as inputs to identify cracks as outputs. This method was tested on several examples and demonstrated effectiveness, even in the presence of noise. N. Gillich *et al.* [9] developed databases with scenarios depicting damage to a cantilever beam, considered crack location and severity in two steps using ANN and RF, and achieved highly accurate results in simulations and experiments.

T. C. Le *et al.* [10] proposed a hybrid approach using Particle Swarm Optimization (PSO) and Support Vector Machines (SVM) for precise damage identification. This method eliminated unnecessary parameters by leveraging PSO's search capabilities and utilized the robust SVM model to predict damage locations and severity, showing superior performance compared to other machine learning models, especially in cases of minor damage. Rathod *et al.* [11] developed models for damage classification based on the behavior of mode conversion versus frequency curves for four wave modes. Their findings highlighted that SVM and RF classifiers excelled in accuracy on the dataset, achieving the lowest error rates. More recently, de Sousa *et al.* [12] introduced a machine learning approach for monitoring structural integrity in beam-like structures, assessing the effectiveness of six algorithms, including SVM, k-nearest neighbors (kNN), Decision Tree (DT), Naive Bayes (NB), and RF. Their methodology achieved up to 100% accuracy with simulated data and 95% accuracy with experimental data, effectively classifying the health condition of the structures.

It is evident that using machine learning algorithms to predict damage is entirely feasible and yields very high accuracy. However, in structural health monitoring, the input data typically represents normal operational conditions or states with minor deviations from the baseline, lacking information from potentially dangerous conditions. This absence makes it challenging to develop a realistic dataset for machine learning models aimed at detecting structural damage. If such data were accessible, it would likely include parameters such as the stress intensity factor range and stress ratio, which are challenging to measure accurately in real-world structures.

As a possible approach, numerical models can be employed to generate training datasets. Machine learning models can then utilize monitoring data from actual structures to predict beam damage. In this study, a RF model was developed to estimate the locations, widths, and depths of saw-cuts in steel beams by analyzing variations in natural frequencies. These frequencies, corresponding to different damage scenarios, were determined using a Finite Element Method (FEM) model. To evaluate the viability of this approach, the natural frequencies without saw-cuts, derived from the Frequency Domain Decomposition (FDD) method, were compared with those obtained from the FEM. Conclusions on the integration of FEM, FDD, and the RF model will be drawn upon completion.

2. Determination of natural frequencies using Frequency Domain Decomposition and Finite Element Method

2.1. Identification of natural frequencies in a steel beam through Frequency Domain Decomposition

Brincker *et al.* [13] introduced the concept of FDD. This approach utilizes singular value decomposition (SVD) to break down the spectral density matrix at each frequency into singular values and corresponding singular vectors. FDD is essentially an enhancement of the traditional frequency domain method, often referred to as the peak picking technique, where natural frequencies are determined by identifying peaks within the spectral density matrix.

The connection between the unknown input $x(t)$ and the observed response output $y(t)$ can be represented by the following equation:

$$[G_{yy}(\omega)] = [H(\omega)]^*[G_{xx}(\omega)][H(\omega)]^T \tag{1}$$

where $[G_{xx}(\omega)]$ is the Power Spectral Density (PSD) matrix of the input, $[G_{yy}(\omega)]$ is the PSD matrix of the responses, $[H(\omega)]^*$ is the complex conjugate matrix of Frequency Response Function (FRF), $[H(\omega)]^T$ is the transpose matrix of FRF.

The FRF can be written in prutial fraction:

$$[H(\omega)] = \sum_{1}^{N} \frac{[R_k]}{j\omega - \lambda_k} + \frac{[R_k]^*}{j\omega - \lambda_k^*} \tag{2}$$

$$\lambda_k = -\sigma_k + j\omega_{dk} \tag{3}$$

where N is the number of modes, λ_k is the pole of the k^{th} mode shape, σ_k is minus the real part of the pole, ω_{dk} is the damped natural frequencies of the k^{th} mode shape.

$[R_k]$ is the residue, defined as follows:

$$[R_k] = \phi_k \cdot \gamma_k^T \tag{4}$$

where ϕ_k is mode shape vector, γ_k is the modal participation vector.

Suppose the input is white noise, its power spectral density is constant or $[G_{xx}(\omega)] = C$. Formula (1) is rewritten as follows:

$$[G_{yy}(\omega)] = \sum_{1}^{N}\sum_{1}^{N}\left[\frac{[R_k]}{j\omega-\lambda_k}+\frac{[R_k]^*}{j\omega-\lambda_k^*}\right]\times C\times\left[\frac{[R_k]}{j\omega-\lambda_k}+\frac{[R_k]^*}{j\omega-\lambda_k^*}\right]^T \quad (5)$$

By multiplying the two partial fraction components and applying the Heaviside partial fraction theorem, the output PSD can be simplified into a pole/residue form after some mathematical manipulations, as shown below:

$$[G_{yy}(\omega)] = \sum_{1}^{N}\frac{[A_k]}{j\omega-\lambda_k}+\frac{[A_k^*]}{j\omega-\lambda_k^*}+\frac{[B_k]}{-j\omega-\lambda_k}+\frac{[B_k^*]}{-j\omega-\lambda_k^*} \quad (6)$$

where $[A_k]$ is the k^{th} residue matrix of the output PSD.

At a specific frequency ω, usually only a few modes, often just one or two, will have a significant impact. Therefore, for a lightly damped structure, the response spectral density can generally be expressed as follows:

$$[G_{yy}(\omega)] = \sum_{k\in Sub(\omega)}\frac{d_k\phi_k\phi_k^T}{j\omega-\lambda_k}+\frac{d_k^*\phi_k^*\phi_k^{*T}}{j\omega-\lambda_k^*} \quad (7)$$

where $k \in Sub(\omega)$ is the set of modes be denoted at a specific frequency.

The FDD method relies on performing a singular value decomposition of the Hermitian response spectral density matrix.

$$[G_{yy}(\omega)] = [U][S][U]^H \quad (8)$$

where $[S]$ denotes a diagonal matrix containing the scalar singular values, $[U]$ represents a unitary matrix with the singular vectors, and $[U]^H$ is a Hermitian matrix.

Using vibration measurement data (acceleration) from the structure, we compute the spectral density matrix $[G_{yy}(\omega)]$ and apply singular value decomposition based on formula (8) to identify the natural frequencies of the structure.

The test involved capturing dynamic responses (acceleration) of steel beam structures at various nodes over time. The resulting vibration data was used to determine the natural frequencies of the structure. The physical parameters of the structure are detailed in Table 1. The testing equipment included the NIcDAQ-9137 and two accelerometers (PCB 352C68 and PCB 353B33). The accelerometers, used to measure beam vibrations (Fig. 1), were connected to the NIcDAQ-9137, which was also connected to a display (Fig. 1). Data from the accelerometers were collected and displayed using the NI Signal Express software. For the measurements, parameters were set up, and vibrations were induced in the structure with a sufficiently large stimulus to ensure it remained in the elastic range. The data collected included acceleration values recorded over time at the locations where the accelerometers were installed.

Following the vibration measurements of the structure, acceleration data at various nodes on the steel girder structure were collected over time. Examples of measurement data are illustrated in Fig. 2. Using the experimental acceleration data, the power spectral density was estimated employing Welch's method, and singular values were resolved using the SVD algorithm, as described in formula (8). The natural frequencies of the structure were identified based on the peaks in the power spectral density function. The identified natural frequencies are presented in Fig. 3.

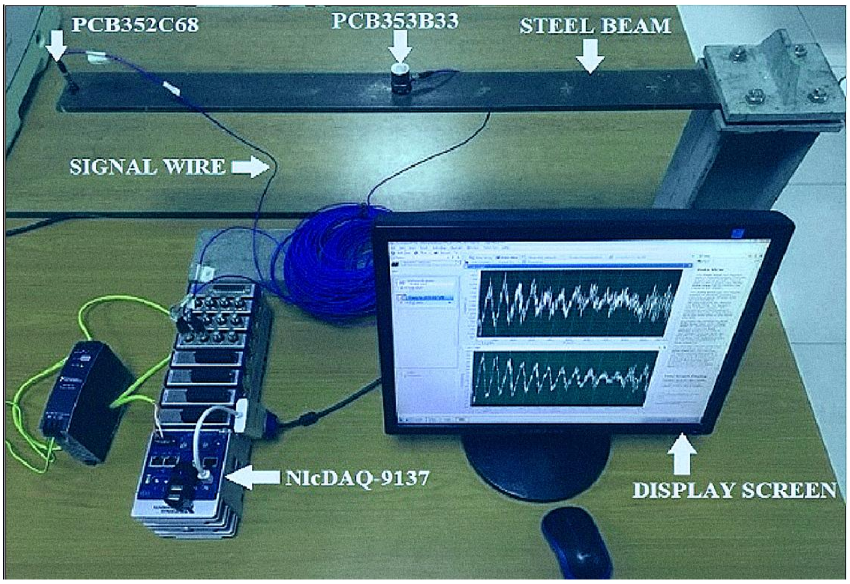


Fig. 1. Diagram of the test structure layout.

Table 1. Specifications of the steel beam structure

No.	Parameter	Value	Unit
1	Length	710	mm
2	Height	8	mm
3	Width	60	mm
4	Density weight	7850	kg/m^3
5	Modulus of elasticity	$2.03 \cdot 10^5$	MPa

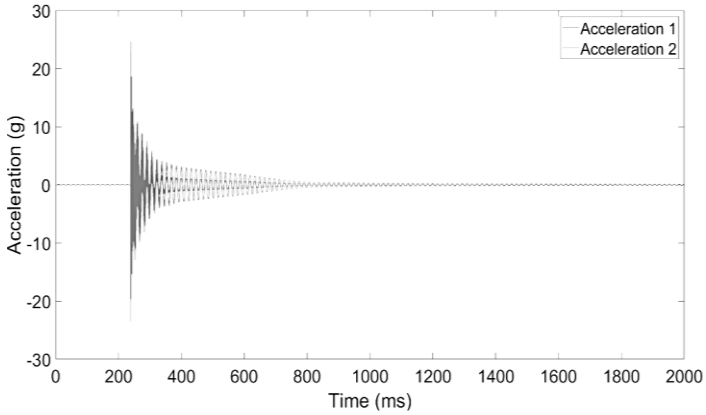


Fig. 2. Acceleration data of the beam.

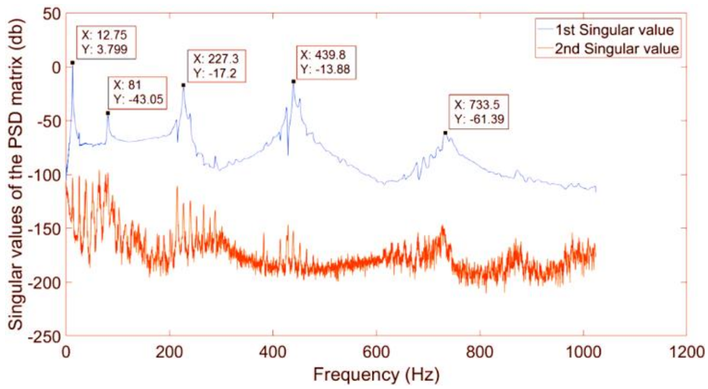


Fig. 3. Spectral power density (SPD).

The structure under examination consists of a steel beam, with its physical properties outlined in Table 1. Table 2 provides a comparison between the natural frequencies derived from the FDD method and their theoretical values. The results indicate that the discrepancies between the FDD method and the theoretical values are minimal.

Table 2. Comparison of natural frequencies obtained via the FDD method with their theoretical counterparts

No.	Mode	Theory [14] (Hz)	Frequency Domain Decomposition (Hz)	Error (%)
1	1	12.9	12.75	1.2
2	2	80.9	81.0	0.1
3	3	226.6	227.3	0.3
4	4	444	439.5	1.01
5	5	734	733.5	0.07

2.2. Determination of natural frequencies in the steel beam using the Finite Element Method

Simulating cracks in laboratory environments presents challenges, even with the use of the FEM. Although some researchers use saw-cuts to create cracks in experimental beams [15, 16], these methods often fail to fully replicate the characteristics of actual cracks. Unlike saw-cuts, real cracks do not alter the beam's mass but only affect its cross-sectional area, complicating accurate predictions of crack geometry. To align FEM simulations with experimental conditions, this study opts to model beam damage using saw-cuts.

A finite element model was developed in Abaqus utilizing three-dimensional elastic beam elements (Fig. 4). The cantilever beam was divided into 34,080 elements. The parameters of the cantilever beam are as shown in Table 1. The results from the model closely matched the measured responses, as detailed in Table 3.

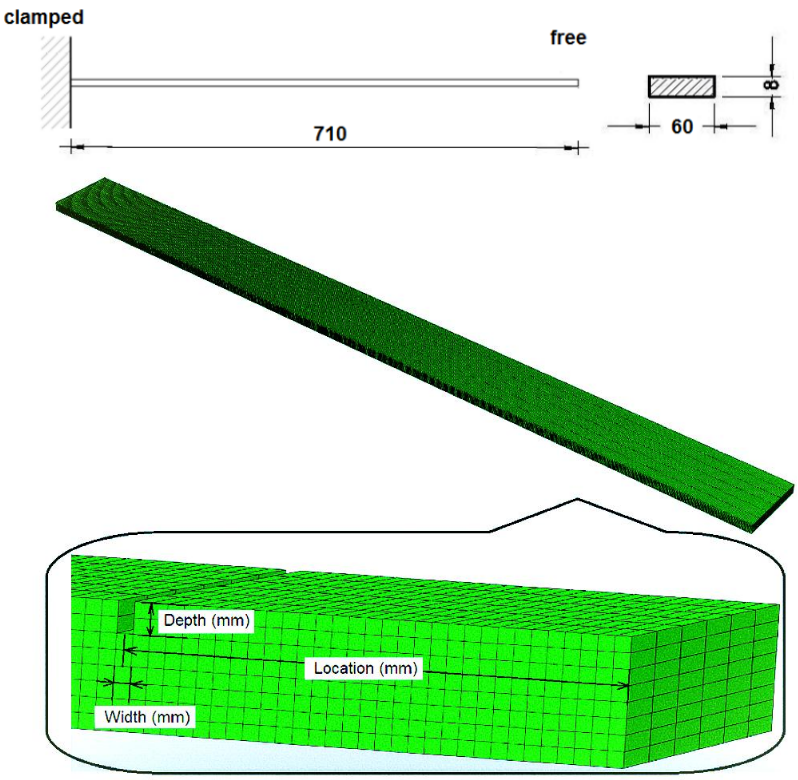


Fig. 4. Finite element model of the beam structure with saw-cut.

To simulate beam damage, elements representing the saw-cut were removed (Fig. 4). A total of 243 damage scenarios were considered, with 214 scenarios (from No. 1 to No. 214) utilized as training data for developing machine learning models, while the remaining 29 scenarios served as testing data. The data for the test set accounted for about 12% of the total dataset and was randomly selected, ensuring no overlap with the training dataset. Notably, its location was distributed across the entire length of the beam. Table 4 presents the damage scenarios and their corresponding natural frequencies, generated through FEM. Table 5 outlines the ranges of variables in the dataset, which are crucial for predictions as they define the boundaries for the models.

Building on the concept of using the FEM method to create the training dataset and the FDD method for construction monitoring, it is essential to minimize the deviation between the frequency results obtained from these two methods. Table 3 demonstrates that the deviation between the results of the two methods is minimal.

Table 3. The physical properties of the test structure

No.	Frequency Domain Decomposition (Hz)	Finite Element Method (Hz)	Error (%)
1	12.75	12.99	1.91
2	81	81.4	0.50
3	227.3	227.83	0.23

Table 4. Database developed by FEM

No.	Location (mm)	Width (mm)	Depth (mm)	Natural frequencies		
				Mode 1 (Hz)	Mode 2 (Hz)	Mode 3 (Hz)
1	10.5	1	1	13.003	81.429	227.89
2	20.5	1	1	13.003	81.426	227.87
3	30.5	1	1	13.003	81.422	227.86
4	40.5	1	1	13.003	81.419	227.84
5	50.5	1	1	13.002	81.416	227.82
…	…	…	…	…	…	…
72	10.5	1	2	13.007	81.454	227.95
73	20.5	1	2	13.007	81.447	227.92
74	30.5	1	2	13.007	81.44	227.89
75	40.5	1	2	13.006	81.433	227.84
76	50.5	1	2	13.006	81.426	227.79
…	…	…	…	…	…	…
143	10	2	1	13.008	81.455	227.95
144	20	2	1	13.007	81.447	227.92
145	30	2	1	13.007	81.441	227.89
146	40	2	1	13.006	81.435	227.86
147	50	2	1	13.006	81.428	227.83
…	…	…	…	…	…	…
239	133.5	1	2	13.003	81.298	226.57
240	133	2	1	13.003	81.365	227.36
241	553.5	1	1	12.965	81.408	227.59
242	553.5	1	2	12.888	81.411	227
243	553	2	1	12.957	81.411	227.55

Table 5. Ranges of variables in the database

No.	Variable	Unit	Count	Min	Max	Type of data
1	Location	mm	243	0	710.5	All of data
2	width	mm	243	0	2	
3	depth	mm	243	0	2	
4	Mode 1	Hz	243	12.783	13.008	
5	Mode 2	Hz	243	80.19	81.455	
6	Mode 3	Hz	243	222.97	227.95	
7	Location	mm	219	0	710.5	Training data
8	width	mm	219	0	2	
9	depth	mm	219	0	2	
10	Mode 1	Hz	219	12.783	13.008	
11	Mode 2	Hz	219	80.19	81.455	
12	Mode 3	Hz	219	224.76	227.95	
13	Location	mm	29	75	674.5	Testing data
14	width	mm	29	1	2	
15	depth	mm	29	1	2	
16	Mode 1	Hz	29	12.804	13.005	
17	Mode 2	Hz	29	80.594	81.413	
18	Mode 3	Hz	29	222.97	227.82	

3. Model development

In this study, the RF model was developed using Python version 3.11.0 along with the Sklearn library. The input variables for the model included the natural frequencies from three different models, while the output variables were the location, width, and depth of the saw-cut.

3.1. Data segmentation and preparation

The dataset was divided into two parts: one for training the model and another for testing. Ranges of variables in the database are shown in Table 5. To maintain consistent focus on all variables during training, preprocessing involved normalizing both input and output variables to a range between 0.0 and 1.0. The scaled value for each variable was computed as follows:

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

where x_{max} and x_{min} denote the maximum and minimum values of the variable x.

3.2. Random forest

The RF model, a widely used machine learning technique, relies on decision trees and is renowned for its effectiveness in handling complex data analysis tasks. Introduced by Breiman [17]. Fig. 5 and Fig. 6 illustrate the unique structures of decision trees and the RF model, respectively. The decision tree model organizes data into a tree-like format, where branches divide different input instances (e.g., x_1, x_2, x_3) and output values are located at the leaf nodes (e.g., R_1, R_2, R_3). This model uses if-then-else rules, which are adjusted based on variables like the input data, the complexity of the tree, and the depth of these rules. However, standalone decision trees are prone to overfitting, which has led to the creation of more sophisticated models like the RF model.

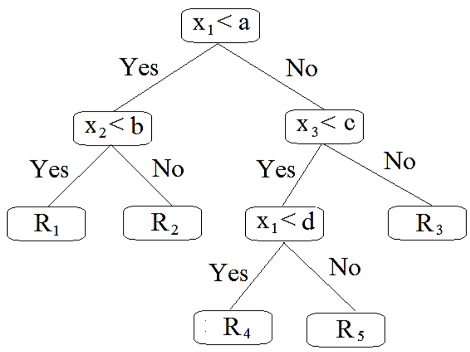


Fig. 5. Regression analysis using decision tree model.

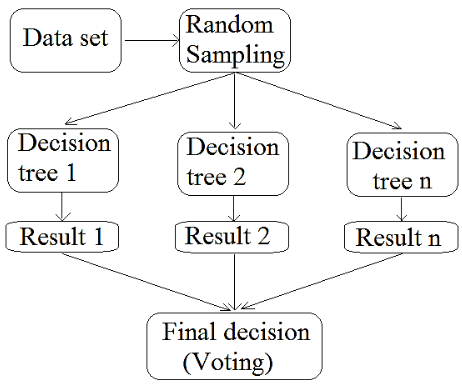


Fig. 6. Visual representation of the random forest model.

The RF model is composed of a collection of decision trees, each developed for training and prediction purposes. Each tree operates independently, using a different subset of the data. The model's final prediction is derived from averaging the outputs of all the individual trees. One of the strengths of this model is its use of the bootstrap technique to randomly construct the trees, ensuring that each tree is built from different subsets of the data. Bootstrap is a statistical technique used to create multiple random subsets of data by sampling with replacement from the original dataset. This method enhances the model's ability to generalize and reduces the risk of overfitting, a common issue with single decision trees.

Developing a robust model requires careful selection of key hyperparameters. These include the number of trees in the forest (*t*), the maximum depth allowed for each tree (*D*), the minimum number of data samples required to split a node (*S*), and the minimum number of samples needed for each leaf node (*L*). While increasing the number of trees (*t*) enhances the model's robustness, it also raises computational demands. Adjusting the values of *D, S,* and *L* is crucial for balancing the model's complexity, computational efficiency, and generalization performance. The final prediction provided by the model is described as follows:

$$y_i = \frac{1}{t} \cdot \sum_{j=1}^{t} f_j(x_i) \tag{10}$$

where t denotes the total number of trees, y_i represents the predicted value for the i_{th} sample, x_i refers to the input vector for the i_{th} sample, and f_j signifies the j_{th} estimator within the forest.

For an in-depth exploration of decision tree construction and the hyperparameters associated with the RF model, refer to Breiman [17].

To achieve optimal performance, an extensive analysis was carried out on key hyperparameters affecting the model's results. This included parameters such as *D, S, L*, and *t*, with the goal of determining the most effective values within their respective ranges.

Table 6 presents a summary of the parameter ranges and values examined in this study. Prior research by Breiman [17] highlights the critical role of hyperparameters such as *D*, *S*, and *L* in shaping the complexity of decision trees and reducing overfitting. Finding the right balance among these parameters is essential, as overly complex models may perform well on training data but may not generalize effectively to new, unseen data. The chosen hyperparameter ranges are intended to limit significant deviations outside of this range, thus minimizing potential negative impacts on the model's performance. This approach helps avoid data leakage, where the model becomes too specialized to the training data and struggles with new, unseen data.

Many factors affect the performance of the RF model; however, two factors - data noise and training data size - have a significant influence. Data Noise: Irrelevant features or errors reduce accuracy. This can be mitigated by cleaning the data and increasing the tree count. Training Data Size: Small datasets lead to underfitting, while large ones increase computational costs. To address this issue, the model's performance was assessed using a 5-fold cross-validation approach, and the accuracy of the predictions was analyzed with respect to variations in hyperparameters.

Table 6. Scope of random forest model hyperparameters

Hyperparameter	**Range**	**Explain**
t	2-100	Number of trees
D	None and 2-20	Max depth
S	2-20	Minimum number for separating trees
L	1-19	Minimum number per leaf

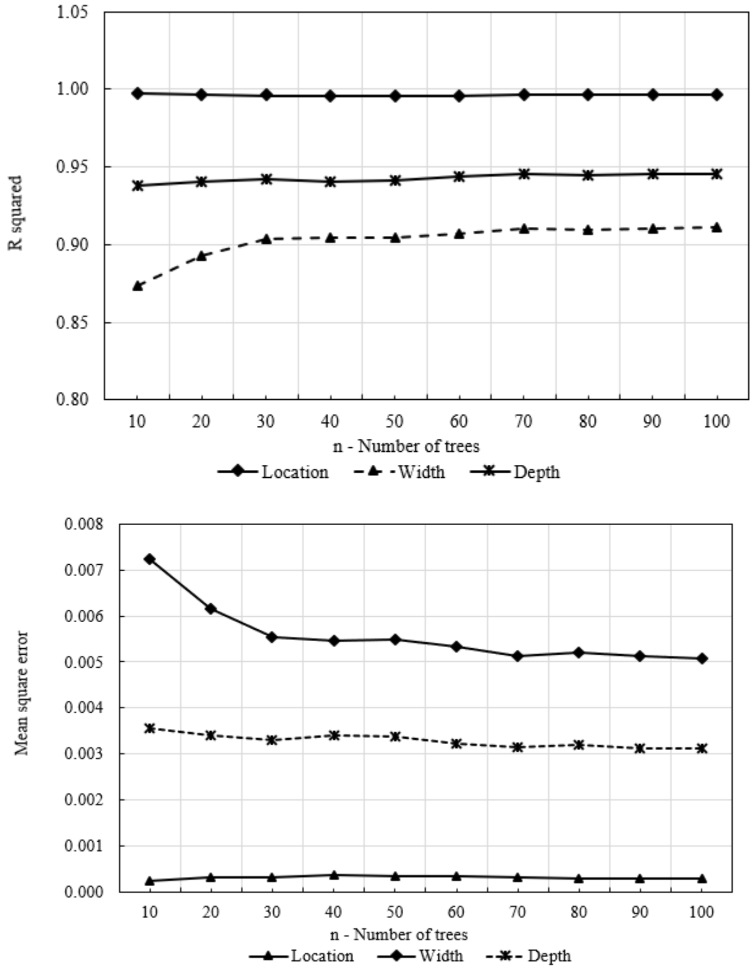


Fig. 7. Impact of tree number on model performance in the training data.

Figure 7 illustrates the effect of tree number on the RF model's accuracy, showing that increasing the number of trees improves performance. Figure 8 displays the results of an evaluation of model accuracy with varying maximum tree depths (*D*) from 2 to 20, including an option for '*None*'. With the number of trees fixed at 100, the results indicate that the best performance is achieved when *D* is set to '*None*'.

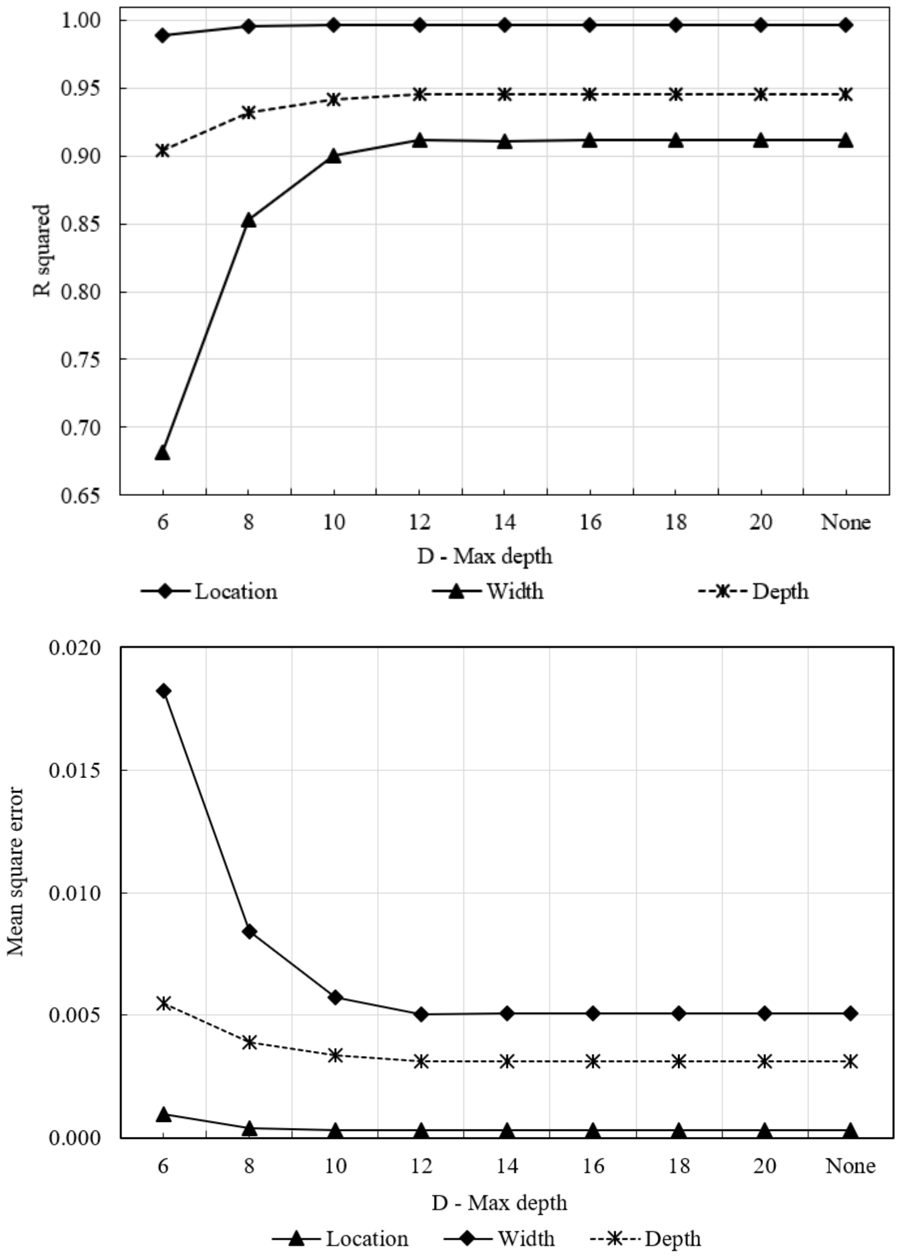


Fig. 8. Impact of max depth on model performance in the training data.

Figure 9 displays the outcomes of an assessment of the model's performance with varying minimum sample sizes needed to split trees (S), ranging from 2 to 20. With the number of trees fixed at 100 and 'None' as the maximum tree depth, the results show that increasing S reduces the model's predictive accuracy. The optimal value for S is found to be 2.

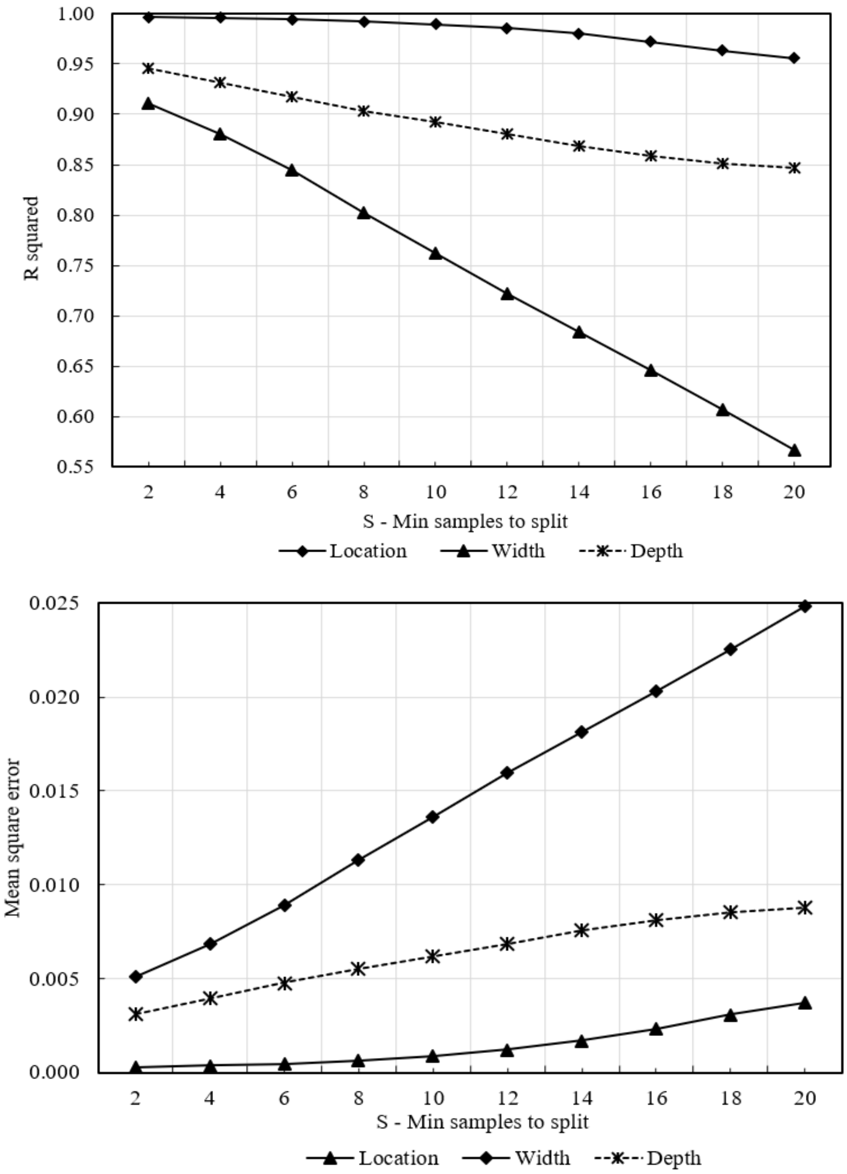
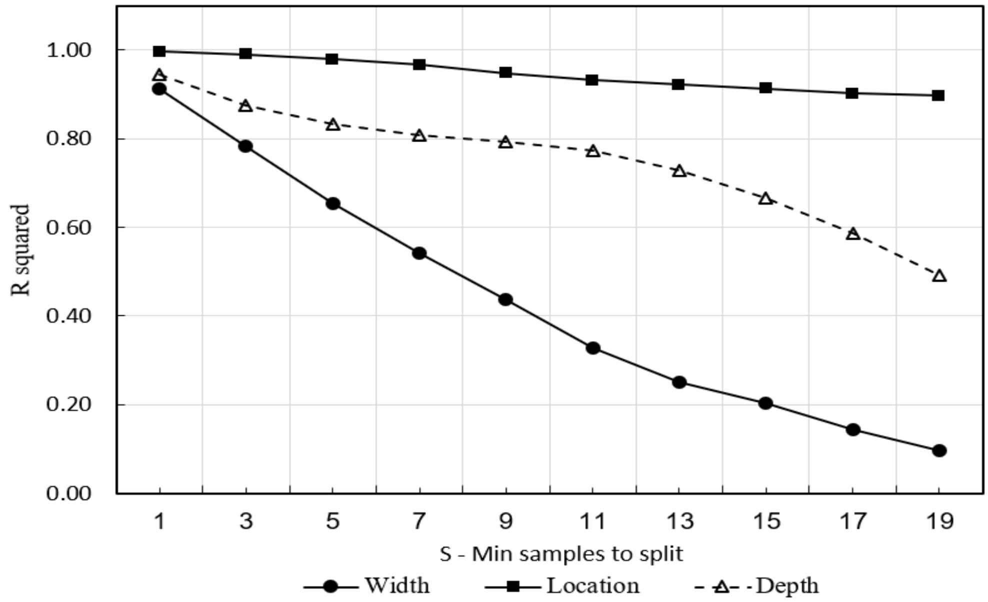


Fig. 9. Impact of min samples to split on model performance in the training data.

Figure 10 illustrates the results of an evaluation of the model's accuracy with varying minimum sample sizes per leaf (*L*), ranging from 1 to 19. The data show that smaller values of *L* yield better predictive performance, with the optimal value being 1.

The results indicate that the most effective setup for the RF model involves using 100 trees, a minimum of 2 samples to split nodes, an unlimited tree depth ("*None*"), and a minimum of 1 sample per leaf. This configuration demonstrated the best accuracy and will be used for subsequent validation and verification of the model.



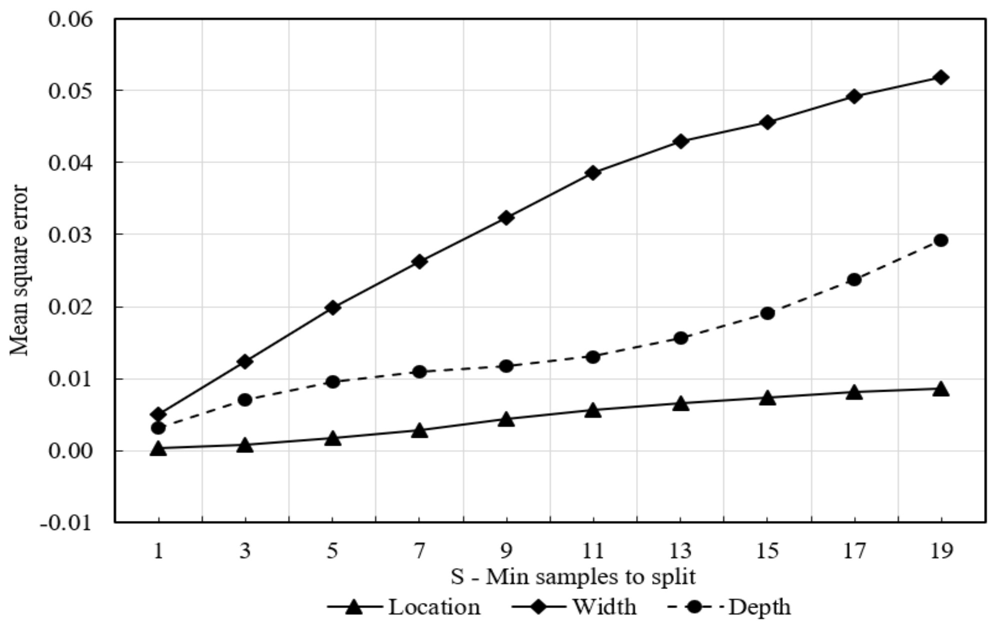


Fig. 10. Impact of minimum leaf samples on model performance in the training data.

4. Results and discussion

After training, the RF model is validated with the test dataset. The testing results showed an R-squared value of 0.996 for location, 0.876 for width, and 0.880 for depth. The mean squared error (MSE) was found to be 0.0003 for location, 0.0313 for width, and 0.0420 for depth. These results reflect the model's performance on completely randomized and new test data. When both R-squared and MSE are near their ideal values (1 and 0, respectively), the model performs well in capturing the relationship between the variables and making precise predictions. This shows that the selected model is quite accurate.

Figure 11-13 display the RF model's performance in predicting the location, width, and depth of the saw-cut, respectively. The model shows the greatest accuracy in forecasting the saw-cut's location, with predicted values closely aligning with the best-fit line. This high accuracy is likely attributed to the more extensive and detailed training data available for predicting the location, compared to the other parameters, which have only two values in the training data (1 mm and 2 mm).

Table 7 and Table 8 display the accuracy of predictions for the saw-cut's location, width, and depth. The location prediction shows the smallest deviation from the actual measurements, with a maximum difference of 9.258%. In comparison, the width and depth predictions have maximum deviations of 42% and 21%, respectively. This observation is consistent with the R-squared values: the testing set yields R-squared values of 0.976 for width and 0.880 for depth, both of which are lower than the value for location prediction. Despite this, all predictions remain highly accurate, with R-squared values above 0.87. Thus, the RF model is effective in accurately predicting the position, width, and depth of the saw-cut.

Figure 11-13 depict the comparison between predicted and actual values for the saw-cut's location, width, and depth using the RF model. It is apparent that the RF model generally underestimated the saw-cut location. For the beam cut width and depth, predictions were often overestimated when the actual measurement was 1 mm and underestimated when it was 2 mm. Consequently, in practice, if predictions from all three models are available, it is advisable to select larger predicted values for the location. For width and depth, if the predicted values are under 2 mm, opt for lower values, whereas higher values should be chosen if the predictions exceed 2 mm.

Table 7. Statistical performance metrics for training and test datasets

Phase	Location		Width		Depth	
	R-squared	***MSE***	***R-squared***	***MSE***	***R-squared***	***MSE***
Training	0.996	0.0003	0.911	0.0051	0.946	0.0031
Testing	0.996	56.78	0.876	0.0313	0.880	0.0420

Table 8. Variation in test data

Absolute deviation (%)	Location	Width	Depth
Maximum	9.258	42	21
Minimum	0.018	0	0

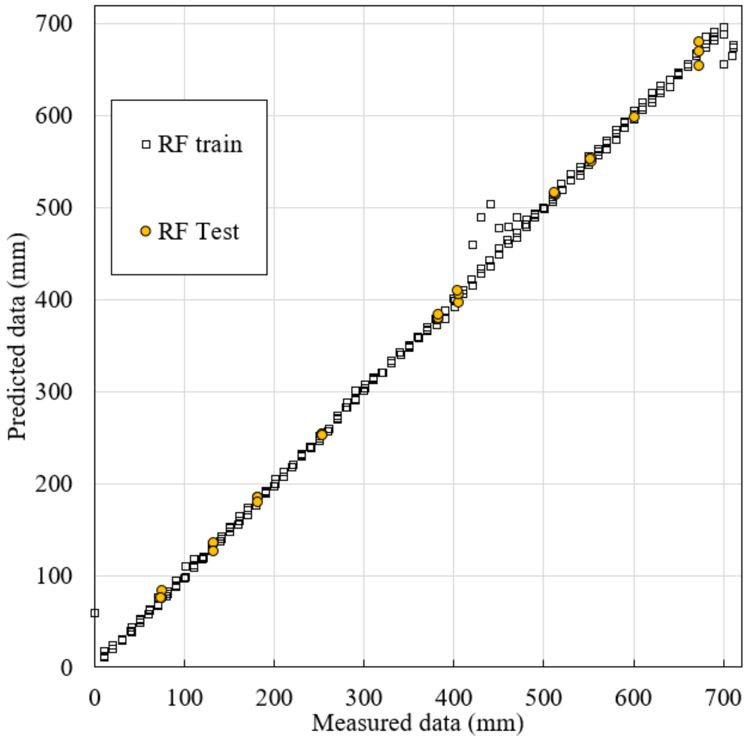


Fig. 11. Comparison of predicted vs. measured location values.

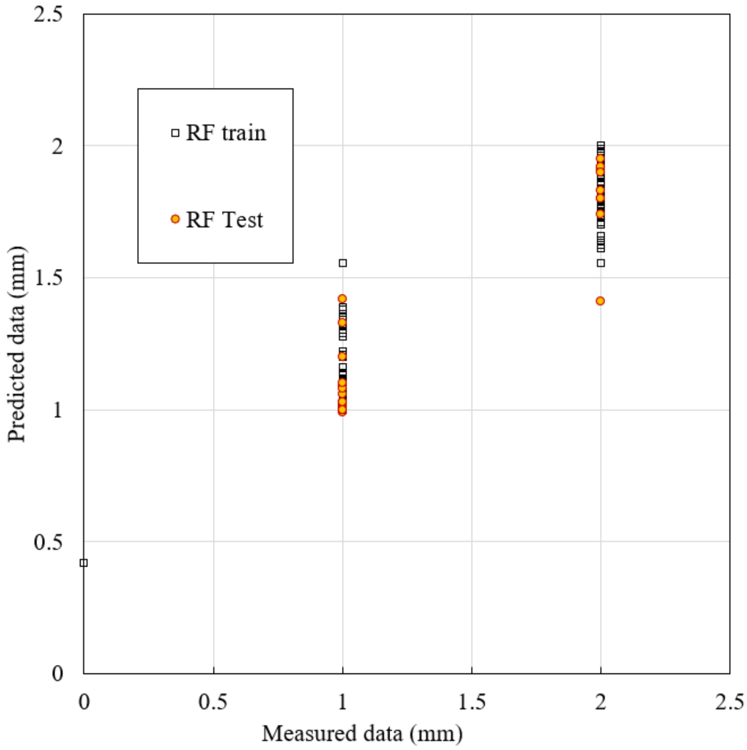


Fig. 12. Comparison of predicted vs. measured width values.

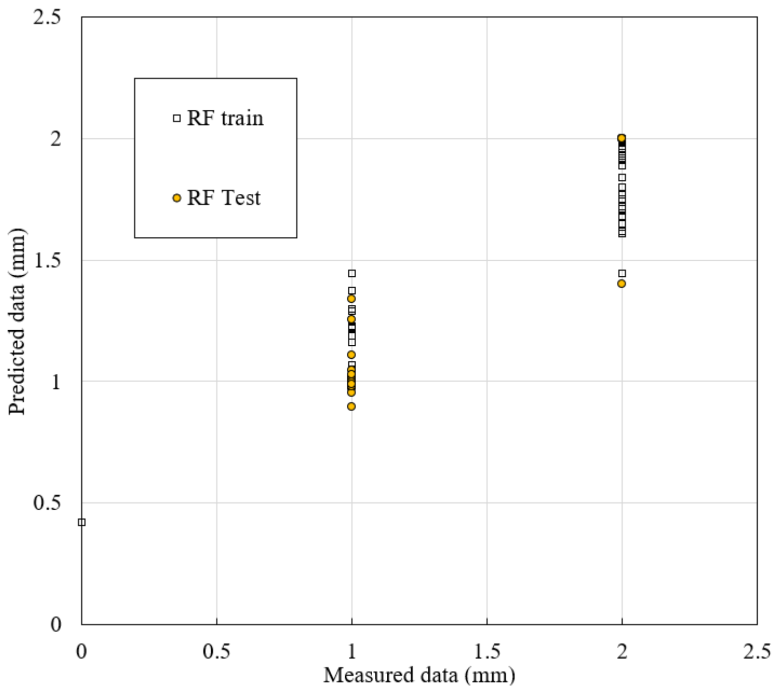


Fig. 13. Comparison of predicted vs. measured depth values.

5. Conclusions

Following a comparison of natural frequencies obtained from the FDD and FEM methods, the use of FEM to produce data for RF models, and the application of the RF model to estimate the saw-cut's location, width, and depth in steel beams based on these frequencies, the key findings of this study are summarized as follows:

- The natural frequencies identified using the FDD method were found to align closely with those obtained from the FEM method. This consistency indicates that these frequencies can be reliably utilized as inputs for the RF model to detect saw-cut damage over time. The FEM method is particularly useful for creating an initial training dataset when monitoring data is sparse. The combination of FEM and FDD methods with the RF model is highly valuable for structural health monitoring. For optimal results, it is essential that the training data generated by FEM includes a comprehensive range of possible damage scenarios.

- The RF model effectively predicted the location, width, and depth of the saw-cut in the beam using natural frequencies. The model tended to underestimate the saw-cut locations compared to the actual measurements. For the beam’s cut width and depth, the predictions were generally higher for actual measurements of 1 mm and lower for actual measurements of 2 mm. Consequently, when using predictions from all three models, opt for the higher predicted values for location. For width and depth, select lower values if predictions are below 2 mm and higher values if they exceed 2 mm.

- The combination of the RF algorithm with FEM and FDD provides an effective tool for detecting and assessing damage in steel beams, helping to reduce risks and optimize maintenance costs. However, a limitation of this approach is the discrepancy between the frequency data obtained from the FDD and FEM methods. Addressing this issue would enhance the accuracy of health monitoring. Additionally, this method can be applied in scenarios with limited experimental data and extended to other types of structures in the future.

References

[1] X. F. Yang, A. S. J. Swamidas, and R. Seshadri, "Crack identification in vibrating beams using the energy method", *Journal of Sound and Vibration*, Vol. 244, Iss. 2, pp. 339-357, 2001. DOI: 10.1006/jsvi.2000.3498

[2] A. S. J. Swamidas, X. Yang, and R. Seshadri, "Identification of cracking in beam structures using Timoshenko and Euler formulations", *Journal of Engineering Mechanics*, Vol. 130, Iss. 11, pp. 1297-1308, 2004. DOI: 10.1061/(ASCE)0733-9399(2004)130:11(1297)

[3] G. R. Gillich, Z. I. Praisach, M. A. Wahab *et al.*, "Free vibration of a perfectly clamped-free beam with stepwise eccentric distributed masses", *Shock and Vibration,* Vol. 2016, 2016. DOI: 10.1155/2016/2086274

[4] Y. L. Zhou, N. M. Maia, R. P. Sampaio, and M. A. Wahab, "Structural damage detection using transmissibility together with hierarchical clustering analysis and similarity measure", *Structural Health Monitoring,* Vol. 16, Iss. 6, pp. 711-731, 2017. DOI: 10.1177/1475921716680849

[5] G. R. Gillich, H. Furdui, M. A. Wahab, and Z. I. Korka, "A robust damage detection method based on multi-modal analysis in variable temperature conditions", *Mechanical Systems and Signal Processing,* Vol. 115, pp. 361-379, 2019. DOI: 10.1016/j.ymssp.2018.05.037

[6] Y. L. Zhou, N. M. Maia, and M. A. Wahab, "Damage detection using transmissibility compressed by principal component analysis enhanced with distance measure", *Journal of Vibration and Control*, Vol. 24, No. 10, pp. 2001-2019, 2018. DOI: 10.1177/1077546316674544

[7] K. Samir, B. Idir, R. Serra, B. Brahim, and A. Aicha, "Genetic algorithm based objective functions comparative study for damage detection and localization in beam structures", in *Journal of Physics: Conference Series*, 2015, Vol. 628, No. 1, IOP Publishing, 012035. DOI: 10.1088/1742-6596/628/1/012035

[8] S. Ghadimi and S. S. Kourehli, "Crack detection of structures using modified extreme learning machine (MELM)", *Inverse Problems in Science and Engineering,* Vol. 25, Iss. 7, pp. 995-1013, 2017. DOI: 10.1080/17415977.2016.1212026

[9] N. Gillich, C. Tufisi, C. Sacarea *et al.*, "Beam damage assessment using natural frequency shift and machine learning", *Sensors,* Vol. 22, No. 3, 2022, 1118. DOI: 10.3390/s22031118

[10] T. C. Le, T. N. Nguyen, S. Khatir *et al.*, "An efficient approach for damage identification based on improved machine learning using PSO-SVM", *Engineering with Computers,* Vol. 38, pp. 3069-3084, 2022. DOI: 10.1007/s00366-021-01299-6

[11] V. T. Rathod, S. Mukherjee, and Y. Deng, "Machine learning enabled damage classification in composite laminated beams using mode conversion quantification", in *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIV*, 2020, Vol. 11380: SPIE, pp. 36-52.

[12] A. A. S. R. de Sousa, J. da S. Coelho, M. R. Machado, and M. Dutkiewicz, "Multiclass supervised machine learning algorithms applied to damage and assessment using beam dynamic response", *Journal of Vibration Engineering & Technologies,* Vol. 11, pp. 2709-2731, 2023. DOI: 10.1007/s42417-023-01072-7

[13] R. Brincker, L. Zhang, and P. Andersen, "Modal identification of output-only systems using frequency domain decomposition", *Smart Materials and Structures,* Vol. 10, No. 3, p. 441, 2001. DOI: 10.1088/0964-1726/10/3/303

[14] W. F. Stokey, "Vibration of systems having distributed mass and elasticity", *Shock and Vibration Handbook*, pp. 7-14, 2002.

[15] M. Seguini, N. Djamel, B. Djilali, S. Khatir, and M. A. Wahab, "Crack prediction in beam-like structure using ANN based on frequency analysis", *Frattura ed Integrità Strutturale*, Vol. 16, No. 59, pp. 18-34, 2022. DOI: 10.3221/IGF-ESIS.59.02

[16] M. Seguini, S. Khatir, D. Boutchicha, D. Nedjar, and M. A. Wahab, "Crack prediction in pipeline using ANN-PSO based on numerical and experimental modal analysis", *Smart Structures and Systems*, Vol. 27, No. 3, pp. 507-523, 2021. DOI: 10.12989/sss.2021.27.3.507

[17] L. Breiman, "Random forests", *Machine Learning*, Vol. 45, pp. 5-32, 2001.

PHÁT HIỆN HƯ HỎNG TRONG DẦM THÉP THÔNG QUA TẦN SỐ TỰ NHIÊN BẰNG MÔ HÌNH RỪNG NGẪU NHIÊN

Vũ Văn Tuấn[1], Đặng Anh Dũng[1], Lâm Hải Đăng[1], Nguyễn Đức Thịnh[1], Trần Trung Đức[1]

[1]*Viện Kỹ thuật công trình đặc biệt, Trường Đại học Kỹ thuật Lê Quý Đôn*

Tóm tắt: Gần đây, các thuật toán học máy (ML) đã được chứng minh là công cụ có hiệu quả cao để dự đoán hư hỏng kết cấu. Tuy nhiên, dữ liệu được sử dụng trong quan trắc tình trạng kết cấu thường trong các điều kiện vận hành bình thường hoặc những sai lệch nhỏ so với trạng thái ban đầu và rất ít dữ liệu thu thập được trong điều kiện nguy hiểm. Các dữ liệu này nếu có thì cũng sẽ liên quan đến các tham số như phạm vi hệ số cường độ ứng suất và tỉ lệ ứng suất, đây là các tham số rất khó đo lường trong kết cấu thực. Hạn chế này khiến việc tạo bộ dữ liệu thực tế để huấn luyện các mô hình ML nhằm phát hiện hư hỏng kết cấu trở nên khó khăn. Trong nghiên cứu này, mô hình Rừng ngẫu nhiên (RF) đã được phát triển để dự đoán vị trí, chiều rộng và độ sâu của vết cắt trên dầm thép dựa trên sự thay đổi tần số tự nhiên. Các tần số tự nhiên này trong các kịch bản thiệt hại khác nhau được xác định bằng phương pháp phần tử hữu hạn (FEM). Tần số tự nhiên của kết cấu ở trạng thái không bị hư hại sẽ được so sánh giữa hai phương pháp (phương pháp phần tử hữu hạn và phân rã miền tần số) để xác định tính chính xác của mô hình FEM. Sau khi huấn luyện, mô hình RF cho kết quả R-squared (trên tập kiểm tra) của vị trí là 0,996; 0,876 với chiều rộng và 0,88 với chiều sâu vết nứt. Tham số đo độ chính xác MSE lần lượt là 0,0003 đối với vị trí, 0,0313 với chiều rộng và 0,042 với chiều sâu. Kết quả chỉ ra rằng việc kết hợp FEM và phân rã miền tần số (FDD) với mô hình RF là rất có tiềm năng trong việc giám sát sức khỏe công trình.

Từ khóa: *Dự đoán vết cắt; rừng ngẫu nhiên; tần số tự nhiên; phân rã miền tần số; phân tích động FEM.*

Received: 30/09/2024; Revised: 23/12/2024; Accepted for publication: 27/12/2024

❑