Hindawi
Structural Control and Health Monitoring
Volume 2023, Article ID 6321012, 20 pages
https://doi.org/10.1155/2023/6321012

Research Article

Time-Domain Structural Damage Identification Using Ensemble Bagged Trees and Evolutionary Optimization Algorithms

Seyed Hossein Mahdavi [1] **and Chao Xu** [2]

[1]*Department of Civil Engineering, Higher Education Complex of Bam, Bam 766138-4744, Iran*
[2]*School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China*

Correspondence should be addressed to Seyed Hossein Mahdavi; sh.mahdavi@bam.ac.ir

Received 23 September 2022; Revised 24 January 2023; Accepted 7 March 2023; Published 15 March 2023

Academic Editor: Jun Li

This paper presents a two-step vibration-based strategy for damage identification of framed structures using ensemble bagged trees known as a well-known supervised machine learning (ML) paradigm in conjunction with evolutionary optimization algorithms. The proposed model incorporates the actual response, wavelet coefficients, and wavelet energy to extract damage-sensitive features from the time-domain of the measured and simulated signals. Unlike available studies in this scope, the key objective of this research is to identify damage with a localization precision down to a single structural member, rather than limiting the evaluation to the group of elements. In order to increase the training performance in contributing to extremely large datasets with numerous class labels, the proposed strategy involves the artificial generation of features. Additionally, a modified genetic algorithm is proposed for fast damage localization. It is deduced that the damage locations are confidently detected within a fast computational time. Subsequently, damage identification is followed by the application of evolutionary optimization algorithms. For comparison purpose, the employment of the water cycle optimization algorithm (WCA) is comparatively investigated with the other three state-of-the-art optimizers, i.e., particle swarm optimization (PSO), imperialist competitive algorithm (ICA), and differential evolution algorithm (DE). The numerical and experimental validation studies evidence satisfactorily reliable identification results with no false detection in dealing with multiple damage scenarios in large-scale and real-world applications. It is concluded that developing the most damage-sensitive features and using the proposed data fusion strategy lead to informative features with a reasonably small size and significantly improve the ML performance.

1. Introduction

Due to damages, the structural performance deteriorates during its service life. Basically, structures are damaged as a result of deterioration, fatigue, overloading, or environmental effects. Damage involves changes in structural material or geometric characteristics. In order to ensure a desirable safety, maintainability, and reliability of structures, it is essential to localize and determine the early stages of damage. In other words, for long-time integrity assessment and maintenance of structures and infrastructures, it is of crucial significance to confidently localize damages as soon as they occur in order to prevent progressive failures and increase safety levels. Over the past two decades, vibration-based structural damage quantification strategies have been rapidly expanded to a variety of structural engineering problems. In general, vibration-based damage identification procedures fall into two main categories. The former category involves time-domain approaches dealing with the abnormal detection of changes in the measured time-history of structural responses compared to those of simulation, while the later one lies on the frequency-domain schemes contributing to the abnormal detection of changes in modal parameters and structural frequencies [1]. For many reasons, the former category seems more reliable than the later one. For instance, the information obtained for higher structural modes is not sufficiently reliable in frequency-domain methods. Another shortcoming of the

frequency-domain procedures is that the modal parameters are calculated based on the linear superposition rule, thus they are limited only to linear systems. In addition, modal parameters are obtained from the governing eigenvalue problem taking into consideration the entire stiffness and mass matrix of the structure, therefore they are not very sensitive to the small changes (damages) in stiffness of structural members. In contrast, time-domain approaches do not require the measurement of modal parameters and frequencies, and instead they directly utilize time-history information composed of all structurally modelled modes [2].

The inverse problem governing structural damage detection has been evaluated with different optimization strategies [3–5]. With the increasing dissimilarity and complexity of data acquired from various sensors, for the assessment of structural condition and the diagnosis of structural damage, an extremely large amount of computational time is required for the use of optimization-based algorithms [6]. Thus, one of the underlying drawbacks of such strategies is revealed, so that they are mostly impractical for online or near real-time condition assessment. In contrast, with the astronomical increase in computer processing power, the broad branch of machine learning algorithms (ML) is utilized for structural damage detection [7, 8]. Despite the fact that the simulation of various damage extents is mostly a computationally intensive task, the improved ML classifiers may be effectively employed and will result in a well-trained model. Once a well-trained model is prepared, structural damage detection can be simply achieved with less computational cost whenever the condition assessment takes place.

Several research streams have been pursued to contribute to time-domain and frequency-domain structural damage detection using the ML paradigm. For instance, the acceleration time-histories are utilized to train an artificial neural network (ANN) for reduction detection of steel members in a bridge [9]. Analogously, an unsupervised scheme is improved to directly process raw sensor data in real-time in order to elaborate abnormal changes and structural novelties [10]. A frequency-domain and unsupervised deep Boltzmann learning algorithm is developed for condition assessment of high-rise buildings based on ambient vibration data [11]. The feature selection is delivered by the implementation of synchrosqueezed wavelet transforms and fast Fourier transforms. Moreover, the applicability of a decision tree ensemble for frequency-domain damage identification problems is investigated [12]. It is observed that the accuracy of the decision tree ensemble as a well-known ML classifier is far better than the other existing learning schemes. This leads to the simulation of a huge dataset originated from frequency-domain features regarding different damage states of the structures.

Committee-based or ensemble algorithms considerably enhance the performance of multiclass ML problems, in which the outcome from multiple trainers is efficiently inferred to gain the best model predictions. Li et al. [13] developed an element-level damage identification system using random forest, known as an ensemble-based ML. The acceleration responses from the measured degrees-of-freedom (DOFs) are concatenated, and principle component analysis (PCA) is implemented to reduce the uncorrelated feature size. Besides, they compared the results with neural network training and deep learning models. From the computational cost point of view, their proposed approach competes with other available ensemble-based classifiers. Moreover, PCA is widely implemented for the aim of data compression and feature size reduction. Practically, it has been more popular to reduce the dimension of time-domain features by taking into account the effect of uncertainties and measurement noise, especially in cases where a long duration of responses is measured [14]. A review of the available literature in this context reveals the popularity of the random forest in dealing with damage detection problems [15–17]. However, the conceptual idea and the key algorithm vary from different ensemble classifiers, and therefore the random forest is more likely to be interpreted as a class of boosted tree classifiers rather than bagged trees [18–20]. Furthermore, structural vibration characteristics have been evaluated through deep learning approaches [21, 22] in constructing efficient damage detection frameworks. For instance, modal parameters, i.e., mode shapes and natural frequencies, have been extracted to constitute the core of a damage quantification strategy in parallel with a deep learning strategy [23]. However, referring back to the main drawback of frequency-domain schemes, one may conclude that more DOF numbers (denoting more sensor locations) are to be measured in order to achieve a precise modal feature extraction.

In addition, a thorough review of the literature discloses that a promising strategy of assessing the safety of framed structures is the application of vibration-based monitoring systems. Taking into account the complexity in structural geometry of large-scale framed structures, these allow observation of the global response for a structure, including damage detection at the local level, classification, and progressive development. In fact, dynamic monitoring systems have proven to be particularly suited for systems whose structural behaviors are strongly influenced by their geometric complexity or the inhomogeneity of their constituent materials. Moreover, because of its nondestructive and noninvasive nature, vibration-based monitoring can be safely applied to damaged structures, which are potentially dangerous under other test conditions. Therefore, the analysis of the time-domain responses can expose local damages or deficiencies amplified or induced by unforeseen events. In other words, the monitoring of a set of appropriately chosen features jointly with identification of the local and global structural weaknesses may reveal the effectiveness of any progression of structural damage. On the other hand, wavelet analysis has attracted tremendous attention in a broad scope of structural health monitoring (SHM) problems. Keeping in mind the significant importance of feature extraction as the core of ML-based methods, the application of wavelet multiresolution analysis demonstrates the appropriateness of this powerful technique in dealing with damage detection problems [24, 25]. In fact, it is deduced that the essence of wavelet analysis lies in time-

frequency-scale analysis, and it operates in the time domain without losing the informative features of frequency contents. This characteristic of wavelet operators provides a comprehensive feature extraction approach towards reaching the most damage-sensitive features with the least environmental noise effect [26, 27]. For instance, a two-step approach is proposed for time-domain input load identification in framed structures using wavelet-based operators to evaluate the third derivation of displacement time-histories as the objective function for metaheuristic optimization strategies. It is observed that the optimization algorithm utilizes the model-based data for impact load localization and then suffers from high computational costs [26]. As a consequence, implementation of a data-driven technique using a powerful machine learning-based classifier for damage localization in large-scale and real-world structures will provide promising results with extremely low computational time involved.

Despite the extensive efforts conveyed for frequency-domain structural damage detection, this paper proposes an efficient strategy for time-domain damage localization and identification using an ensemble bagged tree classifier, considering the structural responses for efficient feature extraction. In that regard, a novel fitness evaluation and data fusion approach are presented based on wavelet operations to extract the wavelet coefficients of the responses as well as response wavelet energies at high detail levels results in the most damage-sensitive features. Unlike available reports, the key finding of this research lies in the identification of damage with a localization precision down to a single structural member, rather than limiting the evaluation to the group of elements. Notwithstanding the sparse literature available for time-domain damage detection in large-scale problems and to be more practical in fast damage localization of large-scaled structures, the idea of generating artificial features based on the structural subdivisions is proposed in order to achieve high-performance data training through machine learning paradigms. Then, the modification of a synthesis optimization-based algorithm originating from the application of a multispecies genetic algorithm for fast damage localization with the resulting accuracy down to a single element level and the water cycle optimization algorithm (WCA) for reliable damage severity identification are presented. Subsequently, the numerical and experimental validity and applicability of the proposed strategy are evaluated.

2. Bagged Trees as the Decision Tree Ensemble

In general, a decision tree is a non-parametric supervised machine learning (ML) strategy. It can be employed to predict the class of output response by learning simple rules obtained for diverse predictors (features). The privilege of decision tree-based classifiers over other classification paradigms in ML is their robustness in dealing with large data sets for the classification of binary as well as multiclass problems. In addition, ensemble methods aim at modifying the predictive performance of given statistical learning or model fitting techniques [27]. The general principle of ensemble approaches is to model a linear combination of some classification methods instead of the implementation of a single fit of the method. Ensemble strategies have become popular since highly accurate classifiers could be obtained by combining the efforts of less accurate learners. As a group of learners is appropriate at some specific tasks, their cooperation will complement each other, resulting in a better performance. The original ensemble method relied on the Bayesian averaging scheme, however, more recent strategies comprise error-correcting output coding, bagging, and boosting algorithms. In other words, bagging and boosting are schemes that generate a diverse ensemble of classifiers by manipulating the training data given to a base learning algorithm. Bagging methods turn out to be variance reduction schemes, whereas boosting methods are primarily reducing the model bias of the base classifier [18, 19]. Some details on the description of such ensemble approaches in the much broader context of other modern statistical strategies may be found in Ref. [27].

It is to be noted that random forest, known as a supervised ML classifier which uses homogenous ensemble techniques, is a very different ensemble method than bagging or boosting. However, from the perspective of prediction, the random forest strategy is about as good as the boosting method [21]. The comparison of different ensemble methods of decision trees is provided in Table 1 [28]. In the present study, due to the essence of the considered problem in dealing with large-scaled structures (known as a multi-label classification problem), the bagged tree classifier is utilized as the family of decision tree ensemble approaches. In fact, each decision tree in the ensemble bagged tree classifier is trained on a different dataset with a replacement from the actual dataset. This strategy is called bagging or bootstrapping. Bagging combines the predictions of various models while their predictions are not correlated with each other. Each tree randomly chooses a set of features from the whole feature set in the training model [13]. Furthermore, in order to prevent overfitting, a crossvalidation strategy is utilized and the dataset is partitioned into k folds (k takes 10) and the accuracy of the trained model is estimated on each fold. The maximum number of splits is taken from the maximum number of samples, which differs from case to case.

In this paper, an efficient data fusion technique is proposed to obtain the most sensitive features to the structural damage in the time-domain, and due to the small size of the feature set, the principle component analysis (PCA) is not considered. The dataset to be trained consists of N samples (including multi-class labels and the associated feature array) and 10 features corresponding to each sample data, which will be elaborated in subsequent sections. As a consequence, each sample of data constitutes an input array of class label, feature set. The number of samples N depends on the scale of the considered application and damage states in model generation. As will be discussed later, in order to enhance the accuracy of the ensemble bagged trees classifier and relatively reduce the number of multiclass labels, an artificial feature selection strategy is implemented.

Table 1: The comparison of different ensemble classifiers of decision trees.

Classifier/ensemble	Description
Decision trees	A decision tree with many leaves that make many fine distinctions between classes is suitable for a maximum number of 100 splits
Bagged trees	A bootstrap-aggregated ensemble of fine decision trees is often very accurate, but can be slow and memory intensive for large data sets
Boosted trees	The model creates an ensemble of medium decision trees using the AdaBoost algorithm. Compared to bagging, boosting algorithms use relatively less CPU time or memory but might need more ensemble members

3. Improved Feature Selection

Fundamentally, the process of feature selection plays an underlying role in data generation prior to any classification strategy. In this study, in order to develop a robust time-domain damage identification strategy, features that have the most sensitivity to the damage during vibration testing and those features with the least effect on the uncertain sources (such as signal-to-noise ratio and sensor placement) are selected for feeding the bagged tree-based data training model. For this aim, an efficient data fusion technique is proposed to extract the most damage-sensitive features with a relatively small dimension. Accordingly, the fitness of the extracted features corresponding to the measured sensors (on an intact structure) over those of the simulated ones is evaluated and the main components of each feature array are constructed.

3.1. Acceleration Time-Histories. Practically, for a specific structural finite element (FE) model, the data generation phase aims to simulate a broad set of multiple damage scenarios, recording the fitness of the simulated dynamic response over the actual structural response corresponding to the intact structure. The proposed methodology herein consists of two main stages. The former step contributes to the acceleration measurement of intact structure on embedded sensors related to the selected degrees-of-freedoms (DOFs). As a result, a set of acquired signals will be available based on the deployed sensor configuration (designated to Acc_meas). In addition, diverse damage scenarios are imposed on the structural elements, and the simulated dynamic responses will be obtained on the FE model to generate a multiclass dataset related to the multiple damage states. As a consequence, for each sample of data, a set of simulated acceleration time-histories (designated to Acc_simul) is produced based on the original sensor placement. In this research, assuming the governing linear behavior, the damage severity array is constructed as $ds \in$ [undamaged (0), damaged $(3:3:30)$]%. Accordingly, different damage states are imposed at the element level by reducing the theoretical stiffness of each element (it is assumed that the element mass is not altered by damage). The updated FE model for dataset generation is formed based on the reduced stiffness values (K) as $K^i = (1 - ds^j) \times K^i$; $i = 1$: number of elements; $j = 1:3:30$ (the components of the damage severity array).

With an emphasis on time-domain considerations, the latter step is to evaluate how well the measured acceleration time-histories to those of simulated for considered class labels related to multiple damage scenarios. In order to enhance the performance of ML-based model training, in referring to feature size reduction, measured and simulated acceleration time-histories are first converted into gray and 2D images with an intensity between 0 (black) and 1 (white). Afterwards, the fitness evaluation contains a comparison of four statistical criteria for evaluating the correlation of the resulting images as follows:

(i) Fitness 1: Mean-squared error between two images (MSE). This criterion measures the error index between two images. It takes larger values for nonsimilar images. Thus, fitness is formed as follows: Fitness = 1/(0.001 + MSE/NSen). NSen is the number of measured DOFs, and 0.001 is a small value to prevent computational complexity.

(ii) Fitness 2: Peak signal-to-noise ratio for one image with another image as the reference (PSNR). This criterion takes larger values as the similarity of two images increases.

(iii) Fitness 3: The structural similarity index for one grayscale image compared to another one as the reference (SSIM). This criterion takes a number between 0 and 1, for nonsimilar images and similar images, respectively.

(iv) Fitness 4: Root mean-squared values of Fitness 1–3.

Subsequently, in order to obtain a fairly uniform feature set, the abovementioned fitness values are normalized to the prescribed values by multiplying in appropriate coefficients, which are decided based on preliminary tests. For instance, 1, 50, 1000, and 1 have been chosen in multiplying fitness values 1 to 4, respectively. The overall representation of each feature array corresponding to each sample of data is depicted in Figure 1. As can be seen in Figure 1, each original feature array is composed of 10 features, and the four components (1 to 4) shown in the figure are referred to as the abovementioned fitness values.

3.2. Wavelet Coefficients of Accelerations. As it was mentioned earlier, in order to accomplish an efficient feature selection strategy, especially for damage-sensitive features, the wavelet coefficients of acceleration time-histories are considered to generate complimentary features. Basically, wavelets are very sensitive to the changes along a signal with the less effect of noise. Essentially, time-history data can be decomposed into scaled and delayed wavelets (referring to

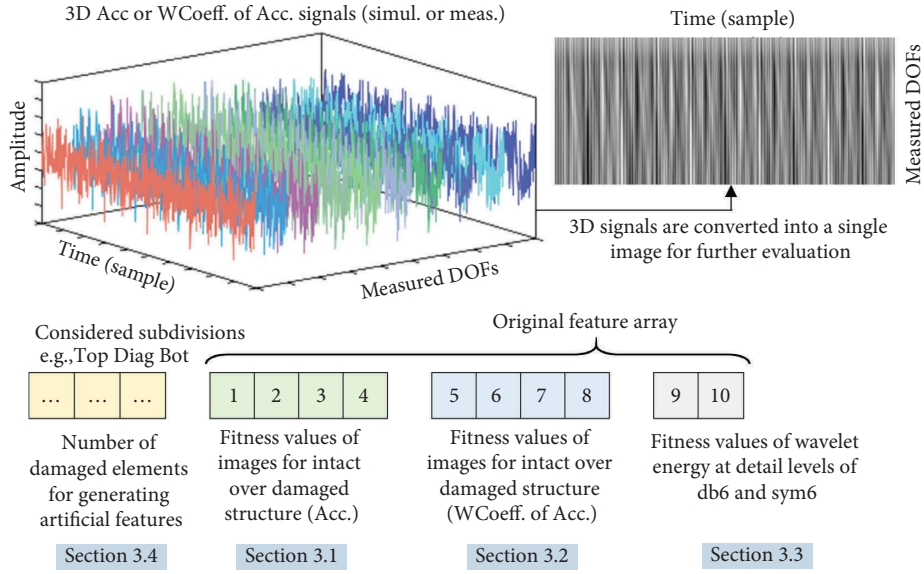

Figure 1: The overall representation of each feature array corresponds to each sample data generation (WCoeff. of Acc., simul., and meas. denote the wavelet coefficients of acceleration, simulated signals for damaged structures, and measured ones for an intact system).

the local wavelet collocations) on its global time intervals. In this regard, the idea of segmentation method (SM) and adaptive collocation points (2M) is discussed in Reference [29] for the use of different wavelet functions. In this study, the application of the first kind of Chebyshev wavelet (FCW) is considered to obtain wavelet coefficients of measured and simulated accelerations (designated to WCoeff. of Acc_meas/simul). The detailed conceptual notations and definitions of adaptive wavelets, i.e., wavelet functions and coefficients, product matrix of integration P, etc., are comprehensively addressed in Reference [29, 30]. Mathematically, the signal $S(t)$ can be decomposed by the truncated series of a wavelet's family as follows:

$$S(t) \cong \sum_{n=1}^{2^{kt-1}} \sum_{m=0}^{M-1} c_{n,m} \psi_{n,m}(t) = C^T \Psi(t), \tag{1}$$

where n, m, and kt represent the wavelet scale, relative order of the Chebyshev polynomial, and transition parameter (in this study, $kt = 2$ is utilized), respectively. C denotes the coefficients vector of the relevant wavelets, i.e., FCW herein, and the corresponding wavelet function vector is designated by $\Psi(t)$ [29, 30]. It is to be pointed out that FCW with $2M = 4$ collocation points is considered as the wavelet basis function with respect to a significantly small sampling rate of 20 samples/s to evaluate WCoeff. of Acc_meas as well as WCoeff. of Acc_simul. Eventually, the same strategy is employed as described in Section 3.1 for fusing time-domain data into new wavelet-based features. Similarly, the explained fitness values 1 to 4 (described in Section 3.1) are investigated for comparison with relevant 2D images and will result in new features 5 to 8 depicted in Figure 1.

3.3. Wavelet Energy of Acceleration Signals. The subsequent damage-sensitive feature used in this research lies on the wavelet energy of acceleration time-histories, both measured and simulated ones. In general, wavelet energy distribution with respect to time and frequency is utilized to recognize the local characteristic variations at various levels and locations [31]. Once the wavelet packet transforms have been implemented on acceleration signals (i.e., $S(t)$ presented in equation (1)), the energies of decomposed signals at different decomposition levels (DL) are used for feature extraction. By the application of orthogonal or semiorthogonal mother wavelets, the signal energy (ES) will be the summation of the j-level (referring to the DL of interest) component energies as follows [31, 32]:

$$E_S \&9; = \int_{-\infty}^{+\infty} S^2(t) \mathrm{d}t = \sum_{i=1}^{2^j} E_j^i, \tag{2}$$

where i and j indicate the modulation and scale parameters, respectively. To be more practical for damage identification in real-world applications in which an insufficient number of sensors and relatively incomplete measured data are available, in this study a simple data fusion strategy is proposed. Accordingly, by implementing two-dimensional wavelet decomposition on the Acc_meas and Acc_simul, the percentage of energy corresponding to the approximation and horizontal, vertical, and diagonal details will be obtained. In this research, the application of both Daubechies wavelets (db6) and Symlet wavelets (sym6) at decomposition level 6 is considered to obtain the wavelet energy (designated as WEnergy) of details [31]. Subsequently, the fitness evaluation is proceeded comparing the difference in wavelet energy of measured and simulated acceleration data and will form new informative feature components 9 and 10 (shown in Figure 1) for sym6 and db6, respectively. The detailed formulation of wavelet energy extraction using wavelet packet decomposition and more conceptual background on wavelet decomposition and multiresolution can be found in References [25, 31–33].

3.4. Generating the Artificial Features. The demanding task after the most damage-sensitive features have been extracted (composed of a fairly small size of 10 components for each sample) is implementing the ensemble bagged trees for model training. The main objective herein is the employment of bagged tree learners in contributing to a multilabel classification problem corresponding to the multiple damage detection strategy in large scale structures. Accordingly, the overall procedure of data generation and data training is tabulated in Table 2.

As a rule of thumb, the complexity of model training dramatically increases as the size of the model to be trained is enlarged. In other words, in dealing with multiple damage scenarios in such structures, the number of samples and relatively multiclass labels are extensively increased due to taking the consequences of damaged states into consideration when generating the sample data.

The proposed method in this study lies in the generation of artificial features in order to circumvent the dimension disaster by decreasing the correlation of extracted features corresponding to different class labels. As a consequence, by defining subclass labels instead of considering original features with numerous class labels governing a multiple damage detection problem, the training processes are performed on artificially generated features with a significantly less number of class labels. Figure 2 schematically shows the proposed method for generating artificial features from original ones. As it is apparent from this figure, the strategy involves the discretization of the large-scale structure into several subdivisions. Afterwards, the original feature array is computed based on the data fusion strategy explained before for the structural responses obtained from the structural dynamic simulation corresponding to damage scenarios imposed on the structure. The numbers of damaged elements fall into the corresponding subdivision number set. In order to efficiently convert the damaged element numbers into a unique class label, a binary representation is utilized based on the global location of the damaged element as well as the local element number. Later, the binary set is converted into a decimal value, forming an updated class label. On the other hand, the original feature array is multiplied by a unique coefficient (α) to generate the most distinguishable features of considered class label. As it can be seen in this figure, α is obtained based on the damaged element number as well as the updated class label. The practical program code development is provided in Figure 2, with the emphasize on the proposed algorithm involving both generating artificial features as well as the reduction of multi-class labels.

For better illustration, the results of the ensemble-bagged tree-based model training for a large-scale problem are depicted in Figure 3. Assuming the 3D truss structure shown in Figure 2 with the existence of two damaged elements, a dataset of size 65014×11 (1 class label plus 10 features for each sample) is originally generated. The number of original class labels considering different sequences of the probable damage states is 8128. In addition, the scatter plots of 4 original features are shown in this figure. As it is evident from this figure, ensemble-bagged tree learners could not gain the predefined convergence, and the ML strategy failed for the original dataset. On the contrary, for artificially generated features, due to the considerably less number of class labels as 55, taking into account the generated unique features based on the proposed strategy, a final precision of 99.8% is achieved on a 10-fold cross validation check. More interestingly, this high performance is accomplished with a desirable CPU computational time involved (about 5 mins).

4. Application of EOAs

Considering the damage detection problem, on one hand, a well-trained model is available for the initial measurement of an intact structure as well as simulated responses for different damage scenarios. On the other hand, a set of original features is obtained for the new measurement of the damaged structure. Taking into account that model training has been performed on artificial features, the proposed strategy in this study involves the application of evolutionary optimization algorithms in conjunction with ensemble bagged trees ML at two essential steps. The former step contributes to a confident damage localization approach (denoting discrete unknown variables) using modified genetic algorithms (GAs), whereas the latter stage aims to accurately identify the damage severities (denoting continuous unknown variables) using WCA based on damage locations, which are already localized on the structure of interest.

4.1. Damage Localization. Basically, the proposed synthesis approach herein involves the application of an ensemble bagged tree classifier in conjunction with a discrete optimization scheme to confidently localize (quantify) multiple damage scenarios very fast at the initial stage. Later, damage severities are precisely identified using the implementation of real-coded evolutionary optimization algorithms. The damage localization approach proposed in this research lies in the evaluation of possible damage sequences and the generation of corresponding artificial features to achieve the most fitted results obtained from the initially trained model. The construction of the improved GA strategy herein is presented in Figure 4(a). Practically, a multispecies binary GA coding system (BGA) is utilized to formulate the design variables of the damage quantification problem. As shown in Figure 4(a), each population is divided into four species. In addition, each individual size is 1 to the number of elements. Each string is regarded to as an element number, which is already fallen into its corresponding subdivision and may take 0 or 1 regarding the associated element number from each subdivision. In this regard, based on the possibility of a damage state on each element, 0 or 1 are assigned to represent these values. Furthermore, random exploration on probable damaged elements and random exploration in the condensed domain are achieved using the second and third species,

Table 2: The overall procedure for training the artificial dataset using bagged tree learners.

Implementation of the ensemble bagged trees for large dataset model training:
(i) The structural input/output data (i.e., externally applied loading and nodal Acc_meas) of the intact structure are measured based on the deployed sensor and force transducer configuration
(ii) Data fusion is applied to obtain grayscale images. *mat2gray (Acc_meas)
(iii) FE dynamic simulation is performed to obtain Acc_simul and relative images
(iv) WCoeff. of Acc data (measured and simulated one) is computed to simulate relative images
(v) WEnergy of signals is obtained at wavelet decomposition of level 6 of sym6 and db6
*[wavelet decomposition vector C, bookkeeping matrix S] = wavedec2 (Acc_meas, 6, "sym6")
*[percentage of wavelet energy for approximation, wavelet energy of details] = wenergy2 (C, S)
(vi) Construct the 1×10 array of features for each sample data by evaluating fitness 1 to 4:
*immse (image of the Acc_meas, image of the Acc_simul)
*psnr (image of the Acc_meas, image of the Acc_simul)
*ssim (image of the Acc_meas, image of the Acc_simul)
(vii) Assign the relative class label for each sample data based on the sequence of damaged elements
(viii) Based on the considered subdivisions in target structure, generate the artificial features
(ix) Train the dataset using the ensemble bagged trees classifier to get the predicted model
*template = templateTree ("MaxNumSplits," 65014)
*classificationEnsemle = fitcensemble (features, class labels, "Method," "Bag," ... "NumLearningCycles," 30, "Learners," template, "ClassNames," [1: updated class labels])
(x) Compute the trained model accuracy on 10 fold cross-validation
*partitionedModel = crossval (trainedClassifier. ClassificatioEnsemble, "KFold," 10)
*[validation prediction, validation Scores] = kfoldPredict (partitionedModel)
*validation accuracy = 1 – kfoldLoss (partitionedModel, "LossFun," "ClassifError")
(xi) Store the predicted model to be utilized iteratively in damage localization step

* *Note*. The as-built functions available in MATLAB software.

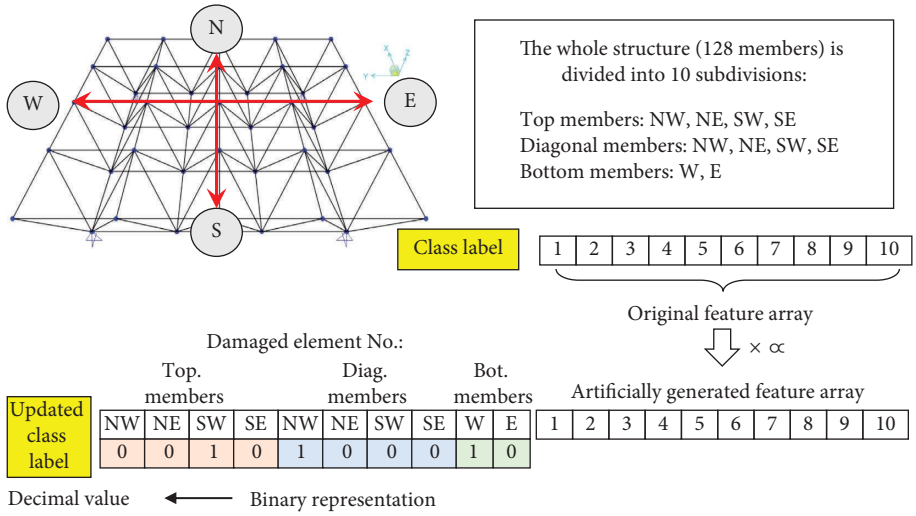

Figure 2: The schematic view of the proposed methodology to generate the artificial class labels and features based on the original dataset.

respectively. Notably, the local optima search is accomplished using the fourth species by focusing on only end-to-end elements in the stiffness matrix.

The general layout of such a multispecies BGA coding procedure is presented to localize impact load locations in Reference [26, 34]. It is observed that, the proposed BGA coding system involves multi-species with enhanced operations to overcome the local optima search capability, which is the main drawback of GA-based algorithms in dealing with such discrete variable problems.

In this study, by selecting the end-to-end elements after the prescribed number of runs, the feasible search domain is reduced to perform the proposed search domain reduction (SDR) technique applicable in the proposed BGA coding. For this aim, the damaged elements designated as the optimal results at each generation are considered, and the connected elements attained from the connectivity matrix of structural configuration are utilized to form the new possible search domain through the proposed SDR approach. Based on each formulated binary individual, prescribed subdivisions and element numbers are decided. Accordingly, the extracted features (i.e., features 1 to 10) obtained from the data fusion of new measured responses could be artificially generated. Eventually, the class label of the current feature array is predicted using the initially trained model. The fitness value to be maximized comprises two underlying

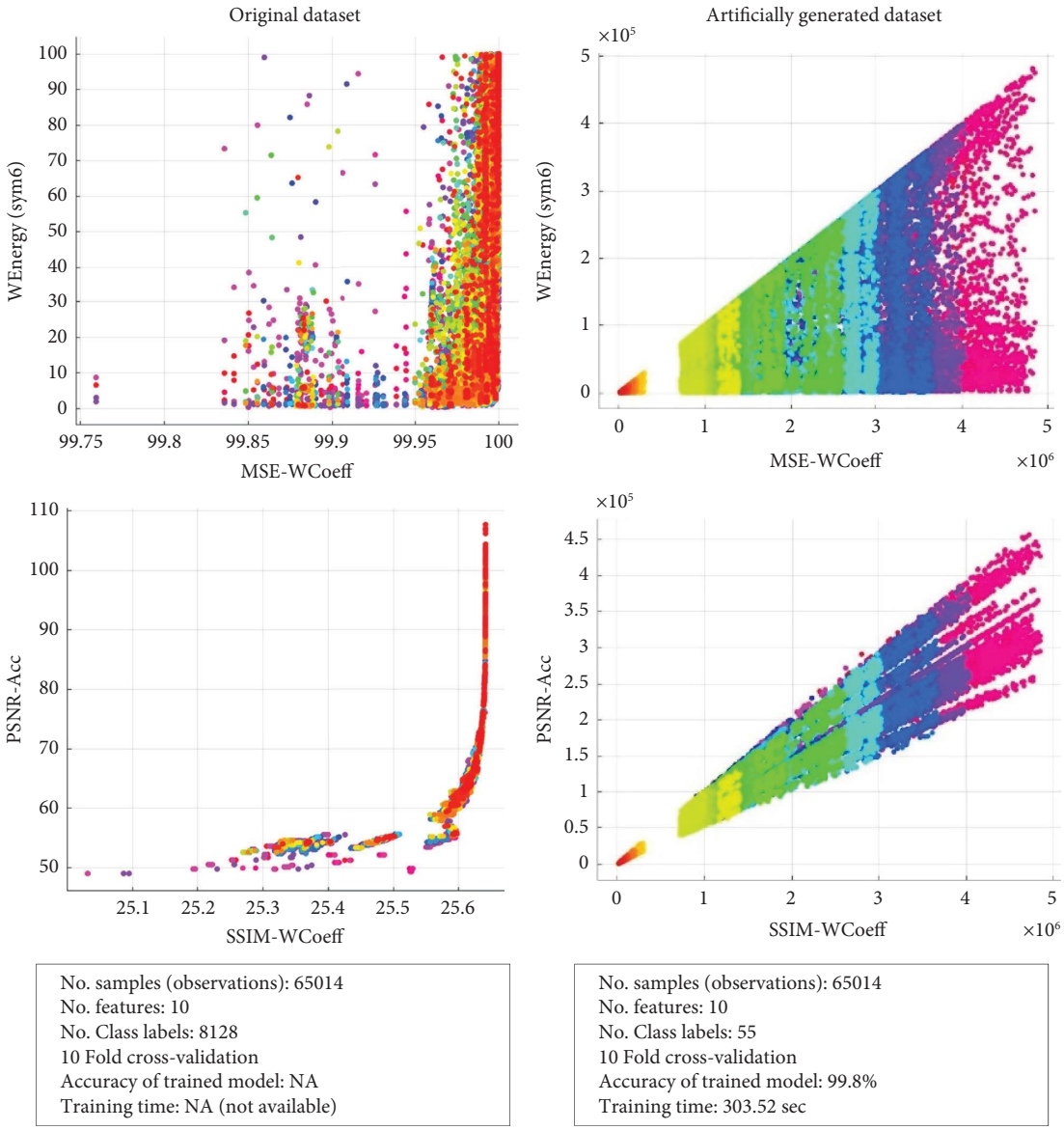

Figure 3: The comparison of scatter plots for different feature selection and the resulted performance for the application of bagged tree model training related to the original dataset and the artificially generated dataset.

criteria. The first involves producing the prescribed number of test data from the recently predicted class label and evaluating the resulting performance from the confusion matrix of the main trained model on artificial features. This fitness evaluation guarantees damage localization at the subdivision level, referring to the exploitation phase of BGA. Furthermore, the second fitness evaluation deals with the comparison of artificial features attained for the measurement of interest against those that were restored in the main dataset denoting the exploration phase of BGA. One may interpret the aforementioned evaluations as maximization of the fitness value as follows:

$$\begin{aligned}
\text{Performance} &= \frac{\text{sum}\left(\text{diag}\left(\text{confusion matrix}\right)\right)}{\text{sum}\left(\text{confusion matrix}\left(:\right)\right)},\\
\text{Fit 1} &= \text{Performance},\\
A &= \left(\text{ArtiDamage Features}\right)_{\text{meas}}\\
&\quad - \left(\text{ArtiDamage Features}\right)_{\text{simul,}}\\
B &= \text{sum}\left(A.^2\right);\ \text{Fit 2} = \frac{1}{\varepsilon + B/10};\ \text{Fitness} = \text{Fit 1} \times \text{Fit 2},
\end{aligned} \tag{3}$$

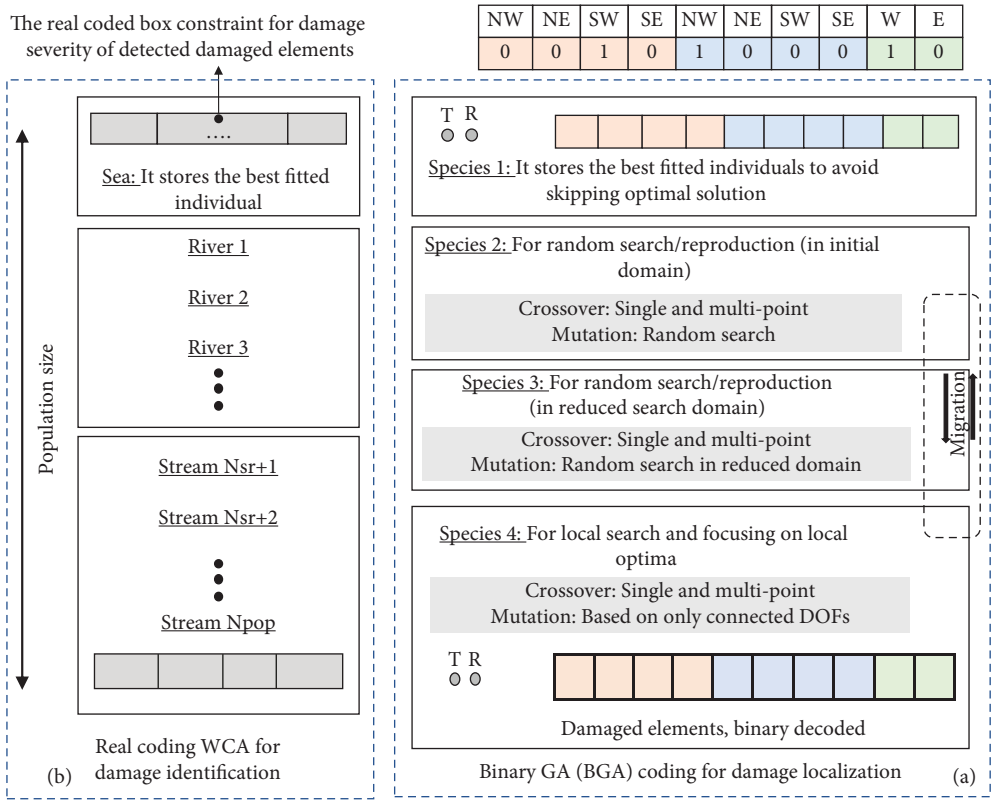

FIGURE 4: Employment of the evolutionary optimization algorithms: (a) multispecies BGA procedure for localization of damaged elements; (b) WCA approach for damage severity identification.

where 10 is the number of extracted features. ArtDamage $\text{Features}_{\text{meas}}$ and ArtDamage $\text{Features}_{\text{simul}}$ denote the artificial damage features corresponding to measured and simulated data, respectively. The ε may have a small value of 0.001 to prevent computational difficulties.

4.2. Damage Identification. Once damage locations have been confidently detected at the element level, a real-coded water cycle algorithm (WCA), known as a metaheuristic and evolutionary optimization algorithm, is adopted to precisely quantify the severity of damages. The schematic construction of each population for the use of real-coded WCA is presented in Figure 4(b). As it is illustrated in this figure, the set of variables shown for each individual contains real values of damage severities associated with element numbers predicted from BGA. Hence, the lower (LL) and upper limits (UL) of each string form the box constraints according to the broad threshold of damage severities. In this research, 3% to 30% of the damage index is treated as LL and UL, respectively. Accordingly, the main idea of WCA is thoroughly addressed in Reference [25, 26, 35]. To be more concise, the real-coded and population-based WCA herein is inspired by the water cycle process in nature and is modelled based on the surface run-off of streams and rivers flowing into the sea. The operation of the raining process is utilized to generate each individual, namely, stream of WCA [25, 26, 35]. The core of cost evaluation for the WCA approach lies in a minimization problem and is formulated as follows:

$$\text{Cost} = (-1) \times \frac{1}{\varepsilon + \sum \left((\text{Damage Fuatures})_m - (\text{Damage Fuatures})_s \right)^2 / \text{DOF}_m}, \tag{4}$$

where Damage Features_m, Damage Features_s, and DOF_m represent the extraction of measured and simulated features and the number of measured DOFs, respectively. The minus sign is utilized to invoke a minimization problem. Overtly, the summation limit varies from 1 to the total number of 10 extracted features. In the present research, in order to well examine the high performance of WCA in contributing with damage identification problems, some of the other state-of-the-art evolutionary optimization algorithms are comparatively investigated. Subsequently, the overall procedure of the proposed damage quantification algorithm is clarified in Tables 2 and 3 for the data training and damage evaluation stages, respectively. In addition, the schematic flowchart of the proposed strategy is depicted in Figure 5 as the summary

Table 3: The overall procedure of the proposed scheme for damage localization and identification.

Damage localization:
(i) The input/output measurement is carried out on the damaged structure for initial sensor and actuator placement as set before
(ii) Compute the 1×10 array of original features for acquired data. The fitness value is evaluated for initially measured Acc data for intact structure compared to those of measured on damaged structure
(iii) Run the proposed multispecies BGA code for each predicted individual as there is the possibility of elements being damaged. LL and UL are sets 1 and the total number of elements, respectively. For this aim, based on the considered subdivisions and predicted damage locations, the obtained array of original features is artificially generated to be compared with the associated class labels in the updated model
(iv) Generate test data of interest in order to formulate the resulted confusion matrix for the predicted class label associated with measured features
(v) Evaluate the cost value as presented in equation (3)
(vi) Continue until the desired convergence is achieved and store the output results as damaged element numbers
Damage severity identification:
(i) Once the damage localization is confidently accomplished, WCA is performed to precisely quantify damage severities for previously detected element locations. In this regard, LL and UL are set 0 and the maximum damage index (i.e., 30%), respectively
(ii) Evaluate the cost value as presented in equation (4) for originally measured features as well as those of simulated for predicted damage severities
(iii) Stop if the desired convergence criterion is accomplished

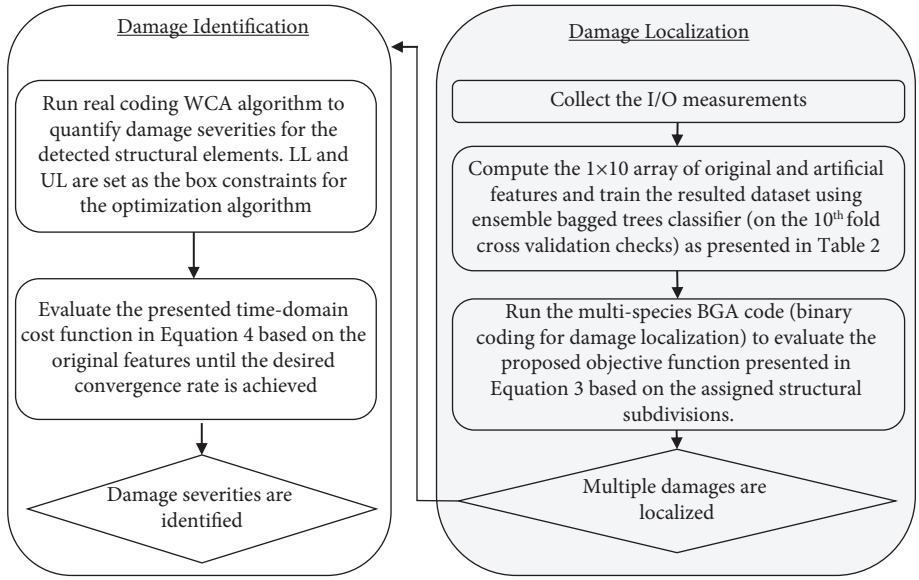

Figure 5: The schematic flowchart for the proposed algorithm of structural damage localization and identification (as presented in Figure 4 and Tables 2 and 3).

of Figure 4 Tables 2 and 3. As it is apparent from Tables 2 and 3, once the model is trained on artificial features, damage localizations are confidently detected at the element level by using the predicted model iteratively through the proposed multi-species BGA algorithm. Eventually, the application of WCA will be performed on the predicted damage locations to accurately identify damage severities.

5. Numerical and Experimental Validations

The appropriateness of the proposed scheme for damage localization and identification is investigated. For both numerical and experimental verification studies, the Newmark constant-average acceleration method at different sampling rates is used to perform the numerical dynamic simulations. The associated lumped mass is utilized for the numerical application. On the contrary, the lumped and consistent masses are considered for damage localization and identification stages of experimental application, respectively. Damping ratios of 2% and 5% are taken into account to perform Rayleigh damping parameters corresponding to the first two modes in the numerical and experimental applications, respectively. A preliminary study is performed to establish suitable values for parameters utilized in the multispecies BGA and WCA strategies. For this aim, the BGA and WCA parameters were reasonably varied to determine the range of values that consistently provided desirable results. In this regard, the BGA parameters utilized in the damage localization stage of numerical application were; 4×50 (species 1–4), 3×10, 0.8, 0.1, 0.05, 3, and 15 as the population size, number of generations (Run = 3), crossover, mutation, and migration rates, number of regenerations, and number of times that reintroduction is taking place, respectively. For

experimental evaluation these parameters were set to 4×20 (species 1–4), 1×10, 0.8, 0.1, 0.05, 1, and 5 as the population size, number of generations (Run = 1), crossover, mutation, and migration rates, number of regenerations, and number of times that reintroduction is taking place, respectively. In addition, $d\_\max = 1e-8$ and N_{Sr} is assumed to be 4 as the number of streams in applying the WCA strategy. Accordingly, the proposed BGA procedure at the first stage (i.e., damage localization) is performed using an efficient parallel computing approach. The results for each anticipated array as the possibility of damage location are simultaneously calculated using twenty CPU cores. Subsequently, the required processing time consumption (computational time involved) is presented for the use of the same hardware environment (Intel i9-7900X CPU at 3.31 GHz, 64 GB of RAM, Operation 64 bit) and a parallel pool of connecting to 20 virtual cores.

5.1. Numerical Validation Study: A 3D Truss Structure. The layout of large-scale and double layered spatial structure considered for the numerical verification study is shown in Figure 6. As it is depicted in Figure 6(a), the considered structural system is composed of 128 rod elements, 41 pinned joints, and 4 hinge supports, resulting in 111 transitional DOFs. Furthermore, the considered structural system is constructed with aluminium pipes. In addition, $A = 11.66\,\text{cm}^2$, $E = 69\,\text{GPa}$, $\overline{m} = 2600\ \text{kg/m}^3$, and $g = 9.81\ \text{m/sec}^2$ are selected as the cross-sectional area for all members, elastic modulus of elasticity, mass per volume, and gravitational acceleration, respectively. The deployed sensor configuration SC1 is highlighted in Figure 6(b). Accordingly, there are 13 sensors (accelerometers) embedded on highlighted truss nodes to measure acceleration (Acc) time-histories along principle directions x, y, and z, resulting in 39 transitional DOFs. The governing eigen-value problem associated to the finite element model of the structure is solved and yields the last two natural frequencies of 776.23 and 797.33 Hz. Therefore, taking into account the Nyquist rule in order to capture the entire structural response, a random multi-sinusoidal wave of 40–800 Hz with maximum amplitude of 1 kN is generated at 2000 sampling rate, and later it is interpolated to match reduced sample rate of 200 Sample/s (S/s). The considered signal on reduced sampling rate comprises the broad-band frequency components; however, it is far flatter than the original signal.

The first 5 sec of the generated signal are applied in the vertical direction of nodes 29, 30, 32, and 33 as the externally applied loading. Furthermore, only output signals are polluted with 5% white Gaussian noise. Eventually, the dynamic response of the intact system is simulated, and the measured Acc_meas data are recorded as available reference data for further evaluations in extracting informative features from the entire sample data.

In general, for damage identifications of such application through ML paradigms, there are two scenarios that may be taken into consideration. The first scenario as the most popular one, is to assume the probability of a unique element or group of elements being damaged or not (i.e., the localization accuracy down to a group of elements). This idea will result in a relatively small-scale model that can be trained with any ML classifier. The second scenario, which is the case of interest in this research, is to assume the probability of each structural element being damaged at different sequences with any other element in referring to multidamage scenarios (i.e., the localization accuracy down to a single element). Consequently, it will result in a significantly large-scaled simulated model to be trained as a multiresponse class problem with bagged trees ensemble herein.

Accordingly, in the numerical evaluation, two damage scenarios (DS) are assumed to be multiple damage cases. In this regard, DS1 denotes 10% of the damage index as the stiffness reduction in element number 60, and DS2 involves 28% and 4% of the damage indexes as the stiffness reduction in element numbers 13 and 76, respectively. Based on the proposed strategy in this study, the simulation of sample data is carried out for various sequences of damaged elements, considering a damage index of $ds \in$ [undamaged (0), damaged (3 : 3 : 30)]% imposed to the theoretical value of each element's stiffness. This step for dataset simulation may interpret as the most computationally demanding task in performing the proposed strategy. Taking into account multiple-damage scenarios and considering imposed damage severities on such a large scale structure will result in a 65014 simulated sample data with 8128 associated multiclass labels related to 10 original features. The computational time involved (CPU time) for the entire process of dynamic simulation and original feature array generation is recorded about 18.42 hours. As presented earlier, the proposed approach lies on training the model originating from the artificial features in order to considerably reduce multiclass labels.

For instance, 55 class labels are the resultant of the proposed scheme for artificial feature generation assigned to 65014 sample data. Subsequently, various ML classifiers are employed for model training of the 65014×10 dataset (including 55 class labels associated with artificial features), and the obtained accuracy and computational time involved after each classifier are comparatively tabulated in Table 4.

The percentile accuracy presented in this table corresponds to the evaluation of a 10-fold cross validation. It is to be pointed out that none of the ML classifiers could deal with such a large scale model of original features and it would be rather inevitable to execute learning tasks on artificial features. The notable observation on the data tabulated in Table 4 lies on the superiority of the ensemble bagged trees classifier in achieving the most precise predicted model in the reasonably fast computational time involved. On the contrary, the least accuracy is recorded for medium tree learners after 16 sec. As it was anticipated, the family of SVM classifiers has produced inaccurate results with far longer computational time involved, which makes the SVM classifiers (in their general format) inapplicable in training such a large scale model.

Once the trained model is provided, it is kept as the reference model for further evaluations whenever SHM of the structure takes place. In this regard, the new

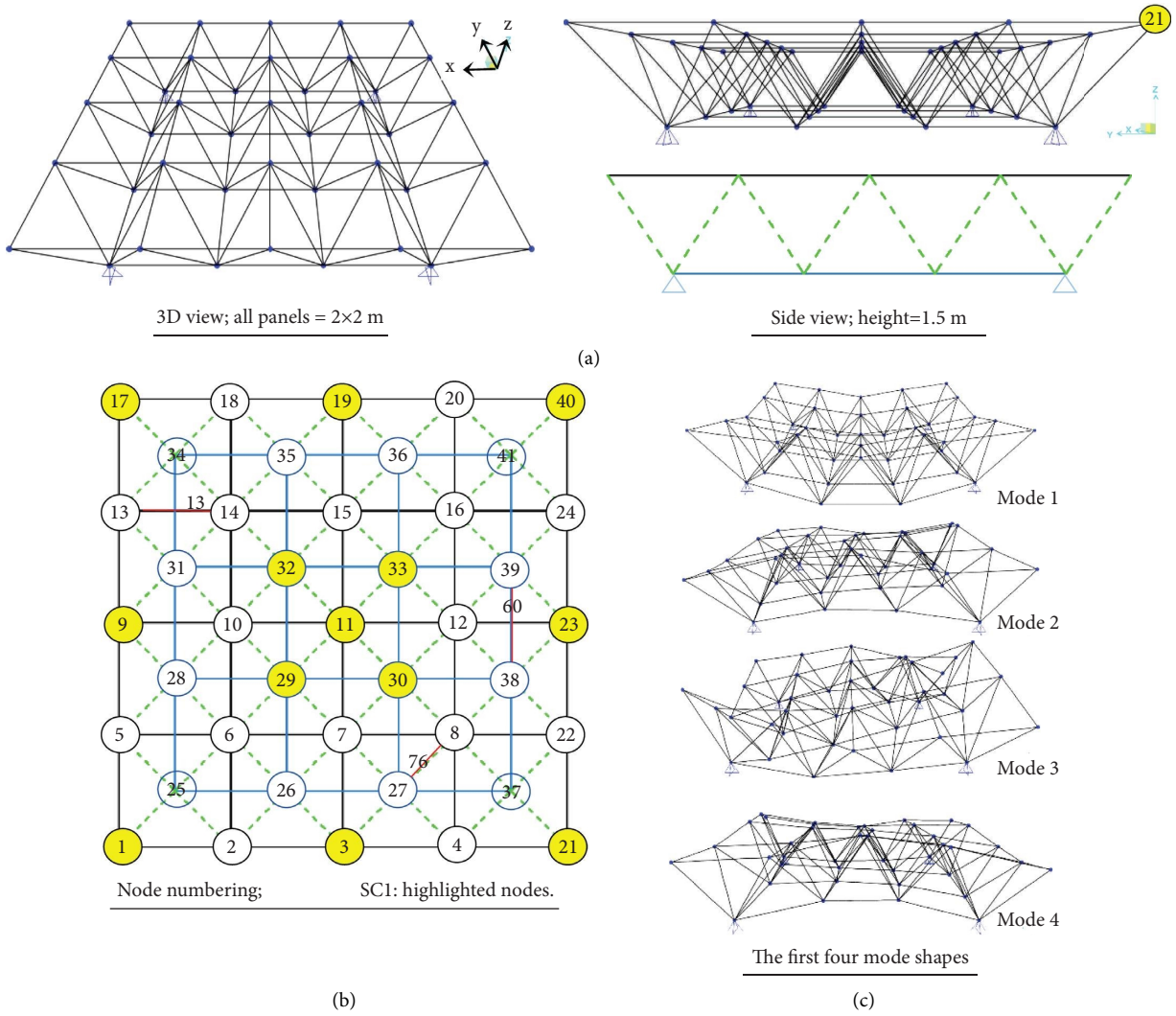

Figure 6: The considered 3D spatial structure: (a) top and side views, (b) sensor configuration (SC1, highlighted joints), joint numbering, corner supports, and damage scenarios (element numbers 13, 34, and 76), and (c) the selected structural mode shapes (i.e., mode shapes 1 to 4).

Table 4: The accuracy and CPU computation time involved for the application of different ML classifiers in training the model of 65014 sample data with associated 55 multiclass labels.

ML classifier	Ense. bagged trees	Ense. boosted trees	Ense. RB. trees	Fine trees	Medium trees	Weighted KNN	Fine KNN	Kernel naive Bay.	Gaus. naive Bay.	Linear SVM	Fine Gaus. SVM	Quadratic SVM	Cubic SVM
Accuracy (%)	99.8	47.3	42.6	26.6	21.3	90.2	89.8	25.6	22.1	22.9	34.3	NA	NA
Time (s)	464	548	692	19	16	65	26	789	25	4203	3415	NA	NA

Note. The CPU time is rounded up in second. Ense., RB., KNN, Bay., Gaus., SVM, NA denote ensemble, RUSBoosted, *K*-nearest neighbors, Bayes, Gaussian, support vector machine, and not available, respectively.

measurement may be carried out on the damaged structure, and the original feature array of 1×10 will be extracted. In the damage localization step, the proposed multi-species BGA is then performed to detect the damage locations. Taking into account the stochastic characteristics of evolutionary optimization strategies, the damage localization and identification code are run 5 times. The history of fitness assessment for damage localization of DS1 and DS2 corresponding to one of the finest results achieved is provided in Figure 7.

The result of the damage localization stage is presented in Table 5 corresponding to DS1 and DS2. It is observed that, due to the comprehensive operations included in the proposed multispecies BGA, the exploitation and exploration phases are thoroughly proceeded. Several attempts have been made for damage localization in such a large-scale structure, and it is concluded that for all cases considered, the damage locations are confidently detected (without false detection). Furthermore, in conducting the evaluation study,

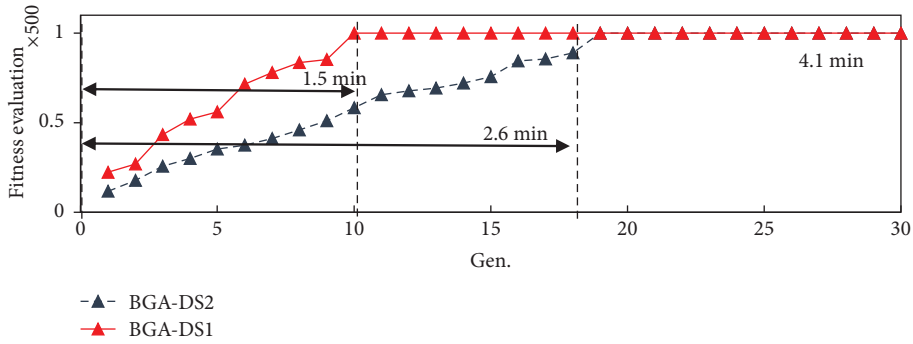

Figure 7: The fitness histories of the proposed multispecies BGA approach for damage localization correspond to DS1 and DS2.

Table 5: Damage localization and identification results correspond to DS1 and DS2.

	Element:	Localization			Identification											
					WCA			PSO			ICA			DE		
		60	13	76	60	13	76	60	13	76	60	13	76	60	13	76
DS1	Error (%)	0	—	—	2.1	—	—	9.8	—	—	15.8	—	—	11	—	—
	Time (min)	4.1	—	—	16.7	—	—	14.1	—	—	10.5	—	—	15	—	—
DS2	Error (%)	—	0	0	—	4.6	12.7	—	7.8	28.2	—	33	64.1	—	21.1	42.5
	Time (min)	—	4.1		—	39.5		—	37.2		—	26.3		—	38.5	

Note. Error values are computed for actual versus identified damage indices.

two scenarios may be considered. The former scenario is to compare the time taken to reach a given convergence rate, while the later one is to compare the best convergence rate that can be achieved in a given time. In this study, the second scenario is considered by setting the best convergence rate as the reference for the sake of comparison. Data shown in Table 5 and Figure 7 demonstrate the satisfactorily reliable results obtained for damage localization of such large-scaled structures using the proposed fitness evaluation of BGA, where the time taken to reach the converged results for DS2 is only 2.6 min. Considering such a large structural system and to prevent progressive failures, it is of crucial significance to detect the existence of damage and its location after an extreme loading event as soon as it occurs. The recorded time consumption reveals the main merit of the proposed strategy in referring to fast damage localization in large-scaled structures. Once damage locations have been confidently detected for DS1 and DS2, the evolutionary optimization algorithms, i.e., WCA, PSO, ICA, and DE, are adopted to identify damage severities. Accordingly, the best cost-history recorded for each optimization strategy (after five repeats) is illustrated in Figure 8 corresponding to DS1 and DS2.

In employing optimization strategies for DS1, 50 and 20 are taken as the population size and the maximum number of generations (iterations), respectively, whereas 50 and 50 are taken as the population size and the maximum number of iterations for DS2, respectively. Damage index of each detected element from the previous stage is treated as unknown variables of optimization. The LL and UL are set as [0, 30]%. In order to get the most reliable results, the optimization codes have been run 5 times and the best results are plotted as the final results. The damage identification results are tabulated in Table 5. The percentile error values presented in this table compare the actual and identified damage indices for different elements. For instance, in a worst situation 12.7% is recorded as relative error in identified damage index of element 76 after the application of WCA (i.e., about 3.5% as the identified damage index compare to 4% as the actual one). Basically, there are two criteria that should be taken into consideration in comparing the cost-histories attained for different optimization schemes shown in Figure 8. These involve the exploration rate in reaching the least cost value (best cost value) as well as the fast convergence rate. As it is shown for DS1 and DS2 in Figure 8, the less cost value is recorded for WCA yielding the less percentile error for damage identification results presented in Table 5. Comparing the results tabulated in Table 5 and the cost-history plot in Figure 8, one may include that the less performance is reported for ICA algorithm. Taking into consideration the above-mentioned 2 criteria, Figure 8 demonstrates the superiority of WCA in dealing with damage identification of such large structures.

In order to well evaluate the performance of different optimization algorithms in dealing with damage identification problem, the comparison of statistical results i.e., the best, mean, worst and standard deviation (SD) results obtained for the relative error values (%) in identified damages corresponding to DS1 and DS2 after 5 repeats is provided in Table 6 for the use of WCA, PSO, ICA, and DE optimization strategies.

Data tabulated in Table 6 reveal the satisfactory reasonable SD results recorded for WCA strategy compared to those of computed for the other optimizers. The less SD

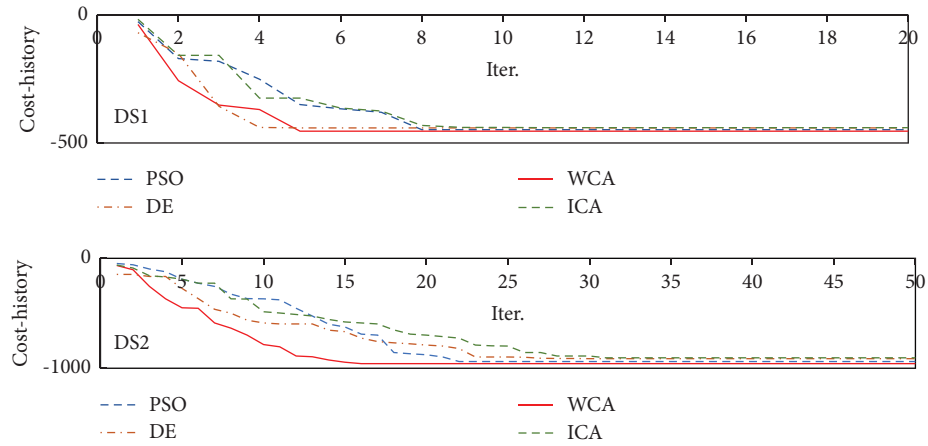

FIGURE 8: The best cost-histories obtained for different optimization strategies in the damage identification step correspond to DS1 and DS2.

TABLE 6: The comparison of statistical results i.e., the best, mean, worst and standard deviation (SD) results obtained for the relative error values (%) in identified damages corresponding to DS1 and DS2 after 5 repeats using WCA, PSO, ICA and DE optimization strategies.

	Element no.	Damage identification											
		WCA			PSO			ICA			DE		
		60	13	76	60	13	76	60	13	76	60	13	76
DS1	Best (error %)	2.1	—	—	9.8	—	—	15.8	—	—	11	—	—
	Mean (error %)	2.5	—	—	10.2	—	—	16.38	—	—	12.1	—	—
	Worst (error %)	3.1	—	—	11.6	—	—	17.9	—	—	14.1	—	—
	SD	0.36	—	—	0.71	—	—	0.78	—	—	1.12	—	—
DS2	Best (error %)	—	4.6	12.7	—	7.8	28.2	—	33	64.1	—	21.1	42.5
	Mean (error %)	—	4.9	13.2	—	8.4	29.4	—	35.8	65.8	—	24.2	45.6
	Worst (error %)	—	5.7	13.9	—	8.9	31.1	—	38.1	66.9	—	27.4	49.1
	SD	—	0.4	0.46	—	0.5	1.14	—	2.03	1.04	—	2.24	2.53

values recorded for WCA guarantee the resulted high performance and reliable results obtained using this algorithm in dealing with damage identification problems.

5.2. Experimental Verification Study: A 2D Steel Pin-Jointed Bridge. To be more practical, the experimental verification of the proposed scheme for damage localization and identification is evaluated in this section. For this purpose, a 2D steel truss structural system is fabricated and tested in the laboratory. It is composed of nine horizontal and vertical elements (constructed with steel hollow sections) and four diagonal members (provided with double steel strips supported with internal stiffeners to prevent buckling phenomena). All truss elements are connected with eight pin connections so that they are free to rotate. The schematic view of the test setup as well as the deployed sensor configuration (denoting measured DOFs), the location of force transducers, the structural subdivisions, and structural characteristics are provided in Figure 9. The considered structure comprises 13 DOFs, indicated by red arrows in this figure. As it is shown in this figure, nine ICP accelerometers (KISTLER magnetic piezoelectric sensors) are embedded in the structure to measure corresponding DOFs at a sampling rate of 5128 Sample/s (as a case of incomplete measurement). Furthermore, a 16-channel digital analyzer (OROS 36) is utilized for acquiring data at different sampling rates.

As a preliminary study, a FEM of the considered structure is modelled in ABAQUS software, assuming the mass density of steel 7850 kg/m^3 and the modulus of elasticity of steel $E = 209$ GPa (known as the theoretical values). The natural frequencies obtained from the FEM of 2D truss vary from about 202 to 2120 Hz corresponding to the first and thirteen modes of the structure. On the other hand, impact tests are carried out on intact structure to extract the natural frequencies related to the as-built structural elements. However, due to the complexity of the structure, the extracted natural frequencies related to the higher modes obtained from the application of the fast Fourier transform (FFT) and the power spectrum of the response were not reliable. This is most probably due to the local vibrations of structural elements, whereby they produce significant energy in the 1800–3000 Hz range and diminish at the higher structural frequencies. It is to be kept in mind that, in planning the damage identification experiment, these frequencies and relatively these modes are not excited. Therefore, a random multisinusoidal loading of 150–1800 Hz with a random magnitude of −50, 50 N is generated for 5 secs at sampling rate of 5128 samples/s in the digital data analyzer (OROS36). Accordingly, the described signal is passed through a power amplifier and then is

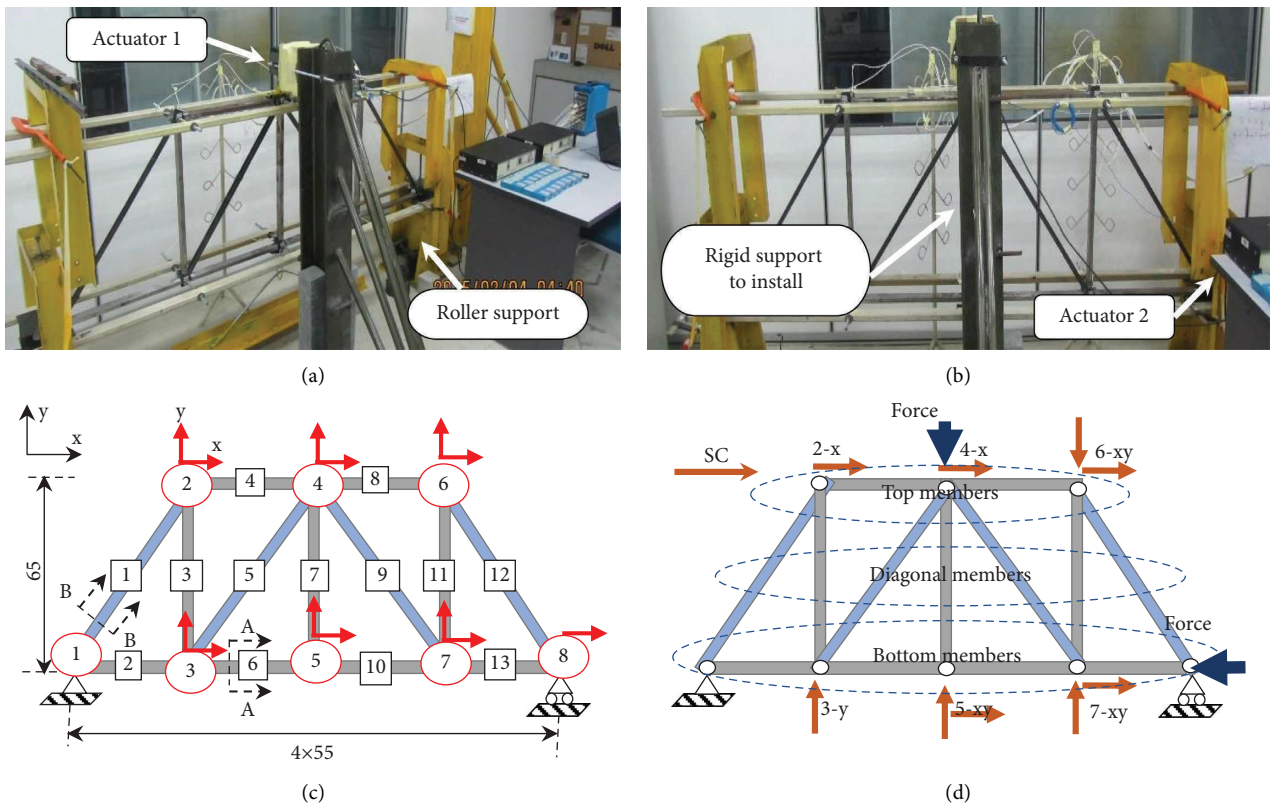

Figure 9: The schematic view of the test setup: (a) the laboratory 2D truss, (b) sensor/actuator placement, (c) layout and dimensions (cm), node and element numbering, and (d) sensor configuration (SC) and structural subdivisions.

applied by two electromagnetic shakers at prescribed actuator locations (downward at 4-y and right-to-left at 8-x). In addition, the measured input and output data are then reduced to 2048 samples/s by interpolating the initial samples to reduced ones so that the numerical simulation is performed for such reduced sampling rates.

Once the test is performed on intact structure, the proposed data fusion technique, as described in Figure 1, is employed to convert 3D responses into sets of related images, which are then kept as the reference data for intact structure. Subsequently, damage scenarios DS1–DS3 with different damage severities, as provided in Table 7, are imposed on structural members, and corresponding data measurements are carried out with respect to each scenario. As it can be seen in Table 7, the large damage severity is induced only for element 4 by cutting its cross section, while the small and moderate damage severities are induced for elements 12 and 5, respectively.

The small and moderate severities are formed by removing one and two internal stiffeners in diagonal members 12 and 5, respectively. In order to well examine the proposed damage identification strategy in estimating the expected reduction in static stiffness due to diverse damage severities, the FE analysis is conducted using ABAQUS software. In this regard, the whole element 4 (intact element) is modelled in the software as a bar element (i.e., node 2 to node 4 shown in Figure 9) under a ∓500 N arbitrary axial force applied to node 4. The resulting model is considered the basis for intact element, utilized in further evaluations. Then, the reduced cross section due to damage is modelled on element 4, applying the same axial force, and the resulting displacement is computed for the damaged element. As a consequence, the damage index of 25.14% is obtained as the expected reduction in static stiffness of member 4 due to the large damage. The same procedure is performed on diagonal members to compute relative damage indices of 18.75% and 14.33% as moderate and small values, respectively.

Table 7: Damage scenarios and corresponding damage indices imposed to the 2D truss.

Element no.	Damage scenarios (DS)		
	DS1	DS2	DS3 (%)
4	25.14%	25.14%	25.14
5	—	—	18.75
12	—	14.33%	14.33

Accordingly, the original dataset and an artificial one are generated in order to tackle the damage identification problem. Notably, this step may be presumed as the most computationally demanding step in recording about 6.23 hours for simulating 14916 samples of features, considering 2048 samples/s in dynamic simulation. However, this computational task is carried out once during the entire service life of the structure. For the sake of comparison, the bagged trees model training is conducted on both original and artificial datasets. The comparison of scatter plots for

different feature selections and the resulted performance for the application of bagged trees model training is illustrated in Figure 10. One of the remarkable observations on this figure lies on the robustness of the proposed algorithm in generating artificial features yielding the less number of multiclass labels and relatively the highest precision for implementation of bagged trees in the model training of such a big dataset.

Afterwards, the proposed multispecies BGA is conducted to localize damage locations. The developed codification is repeated 5 times, and it is concluded that, for all damage scenarios considered, damage locations are confidently detected at element level without no false detections corresponding to different runs. Regardless of diversity in damage scenarios, the computational time involved at this step is recorded at about 1.02 min, corresponding to each case considered. Taking into consideration the difference between the actual (as-built) stiffness of structural elements compared to those of theoretical values in dataset simulation, such outstanding results for damage localization evidence the advantage of the proposed strategy in dealing with real-world applications when the updated results for structural characteristics are not available. Once damage locations have been confidently detected, evolutionary optimization algorithms, i.e., WCA, PSO, DE, and ICA, are comparatively employed to identify the damage severities induced in the detected structural elements. Damage index is treated as unknown variables of optimization. The LL and UL are set as [0, 30]% of theoretical values for stiffness reduction in each element. In this regard, the effective cross-sectional area of each element (A) is considered in order to form $E \times A$ as the resultant of each member's stiffness value. In performing optimization algorithms, the population size is set to 50 and the number of iterations is set to 10, 20, and 50 corresponding to DS1, DS2, and DS3, respectively (these values are kept constant for all optimizers). In order to get the most reliable results, the optimization codes have been repeated 5 times and the best results are considered as the final results. It is to be noted that, the reason for choosing the best results (after 5 repeats) rather than the average, is that an undesirable result may diversely affect the average results so that making that unreliable. Furthermore, the best cost-histories obtained for different optimization strategies in damage identification step corresponding to DS1, DS2, and DS3 are illustrated in Figure 11. As was mentioned earlier, the prescribed convergence criterion is set as the reference for comparison. In other words, the optimization process is stopped when reaches to a predefined convergence rate. Accordingly, the relative errors in identification results and CPU time consumption as an indication for computational cost after each optimization algorithm are tabulated in Table 8.

The satisfactorily high performance of WCA in dealing with the diverse scenarios considered can be deduced from the comparison of cost evaluation histories plotted in Figure 11. For instance, considering DS2, it can be overtly seen that WCA converges at the fourth iteration while DE and ICA have gotten stuck in the exploitation phase. In contrast, the results after PSO have demonstrated an acceptable convergence rate over ICA and DE. However, for a larger optimization problem such as DS3, this optimizer also cannot compete with WCA. Logically, based on the obtained cost values after the implementation of each optimizer, the resulting error values provided in Table 8 seem predictable.

Eventually, in order to achieve reliable results for the experimental verification study, a well-known substructural identification approach [3, 36, 37] is performed for time-domain identification of the intact and 2D truss structure, and the updated values of each member's stiffness are precisely identified [38]. Accordingly, the error measurement for damage severities takes into account the updated stiffness values rather than theoretical ones. More interestingly, it is observed that the identification is consistently giving a larger stiffness value for intact diagonal members. Due to the fact that the axial rigidity of diagonals was computed based on only double strips and the effect of internal stiffeners was ignored, the results were anticipatable in the presence of such internal stiffeners.

It is to be emphasized that, theoretically, it is somehow impossible to calculate the as-built stiffness of diagonals due to the internal stiffeners. As a result, the identified stiffness, value of intact diagonals, including the internal stiffeners is set as the basis for presenting the so-called damage index value. The relative error values in identified damage indices are updated in Table 9 corresponding to the employment of different optimization strategies considering the updated stiffness values of intact structures. Comparing the cost-histories shown in Figure 11 and tabulated data in Table 9, one may deduce that the reliable damage identification results are accomplished by WCA from both the error measurement and computational cost points of view. Finally, taking into account the considered test setup as a real-world application in the vicinity of various uncertain sources (i.e., I/O signal-to-noise ratio, modeling errors, and so forth), it is distinct that the reliable damage identification results are achieved by the proposed strategy in two steps involving the absolutely confident and fast localization and the accurate identification of damage states.

Eventually, the standard deviation (SD) results obtained for the relative error values (%) in identified damages corresponding to DS1-3 after 5 repeats is tabulated in Table 10 for the employment of WCA, PSO, ICA, and DE optimization strategies.

The notable consideration on this table lies in the high values for the standard deviation of results recorded for the DE optimization algorithm. This is due to the inherent formulation of mutation operations in this algorithm, which yields the highest values of errors for the worst results.

5.3. Discussion. The capability and effectiveness of the proposed strategy for structural damage identification were validated numerically and experimentally. It is included that damage locations are confidently detected using the improved multispecies binary genetic algorithm in conjunction with a supervised machine learning paradigm. To achieve this goal, the ensemble bagged trees classifier is utilized for model training and multilabel classification of large-scaled

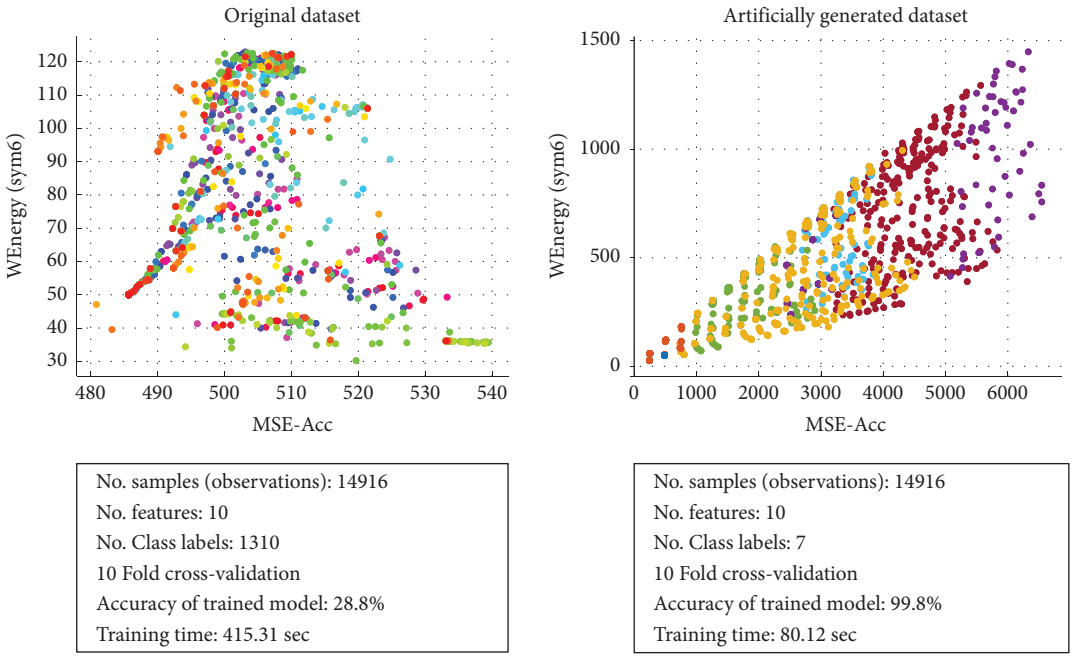

FIGURE 10: The comparison of scatter plots for different feature selection and the resulted performance for the application of bagged tree model training related to the original dataset and the artificially generated dataset of the 2D truss experiment.

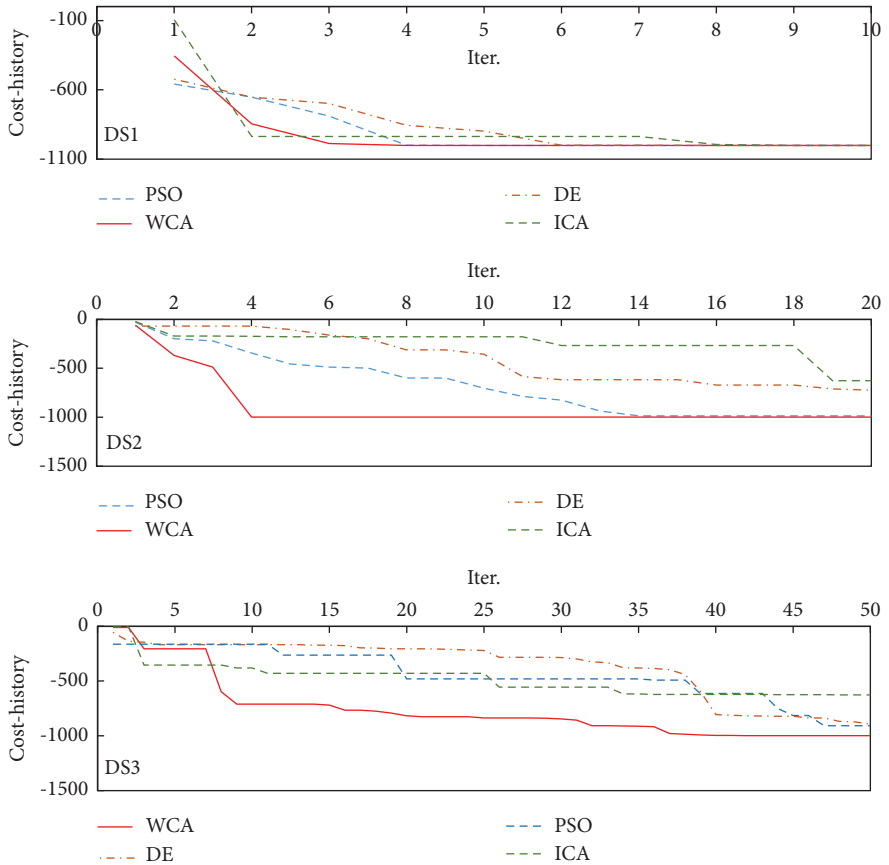

FIGURE 11: The best cost-histories obtained for different optimization strategies in damage identification step corresponding to DS1, DS2, and DS3.

Table 8: Damage identification results corresponding to DS1, DS2, and DS3 related to the theoretical stiffness values.

Element no.:		Identification											
		WCA			PSO			ICA			DE		
		4	12	5	4	12	5	4	12	5	4	12	5
DS1 (10 iter.)	Error (%)	16.9	—	—	19.3	—	—	24.3	—	—	18.6	—	—
	Time (min)	3.60	—	—	3.18	—	—	2.61	—	—	5.29	—	—
DS2 (20 iter.)	Error (%)	17.4	19.9	—	20.5	22.8	—	24.9	28.1	—	20.1	21.9	—
	Time (min)	7.21		—	6.24		—	5.13		—	6.27		—
DS3 (50 iter.)	Error (%)	17.9	21.3	25.6	21.0	23.8	26.1	26.8	29.9	31.0	20.9	24.1	25.9
	Time (min)	18.02			15.81			13.00			16.43		

Note. Error values are computed for actual versus identified damage indices (theoretical stiffness).

Table 9: Damage identification results corresponding to DS1, DS2, and DS3 related to the updated stiffness values.

Element no.		Identification											
		WCA			PSO			ICA			DE		
		4	12	5	4	12	5	4	12	5	4	12	5
DS1	Error (%)	1.2	—	—	2.1	—	—	3.8	—	—	3.1	—	—
DS2	Error (%)	1.6	3.3	—	3.2	4.1	—	13.6	16.3	—	13.1	15.7	—
DS3	Error (%)	1.7	3.6	3.6	7.7	9.4	9.2	17.5	19.7	20.8	17.2	18.3	18.0

Note. Error values are computed for actual versus identified damage indices (updated stiffness).

Table 10: The comparison of standard deviation (SD) results obtained for the relative error values (%) in identified damages corresponding to DS1, DS2, and DS3 after 5 repeats using WCA, PSO, ICA, and DE optimization strategies.

Element no.		Damage identification											
		WCA			PSO			ICA			DE		
		4	12	5	4	12	5	4	12	5	4	12	5
DS1	SD	0.1	—	—	0.4	—	—	0.4	—	—	0.9	—	—
DS2	SD	0.35	0.45	—	0.55	0.75	—	1.15	0.8	—	1.65	1.4	—
DS3	SD	1.03	1.07	0.94	1.34	1.22	1.25	1.4	1.5	1.5	2.2	2.4	2.3

Note. Error values are computed for actual versus identified damage indices (updated stiffness).

datasets. In order to be more applicable in large scale structural systems, an efficient artificial data generation technique is proposed, yielding a significant reduction in multiclass labels corresponding to large datasets. The most effective and damage sensitive features are extracted using a simple data fusion strategy. In this regard, 3D acceleration signals and the resulting wavelet coefficients of signals, which are obtained by the operation of the Chebyshev wavelet on precise collocation points (to capture the entire frequency contents of associated time-history signals), are concatenated into relative 2D images. Accordingly, each feature array is constructed based on the fitness of such time-domain images attained for measured data compared to those for simulated data, taking into account different damage states. Furthermore, the wavelet-based energy of signals is computed and will form the last two components of each feature array. It is concluded that, not only the extracted features have the less effect of signal-to-noise ratio but also the application of the proposed data fusion method dramatically reduces the feature size and will result in a considerable improvement in the performance of data training. In addition, a reasonably fast convergence is achieved by employing bagged tree learners compared to the other classifiers.

On the other hand, the diverse operations of the proposed multispecies genetic algorithm overcome the main shortcoming regarding the local optimal solutions. Therefore, damage locations are confidently detected within an acceptable, fast computational time. This introduces the proposed damage localization scheme as one of the more practical ones for dealing with large-scale problems. It is concluded that, for large-scale structures, the computational performance of the proposed strategy in the model training part is considerably better than that of the other available classifiers. The later step involves the application of evolutionary optimization strategies, i.e., WCA, PSO, ICA, and DE, for structural damage identification. Due to the inherent formulation of WCA, the exploitation and exploration phases are thoroughly proceeded and it is concluded that, even in the case of multiple damage scenarios in large structures, WCA consistently produces the best identification results. For instance, in dealing with 128 unknown structural elements, an average accuracy of 91.4% is achieved by the proposed method using WCA in the damage identification of multiple scenarios with no false detection, compared to 82%, 75.7%, and 68.9% obtained for the application of PSO, DE, and ICA, respectively.

Moreover, the proposed damage detection strategy may be considered as a data-driven scheme. Taking into account the necessity for early damage detection in large scale structures, the underlying strength point lies in the remarkably fast computational time involved and the less effects of uncertain and unpredicted sources, i.e., environmental noise, finite element modeling errors, and so forth.

6. Conclusions

This paper introduces a synthesis approach for time-domain damage localization and identification in structural systems. The main conclusions from the results of this study can be drawn as follows:

(i) Employing the improved multispecies binary genetic algorithm in conjunction with a supervised machine learning-based classification strategy leads to a data-driven technique and yields a cost-effective and reliable damage localization approach with localization accuracy down to a single structural member in dealing with large-scale structures.

(ii) The most damage-sensitive features are extracted using wavelet coefficients and wavelet energies of output signals (at different decomposition levels) through an efficient data fusion technique. It is noted that the extracted features have a lower signal-to-noise ratio.

(iii) The superiority of the proposed data fusion technique lies in its ability to reduce the size of large time-domain features while capturing the entire dynamic behavior of structures regardless of the sensor configurations. In addition, artificial features are generated based on the structural subdomains, and the class labels are accordingly updated. As a consequence, the performance and computational cost of the model training stage are significantly enhanced using ensemble bagged tree classifiers.

(iv) The structural damage identification is satisfactorily accomplished using the water cycle optimization algorithm compared to the other state-of-the-art optimization strategies.

(v) Based on the results obtained for the numerical verification as well as the experimental validation studies, one may deduce the robustness of the proposed approach for damage identification of real-world structural systems. More importantly, the proposed feature extraction and data fusion strategies constitute the main merit of the proposed procedure over the existing methods and make it applicable for vision-based measured data capable of being utilized by remote digital clones in a variety of real-world structural applications.

Data Availability

The data used in this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors wish to acknowledge the financial supports from the Higher Education Complex of Bam, Bam, Iran (Grant no. 81167051400).

References

[1] M. J. Perry, C. G. Koh, and Y. S. Choo, "Modified genetic algorithm strategy for structural identification," *Computers & Structures*, vol. 84, no. 8-9, pp. 529–540, 2006.

[2] Z. Xing and A. Mita, "A substructure approach to local damage detection of shear structure," *Structural Control and Health Monitoring*, vol. 19, no. 2, pp. 309–318, 2012.

[3] J. L. J. Pereira, M. B. Francisco, S. S. D. Cunha Jr, and G. F. Gomes, "A powerful Lichtenberg optimization algorithm: a damage identification case study," *Engineering Applications of Artificial Intelligence*, vol. 97, no. 97, Article ID 104055, 2021.

[4] J. F. Lin, Y. L. Xu, and S. S. Law, "Structural damage detection-orientedmulti-type sensor placement with multi-objective optimization," *Journal of Sound and Vibration*, vol. 422, pp. 568–589, 2018.

[5] G. F. Gomes, F. A. de Almeida, S. S. da Cunha, and A. C. Ancelotti, "An estimate of the location of multiple delaminations on aeronautical CFRP plates using modal data inverse problem," *International Journal of Advanced Manufacturing Technology*, vol. 99, no. 5-8, pp. 1155–1174, 2018.

[6] H. P. Chen, *Structural Health Monitoring of Large Civil Engineering Structures*, Wiley-Blackwell, Hoboken, NY, USA, 2018.

[7] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "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*, vol. 147, Article ID 107077, 2021.

[8] Z. Zhang and C. Sun, "Structural damage identification via physics-guided machine learning: a methodology integrating pattern recognition with finite element model updating," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1675–1688, 2021.

[9] P. J. Chun, H. Yamashita, and S. Furukawa, "Bridge damage severity quantification using multipoint acceleration measurement and artificial neural networks," *Shock and Vibration*, vol. 2015, Article ID 789384, 11 pages, 2015.

[10] R. Almeida Cardoso, A. Cury, F. Barbosa, and C. Gentile, "Unsupervised real-time SHM technique based on novelty indexes," *Structural Control and Health Monitoring*, vol. 26, no. 7, Article ID e2364, 2019.

[11] M. H. Rafiei and H. Adeli, "A novel unsupervised deep learning model for global and local health condition assessment of structures," *Engineering Structures*, vol. 156, pp. 598–607, 2018.

[12] G. Mariniello, T. Pastore, C. Menna, P. Festa, and D. Asprone, "Structural damage detection and localization using decision tree ensemble and vibration data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 9, pp. 1129–1149, 2021.

[13] J. Li, H. Hao, R. Wang, and L. Li, "Development and application of random forest technique for element level structural damage quantification," *Structural Control and Health Monitoring*, vol. 28, no. 3, Article ID e2678, 2021.

[14] Z. Nie, E. Guo, J. Li, H. Hao, H. Ma, and H. Jiang, "Bridge condition monitoring using fixed moving principal component analysis," *Structural Control and Health Monitoring*, vol. 27, no. 6, Article ID e2535, 2020.

[15] P. J. Chun, I. Ujike, K. Mishima, M. Kusumoto, and S. Okazaki, "Random forest-based evaluation technique for internal damage in reinforced concrete featuring multiple nondestructive testing results," *Construction and Building Materials*, vol. 253, Article ID 119238, 2020.

[16] Q. Zhou, Y. Ning, Q. Zhou, L. Luo, and J. Lei, "Structural damage detection method based on random forests and data fusion," *Structural Health Monitoring*, vol. 12, no. 1, pp. 48–58, 2013.

[17] Q. Zhou, H. Zhou, Q. Zhou, F. Yang, and L. Luo, "Structure damage detection based on random forest recursive feature elimination," *Mechanical Systems and Signal Processing*, vol. 46, no. 1, pp. 82–90, 2014.

[18] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*Springer, Berlin/Heidelberg, Germany, 2000.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation*, vol. 9, no. 7, pp. 1545–1588, 1997.

[21] R. F. R. Junior, I. A. D. S. Areias, M. M. Campos, C. E. Teixeira, L. E. B. da Silva, and G. F. Gomes, "Fault detection and diagnosis in electric motors using 1D convolutional neural networks with multi-channel vibration signals," *Measurement*, vol. 190, no. 190, Article ID 110759, 2022.

[22] T. Guo, L. Wu, C. Wang, and Z. Xu, "Damage detection in a novel deep-learning framework: a robust method for feature extraction," *Structural Health Monitoring*, vol. 19, no. 2, pp. 424–442, 2020.

[23] M. Fallahian, E. Ahmadi, and F. Khoshnoudian, "A structural damage detection algorithm based on discrete wavelet transform and ensemble pattern recognition models," *Journal of Civil Structural Health Monitoring*, vol. 12, no. 2, pp. 323–338, 2022.

[24] Z. Chen, Y. Wang, J. Wu, C. Deng, and K. Hu, "Sensor data-driven structural damage detection based on deep convolutional neural networks and continuous wavelet transform," *Applied Intelligence*, vol. 51, no. 8, pp. 5598–5609, 2021.

[25] S. Beskhyroun, T. Oshima, and S. Mikami, "Wavelet-based technique for structural damage detection," *Structural Control and Health Monitoring*, vol. 17, no. 5, pp. 473–494, 2010.

[26] S. H. Mahdavi, F. R. Rofooei, A. Sadollah, and C. Xu, "A wavelet-based scheme for impact identification of framed structures using combined genetic and water cycle algorithms," *Journal of Sound and Vibration*, vol. 443, pp. 25–46, 2019.

[27] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, Berlin, Germany, 2009.

[28] M. Paluszek and S. Thomas, *MATLAB Machine Learning*, Apress, Pune, India, 2016.

[29] S. H. Mahdavi and H. Abdul Razak, "A wavelet-based approach for vibration analysis of framed structures," *Applied Mathematics and Computation*, vol. 220, pp. 414–428, 2013.

[30] S. H. Mahdavi and H. Abdul Razak, "Indirect time integration scheme for dynamic analysis of space structures using wavelet functions," *Journal of Engineering Mechanics*, vol. 141, no. 7, Article ID 04015006, 2015.

[31] A. Silik, M. Noori, W. A. Altabey, and R. Ghiasi, "Selecting optimum levels of wavelet multi-resolution analysis for time-varying signals in structural health monitoring," *Structural Control and Health Monitoring*, vol. 28, no. 8, Article ID e2762, 2021.

[32] J. G. Han, W. X. Ren, and Z. S. Sun, "Wavelet packet based damage identification of beam structures," *International Journal of Solids and Structures*, vol. 42, no. 26, pp. 6610–6627, 2005.

[33] S. A. Broughton and K. Bryan, *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*, John Wiley & Sons, Hoboken, NY, USA, 2018.

[34] S. H. Mahdavi and H. A. Razak, "Optimal sensor placement for time-domain identification using a wavelet-based genetic algorithm," *Smart Materials and Structures*, vol. 25, no. 6, Article ID 065006, 2016.

[35] A. Sadollah, H. Eskandar, A. Bahreininejad, and J. H. Kim, "Water cycle algorithm with evaporation rate for solving constrained and unconstrained optimization problems," *Applied Soft Computing*, vol. 30, pp. 58–71, 2015.

[36] C. G. Koh, L. M. See, and T. Balendra, "Estimation of structural parameters in time domain: a substructure approach," *Earthquake Engineering & Structural Dynamics*, vol. 20, no. 8, pp. 787–801, 1991.

[37] T. N. Trinh and C. G. Koh, "An improved substructural identification strategy for large structural systems," *Structural Control and Health Monitoring*, vol. 19, no. 8, pp. 686–700, 2012.

[38] S. H. Mahdavi, *Structural Health Monitoring Using Adaptive Wavelet Functions*, Doct. Diss. UM, Kuala Lumpur, Malaysia, 2016.