Article

Identification of Bolt Loosening Damage of Steel Truss Structure Based on MFCC-WPES and Optimized Random Forest

Zepu Jiang, Zhiwei Zhu and Debing Zhuo *

School of Civil Engineering and Architecture, Jishou University, Zhangjiajie 427000, China
* Correspondence: zhuodebing2004@jsu.edu.cn

Abstract: In the field of bolt loosening detection, although some progress has been made, there are still challenges such as high operational complexity, single feature extraction methods, and insufficient analysis model performance, especially in large steel truss structures, where there is a lack of efficient and accurate bolt loosening identification solutions. In response to these shortcomings, this article proposes an innovative bolt loosening damage recognition method based on sound signals. This method integrates feature extraction techniques of Mel frequency cepstral coefficients (MFCCs) and wavelet packet energy spectra (WPES), and comprehensively characterizes sound signals by constructing MFCC-WPES combined features. Subsequently, the random forest (RF) algorithm optimized by genetic algorithm was used for feature selection and model training, aiming to improve recognition accuracy and robustness. The experimental results show that this method can not only accurately identify bolt loosening signals in steel truss structure bolt loosening detection, but also has strong identification ability for environmental noise. Compared with traditional methods, the proposed solution in this article shows significant improvements in both performance and practicality, providing a new perspective and solution for the technological advancement of bolt loosening detection in steel truss structures.

Keywords: steel truss; bolt loosening; damage identification; MFCC; random forest

1. Introduction

Bolted connections are a widely used form of structural connection in steel truss structures. When bolted connections are subjected to mechanical actions such as vibration, impact, and creep, their connection state often changes, resulting in loosening, slippage, or even detachment. This can compromise the integrity of the structure and endanger its reliability and safety. Therefore, monitoring the status of bolt connections is of great significance for ensuring structural safety, preventing major safety accidents, and reducing the severity of accidents. Traditional bolt loosening damage detection methods can be divided into two categories: based on modality and based on impedance [1–5]. Both of these methods adopt a contact measurement approach, where the sensor typically needs to be located near the surface of the point being measured and properly attached. This is not easily achievable for a steel truss structure with a large number of connection nodes and bolts [6].

Sound is essentially the propagation of vibrations through the air, and similar to vibration signals, sound can serve as an important source of information reflecting the state of damage. Just as an auto mechanic can diagnose the type of car trouble based on the sound of the engine, railway maintenance workers often use hammers to strike the rails, judging whether there are any loose bolts by the sound. Cableway maintenance workers also frequently determine whether the bolts on the passenger cableway support rods are loose by listening to the sounds. Sound propagation has advantages such as omnidirectional spread, high temporal resolution, and no visibility limitations. At the same time, sound measurement is a non-contact measurement method. The microphone

array measurement device used is relatively low-cost and easy to install, and it can effectively avoid the saturation and failure phenomena that are common with accelerometer measurements [7–9].

Wang et al. [10] proposed a percussion-based cup-lock scaffold loosening detection method using a convolutional bidirectional long short-term memory (CBLSTM) model to classify the percussion acoustic signal features, which provides a new idea for the stability monitoring of scaffolding systems. Next, to effectively prevent the bolted joint from loosening underwater, He et al. [11] proposed a detection method utilizing the knocking sound and KNN algorithm. Liu et al. [12] further investigated the knocking method of bolting loosening monitoring in the presence of noise interference and data insufficiency, and significantly improved the monitoring accuracy by proposing noise interference-resistant speech features and prototype network classifiers. Focusing on bolt loosening monitoring in the earliest stage, Yuan et al. [13] combined variational modal decomposition and residual network (ResNet) to achieve effective monitoring based on intrinsic feature extraction of knocking sound. Chen et al. [14] innovatively used a feature-reduced multiple stochastic convolutional kernel transform (FM-ROCKET) model in underwater bolt flange loosening detection, combining deep learning and shallow learning methods, which significantly improves the effect of detection. Regarding the issue of uneven knocking force, Wang et al. [15,16] proposed a force-adaptive percussion method for assessing bolt loosening, which utilizes sound phase features to mitigate the impact of irregular percussion forces, and also developed a rapid detection method for multiple-bolt loosening in steel beam–column joints based on percussion techniques, enhancing the application potential of percussion-based methods in engineering.

In the field of bolt loosening detection, although scholars have made significant progress through diversified methods and techniques, it still faces a series of urgent problems. Firstly, in terms of motivation, the current mainstream research generally adopts the knocking method. Despite its proven effectiveness, it has encountered practical bottlenecks in real-world applications. Specifically, the tapping operation needs to be carried out directly or in close proximity to the bolt, which is difficult realize in spatially restricted or difficult scenarios. Secondly, there are also limitations in feature extraction. Most studies have focused on MFCC or power spectrum features, which are relatively homogeneous and poor at comprehensively and accurately capturing the complex signal characteristics generated when a bolt is loosened, which in turn may affect the accuracy of the detection results. Furthermore, the analysis models are also insufficient. Existing models either rely on traditional machine learning algorithms such as Bayes or KNN, simple but with limited recognition accuracy, or use deep neural networks that can improve recognition accuracy, but are accompanied by high data requirements, computational resource consumption, tuning difficulty, training time, and deployment complexity. Of particular concern is that there is a gap in current research and a lack of targeted solutions for the identification of loose bolts in large steel truss structures, such as transmission towers and cableway towers. In view of the above status quo and deficiencies, this paper focuses on the damage detection of bolted connections in steel truss structures, and innovatively proposes a recognition method based on acoustic signals. The method integrates MFCC-WPES feature extraction and Random Forest algorithm, aiming to overcome the limitations of existing techniques. In order to verify the effectiveness and practicality of the method, this paper also designs and implements an experiment on the loosening of bolts in steel joist structures, aiming to contribute to the technological advancement in this field.

2. Principles and Methods

2.1. Signal Preprocessing and Feature Extraction

2.1.1. Preprocessing

Most sound signals are non-stationary, but they typically exhibit short-term stationarity characteristics. Therefore, by applying frame-based windowing operations, these signals can be segmented into short segments and processed as stationary signals.

Assuming $y(n)$ represents the discrete sequence of sound signals received by the microphone, $w(n)$ represents the window function, N represents the total number of frames after segmentation, L represents the frame length, and *inc* represents the frame shift, then the *i*th frame signal $y_i(n)$ can be represented as:

$$y_i(n) = \omega(n) * y((i-1) * inc + n), 1 \le i \le N \tag{1}$$

The short-time energy of the *i*th frame signal is denoted by e_i, and the expression is as follows:

$$e_i = \sum_{n=0}^{L-1} y_i^2(n) \; 1 \le i \le N \tag{2}$$

In general, the frame length L is typically set to 10 to 30 milliseconds. The ratio of frame shift to frame length is typically taken between 0 and 1/2.

2.1.2. MFCC Feature Extraction

MFCC is a commonly used feature in the field of radio recognition. They are primarily extracted based on Mel filters, which are designed to mimic the characteristics of the human auditory system. Due to the human ear's sensitivity to low-frequency sounds and relative insensitivity to high-frequency ones, Mel filters are characterized by being dense in the low-frequency region and sparse in the high-frequency region [17,18]. The approximate expression for the conversion between Mel frequency and linear frequency is as follows:

$$mel(f) = 2595.lg\left(1 + \frac{f}{700}\right) \tag{3}$$

In the equation, f represents the linear frequency, measured in Hertz (Hz).

The process of MFCC feature extraction includes preprocessing, Fast Fourier Transform (FFT), Mel filtering, logarithm operation, Discrete Cosine Transform (DCT), and dynamic feature extraction, as illustrated in Figure 1.
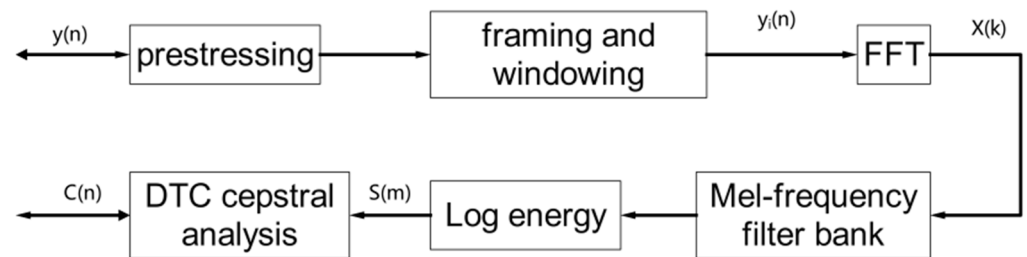

Figure 1. The extraction process of MFCC.

Firstly, the sound signal is segmented into frames and windowed. Then, FFT is applied to each frame of the sound signal to obtain its spectrum. Next, the spectrum is passed through a series of Mel filters to filter out energy within specific frequency ranges. The outputs of the filters are then subjected to logarithmic operation to enhance the features of low-energy frequency bands. Finally, the logarithmic energy vectors obtained from the Mel filterbanks are subjected to Discrete Cosine Transform (DCT). By setting the number of outputs of the DCT, different dimensional MFCC features can be obtained.

$$C(n) = \sum_{m=0}^{N-1} s(m)cos\left(\frac{\pi n(m-0.5)}{M}\right), n = 0, 1, 2..., L \tag{4}$$

In the equations, M represents the number of filters; N denotes the number of points in the FFT; L represents the order of the MFCC features; $s(m)$ represents the logarithmic energy output of each filterbank.

2.1.3. WPES Feature Extraction

The structural damage information can typically be characterized by the distribution of Wavelet Packet Energy Spectrum (WPES) [19,20].

The original signal undergoes a wavelet packet decomposition into n levels, yielding 2^n signal frequency components $\{Y_1, Y_2, \ldots, Y_{2^n}\}$. Each component Y_i ($i = 1, 2, \ldots, 2^n$) can be represented as:

$$Y_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,L}] \tag{5}$$

In this equation, y represents the amplitude of the reconstructed signal at the discrete points, L corresponds to the number of samples in a frame, i.e., the frame length, and i is the frequency band index. Therefore, the energy E_i of the ith component Y_i is given by:

$$E_i = \parallel Y_i \parallel_2^2 = y_{i,1}^2 + y_{i,2}^2 + \cdots + y_{i,L}^2 \tag{6}$$

Subsequently, according to Equation (6), a feature vector can be constructed with each frequency band energy as its element:

$$E = [E_1, E_2, \ldots, E_{2^n}] \tag{7}$$

To better illustrate the energy difference between intact and damaged signal samples, this paper employs a dimensionless energy ratio R as the feature vector:

$$R = [R_1, R_2, \ldots, R_{2^n}] \tag{8}$$

where the expression for the energy ratio R_i for each frequency band is:

$$R_i = \frac{E_i}{\sum_{i=1}^{2^n} E_i} \tag{9}$$

2.2. Principles of Optimized Random Forest Algorithm

2.2.1. Principles of Random Forest Algorithm [21–23]

Definition 1. *A Random Forest is an ensemble classifier consisting of a group of decision tree classifiers $\{h(X, \theta_k), k = 1, 2, \ldots, K\}$. Each tree is formed by a random vector $\{\theta_k\}$, which is independently and identically distributed. K denotes the number of decision trees in the random forest. For a given independent variable X, each decision tree classifier votes to decide the optimal classification result.*

Essentially, a random forest is a classifier that aggregates multiple decision trees; if a decision tree is considered an expert in the classification task, then a random forest is a collection of such experts collaboratively classifying a given task.

The steps to generate a random forest are as follows:

Step 1. From the original training dataset, apply the bootstrap method to randomly draw K new bootstrap samples with replacement, constructing K classification regression trees. Each time, the samples not drawn form the K out-of-bag (OOB) datasets.

Step 2. If there are n features, randomly select m_{try} features ($m_{try} \leq n$) at each node of each tree. Compute the information contained in each feature and select the one with the highest classification capability for node splitting.

Step 3. Grow each tree to the maximum extent without pruning.

Step 4. Combine the generated trees to form a random forest. Classify new data using this forest, with the classification result determined by the majority vote of the tree classifiers.

Definition 2. *Given a set of classifiers h1(X), h2(X), . . ., hk(X), where each classifier's training set is randomly sampled from the original dataset (Y, X) that follows a random distribution, the margin function is defined as:*

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \tag{10}$$

where $I(\cdot)$ is the indicator function.

The margin function measures the extent to which the average number of correct classifications exceeds the average number of incorrect classifications. The larger the margin value, the more reliable the classification prediction.

Definition 3. *The generalization error is defined as:*

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \tag{11}$$

where P covers the space of X and Y.

In a random forest, when there are enough decision tree classifiers, $h_k(X) = h(X, \theta_k)$ follows the strong law of large numbers.

Theorem 1. *As the number of decision trees in the random forest increases, for all sequences $\theta_1, \theta_2, \cdots, \theta_k$, PE^* almost surely converges to:*

$$P_{X,Y} \left\{ P_\theta(h(X, \theta) = Y) - \max P_\theta \left(h(X, \theta) = j \right) < 0 \right\} \tag{12}$$

Theorem 1 indicates that the random forest does not suffer from the overfitting problem as the number of decision trees increases, but it may have a certain degree of generalization error within limits.

2.2.2. Optimized Strategies

(1) Feature Importance Evaluation and Feature Selection

Evaluating the importance of various features helps identify the most influential features for the prediction target. This paper adopts the variable importance measure based on out-of-bag (OOB) data classification accuracy.

Definition 4. *The variable importance measure based on OOB data classification accuracy is defined as the average reduction in classification accuracy before and after a slight perturbation of the OOB data's independent variable values.*

Suppose there are bootstrap samples $b = 1, 2, \ldots, B$, where B represents the number of training samples. The variable importance measure $\overline{D_j}$ of feature X_j is calculated as follows:

Step 1. Set $b = 1$, create a decision tree T_b on the training sample, and mark the OOB data as L_b^{oob}.

Step 2. Use T_b to classify L_b^{oob} data, and count the number of correctly classified instances, denoted as R_b^{oob}.

Step 3. For feature X_j, $j = 1, 2, \ldots, N$, perturb the value of X_j in L_b^{oob}, and denote the perturbed dataset as L_{bj}^{oob}. Use T_b to classify L_{bj}^{oob} data and count the number of correctly classified instances, denoted as R_{bj}^{oob}.

Step 4. For $b = 2, 3, \ldots, B$, repeat steps 1–3.

Step 5. Feature importance calculation: The variable importance measure $\overline{D_j}$ of feature X_j is calculated using the following formula:

$$\overline{D_j} = \frac{1}{B} \sum_{i=1}^{B} \left(R_b^{oob} - R_{bj}^{oob} \right) \tag{13}$$

Step 6. Feature Ranking: Rank all features in descending order based on their importance measures.

Step 7. Feature Selection: Select the top-ranked features for model training and prediction.

(2) Random Forest Parameter Optimization

The performance of the random forest model is mainly influenced by the number of trees n, the maximum depth of each decision tree d, the minimum number of samples required to split an internal node s, and the maximum number of leaf nodes m. Traditional methods use grid search for parameter optimization, which involves calculating the objective function value for each parameter combination by traversing each grid point. This approach often becomes trapped in local optima and has high computational complexity. To improve the efficiency and accuracy of parameter optimization, this paper employs a genetic algorithm for adaptive parameter optimization.

As a classical intelligent algorithm (such as Harmony search algorithm [24,25], cuckoo search algorithm [26,27], and Sparrow Search Algorithm [28]), the Genetic Algorithm (GA) is considered one of the most classical heuristic algorithms, with perfect performance. The GA combines mathematical analysis and computer simulation methods, transforming the optimization problem-solving process into mechanisms similar to Darwin's theory of natural selection and genetic mechanisms like chromosome gene crossover and mutation. Compared to some conventional optimization algorithms, this approach can often achieve better results faster, making it suitable for solving complex combinatorial optimization problems [29]. The steps for optimizing the random forest algorithm based on the genetic algorithm are as follows.

Step 1. Initialization: Randomly generate a set of parameter combinations containing n, d, s, and m to form the initial population.

Step 2. Evaluation and Selection: Apply the random forest algorithm to each parameter combination in the initial population. Use classification accuracy as the evaluation criterion, calculate the classification accuracy of each parameter combination, and select the best parameter combination. Classification accuracy is defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP (true positive) represents correct positive predictions; TN (true negative) represents correct negative predictions; FP (false positive) represents incorrect positive predictions; FN (false negative) represents incorrect negative predictions.

Step 3. Genetic Operations and Iterations: Perform crossover operations on the selected excellent parameter combinations to generate new potential excellent parameter combinations. Optionally introduce mutation operations to increase population diversity and prevent local optima. Form a new generation population with the newly generated parameter combinations and the excellent individuals from the parent generation. Repeat evaluation, selection, and genetic operations until the preset maximum number of iterations or until the fitness improvement threshold is met.

Step 4. Output Results: Select the parameter combination with the highest fitness (highest accuracy) from the final population. Output this optimal parameter set as the best configuration for the random forest algorithm.

2.3. Identification Process of Bolt Loosening Damage in Steel Truss Structures Utilizing MFCC-WPES and Optimized Random Forest

The structural diagram of the steel truss bolt loosening damage identification method proposed in this article is shown in Figure 2. Firstly, the sound signals collected under the set working conditions are subjected to preprocessing through framing, windowing, and feature extraction. Mel Frequency Cepstral Coefficient (MFCC) and Wavelet Packet Energy Spectrum (WPES) features are extracted from the signals. Subsequently, the MFCC and WPES features are jointly used to construct the MFCC-WPES combined feature training dataset. The obtained combined features are then utilized to train a Random Forest (RF) classifier for sensitive feature selection and damage signal identification. Utilizing random forest for feature selection generally involves two main processes: feature importance evaluation and feature elimination. After obtaining the importance of each feature and arranging them in descending order, a certain proportion of redundant features are eliminated to obtain a new feature set. This new set is then inputted into the random forest model for training. Concurrently, a genetic algorithm is employed to optimize the parameters n, d, s, and m of the random forest model; this optimization aims to find the best parameter combination, ultimately resulting in the optimal recognition model. To enhance the efficiency of model testing, a preliminary assessment is conducted using short-time energy. Signal frames with short-time energy greater than a predetermined threshold T are selected for analysis. These selected frames are then inputted into the random forest model for testing. This approach ultimately achieves the goal of quickly identifying the sound signals indicating loose bolt damage.
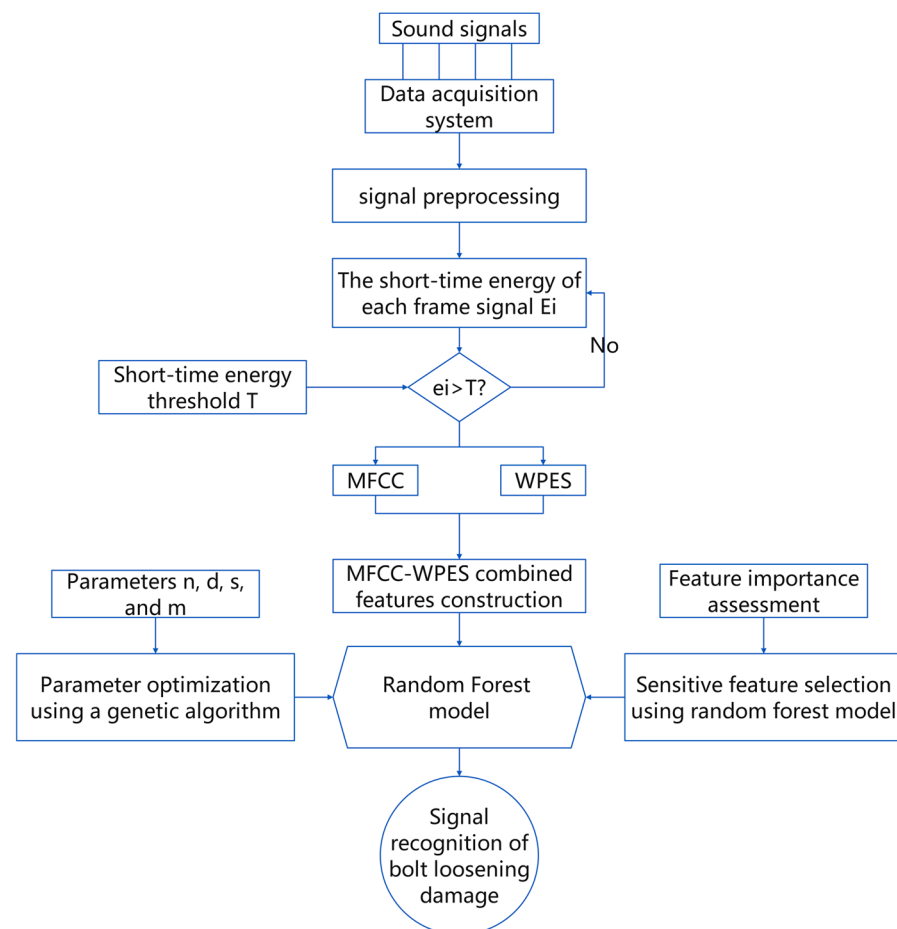

Figure 2. Structural diagram of bolt loosening identification method for steel truss structure.

3. Steel Truss Experiment

3.1. Experiment Introduction

The experimental model is a hexagonal steel truss structure with bolted connections located on an outdoor floor, as shown in Figure 3. The model has a side length of 1.5 m and a height of 6 m, divided into four layers. The distance from the ground to the bottom layer is 0.5 m, while the height of the other layers is 1.8 m. The model columns are made of Q235 seamless steel pipes with a diameter of 16 cm. All horizontal and diagonal members are made of equal angle steel L40 × 4. Each end of the members is connected to the column node plate using two M20 bolts. Each column is connected to the ground using four anchor bolts. There are six nodes on each level of the model, with a total of twenty-four connection nodes across four levels. For the convenience of subsequent analysis, all nodes are uniformly numbered, as shown in Figure 3f. For example, "J_{i4}" represents the fourth node on the ith level.
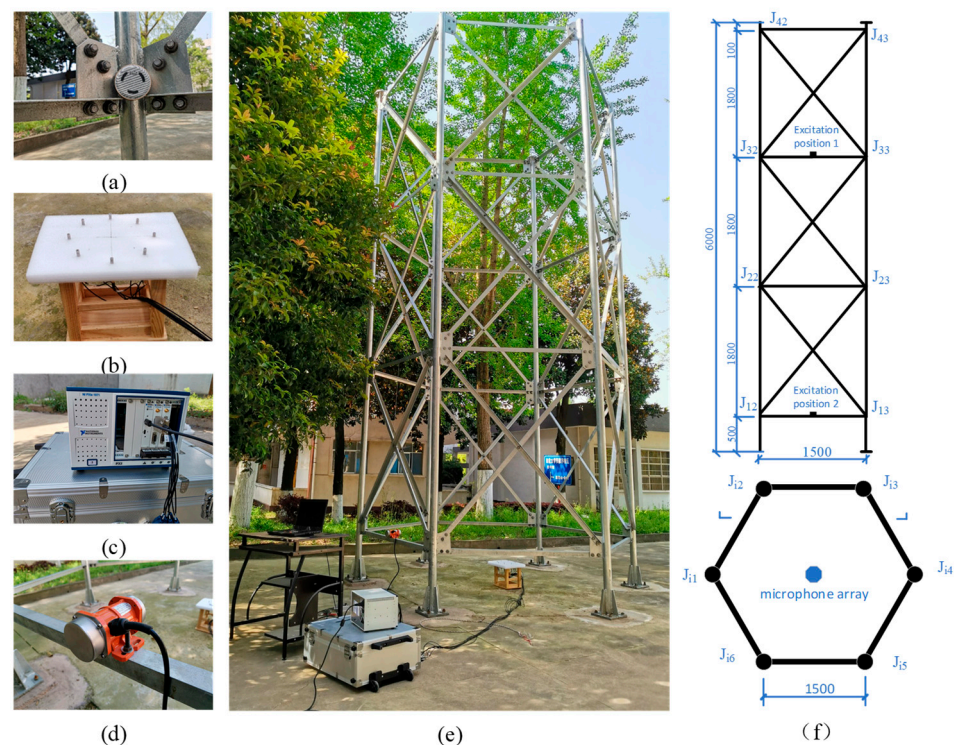

Figure 3. Experimental model and test system composition: (**a**) node and sound source, (**b**) 8-element ring microphone array, (**c**) PXI dynamic signal test analyzer, (**d**) small eccentric vibrator, (**e**) experimental model, and (**f**) dimensions of experimental model.

The experimental data acquisition system consists of sound pickup devices, data acquisition equipment, and a computer. The sound pickup device consists of an eight-element circular microphone array placed at the central position of the bottom of the model. The array has a diameter and height of 20 cm, with MPA416 omnidirectional microphones used. The data acquisition module utilizes a PXI dynamic signal analyzer, as shown in Figure 3b,c.

Experimental steps: (1) Simulation of damage condition and noise condition: Loose bolts at different positions and degrees are simulated to mimic the condition of loose bolt damage, as shown in Table 1. A Bluetooth speaker controlled by a smartphone placed at node J_{12} at the bottom layer of the model simulates seven common environmental sounds: the cicada sound, the construction noise, the wind sound, the thunder sound, birdsong, the rain sound, and speech, as shown in Figure 3a. (2) Model excitation: A small eccentric vibrator was used to excite the model to simulate environmental vibration. By adjusting the motor frequency, significant vibration of the model was achieved. When testing the

nodes of the first and second layers, the vibrator was fixed at the crossbar position between nodes J_{12} and J_{13}. When testing the nodes of the third and fourth layers, the vibrator was fixed at the crossbar position between nodes J_{32} and J_{33}, as shown in Figure 3d–f. (3) Signal acquisition: We used a PXI dynamic signal analyzer to collect sound signals from both the bolt loosening damage condition and seven different ambient noise conditions, with a signal sampling rate of 20 kHz.

Table 1. Simulation of bolt loosening damage.

Loose Position	J_{15}	J_{15}	J_{15}	J_{15}	J_{15}	J_{24}	J_{24}	J_{36}	J_{36}	J_{42}
Number of looseness	1	1	1	2	3	1	2	1	2	1
Degree of looseness	0.5 turn	1 turn	1.5 turn	0.5 turn	0.5 turn	0.5 turn	0.5 turn	0.5 turn	0.5 turn	0.5 turn

3.2. Sound Signal Feature Extraction

The collected sound signals from each condition are preprocessed by framing and windowing. The framing parameters are set as follows: frame length of 512 samples, frame shift of 256 samples, and Hanning window function. When the length of the sound signal is set to 1 s, each segment of the signal consists of 77 frames of valid signals. Figures 4–6, respectively, show a typical waveform of loose bolt damage signal, waveforms of two types of strong environmental interference noise (cicada sound and rain sound), short-time energy, and the distribution of MFCC and wavelet packet energy bands of the frame signal with maximum short-time energy.
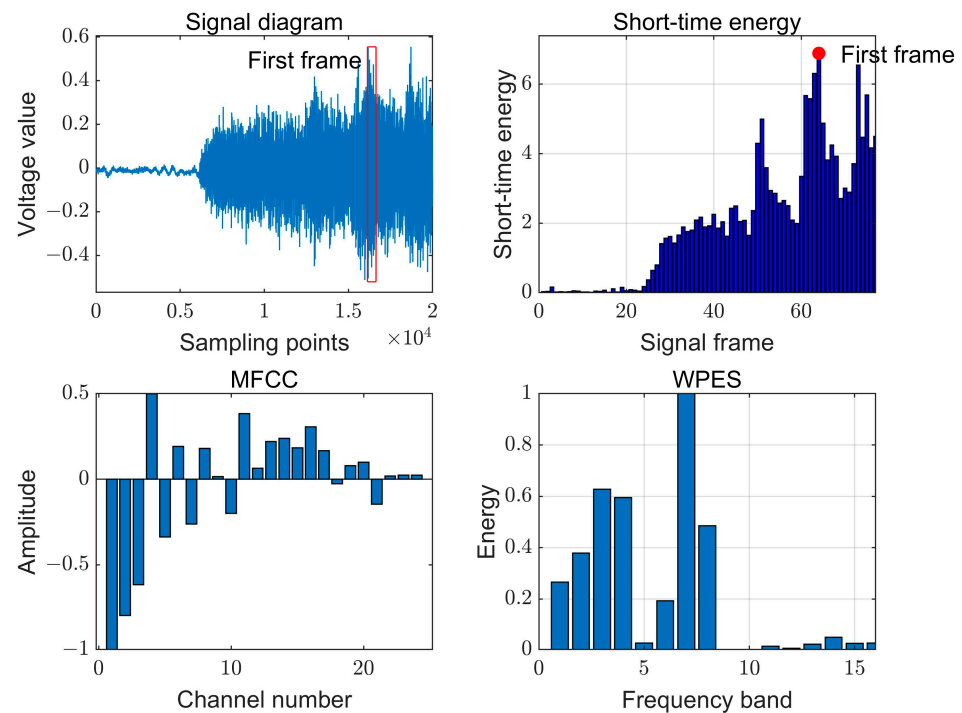

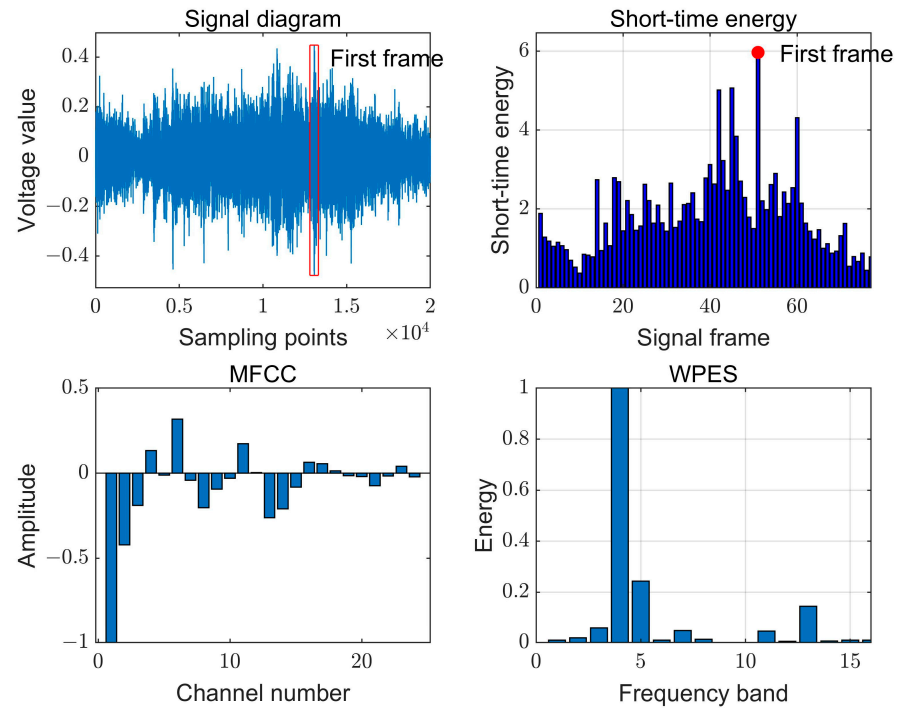**Figure 4.** Bolt loosening signal map and feature extraction results.

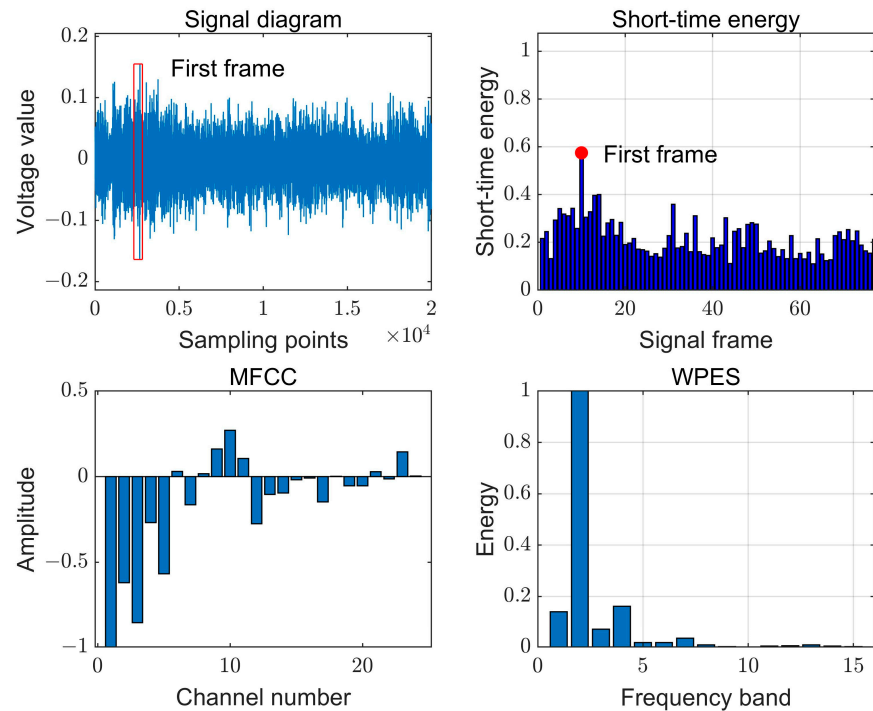**Figure 5.** Cicada chirping signal map and feature extraction results.

Figure 6. Rain sound signal map and feature extraction results.

From Figure 4, it is evident that the short-time energy of the sound signal emitted from the loose bolt is generally higher than that of the exciter signal (in the first 0.3 s segment). Therefore, to exclude the influence of the exciter signal on the recognition results during specific analysis, an initial judgment is made using short-time energy. Frames with short-time energy greater than the warning threshold T are selected, and the signal analysis is performed sequentially in descending order (in this experiment, T is set to 0.5). Compared to the loose bolt damage signal, there are significant differences in the distribution patterns of MFCC and wavelet packet energy bands for the two types of typical environmental

interference noise. From the distribution of MFCC, it can be observed that the amplitudes of the three types of signals are both positive and negative. When the channel number is relatively small, the amplitude is negative and relatively large. As the channel number increases, the absolute value of the amplitude generally decreases. From the perspective of wavelet packet band energy distribution, the sound of cicadas and the sound of rain have a relatively narrow distribution range. The former is mainly distributed in the fourth frequency band, while the latter is primarily distributed in the second frequency band. In contrast, the signal of bolt loosening has a relatively wide distribution, spreading across the first to fourth and sixth to eighth frequency bands. Overall, the MFCCs of various signals and the energy distribution of wavelet packet frequency bands are not only identical, but it is also difficult to directly discern their patterns through manual judgment, thus distinguishing the sound signals of bolt loosening under different working conditions.

3.3. Signal Recognition and Analysis

3.3.1. Dataset Construction

In this experiment, sound signals from a total of eight conditions, including loose bolt damage and environmental noise mentioned above, were collected. For each condition, typical signal lengths totaling 10 s were collected. After framing, 780 frames of signals were obtained. The first 400 frames of signals with relatively high short-time energy were selected as sample data. The sample set was divided into training and validation sets using a random split ratio of 7:3. Therefore, the total number of samples for the eight conditions is 3200, with 2240 samples in the training set and 960 samples in the validation set. The eight operating conditions are identified by the numbers 1–8, as shown in Table 2.

Table 2. Sound sample information.

Sample Type	Label	Training Samples	Testing Samples
Cicada sound	1	280	120
Construction noise	2	280	120
Wind sound	3	280	120
Thunder sound	4	280	120
Birdsong	5	280	120
Rain sound	6	280	120
Bolt loosening sound	7	280	120
Speech	8	280	120
Total		2240	960
Total number of samples		3200	

Following the analysis process outlined above, the collected samples were first analyzed in the time domain to obtain the short-time energy temporal features of each frame signal. Then, MFCC analysis was performed with 24 Mel filters, resulting in 24 MFCC features for each frame signal. Furthermore, wavelet packet analysis was conducted using the db10 wavelet basis function with a decomposition level of 4, resulting in $2^4 = 16$ wavelet packet energy features. The MFCC and WPES features were concatenated to obtain the MFCC-WPES combined features. Therefore, the original feature set contains a total of $24 + 16 = 40$ features. The corresponding input dataset for the classifier is a 3200×40 matrix.

3.3.2. Feature Selection and Parameter Optimization

According to Section 2.2, a Random Forest model is constructed, and four parameters are optimized through a genetic algorithm: the number of trees n, the maximum depth of each decision tree d, the minimum number of samples required to split an internal node s, and the maximum number of leaf nodes m. The optimization ranges for these parameters are shown in Table 3. A population consisting of 100 individuals is created, with a crossover probability of 0.5, a mutation probability of 0.2, and a number of generations of 50. The

fitness variation curve during the optimization process is shown in Figure 7. The blue curve represents the average fitness, which rises rapidly in the initial stage and stabilizes after about 10 generations, ultimately settling at a high level of approximately 0.99. This indicates that the overall fitness of the population has significantly improved during the evolutionary process. After optimization, the final values are n of 114, d of 60, s of 6, and m of 235.

Table 3. Optimization range and results of parameters.

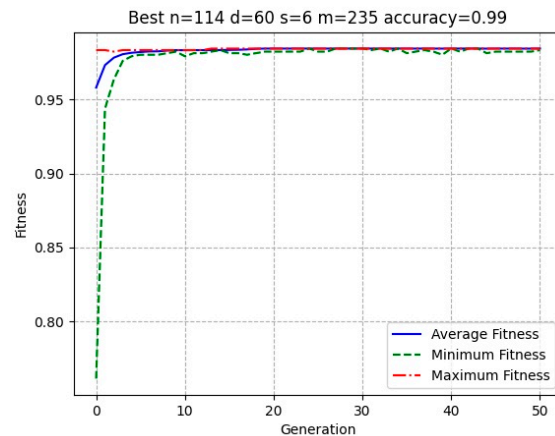| Parameter	n	d	s	m
Optimization scope	[50, 500]	[1, 200]	[2, 200]	[10, 500]
Optimal parameters	114	60	6	235

Figure 7. Genetic algorithm parameter optimization.

After obtaining the optimal parameters, random forest was used to evaluate the feature importance by calculating the importance index for all 40 features. Figure 8a shows the results, where the blue bars represent MFCC features, and the orange bars represent wavelet packet energy features. It can be observed that the contribution of MFCC features to the results is mainly concentrated in the first 12 features, while the contribution of the latter 12 features is relatively small. In contrast, the contribution of wavelet packet energy features is distributed relatively evenly across all frequency bands. After obtaining the importance indicators of each feature, all features are sorted by their contribution degree from largest to smallest, as shown in Figure 8b.
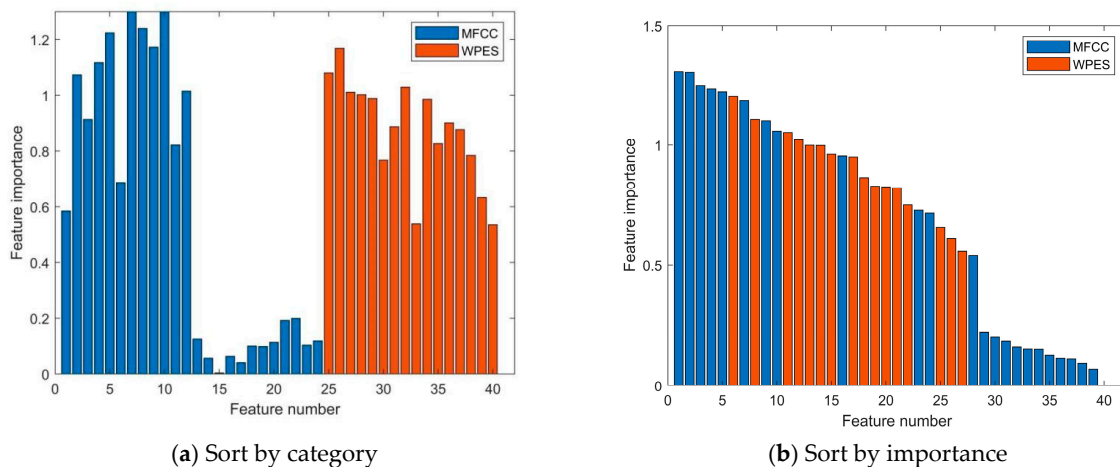

(**a**) Sort by category

(**b**) Sort by importance

Figure 8. Feature importance assessment and ranking.

To analyze the impact of the number of features on the model recognition results, the top-ranked features based on their importance index were sequentially selected and input into the random forest model for training. The training results are shown in Figure 9. It can be observed that as the number of input features increases, the change in recognition accuracy can be roughly divided into three stages: rapid increase (1–6), slow increase (7–15), and stabilization (16–40). When the feature index is 1, i.e., only the first feature is selected, the accuracy is less than 30%. Subsequently, the recognition accuracy rapidly increases with the increase in feature index. When the feature index is 6, the accuracy reaches 94.6%, and when the feature index is 15, it basically reaches the maximum value of 98.9%. After that, there is no significant increase in accuracy with the increase in the number of features. When using all 40 features for training, the accuracy is 98.6%, indicating that the features ranked lower have little impact on the recognition results and can be considered redundant features.
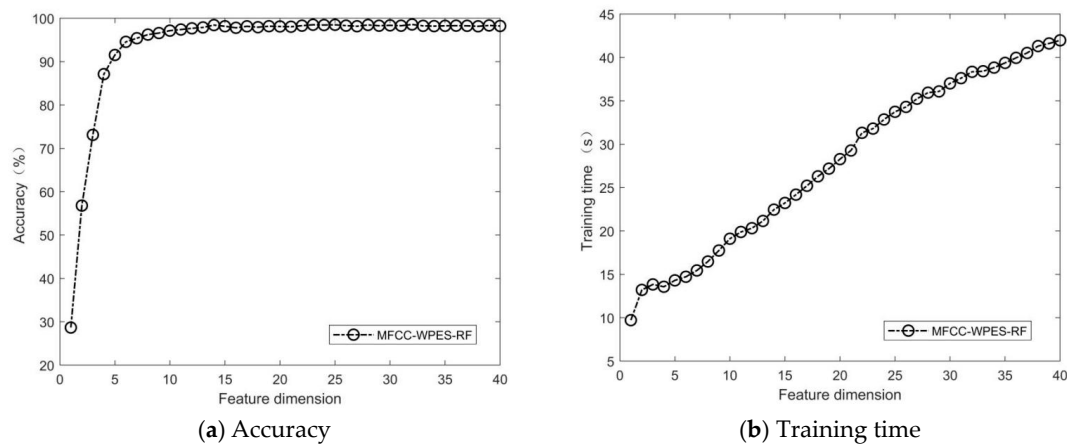

(**a**) Accuracy   (**b**) Training time

Figure 9. The influence of feature dimension on accuracy and training time.

Furthermore, regarding training time, as shown in Figure 9b, it roughly scales linearly with the total number of input features to the model. When only the first feature is selected, the training time is only 10 s. It takes 21 s to train using the top 15 features, while training with all features requires 42 s. Therefore, considering both the model recognition accuracy and training time cost, and aiming to minimize feature redundancy, this study ultimately selects only the top 15 features based on their importance index to input into the random forest model for training and recognition. As shown in Figure 8b, these fifteen features include eight MFCC features and seven WPES features.

3.3.3. Analysis of Results and Discussion

To validate the effectiveness of the model obtained after the aforementioned training, a test set was used to evaluate the model. The construction method of the test set is shown in Table 2. This paper employs four metrics: Confusion Matrix, Precision, Recall, and F1 score, to measure the classification accuracy of the model on the test samples. The Confusion Matrix, typically represented as an n × n matrix, is a standard format for accuracy evaluation, also known as an error matrix. The definition of accuracy is provided in Formula 14, while the definitions and calculation methods of Precision, Recall, and F1 score are presented in Table 4. The final analysis results are shown in Figure 10 and Table 5.

In the classification results in Figure 10, it can be seen that for the 960 samples in the test set, the overall recognition accuracy of the 15-dimensional combined features reached 98.75%. Specifically, among the one hundred and twenty samples in each category, except for four samples in both the third and sixth categories being misclassified, the other categories can basically be accurately identified. Among them, only one sample in the second, fourth, fifth, and eighth categories was misclassified, while the samples in the first

and seventh categories were fully recognized without any misclassification. This indicates that the classifier can correctly identify the sample categories in most cases.

Table 4. Precision, Recall, and F1 score.

| Indicator	Formula	Meaning of Indicator
Precision	$\text{Precision} = \frac{TP}{TP+FP}$	The proportion of correct predictions among all positive predictions made by the model
Recall	$\text{Recall} = \frac{TP}{TP+FN}$	The proportion of positive predictions made correctly by the model among all true positives
F1 score	$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	The harmonic mean of precision and recall
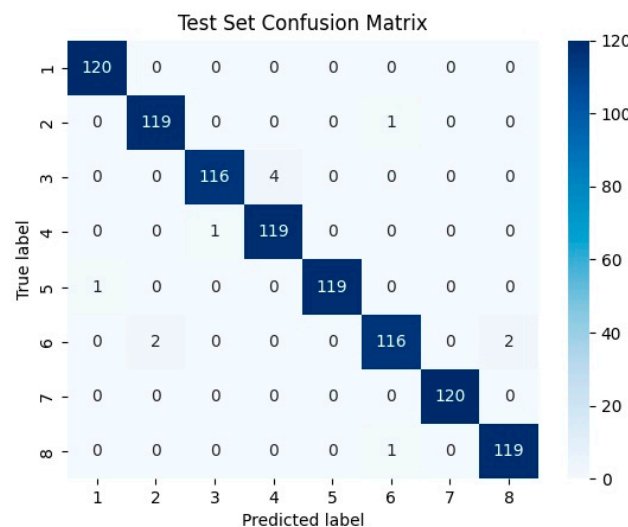
Figure 10. Confusion matrix.

Table 5. Recognition result.

Label	1	2	3	4	5	6	7	8
Precision	0.99	0.98	0.99	0.97	1	0.98	1	0.98
Recall	1	0.99	0.97	0.99	0.99	0.97	1	0.99
F1 score	1	0.99	0.98	0.98	1	0.97	1	0.99

In addition, as can be seen from the recognition results in Table 5, the precision is above 0.98 for all categories except for the slightly lower precision of 0.97 in the fourth category. Notably, the precision for the seventh category of bolt loosening sound signals reaches 1, indicating that all samples predicted as the seventh category are truly from the seventh category. The recall rates are high for all categories except for the slightly lower rates of below 0.99 in the third and sixth categories. Specifically, the recall rates for the first, seventh, and eighth categories all reach 1, meaning that all true samples from these categories are correctly predicted by the model. The F1 score are above 0.98 for all categories, except for a slightly lower score of 0.97 in the sixth category. Notably, the F1 score for the first, fifth, and seventh categories all reach 1, indicating excellent performance of the model in these categories.

Meanwhile, in this paper, in addition to using combined features for training recognition, independent training recognition using 24-dimensional MFCC features and 16-dimensional WPES features is also attempted separately. The specific data on recognition accuracy, precision, recall, and F1 score are shown in Figure 11 and Table 6. As can be clearly seen from Figure 11, the recognition accuracy using only MFCC features is 96.2%, while the recognition accuracy using only WPES features is 97.7%. Both of these accuracies

are lower than the accuracy achieved by using the 15-dimensional combined features for recognition, which reaches a high level of 98.8% on the test set. Furthermore, according to the recognition results of precision, recall, and F1 score for each category in Table 6, the recognition effect of using the MFCC-WPES combined features is also significantly better than that of using these two types of features alone.
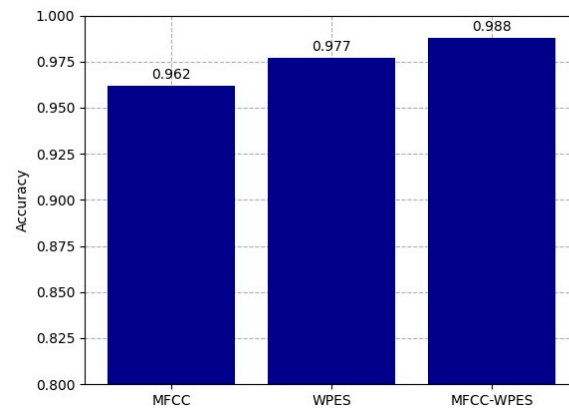

Figure 11. Comparison chart of recognition accuracy of different training features.

Table 6. Recognition results of different training features.

Indicator	Feature	1	2	3	4	5	6	7	8
	MFCC	1.00	0.93	0.91	0.97	1.00	0.96	1.00	0.92
Precision	WPES	1.00	0.94	0.98	0.98	0.99	0.97	1.00	0.96
	MFCC-WPES	0.99	0.98	0.99	0.97	1.00	0.98	1.00	0.98
	MFCC	1.00	0.96	0.97	0.91	1.00	0.93	1.00	0.92
Recall	WPES	1.00	0.97	0.98	0.97	1.00	0.96	0.99	0.93
	MFCC-WPES	1.00	0.99	0.97	0.99	0.99	0.97	1.00	0.99
	MFCC	1.00	0.94	0.94	0.94	1.00	0.95	1.00	0.92
F1 score	WPES	1.00	0.96	0.98	0.98	1.00	0.96	1.00	0.95
	MFCC-WPES	1.00	0.99	0.98	0.98	1.00	0.97	1.00	0.99

To evaluate the performance of the RF model, we tried various machine learning models, including Bayes, KNN, SVM, BP neural network, and XGBoost. These models were trained using the MFCC-WPES combined features, and their recognition performance was compared and analyzed. The specific recognition results are shown in Figure 12 and Table 7. The results clearly indicate that the XGBoost, SVM, and BP models exhibit relatively high recognition accuracy, all reaching approximately 0.98. However, they still perform below the accuracy of the RF model. In comparison, the KNN and Bayes models have lower accuracy, especially Bayes with only 0.903, which may suggest that these two models are not the best choices for handling the current task. Furthermore, from the per-class recognition results presented in Table 7, we can also observe that, except for slightly lower precision and recall in a few categories, the RF model outperforms other classifiers in F1 score for all categories.

The analysis results above indicate that the combined MFCC-WPES features adopted in this paper can more comprehensively represent sound signals, and their recognition performance is significantly better than traditional MFCC and WPES. Meanwhile, when compared with several other machine learning models, the optimized Random Forest (RF) model exhibits better performance in key performance indicators, such as recognition accuracy and F1 score. Ultimately, this method not only achieves accurate identification of bolt loosening damage signals but also demonstrates good discrimination ability for different types of environmental noise signals.
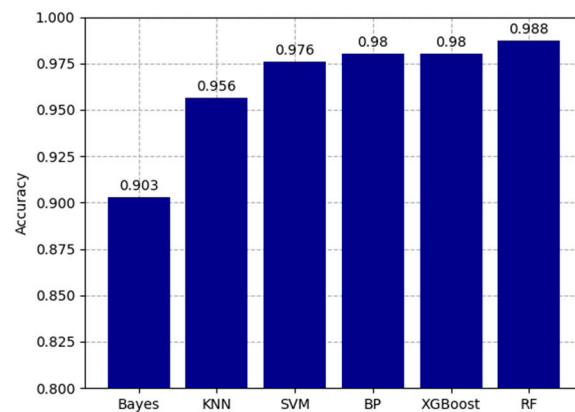
Figure 12. Comparison chart of recognition accuracy of different training models.

Table 7. Recognition results of different training models.

Indicator	Model	1	2	3	4	5	6	7	8
Precision	Bayes	1.00	0.96	0.91	0.84	0.96	0.92	0.92	0.77
	KNN	1.00	0.87	0.89	0.99	1.00	0.99	1.00	0.93
	SVM	1.00	0.96	0.95	0.98	0.99	0.97	1.00	0.93
	BP	1.00	0.93	0.95	0.95	1.00	0.99	1.00	0.99
	XGBoost	1.00	0.98	0.97	0.97	0.98	0.98	1.00	0.95
	RF	0.99	0.98	0.99	0.97	1.00	0.98	1.00	0.98
Recall	Bayes	1.00	0.68	0.84	0.91	0.97	0.90	0.96	0.97
	KNN	1.00	0.94	0.99	0.88	1.00	0.94	1.00	0.90
	SVM	1.00	0.97	0.98	0.95	1.00	0.96	0.99	0.93
	BP	1.00	1.00	0.95	0.95	1.00	0.97	1.00	0.94
	XGBoost	1.00	0.98	0.97	0.97	1.00	0.96	0.98	0.97
	RF	1.00	0.99	0.97	0.99	0.99	0.97	1.00	0.99
F1 score	Bayes	1.00	0.80	0.87	0.88	0.96	0.91	0.94	0.86
	KNN	1.00	0.90	0.94	0.93	1.00	0.97	1.00	0.92
	SVM	1.00	0.97	0.97	0.97	1.00	0.97	1.00	0.93
	BP	1.00	0.96	0.95	0.95	1.00	0.98	1.00	0.97
	XGBoost	1.00	0.98	0.97	0.97	0.99	0.97	0.99	0.96
	RF	1.00	0.99	0.98	0.98	1.00	0.97	1.00	0.99

In practical applications, the significance of the method proposed in this paper is mainly reflected in the following aspects compared with existing methods. Firstly, the excitation method using a small vibration exciter significantly simplifies the operation process and improves applicability compared with the traditional method of knocking bolts individually. Secondly, compared with traditional contact vibration signal measurement, the process of sound signal acquisition is simpler and faster. In practical applications, only ordinary recording equipment or even a mobile phone is required, greatly reducing the implementation threshold. Finally, the random forest recognition model adopted in this paper not only has a high accuracy rate but also exhibits strong robustness to noise, enabling effective processing of high-dimensional data. Additionally, it boasts low training costs and simple deployment, making it highly suitable for application in mobile devices and embedded systems.

However, it must be pointed out that this research also has certain limitations and challenges. Due to experimental constraints, the environmental noise collected in this paper is primarily simulated through a Bluetooth speaker, which differs significantly from the complex and varied noise environments in the real world. In real-world settings, noise is more diverse and often involves mixtures of multiple types of noise, posing greater challenges for accurate sound signal recognition. Therefore, future research should further explore and address these issues to enhance the generalization ability and robustness of the model in practical applications.

4. Conclusions

This study addresses the existing issues and limitations of traditional methods in bolt loosening detection for steel truss structures. It focuses on acoustic signal recognition technology and proposes an innovative bolt loosening damage identification strategy based on MFCC-WPES feature fusion and an optimized Random Forest algorithm. The research process involves frame segmentation, windowing preprocessing of acoustic signals, feature extraction using MFCC and WPES, fusion of these features to construct an MFCC-WPES combined feature training set, and training an RF classifier for sensitive feature selection and effective damage signal identification. A genetic algorithm is employed to optimize key parameters of the Random Forest model, aiming to construct an identification model with optimal performance. The results demonstrate that the proposed method accurately identifies bolt loosening damage signals and exhibits strong robustness to environmental noise. Compared to methods relying on single features and traditional machine learning models, this method shows significant advantages in core performance indicators.

Future research directions will focus on testing the model under more complex environments and a wider range of damage conditions, with plans to conduct field measurements on large-scale truss steel structures under actual working conditions to further validate and optimize the practical application effectiveness of the proposed method.

Author Contributions: Conceptualization, D.Z.; methodology, D.Z.; software, Z.Z.; validation, Z.J.; formal analysis, Z.J.; investigation, Z.J.; resources, Z.Z.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, D.Z.; visualization, Z.Z.; supervision, D.Z.; project administration, D.Z.; funding acquisition, D.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tao, H.; He, G.S.; Yang, J.X.; Ai, P.W. Status identification of transmission tower bolts based on MIMU. *J. Vib. Shock* **2023**, *42*, 98–104.
2. Cao, Z.F.; Tan, Z.; Jiang, D.; He, D.D.; Fei, Q.G. Loosening discrimination of a bolted connector under high-temperature vibration environment based on time-frequency analysis. *J. Vib. Shock* **2019**, *38*, 205–210.
3. Wang, J.L.; Sheng, J.J.; Jiang, J.Y.; Wen, J.P.; Xiao, S.F.; Zhang, Z.S. Identification method of bolt looseness using STFT-IP time-frequency feature extraction technology. *Noise Vib. Control* **2021**, *41*, 108–112.
4. Zhang, M.Z.; Wang, L.; Tian, X.H. Bolt loosening state monitoring based on inner product matrix and convolutional autoencoder. *Eng. Mech.* **2022**, *39*, 222–231.
5. Shao, J.H.; Wang, T.; Wang, Z.O.; Wei, D.L.; Li, Y.R. Bolt looseness detection using piezoelectric impedance frequency shift method. *China Mech. Eng.* **2019**, *30*, 1395–1399+1408.
6. Zhuo, D.B.; Cao, H. Damage identification of bolt connection in steel truss structures by using sound signals. *Struct. Health Monit.* **2022**, *21*, 501–517. [CrossRef]
7. Wang, S.P.; Yang, P.; Sun, H. Indoor sound-position fingerprint method based on scenario analysis. *J. Beijing Univ. Technol.* **2017**, *43*, 224–229.
8. Zhao, S.T.; Zhang, P.; Shen, L.; Guo, J. Vibration and acoustic joint mechanical fault diagnosis method of high voltage circuit breakers. *Trans. China Electrotech. Soc.* **2014**, *29*, 216–221.
9. Benko, U.; Petrovcic, J.; Juricic, D.; Tavčar, J.; Rejec, J. An approach to fault diagnosis of vacuum cleaner motors based on sound analysis. *Mech. Syst. Signal Process.* **2005**, *19*, 427–445. [CrossRef]
10. Wang, F.; Song, G. Looseness detection in cup-lock scaffolds using percussion-based method. *Autom. Constr.* **2020**, *118*, 103266. [CrossRef]

11. He, S.H.; Chen, J.A.; Chen, Z.; Song, G. An exploratory study of underwater bolted connection looseness detection using percussion and a shallow machine learning algorithm. *Acta Mech. Sin.* **2023**, *39*, 1–7. [CrossRef]

12. Liu, P.T.; Wang, X.P.; Chen, T.N.; Wang, Y.; Mao, F.; Liu, W. Research on a percussion-based bolt looseness identification method based on phase feature and convolutional neural network. *Smart Mater. Struct.* **2023**, *32*, 035010. [CrossRef]

13. Yuan, R.; Lv, Y.; Xu, S.J.; Li, L.; Kong, Q.; Song, G. ResNet-integrated very early bolt looseness monitoring based on intrinsic feature extraction of percussion sounds. *Smart Mater. Struct.* **2023**, *32*, 034002. [CrossRef]

14. Chen, J.; Chen, Z.; Zhu, W.H.; Song, G. Underwater bolted flange looseness detection using percussion-induced sound and Feature-reduced Multi-ROCKET model. *Struct. Health Monit.* **2024**, *23*, 495–511. [CrossRef]

15. Wang, S.Y.; Zhou, Y.; Kong, Q.Z. A force-adaptive percussion method for bolt looseness assessment. *J. Civ. Struct. Health Monit.* **2024**, *14*, 831–841. [CrossRef]

16. Wang, S.Y.; Kong, Q.Z.; Zhou, Y. Percussion-based bolt loosening detection method in a single-row multi-bolted steel beam-column joint. *Earthq. Eng. Eng. Dyn.* **2024**, *44*, 95–103.

17. Gao, L.; Liu, Z.K.; Zhang, H.Y.; Wei, X.Y.; Zhang, K. Percussion inspection voice recognition of railway tunnel voids based on ensemble methods. *J. Vib. Shock* **2022**, *41*, 58–63+83.

18. Yang, L.; Quan, W.; Li, L.; Wang, J.Q.; Gao, F.; Yu, X. Heterodyne detection technique of nuclear power electrical equipment based on ITD-MFCC and convolutional neural network. *Noise Vib. Control* **2023**, *43*, 122–128+207.

19. Qian, X. Damage identification of wind-break wall structures based on the further updated wavelet packet frequency bands energy ratio spectrum. *Shock. Vib.* **2022**, *2022*, 8200199.

20. Yu, Y.; Dackermann, U.; Li, J.; Niederleithinger, E. Wavelet packet energy-based damage identification of wood utility poles using support vector machine multi-classifier and evidence theory. *Struct. Health Monit.* **2019**, *18*, 123–142. [CrossRef]

21. Zhou, J.L.; Zhou, Z.; Cui, L. Loudspeaker abnormal sound classification using variational modal decomposition and the random forest feature selection algorithm. *J. Vib. Shock* **2022**, *41*, 277–283.

22. Zhe, N.; Liu, C.X.; Yang, T.B.; He, P.; Jian, J.; Wang, G. Application of VMD and random forest in reactor metal impact signal recognition. *J. Vib. Shock* **2021**, *40*, 102–107.

23. Ding, X.; Xu, J.; Teng, W.; Liu, Y.B. Failure prognostic prediction of direct drive wind turbines based on algorithms of random forest and adaptive network-based fuzzy inferer system. *Noise Vib. Control* **2018**, *38*, 209–2014.

24. Ye, S.Q.; Zhou, K.Q.; Zain, A.M.; Wang, F.; Yusoff, Y. A modified harmony search algorithm and its applications in weighted fuzzy production rule extraction. *Front. Inf. Technol. Electron. Eng.* **2023**, *24*, 1574–1590. [CrossRef]

25. Qin, F.; Zain, A.M.; Zhou, K.Q. Harmony search algorithm and related variants: A systematic review. *Swarm Evol. Comput.* **2022**, *74*, 101126. [CrossRef]

26. Zhang, C.X.; Zhou, K.Q.; Ye, S.Q.; Zain, A.M. An improved cuckoo search algorithm utilizing nonlinear inertia weight and differential evolution for function optimization problem. *IEEE Access* **2021**, *9*, 161352–161373. [CrossRef]

27. Ye, S.Q.; Zhou, K.Q.; Zhang, C.X.; Zain, A.M.; Ou, Y. An improved multi-objective cuckoo search approach by exploring the balance between development and exploration. *Electronics* **2022**, *11*, 704. [CrossRef]

28. Zhang, X.Y.; Zhou, K.Q.; Li, P.C.; Xiang, Y.H.; Zain, A.M.; Sarkheyli-Hagele, A. An improved chaos sparrow search optimization algorithm using adaptive weight modification and hybrid strategies. *IEEE Access* **2022**, *10*, 96159–96179. [CrossRef]

29. Qian, H.B.; Li, Y.L. A SVM algorithm based on convex hull sparsity and genetic algorithm optimization. *J. Chongqing Univ. (Nat. Sci. Ed.)* **2021**, *44*, 29–36.