ORIGINAL ARTICLE

Improving Ultrasonic Testing by Using Machine Learning Framework Based on Model Interpretation Strategy

Siqi Shi[1], Shijie Jin[1*], Donghui Zhang[2], Jingyu Liao[2], Dongxin Fu[2] and Li Lin[1*]

Abstract

Ultrasonic testing (UT) is increasingly combined with machine learning (ML) techniques for intelligently identifying damage. Extracting significant features from UT data is essential for efficient defect characterization. Moreover, the hidden physics behind ML is unexplained, reducing the generalization capability and versatility of ML methods in UT. In this paper, a generally applicable ML framework based on the model interpretation strategy is proposed to improve the detection accuracy and computational efficiency of UT. Firstly, multi-domain features are extracted from the UT signals with signal processing techniques to construct an initial feature space. Subsequently, a feature selection method based on model interpretable strategy (FS-MIS) is innovatively developed by integrating Shapley additive explanation (SHAP), filter method, embedded method and wrapper method. The most effective ML model and the optimal feature subset with better correlation to the target defects are determined self-adaptively. The proposed framework is validated by identifying and locating side-drilled holes (SDHs) with 0.5λ central distance and different depths. An ultrasonic array probe is adopted to acquire FMC datasets from several aluminum alloy specimens containing two SDHs by experiments. The optimal feature subset selected by FS-MIS is set as the input of the chosen ML model to train and predict the times of arrival (ToAs) of the scattered waves emitted by adjacent SDHs. The experimental results demonstrate that the relative errors of the predicted ToAs are all below 3.67% with an average error of 0.25%, significantly improving the time resolution of UT signals. On this basis, the predicted ToAs are assigned to the corresponding original signals for decoupling overlapped pulse-echoes and reconstructing high-resolution FMC datasets. The imaging resolution is enhanced to 0.5λ by implementing the total focusing method (TFM). The relative errors of hole depths and central distance are no more than 0.51% and 3.57%, respectively. Finally, the superior performance of the proposed FS-MIS is validated by comparing it with initial feature space and conventional dimensionality reduction techniques.

Keywords Ultrasonic testing, Machine learning, Feature extraction, Feature selection, Shapley additive explanation

*Correspondence:
Shijie Jin
jinshijie@dlut.edu.cn
Li Lin
linli@dlut.edu.cn
[1] NDT & E Laboratory, Dalian University of Technology, Dalian 116024, China
[2] China Nuclear Industry 23 Construction Co., Ltd., Beijing 101300, China

1 Introduction

Recently, the demand for defect characterization and damage identification in materials/structures has been growing in various industrial applications, such as aerospace, nuclear, oil and gas, to ensure high performance and safety [1]. To this end, ultrasonic nondestructive testing (UT) is widely used owing to low cost, low power consumption and no change to materials/structures [2]. The scattered/diffracted/reflected waves are employed by UT to detect and characterize unknown defects by signal

Shi *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:127

Page 2 of 13

processing [3–5] and imaging processing techniques [6–8]. With the development of computer science and artificial intelligence, data-driven machine learning (ML) techniques have been adopted in UT area to facilitate signal interpretation [9, 10]. ML provides a powerful tool to find and establish the complicated nonlinear relationship between observed UT data and physical properties of the probed structures owing to its advantages in high speed and strong fitting ability [11]. Compared to manual interpretation, ML eliminates the influence of subjective factors and realizes the intelligent identification of defects [12–14]. For instance, Yuan et al. [15] proposed a neural network model to identify echoes from defects in B-scans of train wheels and the accuracy of defect recognition was improved to 92%. It should be noted that the performance of ML techniques is primarily dependent on the features extracted from UT data [16]. Original UT data contain a number of invalid and redundant features. Applying raw UT data directly in defect characterization increases the complexity and computational time of ML models [12, 16]. Therefore, the extraction and selection of defect features with meaningful information are crucial for improving defect characterization accuracy and computational efficiency [17].

Current researches on UT combined with ML typically focus on the extraction of sensitive features from the time domain, frequency domain and time-frequency domain by statistical techniques [10], Fourier transform [18], wavelet transform [19] or empirical mode decomposition (EMD) [20]. Then, the appropriate ML model, such as support vector machine (SVR) [21], artificial neural network (ANN) [22] or extreme learning machine (ELM) [23], is established by using these features to predict defect parameters. On this basis, feature selection methods preserve important information and remove redundant features without changing the physical meaning of original feature set, reducing the overfitting of the prediction model.

Ma et al. [24] developed a back propagation neural network optimizing Gaussian process regression (BP-GPR) algorithm to predict the porosity of thermal barrier coating. The features extracted from ultrasonic reflection coefficient amplitude spectrum were optimized by combining BP neural network and high determination coefficient rule. The predictive accuracy of BP-GPR was 32% and 48% higher than that predicted only by BP neural network or GPR algorithm, respectively. Bai et al. [25] extracted the scattering matrix from array data to characterize the sizes and orientation angles of small defects. A dimensionality reduction approach based on locality preserving projection was proposed to separate the scattering matrices of unfavorably oriented defects. In addition, the filter method [26], embedded method

[27] and wrapper method [28] used in the field of fault diagnosis are typically implemented according to the divergence or correlation of features to optimize feature space. Nazir et al. [29] monitored the tool conditions of ultrasonic metal welding via sensor fusion and ML. The filter method, embedded method and wrapper method were employed to select ten sensitive features from the initial feature space containing 97 features. The classification accuracies of tool conditions for training and testing datasets were both close to 100%. Besides, dimensionality reduction techniques, such as principal component analysis (PCA) [30] and factor analysis (FA) [31], can also be used for feature selection by fusing the high-dimensional feature set to the significant lower-dimensional features [16]. Lv et al. [32] adopted the noncontact laser ultrasonic technique and the identified ML algorithm to quantify the widths and depths of subsurface defects simultaneously. PCA was applied to reduce mutually correlated features and improve detection accuracy. The highest recognition rate of subsurface defects was 98.48%.

However, the applications of ML methods in UT still face some challenges, e.g., the hidden physics behind ML and the unexplained contribution of each feature [11]. The intrinsic black-box character of ML models induces the reduction of the generalization performance and applicability [33, 34], and the lack of knowledge is a barrier to the deployment of ML in UT area. The higher the interpretability of an ML model, the easier it is for someone to comprehend certain decisions or predictions made, and the interpretability is strongly reliant on the contribution of each feature [35]. For example, Xu et al. [34] proposed an explainable ensemble tree model to identify pipeline leakage scenarios. The optimized feature space of each pipe leakage state was summarized and analyzed by Shapley additive explanation (SHAP). While retaining the advantages of ML, the method overcomes the problem that the correlation between the results brought by black-box character and the feature space cannot be analyzed. Consequently, to obtain an interpretable ML model, various signal processing techniques should be applied to extract multi-domain features for comprehensively mining the useful information and intrinsic properties from UT data. Then, the sensitive features in the initial feature space are selected by establishing the relation of features and predicted results based on the model interpretation strategy to eliminate the deficiencies, such as dependence on expert experiences and poor universal applicability of features.

In this paper, a generally applicable ML framework based on model interpretation strategy is proposed by combining the UT methods, signal processing techniques and ML algorithms for improving defect characterization accuracy and computational efficiency. The outline of this

Shi *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:127

Page 3 of 13

paper is organized as follows. Section 2 gives an overview of the proposed ML framework. In Section 3, two illustrative examples are given to show the effectiveness of the proposed framework by identifying and locating the side-drilled holes (SDHs) with subwavelength spacing. In Section 4, some comparisons are conducted to highlight the superiority of the proposed feature selection method. Conclusions are drawn in the final section.

2 Generally Applicable ML Framework

The input pulse used in UT is transmitted into the material under test, and the presence of material discontinuities or defects gives rise to scattered/diffracted/reflected signals [36]. On this basis, the identification and characterization of defects are carried out by appropriate signal processing techniques. ML has the ability to obtain the complicated relationship between observed data and physical properties of the probed structure by adaptive learning and training, as shown in Figure 1, having been widely used in UT area. ML maps inputs (or features) to outputs (or target variables) during training to produce a model that accurately predicts the outputs of previously unseen input data [37]. However, the implementation process requires expertise to extract and select appropriate features from UT data as model inputs, determine the ML algorithm and find a suitable set of model hyperparameters.

The environmental noises accompanying measured UT signals obstruct damage diagnosis. The acquired signals are preprocessed firstly by filtering, smoothing and normalization to suppress noise. However, original UT signals contain invalid and redundant information. To reduce the complexity of ML models, various signal processing techniques are conducted on the preprocessed signals to deeply mine and extract the effective multi-domain features (e.g., time-domain, frequency-domain and time-frequency domain features). Every raw UT signal is transformed into a set of features with physical and statistical meaning related to the target defect, and an initial high-dimensional feature space is constructed.

Next, a feature selection method based on model interpretable strategy (FS-MIS) is proposed to self-adaptively obtain the optimal feature subset with more physically interpretable. Filter method and embedded method [38] are used to perform feature preselection from the initial feature space by considering two aspects: (1) whether the feature diverges or converges; and (2) the correlation between the feature and the target. Moreover, the optimal ML model depends on the issue to be addressed and is determined by evaluating the predictive capability of several commonly used models in complex and nonlinear problems. Support vector regression (SVR) is a powerful learning model to minimize structural risk with better generalization capability based on statistical theory [21]. Gradient boosted regression (GBR) is an ensemble learning algorithm that promotes a series of weak learners to strong learners through iterative calculations [39]. As an extended variant of the bagging mode in ensemble learning, random forest regression (RFR) introduces random attribute selection in the training process of the decision tree to implement with powerful performance in prediction and regression [39]. The extreme gradient boosting (XG-Boost) model uses a second-order Taylor expansion to extend the loss function and add a regularization term, having the advantages of low computational complexity, fast running speed and high accuracy [40]. Back-propagation neural network (BPNN) is a multi-layer feedforward neural network based on the error back propagation algorithm [41]. By continuously adjusting the weight values of the network, the final network outputs are as close as possible to the expected outputs to achieve the purpose of training. The hyperparameters of the aforementioned ML models are determined by grid search [42].

Two statistical indexes, mean squared error (*MSE*) and determination coefficient (R^2), are introduced to evaluate the model performance. The smaller *MSE* and the larger R^2 indicate better reliability and predictive accuracy.

$$MSE(T,H) = \frac{1}{m} \sum_{i=1}^{m} (T_i - H_i)^2, \tag{1}$$

$$R^2(T,H) = \frac{\sum\limits_{i=1}^{m} (T_i - \overline{T})(H_i - \overline{H})}{\sqrt{\sum\limits_{i=1}^{m} (T_i - \overline{T})^2 (H_i - \overline{H})^2}}, \tag{2}$$

where *m* represents the number of samples; T_i and H_i are respectively the expected and predicted values, and $\overline{T}$ and $\overline{H}$ are the averages of expected and predicted values, respectively.

It is difficult to understand the model decisions and the influences of features due to the intrinsic black-box character of ML models [2]. Therefore, SHAP [43] is incorporated to explore the importance of each feature on the predicted results and self-adaptively sort out highly
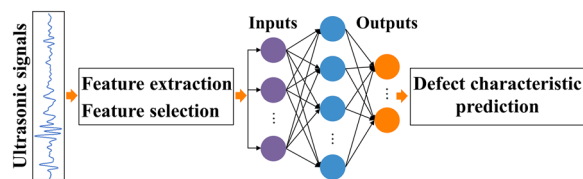

Figure 1 Schematic diagram of UT combined with ML

Shi *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:127

Page 4 of 13

sensitive features with more defect information, further reducing the dimension of feature space. SHAP is a model interpreter, a concept in game theory [33]. The SHAP value ϕ_P of feature P is the average of its marginal contributions across all possible permutations and combinations considered [33].

$$\phi_P = \frac{1}{|n|!} \sum_{V \subseteq n \setminus \{P\}} |V|!(|n| - |V| - 1)![F(V \cup \{P\}) - F(V)], \tag{3}$$

where $F(V)$ corresponds to the output of the ML model to be explained using a set V of features, and n is the complete set of all features.

Noteworthy, features that appear irrelevant to the target singly may become highly relevant by taking with others [28]. The impact of feature combination should also be considered. Hence, the wrapper method is utilized to assess the potential feature subsets [38, 44]. Multiple combinations of the available features are tested, and the feature subset presenting the best performance is finally chosen [38]. In this paper, the optimal feature subset is determined by comparing the predictive performance of the feature subsets obtained by sequential forward selection (SFS) [45] and sequential backward selection (SBS) [46].

Finally, the ML model trained by the feature subset selected with FS-MIS is determined whether it is optimal according to the predictive accuracy. If the outputs deviate greatly from the expected values, the initial features will be re-extracted. Repeat the above processes until the most effective ML model and the optimal feature subset highly correlated to the target characteristics are acquired. Overall, Figure 2 shows the proposed ML framework, which can be applied to different UT scenarios for locating and characterizing defects quantitatively.

3 Experiments

3.1 Specimens and Experimental Details

To evaluate the superior performance of the proposed framework, the experiments were conducted on six 180 mm × 95 mm × 15 mm 6061 aluminum alloy specimens containing adjacent SDHs. The longitudinal wave velocity was 6300 m/s, and the corresponding wavelength λ in aluminum alloy was about 2.8 mm at 2.25 MHz inspection frequency. As schematically illustrated in Figure 3a, the central distances of the SDHs in specimens are 1.40 mm (0.5λ) ~ 2.80 mm (1.0λ) with a step of 0.28 mm (0.1λ), and the diameter and central depth of the SDHs are 1.0 mm and 50 mm, respectively.

The full matrix capture (FMC) technique is introduced to capture all the possible independent information from the array elements and provide plenty of flexibility for post-processing [47]. For an array with N elements, N^2

signals are obtained by FMC. Figure 3a shows the ultrasonic path from the ith element (with coordinates (x_i, 0)) to the jth element (with coordinates (x_j, 0)) through a potential scatterer located at coordinates (x_{ref}, z_{ref}), and $y_{ij}(t)$ denotes the corresponding A-scan signal. The Eddyfi M2M PANTHER and a linear array probe (64 elements, 0.6 mm pitch and 2.25 MHz central frequency) are employed to acquire FMC data with 100 MHz sampling frequency from the top and bottom surfaces of each specimen, as shown in Figure 3b. Therefore, the actual to-be-measured SDH depths included 45 mm and 50 mm. To reduce data redundancy, only the A-scan signals transmitted and received by the left 32 elements were considered according to the symmetry and reciprocity of the inspection model. Therefore, 12288 time-domain signals corresponding to 12 FMC datasets were obtained from experiments.

The representative time traces of the scattered waves are plotted in Figure 4a, where the pulses from two SDHs are overlapped due to the low time resolution [3]. The time resolution depends on the spatial pulse length (SPL) of the probing signal, and the theoretical resolution limit in UT is equal to half the SPL [48]. The SPL in this study was about 1.08 μs, so the resolution limit was 0.54 μs. Taking the SDHs with 0.5λ central distance in 45 mm and 50 mm depths as examples, the pulse-echoes in 2048 signals were strongly coupled, since the calculated interval of the times of arrival (ToAs) of scattered waves ranged from 0.0048 to 0.19 μs. It is desirable to improve the time resolution of each A-scan signal in the FMC datasets for accurately locating the SDHs.

Moreover, post-processing imaging techniques, such as the total focusing method (TFM), can be performed on the FMC data to obtain high-resolution ultrasonic images [49]. TFM is a delay-and-sum beamforming algorithm, in which the array signals are synthetically focused on each point in the region of interest [50].

As shown in Figure 3a, the delay law is calculated based on the ray path from each array element to point Q, and the corresponding intensity $I(x_{ref}, z_{ref})$ is given by

$$I\left(x_{ref}, z_{ref}\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij}(t_{ij}(x_{ref}, z_{ref})), \tag{4}$$

where t_{ij} represents the travel time from the ith element through focus point Q to the jth element.

The TFM images of the SDHs with 0.5λ central distance in 45 mm and 50 mm depths are presented in Figure 4b. It is challenging to distinguish and locate the SDHs with subwavelength spacing due to the diffraction limit [51]. Focusing on the above two basic issues in UT, the proposed ML framework based on model interpretation strategy is applied to ultrasonic signal analysis and image
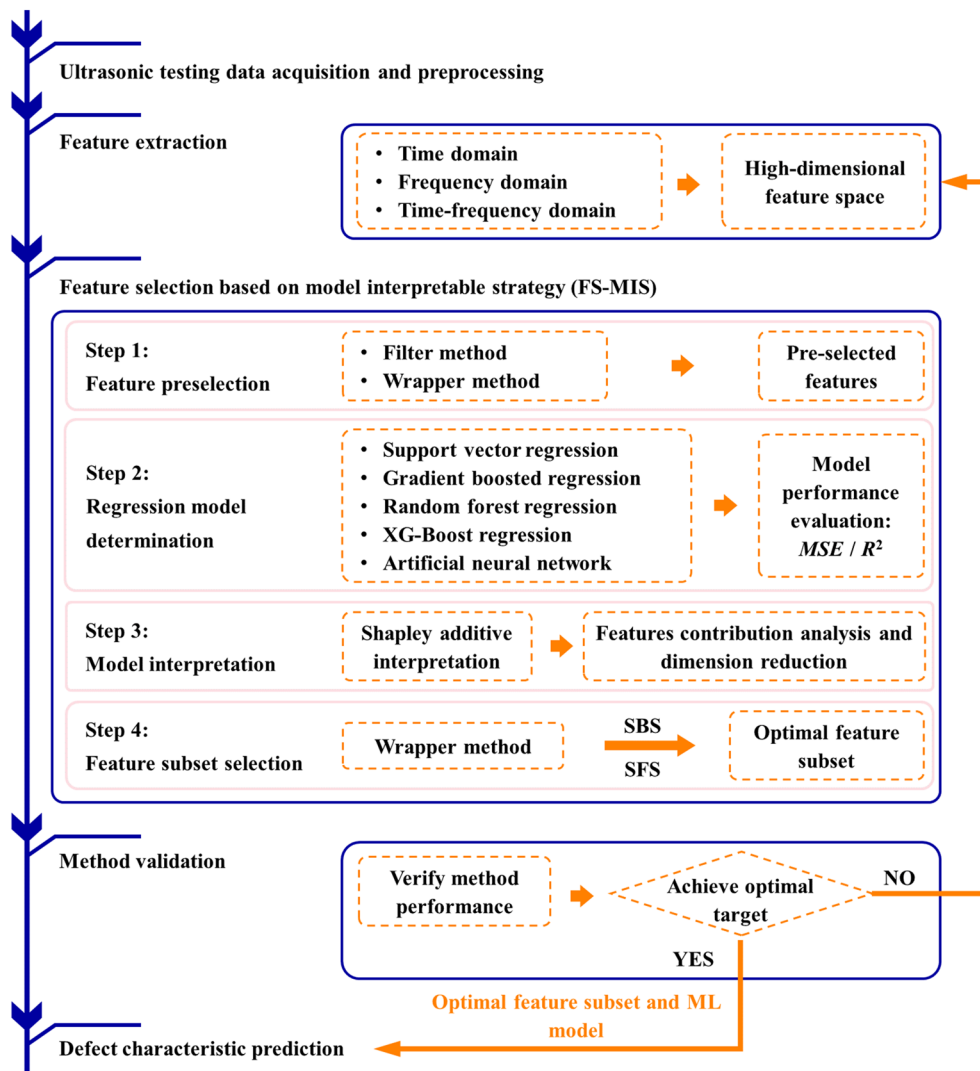
Figure 2 Machine learning framework based on model interpretation strategy for improving UT

processing to simultaneously improve the time and imaging resolutions and verify the performance.

3.2 Construction of Feature Space

Considering that the key to improving time and imaging resolutions is to decouple the overlapped pulse-echoes from two closely spaced scatterers [49], the outputs of the ML model adopted the corresponding ToAs t_1 and t_2. As given by Eq. (5), the predicted ToAs of the scattered waves are assigned to the corresponding original signal to decouple the overlapped pulse-echoes. If and only if $t = t_1$ or t_2, the signal amplitude is 1; otherwise, the amplitudes are all equal to 0. The schematic diagrams of the raw and decoupled time-domain signals are shown in Figure 5.

$$\hat{y}_{ij}(t) = \begin{cases} 1, & t = t_1 \text{ and } t_2, \\ 0, & t \neq t_1 \text{ or } t_2. \end{cases} \tag{5}$$

The initial feature space was established by extracting 82 features from each A-scan signal in the FMC datasets based on various signal processing techniques. There were 21 statistical features associated with signal amplitude and time information extracted in the time domain, including peak value, ToA of peak value, root-mean-square, peak-to-peak value, variance and skewness [52], etc. Shannon entropy [53] is a measurement of uncertainty and depicts the distribution and variation of UT signals. The entropy at given scales of the UT signal from SDHs always varies with central distance and can be
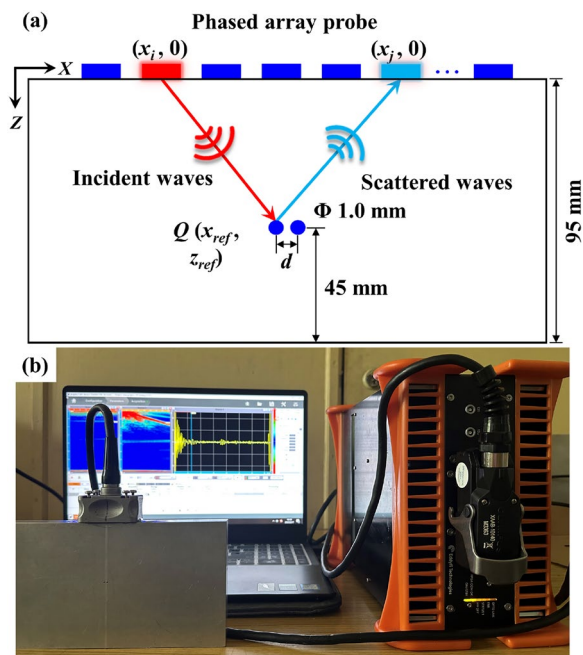
Figure 3 Schematic diagram of aluminum alloy specimens and experimental equipment: **a** schematic diagram of aluminum alloy specimens; **b** experimental equipment
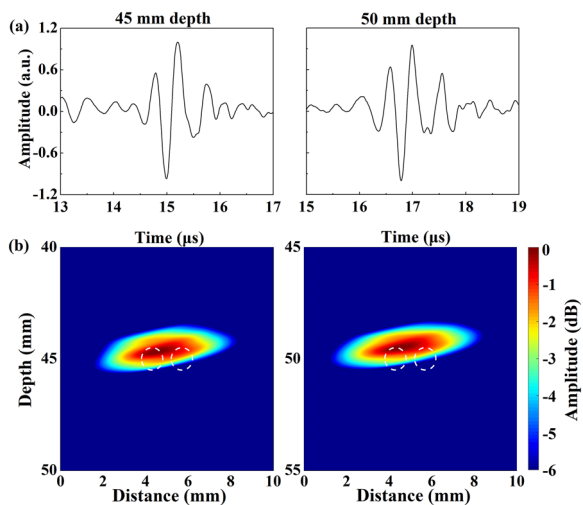

Figure 4 Typical experimental A-scan signals and TFM images for the SDHs with 0.5λ central distance and different depths: **a** A-scan signals; **b** TFM images

considered as another important feature for defect characterization [54].

Frequency domain analysis extracts the features advantageous in defect identification [16]. For example, the intervals between the extreme values in the frequency spectrum are related to the path/time difference of the scattered waves from adjacent defects [5]. A total of 22 features were extracted from the frequency spectrum
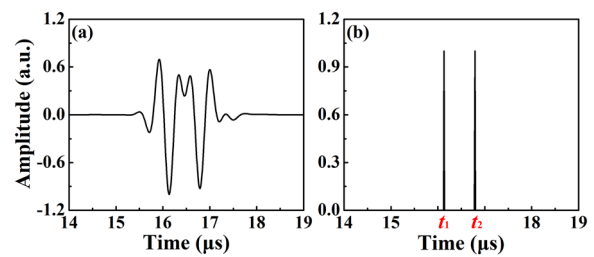

Figure 5 Schematic diagrams of the raw and decoupled time-domain signals: **a** raw signal $y_{ij}(t)$, **b** decoupled signal $\hat{y}_{ij}(t)$

obtained by fast Fourier transform (FFT), such as maximum amplitude, mean square frequency, − 6 dB bandwidth, resonant frequency, gravity frequency and frequency variance [10], etc. In addition, autoregressive (AR) spectrum extrapolation has the ability to extend the effective frequency band and compress time-domain pulse width to improve time resolution [55]. To this end, AR spectrum extrapolation was implemented on each A-scan signal. The AR parameters were determined by knowledge-based methods [49], and the AR coefficients were extracted as frequency-domain features [56, 57].

In time-frequency domain analysis, wavelet packet transform (WPT) with 'DB5' mother wavelet and 4 deposition layers was used to decompose each A-scan signal into 16 frequency band signals. The Shannon entropy of each frequency band and the energy ratio in total energy were extracted as the time-frequency domain features, resulting in a total of 32 features. Furthermore, as an adaptive time-frequency analysis method, EMD [58] was introduced to decompose UT signals into a finite number of stationary intrinsic mode functions (IMFs). The largest eigenvalue of the covariance matrix constructed by all IMFs (except the residual IMF) [20], along with the normalized energy and energy moment of the first three IMFs, are adopted as the time-frequency features.

3.3 Selection of Features and Regression Model
Feature selection has significant influences on the predictive accuracy of ML models. Determining suitable features can reduce the complexity and overfitting, alleviate the effect of the curse of dimensionality and improve the generalization capability and interpretability [26]. In this paper, FS-MIS was proposed by integrating SHAP, filter method, embedded method and wrapper method to reduce the dimension of initial feature space and make feature selection more physically interpretable. The optimal ML model was determined simultaneously in this process, and the sensitive features with minimum redundancy and maximum relevance to target defects were selected self-adaptively.

Firstly, the filter method was implemented to select features. The features whose variances dissatisfied the threshold of 0.05 were removed, and 66 features were retained, since the feature with low variance is not beneficial to the discrimination of different samples [29].

Mutual information (MI) was used to measure the linear or nonlinear relationship between each feature and ToAs. The irrelevant features with the maximal information coefficient (MIC) equal to 0 were removed from the feature space.

$$\text{MIC: } I(A; B) = \int \int p(a,b) \lg \left(\frac{p(a,b)}{p(a)p(b)} \right) \text{d}a\text{d}b, \tag{6}$$

where $p(a)$ and $p(b)$ are respectively the probability of input a and output b, and $p(a, b)$ is the joint distribution probability of a and b.

Embedded method integrates the feature selection and the training of the learner, which are completed in the same optimization process. A total of 20 important features higher than the average weight were determined by random forest method [59], as shown in Table 1.

A total of 12288 A-scan signals (12 FMC datasets) were acquired by experiments and randomly divided into 80% training data and 20% testing data. SVR, GBR, RFR, XG-boost and BPNN were adopted to establish regression models. The hyper-parameters of each model were found by grid search. Ten-fold cross-validated-average MSE and R^2 were calculated to evaluate the accuracy of the above models. As shown in Figure 6, the BPNN model has the best overall performance with the lowest MSE and highest R^2, since it has a strong ability for data mining and solving inverse problems with highly nonlinear correlations [10] and the mapping between input and output
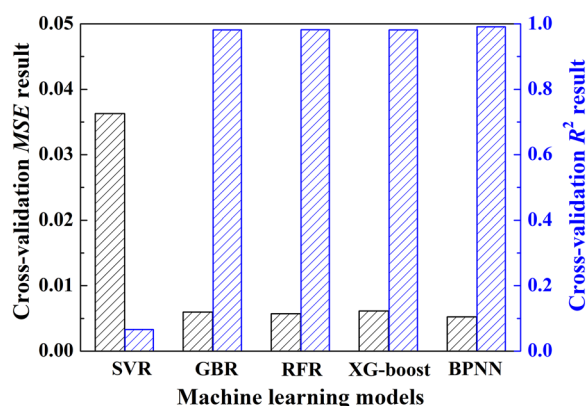

Figure 6 Performance metrics of different ML models

data can be obtained by adaptive training with sufficient samples. Consequently, BPNN was chosen as the optimal ML model in the following parts.

In addition, strong correlations may exist among the 20 selected features. If one feature provides enough information, the other highly relevant features no longer provide additional contributions. Pearson correlation coefficient was calculated to select relevant features for overcoming the influence of multicollinearity. Figure 7a shows the correlation degree between features. The grids with crossed horizontal and vertical coordinates represent the Pearson correlation coefficient scores. The darker color indicates a higher correlation degree between the two features. The 20 features were divided into six groups of relevant features (the absolute value of Pearson correlation coefficient > 0.9) and five independent features (P_{10}, P_{11}, P_{12}, P_{13} and P_{20}). Subsequently, SHAP was incorporated to analyze the importance of each feature on the outputs in BPNN model. Figure 7b depicts the stacks of the mean absolute SHAP values of each feature for

Table 1 Indexes and implications of 20 important features

| Feature index	Feature name	Feature index	Feature name
P_1	Maximum amplitude of time-domain signal	P_{11}	The low frequency corresponding to the − 6 dB bandwidth of frequency spectrum
P_2	ToA of maximum amplitude	P_{12}	The high frequency corresponding to the − 6 dB bandwidth of frequency spectrum
P_3	Temporal kurtosis	P_{13}	Resonant frequency
P_4	Minimum amplitude of time-domain signal	P_{14}	Spectral mean
P_5	ToA of minimum amplitude	P_{15}	Spectral variance
P_6	Temporal mean	P_{16}	Spectral skewness
P_7	Crest factor	P_{17}	Spectral kurtosis
P_8	Temporal margin index	P_{18}	Energy ratio of 1st wavelet sub-band
P_9	Spectral peak value	P_{19}	Energy ratio of 2nd wavelet sub-band
P_{10}	Center frequency	P_{20}	Energy ratio of 3rd wavelet sub-band
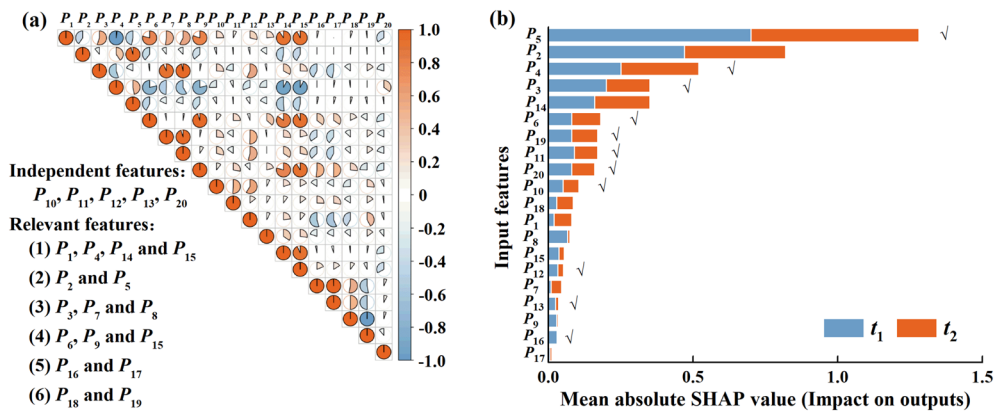
Figure 7 Results of feature selection by Pearson correlation and SHAP: (**a**) Pearson correlation coefficients, (**b**) mean absolute SHAP values of 14 features for different outputs (t_1 and t_2)

two outputs (t_1 and t_2), and the higher sum indicates the greater impact during the prediction process [35]. It can be seen that the importance of features is different. For each group of the relevant features, the features with higher SHAP values (P_4, P_5, P_3, P_6, P_{16} and P_{19}) were selected and retained together with the other five independent features, resulting in a total of 11 features.

To further reduce the redundancy of feature space, it is necessary to consider the contribution of feature combination. Two greedy wrapper methods (SFS and SBS) were adopted to select the optimal feature subset from the 11 features. A total of 12288 A-scan signals were split into the training set and testing set at a ratio of 8:2, and the ten-fold cross-validated-average *MSE* and R^2 were used to test the predictive accuracy of different feature subsets.

(1) SFS starts with an empty set and iteratively selects one feature at a time until no improvement in predictive accuracy can be achieved. As shown in Figure 8a, the feature set ***P1*** = (P_3, P_4, P_5, P_6, P_{11}, P_{12}, P_{13}, P_{19}, P_{20}) has the smallest *MSE* = 0.0050 and the largest R^2 = 0.99197.

(2) SBS starts with the set of all features and progressively eliminates the least promising one. This process stops if the performance of the learning algorithm drops below a given threshold. As shown in Figure 8b, the features set ***P2*** = (P_3, P_4, P_5, P_6, P_{11}, P_{13}, P_{16}, P_{19}, P_{20}) has the smallest *MSE* = 0.0049 and the largest R^2 = 0.99198.

The two feature subsets determined by SFS and SBS both contained nine features, of which only one feature was different. Considering that the *MSE* and R^2 of ***P1*** and ***P2*** were almost the same, the feature subset ***P2*** with relatively good performance was selected as the optimal feature subset in this study. The results
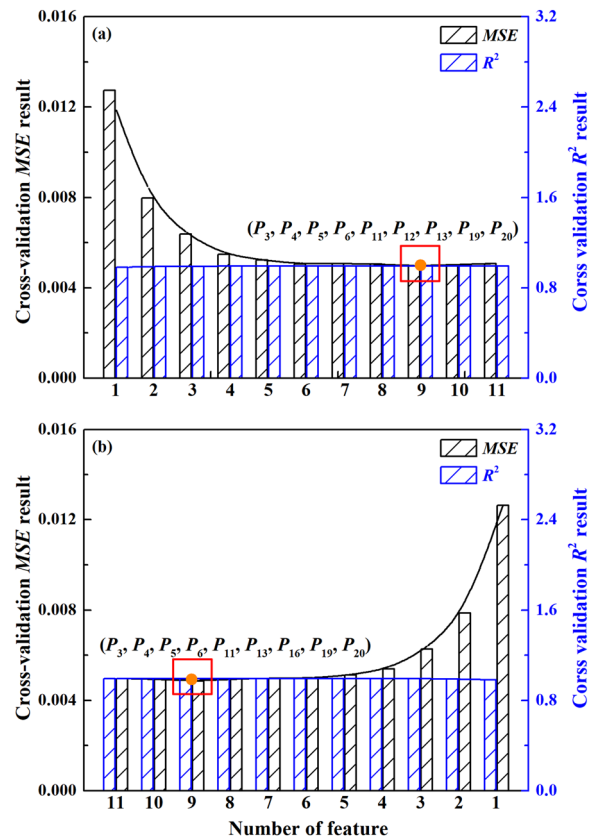

Figure 8 Results of feature selection by different wrapper methods: (**a**) SFS, (**b**) SBS

demonstrated that time-domain features, frequency-domain features and the features obtained by wavelet decomposition were identified as the most significant features for predicting ToAs.

3.4 Experimental Results and Analysis

3.4.1 Enhancement of Time Resolution in UT

As mentioned in Section 3.1, 12288 A-scan signals (12 FMC datasets) were acquired from the aluminum alloy specimens, where the central distances of SDHs were varied from 0.5λ to 1.0λ. Taking the 0.5λ central distance SDHs in 45 mm and 50 mm depths as examples, some typical UT signals captured by different transmitter-receiver pairs are presented in Figures 9a, c, respectively. The pulse-echoes from SDHs are overlapped, and it is challenging to extract the ToA of the respective scattered wave. Nine features (P_3, P_4, P_5, P_6, P_{11}, P_{13}, P_{16}, P_{19}, P_{20}) were extracted from each A-scan signal to construct the feature set used as the inputs of BPNN model. Meanwhile, the ToAs (t_1 and t_2) of the scattered waves from adjacent SDHs were set as the outputs. The 10240 signals collected from the SDHs with $0.6\lambda \sim 1.0\lambda$ central distance were employed to train the model for obtaining the optimized weights and biases, while the remaining 2048 signals corresponding to 0.5λ central distance SDHs were used to test the model.

The calculated R^2 and *MSE* are respectively equal to 0.99 and 0.0055, indicating that the trained BPNN model has excellent predictive accuracy and generalization capability [21].

Figures 10a, b present the predicted ToAs of the testing data. The discrete points are well located around the solid line with a slope of 1, indicating that the predicted values are approximately the same as the expected values.
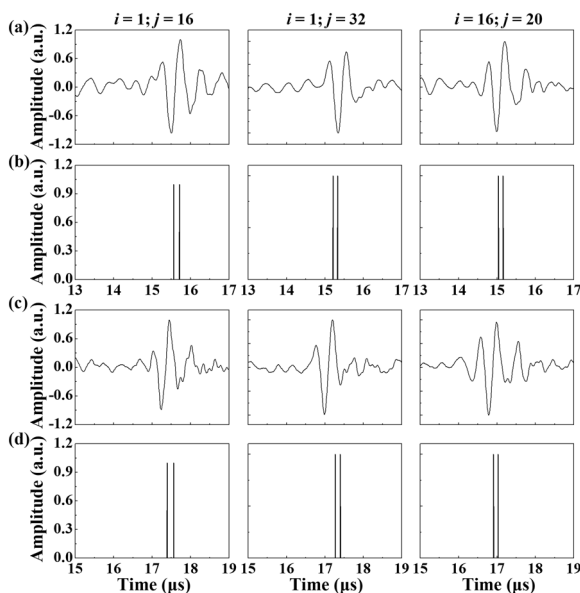

Figure 9 Raw and decoupled signals from different transmitter-receiver pairs for the SDHs with 0.5λ central distance and different depths: (**a**) Raw signals-45 mm, (**b**) Decoupled signals-45 mm, (**c**) Raw signals-50 mm, (**d**) Decoupled signals-50 mm
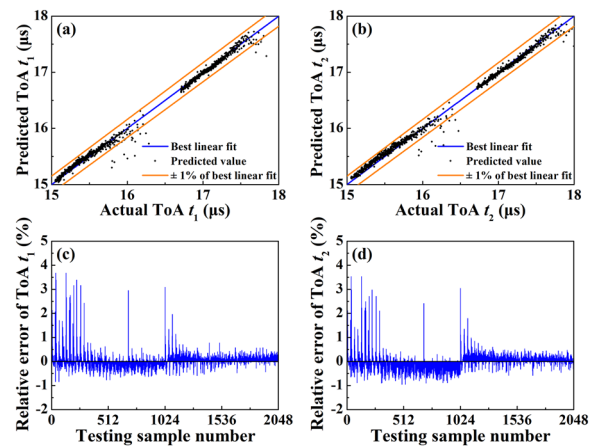

Figure 10 Comparison of actual value and predicted value: (**a**) t_1, (**b**) t_2; and relative error between actual value and predicted value: (**c**) t_1, (**d**) t_2

The band lines in the figures show that about 98% of predicted values are within 1% deviation from the expected values. Figures 10c, d show the relative errors of the predicted ToAs, which are all below 3.67% with an average error of 0.25%. Such low errors suggest that the proposed ML framework based on model interpretation strategy effectively separates the overlapped UT signals and improves the time resolution, i.e., $t_2 - t_1$.

3.4.2 Enhancement of Imaging Resolution in UT

The predicted ToAs presented in Section 3.4.1 were applied to reconstruct new FMC datasets containing decoupled signals for TFM imaging. As shown in Figure 11a, the SDHs with 0.5λ central distance at different depths are identified from the delay-and-sum images. The relative measurement errors of hole depths and central distances are no more than 0.51% and 3.57%, respectively.

Two key parameters, i.e., the peak to central intensity difference (τ) and the array performance indicator (API) [60], were introduced to describe the TFM images quantitatively. The smaller τ and API values refer to better imaging performance. Figure 11b presents the cross-sections taken through the centers of the SDHs in TFM images with raw FMC datasets and reconstructed high-resolution FMC datasets. The API values for the latter are reduced by 92.71% and 87.39% compared to those for the former. It is difficult to determine τ values from the original TFM images. In contrast, the τ values for the TFM images with reconstructed FMC datasets are -17.32 dB and -16.42 dB, less than -6 dB. The experimental results demonstrate that the proposed framework is suitable for determining the optimal ML model and feature subset, accurately predicting the ToAs of the scattered
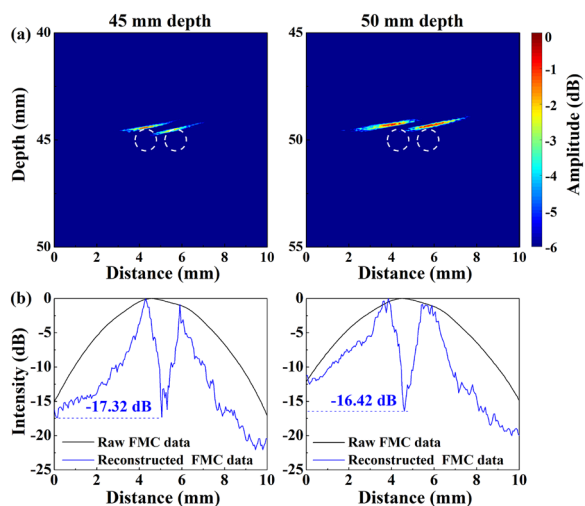
Shi *et al. Chinese Journal of Mechanical Engineering* (2023) 36:127

Page 10 of 13

Figure 11 TFM images of the SDHs with 0.5λ central distance at different depths based on the proposed framework: (**a**) TFM images, (**b**) Cross-sections taken through scatterers

Figure 12 Performance metrics of BPNN model with different feature selection methods

waves from adjacent defects. The imaging resolution can be improved to subwavelength-scale by combining the proposed framework and TFM, breaking the diffraction limit and highlighting the target characteristics with accurate location.

4 Discussion

The proposed FS-MIS was validated by comparing it with four commonly used feature selection methods, including PCA, FA, kernel principal component analysis (KPCA) and independent component analysis (ICA). PCA is a linear dimensionality reduction technique representing the maximum variance in the data [30]. KPCA is a nonlinear PCA developed with the kernel method by transforming the input features into a high-dimensional space through the nonlinear mapping function and performing PCA to achieve feature fusion and dimension reduction [61]. FA describes the variability among the original features in terms of fewer variable factors [31]. The original features are modeled as the linear combinations of factors plus error. ICA is a statistical and computational technique for revealing hidden information underlying feature set [31]. The original features in ICA are transformed into new features which are mutually statistically independent [62]. In a word, these four methods integrate the high-dimensional initial feature space to significant low-dimensional features.

The mentioned feature selection methods were used to reduce the dimensionality of the initial feature space. The first two eigenvalues in PCA exhibited the maximum cumulative proportion variation equal to 0.99 and we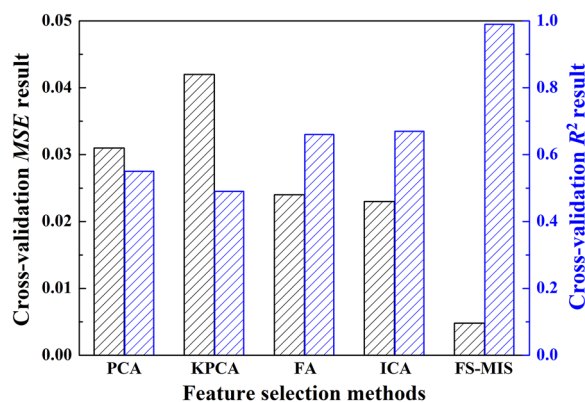re chosen for evaluation. The first five principal component features were obtained by KPCA with the polynomial kernel method. Ten factors were selected by FA according to the variance percentage. FastICA algorithm was applied to ICA, and five independent components were extracted from the initial feature space. The feature sets determined by the aforementioned methods were used independently as the inputs to predict the ToAs in BPNN model. The dataset with 12288 A-scan signals was randomly split into the training set and testing set at a ratio of 8:2. The ten-fold cross-validated-average MSE and R^2 were employed to test the predictive performance of each feature set. As shown in Figure 12, FS-MIS has the lowest MSE (0.0048) and the highest R^2 (0.99), i.e., the best overall performance compared to other unsupervised techniques (PCA, KPCA, FA and ICA). The unsupervised dimensionality reduction is implemented based on the features rather than the effect of each feature and feature combination on the targets. In contrast, the proposed FS-MIS method has the capability to self-adaptively obtain the optimal feature subset by integrating SHAP, filter method, embedded method and wrapper method, quantitatively analyzing the contributions of each feature and feature combination.

To demonstrate the advantages of FS-MIS method in improving computational efficiency, we compared the performance of the BPNN models trained with nine features selected by FS-MIS and all 82 initial features. For the 12288 experimental signals in Section 3, 10240 signals corresponding to the SDHs with 0.6λ ~ 1.0λ central distances were employed for training the model, and the remaining 2048 signals corresponding to the SDHs with 0.5λ central distance were adopted to test the model. On this basis, 82 features extracted from each A-scan signal were used as the inputs to predict the ToAs. The statistical indexes R^2 and MSE were equal to 0.99 and 0.0092, respectively. Compared to the

evaluation results with nine features, the performance of the trained model with 82 features is still at a high level, and the predictive accuracy falls slightly. However, the training time is up to 167.25 s, while that of nine features is only 5.41 s. The results demonstrate that the proposed FS-MIS method is beneficial to improve computational efficiency with high predictive accuracy.

As given by Eq. (5), the predicted ToAs using 82 features were also employed to decouple the overlapped pulse-echoes. Figure 13 shows the relative errors of the testing dataset between predicted ToAs and expected values, where the average error is 0.37% and is increased by 0.12% compared to Figures 10c, d. Subsequently, the predicted ToAs were employed to reconstruct high-resolution FMC datasets, and TFM imaging was conducted by delay-and-sum beamforming. As illustrated in Figure 14a, the SDHs with 0.5λ central distance in 45 mm and 50 mm depths are resolved, but the maximum measurement errors of hole depths and central distance were 0.71% and 59.59%, much larger than those observed in Figure 11a. Figure 14b presents τ values and API values of the TFM images obtained by different feature sets. Compared to the TFM images combined with 82 features, the τ and API values corresponding to nine features are reduced significantly. The experimental results demonstrate that the feature subset selected by FS-MIS excellently describes the intrinsic property of UT signals and accurately predicts the ToAs of the scattered waves from adjacent defects. The proposed ML framework based on model interpretation strategy is beneficial to improving the accuracy of defect characterization and calculation efficiency to
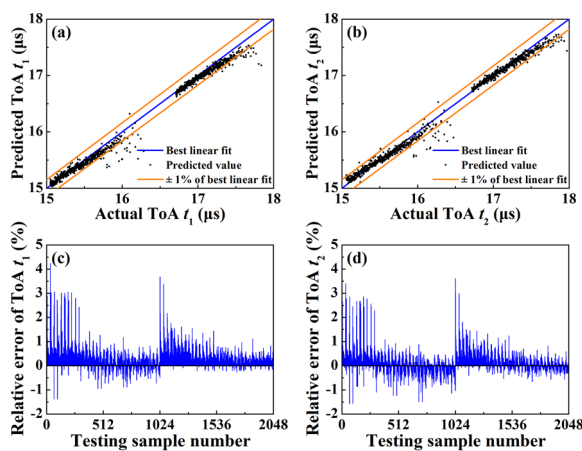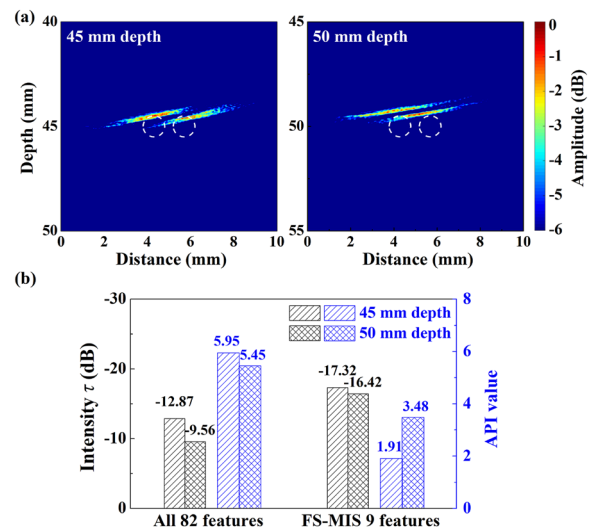

Figure 14 TFM images of the SDHs with 0.5λ central distance based on all 82 features and the performance indicators of the TFM images combined with different feature sets: (**a**) TFM images, (**b**) Performance indicators

meet the requirements of nondestructive testing and evaluation.

5 Conclusions and Further Work

(1) A generally applicable ML framework for UT based on model interpretation strategy is proposed to improve the accuracy and efficiency of defect characterization. Signal processing techniques are conducted to extract multi-domain features from the UT signals and construct typical feature space. FS-MIS method is developed to self-adaptively determine the optimal feature subset showing better correlation with the target defects and make the feature selection more physically interpretable.

(2) The experimental results indicate that the proposed framework has the capability to decouple the overlapped pulse-echoes from the SDHs with 0.5λ central distance and improve the time resolution of UT signals. The relative errors of the predicted ToAs are all below 3.67% with an average error of 0.25%. On this basis, the ultrasonic imaging resolution is enhanced to 0.5λ by combining TFM. The relative measurement errors of hole depths and central distance are no more than 0.51% and 3.57%, respectively.

(3) FS-MIS is adopted to visualize the contributions of each feature and feature combination on targets by integrating the SHAP, filter method, embedded method and wrapper method. Compared to the initial feature space and the features determined by

Figure 13 Comparison of actual value and predicted value trained by 82 features: (**a**) t_1, (**b**) t_2, and relative error between actual value and predicted value: (**c**) t_1, (**d**) t_2

Shi *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:127

Page 12 of 13

conventional dimensionality reduction techniques, the feature subset selected by FS-MIS is beneficial to improving the predictive accuracy and computational efficiency of ML models.

(4) In future work, more diverse datasets corresponding to the defects with various sizes, shapes and locations will be incorporated for accurately detecting and characterizing unknown damage. In addition, we will also explore the comprehensive impact of structural noise originating from grain boundaries and structural features in multi-phase materials on the predictive performance of the ML framework.

Author Contributions
SS wrote the draft manuscript and conducted experiment; LL and SJ in charge of the whole trial; LL and SJ checked and improved the manuscript in writing. DZ, JL and DF gave some advices on the manuscript. All authors read and approved the final manuscript.

Authors' Information
Siqi Shi, born in 1995, is currently a PhD candidate at *NDT & E Laboratory, Dalian University of Technology, China*. Her research interests include ultrasonic signal processing and machine learning.

Shijie Jin, born in 1984, is currently an associate professor at *NDT & E Laboratory, Dalian University of Technology, China*. His research interest is nondestructive testing and evaluation for materials.

Donghui Zhang, born in 1974, is currently a center manager at *China Nuclear Industry 23 Construction Co., Ltd., China*. His research interests include construction of nuclear engineering and nondestructive testing.

Jingyu Liao, born in 1985, is currently a director of research at *China Nuclear Industry 23 Construction Co., Ltd., China*. Her research interests include construction of nuclear engineering and nondestructive testing.

Dongxin Fu, born in 1996, is currently a verification service engineer at *China Nuclear Industry 23 Construction Co., Ltd., China*. Her main research interest is nondestructive testing and evaluation for materials.

Li Lin, born in 1970, is currently a Professor at *NDT & E Laboratory, Dalian University of Technology, China*. Her main research interest is nondestructive testing and evaluation for materials.

Availability of Data and Materials
All data generated or analyzed during this study are included in this published article.

Declarations

Competing Interests
The authors declare no competing financial interests.

References
[1] Z Wang, Z C Fan, X D Chen, et al. Modeling and experimental analysis of roughness effect on ultrasonic nondestructive evaluation of micro-crack. *Chinese Journal of Mechanical Engineering*, 2021, 34: 114.

[2] A L Bowler, M P Pound, N J Watson. A review of ultrasonic sensing and machine learning methods to monitor industrial processes. *Ultrasonics*, 2022, 124: 106776.

[3] J Chen, E Y Wu, H T Wu, et al. Enhancing ultrasonic time-of-flight diffraction measurement through an adaptive deconvolution method. *Ultrasonics*, 2019, 96: 175-180.

[4] X Sun, L Lin, Z Y Ma, et al. Enhancement of time resolution in ultrasonic time-of-flight diffraction technique with frequency-domain sparsity-decomposability inversion (FDSDI) method. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2021, 68(10): 3204-3215.

[5] S J Jin, B Zhang, X Sun, et al. Reduction of layered dead zone in Time-of-Flight Diffraction (TOFD) for pipeline with spectrum analysis method. *Journal of Nondestructive Evaluation*, 2021, 40(2): 48.

[6] C G Fan, L Yang, Y Zhao. Ultrasonic multi-frequency time-reversal-based imaging of extended targets. *NDT & E International*, 2020, 113: 102276.

[7] J Rao, J Yang, M Ratassepp, et al. Multi-parameter reconstruction of velocity and density using ultrasonic tomography based on full waveform inversion. *Ultrasonics*, 2020, 101: 106004.

[8] X Sun, L Lin, S J Jin. Resolution enhancement in ultrasonic TOFD imaging by combining sparse deconvolution and synthetic aperture focusing technique (Sparse-SAFT). *Chinese Journal of Mechanical Engineering*, 2022, 35: 94.

[9] H Lee, B Koo, A Chattopadhyay, et al. Damage detection technique using ultrasonic guided waves and outlier detection: Application to interface delamination diagnosis of integrated circuit package. *Mechanical Systems and Signal Processing*, 2021, 160: 107884.

[10] K X Zhang, G L Lv, S F Guo, et al. Evaluation of subsurface defects in metallic structures using laser ultrasonic technique and genetic algorithm-back propagation neural network. *NDT & E International*, 2020, 116: 102339.

[11] J Tong, M Lin, X Wang, et al. Deep learning inversion with supervision: A rapid and cascaded imaging technique. *Ultrasonics*, 2022, 122: 106686.

[12] S J Farley, J F Durodola, N A Fellows, et al. High resolution non-destructive evaluation of defects using artificial neural networks and wavelets. *NDT & E International*, 2012, 52: 69-75.

[13] F Nafiah, A Sophian, M R Khan, et al. Quantitative evaluation of crack depths and angles for pulsed eddy current non-destructive testing. *NDT & E International*, 2019, 102: 180-188.

[14] W Xu, X Li, J Zhang, et al. Ultrasonic signal enhancement for coarse grain materials by machine learning analysis. *Ultrasonics*, 2021, 117: 106550.

[15] M Yuan, J Li, Y Liu, et al. Automatic recognition and positioning of wheel defects in ultrasonic B-scan image using artificial neural network and image processing. *Journal of Testing and Evaluation*, 2020, 48(1): 308-322.

[16] S Buchaiah, P Shakya. Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. *Measurement : Journal of the International Measurement Confederation*, 2022, 188: 110506.

[17] C Yang, B Hou, B Ren, et al. CNN-based polarimetric decomposition feature selection for PolSAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 8796-8812.

[18] J Liu, G Xu, L Ren, et al. Defect intelligent identification in resistance spot welding ultrasonic detection based on wavelet packet and neural network. *The International Journal of Advanced Manufacturing Technology*, 2017, 90: 2581-2588.

[19] Y Wang. Wavelet transform based feature extraction for ultrasonic flaw signal classification. *Journal of Computers*, 2014, 9(3): 725-732.

[20] M Mousavi, M S Taskhiri, D Holloway, et al. Feature extraction of wood-hole defects using empirical mode decomposition of ultrasonic signals. *NDT & E International*, 2020, 114: 102282.

[21] L Lin, W Zhang, Z Y Ma, et al. Porosity estimation of abradable seal coating with an optimized support vector regression model based on multi-scale ultrasonic attenuation coefficient. *NDT & E International*, 2020, 113: 102272.

[22] D W Huang, S H Tang, D J Zhou, et al. NOx emission estimation in gas turbines via interpretable neural network observer with adjustable intermediate layer considering ambient and boundary conditions. *Measurement*, 2022, 189: 110429.

[23] L C Silva, E F Simas Filho, M C S Albuquerque, et al. Segmented analysis of time-of-flight diffraction ultrasound for flaw detection in welded steel plates using extreme learning machines. *Ultrasonics*, 2020, 102: 106057.

[24] Z Y Ma, W Zhang, Z B Luo, et al. Ultrasonic characterization of thermal barrier coatings porosity through BP neural network optimizing Gaussian process regression algorithm. *Ultrasonics*, 2020, 100: 105981.

[25] L Bai, M Liu, N Liu, et al. Dimensionality reduction of ultrasonic array data for characterization of inclined defects based on supervised locality preserving projection. *Ultrasonics*, 2022, 119: 106625.

[26] K Zhang, Y Li, P Scarf, et al. Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks. *Neurocomputing*, 2011, 74(17): 2941-2952.

[27] C Lin, H Chen, Y Wu. Study of image retrieval and classification based on adaptive features using genetic algorithm feature selection. *Expert Systems with Applications*, 2014, 41(15): 6611-6621.

[28] I A Gheyas, L S Smith. Feature subset selection in large dimensionality domains. *Pattern Recognition*, 2010, 43(1): 5-13.

[29] Q Nazir, C Shao. Online tool condition monitoring for ultrasonic metal welding via sensor fusion and machine learning. *Journal of Manufacturing Processes*, 2021, 62: 806-816.

[30] H Abdi, L J Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(4): 433-459.

[31] D Salas-Gonzalez, J M Gorriz, J Ramirez, et al. Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. *Medical Physics*, 2010, 37(11): 6084-6095.

[32] G L Lv, S F Guo, D Chen, et al. Laser ultrasonics and machine learning for automatic defect detection in metallic components. *NDT & E International*, 2023, 133: 102752.

[33] R Rodríguez-Pérez, J Bajorath. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 2020, 34(10): 1013-1026.

[34] W N Xu, S D Fan, C P Wang, et al. Leakage identification in water pipes using explainable ensemble tree model of vibration signals. *Measurement*, 2022, 194: 110996.

[35] T Ye, M Dong, Y Liang, et al. Modeling and optimization of the NOx generation characteristics of the coal-fired boiler based on interpretable machine learning algorithm. *International Journal of Green Energy*, 2021: 1-15.

[36] L Bai, F Le Bourdais, R Miorelli, et al. Ultrasonic defect characterization using the scattering matrix: A performance comparison study of Bayesian inversion and machine learning schemas. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2021, 68(10): 3143-3155.

[37] Z Ge, Z Song, S X Ding, et al. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, 2017, 5: 20590-20616.

[38] E M Mahjoub, Y Slah, R José, et al. Feature selection techniques for identifying the most relevant damage indices in SHM using Guided Waves. *8th European Workshop on Structural Health Monitoring (EWSHM 2016)*, 2016.

[39] Y Peng, H Liu, X Li, et al. Machine learning method for energy consumption prediction of ships in port considering green ports. *Journal of Cleaner Production*, 2020, 264: 121564.

[40] M G Li, F Wang, X J Jia, et al. Multi-source data fusion for economic data analysis. *Neural Computing and Applications*, 2021, 33: 4729-4739.

[41] D Rumelhart, G E Hinton, R J Williams. Learning representations by back propagation errors. *Nature*, 1986, 323(6088): 533-536.

[42] S Bedi, A Samal, C Ray, et al. Comparative evaluation of machine learning models for groundwater quality assessment. *Environmental Monitoring and Assessment*, 2020, 192(12): 1-23.

[43] L S Shapley. *A value for n-person games*. Contributions to the Theory of Games, Princeton Univ Press, Princeton, NJ, USA, 1953: 307-317.

[44] W C Zhao, C Zheng, B Xiao, et al. Composition refinement of 6061 Aluminum alloy using active machine learning model based on Bayesian optimization sampling. *ACTA Metallurgica Sinica*, 2021, 57 (6): 797-810. (in Chinese)

[45] A W Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 1971, 20(9): 1100-1103.

[46] S F Cotter, K Kreutz-Delgado, B D Rao. Backward sequential elimination for sparse vector subset selection. *Signal Processing*, 2001, 81: 1849-1864.

[47] H Zhou, Z Han, D Du. An improved ultrasonic imaging method for Austenitic welds based on grain orientation distribution inversion algorithm. *Journal of Nondestructive Evaluation*, 2020, 39(3): 54.

[48] S K Shastri, S Rudresh, R Anand, et al. Axial super-resolution in ultrasound imaging with application to non-destructive evaluation. *Ultrasonics*, 2020, 108: 106183.

[49] S Q Shi, L Lin, Z B Luo, et al. Resolution enhancement of ultrasonic imaging at oblique incidence by using WTFM based on FMC-AR. *Measurement*, 2021, 183: 109798.

[50] X Y Zhao, Z M Ma, J Y Zhang. Simplified matrix focusing imaging algorithm for ultrasonic nondestructive testing. *Chinese Journal of Mechanical Engineering*, 2022, 35: 19.

[51] N Laroche, S Bourguignon, E Carcreff, et al. An inverse approach for ultrasonic imaging from full matrix capture data. Application to resolution enhancement in NDT. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020, 67(9): 1877-1887.

[52] H Yun, R Rayhana, S Pant, et al. Nonlinear ultrasonic testing and data analytics for damage characterization: A review. *Measurement*, 2021, 186: 110155.

[53] M Meng, Y J Chua, E Wouterson, et al. Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. *Neurocomputing*, 2017, 257: 128-135.

[54] Z Gao, Y Liu, Q Wang, et al. Ensemble empirical mode decomposition energy moment entropy and enhanced long short-term memory for early fault prediction of bearing. *Measurement : Journal of the International Measurement Confederation*, 2022, 188: 110417.

[55] F Honarvar, H Sheikhzadeh, M Moles, et al. Improving the time-resolution and signal-to-noise ratio of ultrasonic NDE signals. *Ultrasonics*, 2004, 41(9): 755-763.

[56] P Li, Z Q Lang, L Zhao, et al. System identification-based frequency domain feature extraction for defect detection and characterization. *NDT & E International*, 2018, 98: 70-79.

[57] G R B Ferreira, M G de Castro Ribeiro, A C Kubrusly, et al. Improved feature extraction of guided wave signals for defect detection in welded thermoplastic composite joints. *Measurement*, 2022, 198: 111372.

[58] M Mousavi, A H Gandomi. Wood hole-damage detection and classification via contact ultrasonic testing. *Construction and Building Materials*, 2021, 307: 124999.

[59] S T Yang, L J Gu, X F Li, et al. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote Sensing*, 2020, 12(19): 3119.

[60] C G Fan, M H Caleap, M C Pan, et al. A comparison between ultrasonic array beamforming and super resolution imaging algorithms for non-destructive evaluation. *Ultrasonics*, 2014, 54(7): 1842-1850.

[61] J Shen, F Xu. Method of fault feature selection and fusion based on poll mode and optimized weighted KPCA for bearings. *Measurement*, 2022, 194: 110950.

[62] J Lee, C Yoo, I Lee. Statistical process monitoring with independent component analysis. *Journal of Process Control*, 2004, 14: 467-485.