



# OPEN Machine learning–based method for analyzing stress distribution in a ship

Bowen Jin<sup>1</sup>, Ji Zeng<sup>1✉</sup>, Liuqi Xu<sup>2</sup>, Lei Chen<sup>2</sup> & Yiting Zhan<sup>1</sup>

Ship structural health monitoring (SHM) systems are essential for ensuring operational safety. Monitoring stress distribution is a critical function of such systems. Analysis of stress data collected by pressure sensors can enhance understanding of SHM performance and facilitate the prevention of structural failures. In this study, we propose a machine learning–based computational method to analyze the stress distribution of a crane ship. The method first employed extreme gradient boosting (XGBoost) to evaluate relationships among stress data from multiple pressure sensors, and then used multilayer perceptron (MLP) to construct regression models. For each pressure sensor, an optimal MLP regression model was established to infer its stress values from those of related sensors. These models demonstrated strong fitting performance on both training and independent test datasets. Furthermore, the method identified key pressure sensors whose measurements were particularly important for recovering stresses at most monitoring points. These findings improve interpretability of the method and provide insights into the mechanisms of the monitoring system. The robustness and rationality of the approach were further examined through ablation tests, confirming its effectiveness for ship SHM applications.

**Keywords** Stress distribution, Crane ship, Machine learning, XGBoost, Multilayer perceptron

Ship structural health monitoring (SHM) systems are of critical importance, particularly for specialized vessels such as crane ships. Designing an SHM system that delivers reliable monitoring requires and decisions on numerous specifications, including sensor types, quantities, and placements, as well as data transfer, storage, and analysis techniques. However, the complexity of hull structures and the variability of marine environments present substantial challenges for system topology and data acquisition. The deployment of numerous sensors across wide areas can result in issues such as signal transmission delays, data conflicts, and insufficient real-time performance. Meanwhile, temperature fluctuations, vibration interference, and systematic errors may distort monitoring data, thereby compromising the accuracy of subsequent analyses. A well-functioning SHM system mitigates equipment failures and safety risks caused by structural deficiencies, ensuring personal safety, asset protection, ecological security, and effective vessel operation. One of the principal functions of an SHM system is to monitor stress distribution across the hull, which is achieved by deploying pressure sensors at key structural locations. Data from these sensors form the primary input for stress measurement and modeling, inform sensor placement strategies, and must be utilized for stress correction. In this study, feature attribution and ablation analyses are applied to identify sensors with disproportionately high influence and to quantify performance degradation under sensor loss, thereby providing evidence-based guidance for sensor prioritization, maintenance, and resilient operation. Stress values detected by pressure sensors are therefore considered essential data for analyzing and optimizing SHM systems.

Historically, pressure sensor outputs were displayed on screens, where crew members relied on empirical thresholds to assess ship structural health and trigger alarms. Advances in computer science have enabled more sophisticated data processing, allowing stress measurements to be exploited for deeper knowledge extraction. For example, Mincheul et al. estimated local ice loads from shear forces acting on transverse frames<sup>1</sup>, and Tak-Kee et al. employed a finite element model to derive influence coefficients for converting measured hull strains into local ice pressures<sup>2</sup>. In practice, however, the limited number of pressure sensors installed on vessels hinders comprehensive stress monitoring. Certain structural locations are also unsuitable for direct sensor deployment, leaving stress values at those points unmeasured. Furthermore, SHM systems must incorporate fault tolerance: if one pressure sensor fails, the system should recover the corresponding stress values. Stress recovery is closely

<sup>1</sup>College of Ocean Science and Engineering, Shanghai Maritime University, Shanghai 201306, People's Republic of China. <sup>2</sup>College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China. ✉email: zengji@shmtu.edu.cn

related to these issues, as it enables reconstruction of full-field stress distributions and inference of stresses at unmonitored points. Several approaches have been proposed to address this problem, including modal methods<sup>3</sup>, Ko's displacement theory<sup>4</sup>, and inverse finite element methods (iFEM)<sup>5–11</sup>. Among these, iFEM has gained wide application in stress recovery research. For example, Wei et al. developed an iFEM-based framework to construct a real-time digital twin of ship hull structures, where measured strains were assimilated online to reconstruct the full deformation field, subsequently visualized in a VR environment<sup>10</sup>. Similarly, Shen et al. employed a radial basis function approach to solve the inverse prediction problem and optimized sensor placement schemes to support SHM<sup>11</sup>.

In recent years, artificial intelligence (AI) algorithms, particularly machine learning (ML) and deep learning (DL), have opened new avenues for addressing the stress recovery problem. Hassani et al. discussed the use of optimization algorithms in SHM and confirmed the role of AI in analyzing the large volumes of data generated by monitoring systems<sup>12</sup>. Karvelis et al. framed ship SHM as a classification problem and employed a discrete cosine transform combined with a deep neural network to construct a classification model<sup>13</sup>. Sun et al. developed a data-driven approach to recover the full-field stress distribution in ship hull structures<sup>14</sup> and later designed an end-to-end DL method for reconstructing full-field stress distributions<sup>15</sup>. Although these methods achieved strong performance, they did not identify the key factors underlying stress recovery. Such approaches are useful tools for reconstructing stress distributions but provide limited guidance for solving essential SHM problems, including optimal pressure sensor deployment and identification of the most critical monitoring points. This limitation reduces their capacity to advance fundamental understanding in ship engineering and support its development. Similar issues are evident in other SHM contexts, such as the connector structures of offshore platforms<sup>16</sup> and the Shenzhen Bay stadium<sup>17</sup>. Current research tends to emphasize developing powerful computational techniques rather than uncovering the essence of the problem. ML and DL algorithms are highly effective in prediction and regression tasks across diverse fields, largely due to their ability to analyze large-scale datasets and extract meaningful patterns. However, these algorithms cannot address all challenges. Most ML and DL models operate as black boxes, relying on complex iterative processes that limit interpretability and obscure the essential information underlying a problem. Designing efficient computational methods based on ML and DL is valuable, but uncovering the mechanisms behind the problem with the aid of these algorithms is equally important, particularly in fields where theoretical development is critical. For stress recovery, ML and DL can serve as excellent tools, but extracting the underlying principles of the problem is essential for guiding the design of more powerful and resilient SHM systems. Few studies, however, have addressed this aspect.

In this study, we propose an ML-based computational approach to analyze stress distribution in ships. The contributions are twofold. First, we constructed optimal regression models to infer stress at monitoring points using ML algorithms. This contribution is comparable to those reported in previous ML- and DL-based studies. Second, we extracted key monitoring points for the entire system by analyzing the components of the optimal regression models, thereby enhancing the interpretability of the method. To our knowledge, this aspect has received little attention in prior research. Stress data were collected from a crane ship to build one training dataset and one independent test dataset. In the training dataset, stress values measured by a given pressure sensor were treated as the target, while stress values from other sensors served as explanatory variables. We adopted extreme gradient boosting (XGBoost)<sup>18</sup>, a powerful feature selection algorithm, to evaluate the relationships between targets and variables and to generate feature rankings. Incremental feature selection (IFS)<sup>19</sup>, incorporating a multilayer perceptron (MLP) as the regression algorithm, was then applied to the ranked features. An optimal regression model was constructed for each pressure sensor, enabling inference of its stress values based on readings from related sensors. These models achieved high performance on both training and independent test datasets. Moreover, we identified the key pressure sensors in the monitoring system, offering new insights into the mechanisms underlying stress distributions measured by SHM systems.

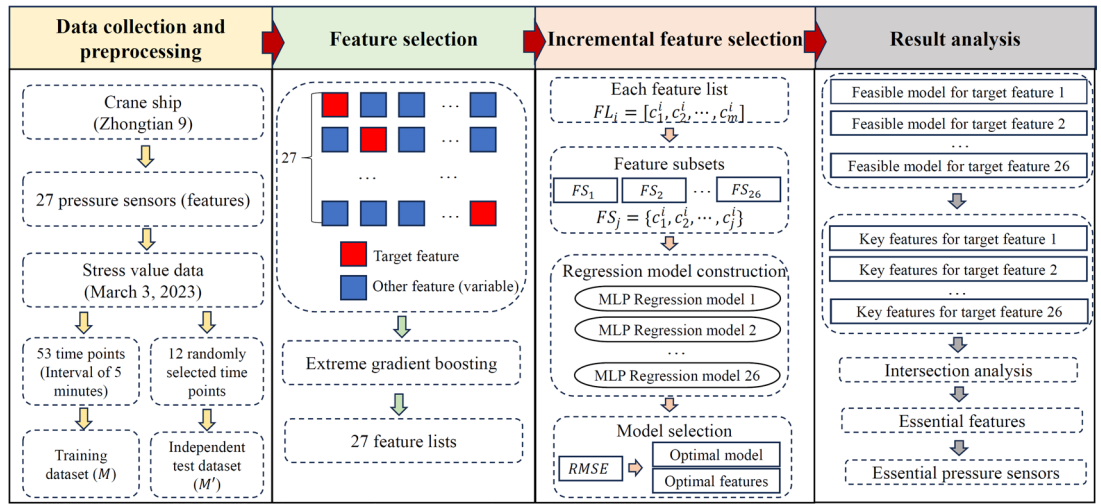
## Materials and methods

In this study, we designed a ML-based method for analyzing stress values on a crane ship. The objectives were twofold: (1) to construct a regression model for each pressure sensor to infer its stress values, and (2) to identify the essential pressure sensors of the ship SHM system. The overall workflow is illustrated in Fig. 1 and comprises four stages. In the first stage, stress data were collected from a crane ship, and preprocessing was performed to construct two well-defined datasets: a training dataset and an independent test dataset. In the second stage, each pressure sensor was treated in turn as a target feature, while the remaining sensors served as explanatory variables. For each target feature, an optimal regression model was built from these variables. XGBoost was adopted to evaluate the relationships between the target feature and the variables, yielding a ranked feature list that ordered variables by decreasing importance. In the third stage, each feature list was analyzed using IFS, with MLP regression models constructed and evaluated across all possible feature subsets. This process produced the optimal regression model and optimal feature subset for each target. Finally, in the fourth stage, the optimal models were further analyzed to extract feasible models for each target. Intersection analysis of the key features from these feasible models enabled the identification of essential features (pressure sensors) that are most critical for monitoring stress in the crane ship. Detailed descriptions of each step are provided below.

## Dataset and problem description

Stress data were retrieved from a 1600-ton crane ship, “Zhongtian 9” (Fig. 2). The vessel is equipped with 27 identical pressure sensors, deployed in four monitoring areas. Specifically, four sensors were installed beneath the main deck, while 23 were installed at the ship bottom—five at midship and nine each at the port and starboard sides. The distribution and identifiers of these sensors are summarized in Table 1.

Stress values were collected on March 3, 2023, yielding 65,535 data records that included monitoring time, sensor identifier, and stress value. These records were first grouped by sensor. Each sensor captured 2,427 or



**Fig. 1.** Flowchart to show the computational method. The method contains four stages. In the first stage, stress value data is collected from a crane ship, which is preprocessed to generate one training dataset and one independent test dataset. In the second stage, XGBoost is adopted to yield 27 feature lists. The incremental feature selection method is used in the third stage for analyzing the feature lists, yielding the optimal models and features. In the last stage, we further extract the feasible models for different target features and analyze the key features of these models to extract essential features and corresponding pressure sensors.



**Fig. 2.** Picture of the crane ship “Zhongtian 9”.

2,428 values at different time points. Because stress values recorded by a sensor within a short interval were highly similar, homogenization could occur if all values were retained. To mitigate this, a sampling interval of five minutes was adopted. Accordingly, each sensor contributed 53 stress values at different time points. The selected values from all 27 sensors were arranged into a  $53 \times 27$  matrix, denoted by  $M$ . Each row of  $M$  indicated the stress values of 27 sensors at one monitoring time, whereas each column corresponded to stress values detected by a single sensor across 53 time points.

In this formulation, each row of  $M$  was treated as a sample and each column as a feature. The system therefore consisted of 53 samples,  $s_1, s_2, \dots, s_{53}$ , each described by 27 features,  $F_1, F_2, \dots, F_{27}$ . The aim

Monitoring area	Number of sensors	Identifier
Main deck	4	MD1, MD2, MD3, MD4
Ship bottom (midship)	5	BM1, BM2, BM3, BM4, BM5
Ship bottom (port)	9	BP1, BP2, BP3, BP4, BP5, BP6, BP7, BP8, BP9
Ship bottom (starboard)	9	BS1, BS2, BS3, BS4, BS5, BS6, BS7, BS8, BS9

**Table 1.** Distribution of pressure sensors.

was to model the relationships among features by constructing a regression model for each feature using the others as explanatory variables. For each feature  $F_k$ , a regression model  $RM_k$  was defined by the following optimization problem:

$$\begin{aligned} & \text{Min } ||F_k - \widehat{F}_k|| \\ & \text{subject to } \widehat{F}_k = RM_k(F_1, F_2, \dots, F_{k-1}, F_{k+1}, \dots, F_{27}) \end{aligned} \quad (1)$$

where  $\widehat{F}_k$  denotes the predicted value of  $F_k$  and the norm  $||F_k - \widehat{F}_k||$  measures the discrepancy between observed and predicted values. These models were trained on the dataset  $M$ , which was designated as the training dataset.

To fully evaluate the regression models, an independent test dataset was constructed. Twelve monitoring time points not included in the training dataset were randomly selected from the original 65,535 records. The stress values of the 27 sensors at these time points formed the independent test dataset, organized into another matrix denoted as  $M'$ .

### Extreme gradient boosting

In this study, we adopted XGBoost to assess relationships among features (pressure sensors). XGBoost is a classic and robust feature selection algorithm<sup>18</sup> with broad applications in analyzing complex systems<sup>20–27</sup>.

XGBoost was originally designed for classification and regression tasks. It is based on a gradient boosting framework that enhances predictive performance by integrating multiple decision trees. Compared with traditional methods, XGBoost incorporates regularization, parallel computing, and automatic handling of missing values, thereby improving efficiency and reducing overfitting. It is particularly effective for large-scale, high-dimensional data. As a tree ensemble method, XGBoost builds additive functions in which the outputs of successive trees are aggregated until a predefined number of trees is reached. Suppose that there are  $K$  decision trees. The prediction of XGBoost is defined as:

$$\widehat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (2)$$

where  $f_k(x_i)$  is the output of the  $k$ -th decision tree for input  $x_i$ , which is the weight of one leaf in this tree. The loss function to optimize the ensemble is:

$$L = \sum_i l(y_i, \widehat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

where  $l(y_i, \widehat{y}_i)$  is a differentiable convex loss function between prediction  $\widehat{y}_i$  and target  $y_i$ . The regularization term is defined as  $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda ||w||^2$  ( $T$  stands for the number of leaves in the  $k$ -th tree and  $w$  represents the weights of the leaves). The first term of the loss function measures prediction error, whereas the second controls model complexity to avoid overfitting.

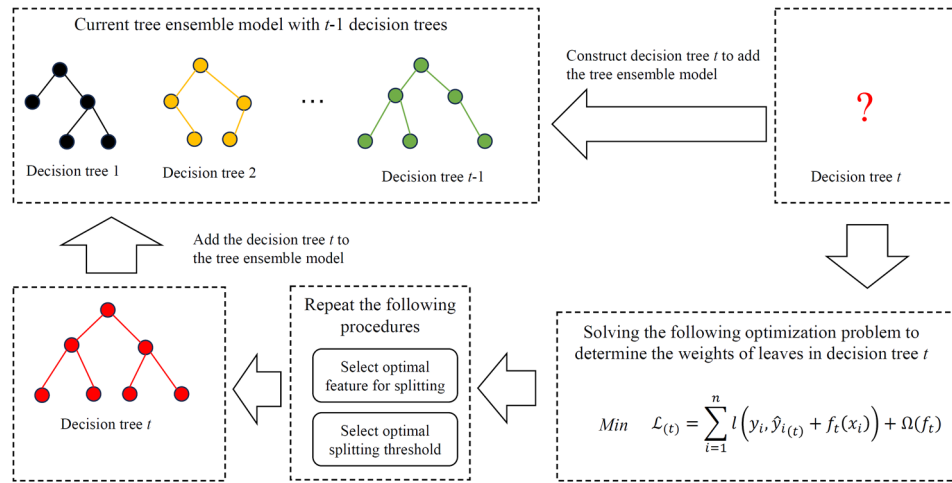
XGBoost is built by iteratively adding a decision tree to the current ensemble tree model. Suppose the current ensemble tree model contains  $t-1$  trees. To construct the  $t$ -th tree, XGBoost determines the weights of leaves of the new tree by solving the following optimization problem:

$$\text{Min } \mathcal{L}(t) = \sum_{i=1}^n l(y_i, \widehat{y}_{i(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (4)$$

where  $\widehat{y}_{i(t-1)}$  denotes the prediction of the ensemble before adding the  $t$ -th tree, and  $f_t$  is the output of the candidate new tree. The tree is built by repeatedly selecting the optimal splitting features and thresholds. Once constructed, the tree is added to the ensemble. This process repeats until the predefined number of trees is reached. Figure 3 illustrates this procedure.

Beyond its predictive capabilities, XGBoost can also be used for feature selection. Because XGBoost builds an ensemble of decision trees, the importance of a feature can be quantified by how frequently it is chosen to split nodes across all trees. Features that appear more frequently are considered more influential. Thus, XGBoost provides a ranked feature list based on feature importance.

In this study, XGBoost was employed for feature selection. Each of the 27 features was treated in turn as the target, with the remaining 26 provided as explanatory variables. This process produced a feature list for each target, in which the 26 variables were ranked by their relevance to the target feature. Since the 27 features correspond to stress values measured by 27 pressure sensors, this procedure enabled extraction of inter-sensor



**Fig. 3.** Flowchart to show the addition of a new decision tree to the current ensemble tree model in XGBoost. The weights of leaves in the new decision tree are determined by solving an optimization problem. Then, the new decision tree is constructed by repeatedly selecting optimal feature for splitting and optimal splitting threshold.

relationships, forming the basis for inferring stress values at one monitoring point from those at others. For a given feature  $F_i$ , the feature list generated by XGBoost was denoted as  $FL_i$ . The XGBoost package was obtained from <https://xgboost.readthedocs.io/en/stable/> and executed using default parameters.

**Incremental feature selection**

XGBoost provides a ranking of feature importance but does not specify which features should be selected for modeling or analysis. To address this, we employed IFS<sup>19</sup> to refine the feature lists. Suppose the feature list for target feature  $F_i$  is  $FL_i = [c_1^i, c_2^i, \dots, c_m^i]$ , where  $m$  is the number of explanatory features ( $m = 26$  in this study). IFS constructs a series of feature subsets, each containing the top features in the ranked list. The  $j$ -th feature subset is defined as  $FS_j = \{c_1^i, c_2^i, \dots, c_j^i\}$ . For each subset, the samples are represented by the selected features, and a regression model is built using these features to predict the target. Model performance is evaluated using a metric such as the root mean square error (RMSE). The model with the best performance is selected as the optimal model, and the features in this model are designated as optimal features. These features are considered to have particularly strong relationships with the target, as they enable the most accurate inference of its values.

**Multilayer perceptron**

Within the IFS procedure, a regression model must be constructed for each feature subset. In this study, we employed a MLP to build regression models, as relationships among stress values from different pressure sensors may be nonlinear, and MLPs can capture such nonlinearities.

An MLP is a feedforward neural network composed of an input layer, one or more hidden layers, and an output layer. Information is transmitted between nodes through weighted connections, with nonlinear transformations introduced by activation functions. For an input vector  $x^{(0)}$ , the transformation through the  $j$ -th hidden layer is:

$$x^{(j)} = \sigma(W^{(j)}x^{(j-1)} + b^{(j)}) \tag{5}$$

where  $x^{(j)}$  is the output of the  $j$ -th hidden layer,  $W^{(j)}$  and  $b^{(j)}$  are the trainable weight matrix and bias in the  $j$ -th hidden layer, and  $\sigma$  is an activation function. If there are  $l$  hidden layers, the output layer processes the  $x^{(l)}$  using the following equation:

$$\hat{y} = W'x^{(l)} + b' \tag{6}$$

where  $\hat{y}$  is the output of the MLP,  $W'$  and  $b'$  are the trainable weight matrix and bias in the output layer. To quantify regression performance, we adopted the widely used loss function for regression, RMSE loss, defined as:

$$\mathcal{L}_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{7}$$

where  $n$  is the number of samples,  $y_i$  and  $\hat{y}_i$  are the observed and predicted values of the  $i$ -th sample, respectively. The RMSE loss reflects the degree of deviation between observed and predicted values. In this study,  $y_i$  corresponds to the stress measured by a target pressure sensor, and  $\hat{y}_i$  to the predicted stress value. A smaller

RMSE indicates better performance. Using this loss can help the MLP to extract optimal trainable parameters through the back propagation, which were implemented by Adam optimizer<sup>28</sup>.

We implemented an MLP with three hidden layers containing 64, 32, and 16 neurons, respectively. These configurations were selected after testing several alternatives and optimizing through grid search. ReLU was used as the activation function in all hidden layers.

### Model performance

Multiple metrics were used to evaluate regression models. While RMSE (Eq. 7) served as the primary evaluation metric, we also employed mean absolute error (MAE) and the coefficient of determination  $R^2$  for a more comprehensive assessment:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

where  $n$ ,  $y_i$ , and  $\hat{y}_i$  are same as those in Eq. 7, and  $\bar{y}$  is the mean of the observed values. RMSE and MAE measure error magnitude, with smaller values indicating better performance.  $R^2$  reflects the proportion of variance explained by the model, with values closer to 1 indicating stronger predictive capability.

## Results and discussion

### Results of XGBoost

As illustrated in Fig. 1, XGBoost was first applied to analyze the relationships among features (pressure sensors). Using the training dataset organized in matrix  $M$ , each feature (one column in  $M$ ) was treated sequentially as the target, while the remaining features served as explanatory variables. XGBoost was then applied to produce a ranked feature list for each target. Because the dataset contained 27 features in  $M$ , 27 feature lists were obtained and are provided in Table S1. In this table, each column corresponds to a feature list for one target feature, identified by the pressure sensor label. For example, the first column (excluding the “Rank” header) represents the feature list generated when BM1 was the target. In this case, the stress values from BP7 showed the closest relationship with BM1, as BP7 occupied the highest rank. MD1 and BP6 followed as the second- and third-ranked related features, respectively. These feature lists served as the basis for subsequent analyses using the IFS method.

### Results of the IFS method

The IFS method was applied to each feature list. For every target, 26 MLP regression models were constructed using progressively larger subsets of top-ranked features. Their RMSE values are provided in Table S2. To facilitate interpretation, IFS curves were plotted for each target, with RMSE on the y-axis and the number of features used on the x-axis (Fig. 4).

For sensors located under the main deck, the IFS curves are shown in Fig. 4(A). The lowest RMSE values for MD1, MD2, MD3, and MD4 were 0.219, 0.358, 0.315, and 0.356, respectively. These results were achieved using 25, 19, 14, and 14 features. The corresponding feature subsets were selected as the optimal features, and the regression models built from them were designated as the optimal models. Table 2 lists the MAE and  $R^2$  values for these models, all of which exhibited  $MAE < 0.25$  and  $R^2$  values approaching 1, confirming their strong predictive performance.

For the five midship sensors located at the ship bottom, the IFS curves are shown in Fig. 4(B). The lowest RMSE values for BM1, BM2, BM3, BM4, and BM5 were 0.449, 0.257, 0.286, 0.245, and 0.439, respectively, obtained using 15, 25, 6, 22, and 23 features. These models likewise demonstrated low MAE and high  $R^2$  values (Table 2), further validating their robustness.

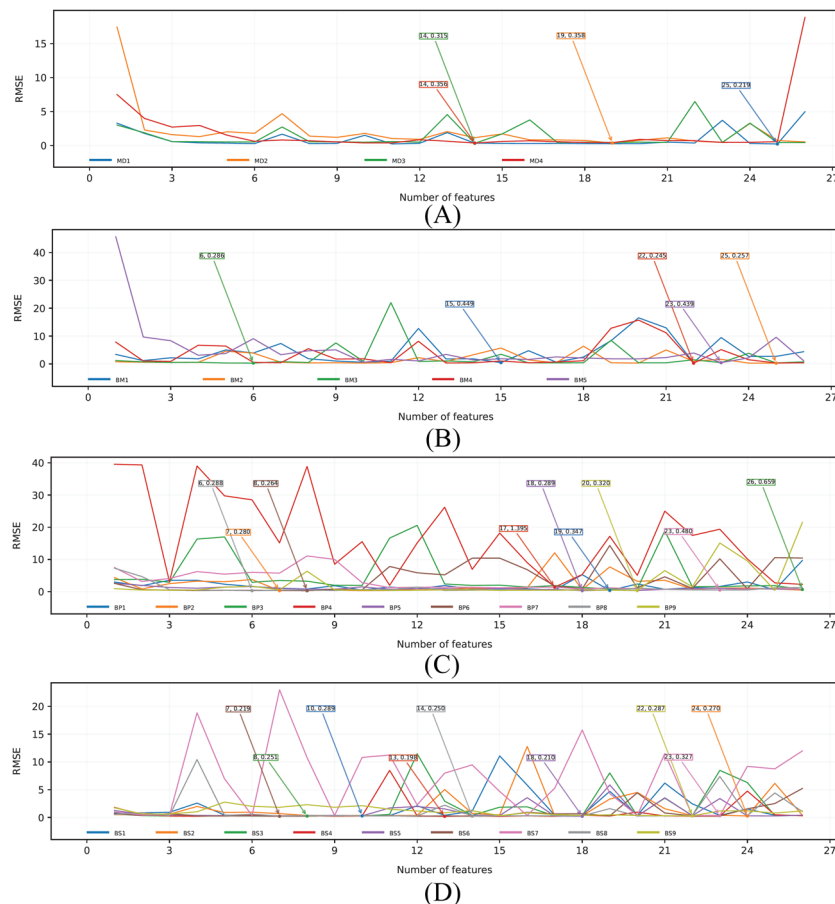
For the nine portside bottom sensors, the IFS curves are shown in Fig. 4(C). The optimal models required 19, 7, 26, 17, 18, 8, 23, 6, and 20 features, yielding RMSE values of 0.347, 0.280, 0.659, 1.395, 0.289, 0.264, 0.480, 0.288, and 0.320, respectively. Except for BP4, which showed an RMSE above 1.0, all models achieved RMSE values below 0.7, indicating reliable performance. MAE and  $R^2$  values (Table 2) further confirmed the effectiveness of these models.

Finally, for the nine starboard bottom sensors, the IFS curves are presented in Fig. 4(D). The optimal models used 10, 24, 8, 13, 18, 7, 23, 14, and 22 features, and yielded RMSE values of 0.289, 0.270, 0.251, 0.198, 0.210, 0.219, 0.327, 0.250, and 0.287, all below 0.4. This consistently demonstrates high performance, which was corroborated by MAE and  $R^2$  values in Table 2.

In total, 27 optimal MLP regression models were obtained. The mean RMSE, MAE, and  $R^2$  values across all models were 0.354, 0.239, and 0.993, respectively (Table 2), highlighting the strong overall fitting ability of the models. Each optimal model enables inference of stress values at a specific monitoring point based on stress values recorded by a subset of highly related sensors.

### Performance of the optimal models on the independent test dataset

As described in section “Dataset and problem description”, an independent test dataset was constructed and organized in matrix  $M'$ . This dataset was used to evaluate the generalization performance of the optimal models obtained in section “Results of the IFS method”. The results, expressed as RMSE, MAE, and  $R^2$ , are summarized in Table 3. Overall, RMSE values increased for all 27 pressure sensors. Four sensors (BP1, BP7, MD4, and



**Fig. 4.** IFS curves for 27 pressure sensors. **(A)** IFS curves for four pressure sensors under the main deck; **(B)** IFS curves for five pressure sensors at ship bottom (midship); **(C)** IFS curves for nine pressure sensors at ship bottom (port); **(D)** IFS curves for nine pressure sensors at ship bottom (starboard). The lowest RMSE and the corresponding number of features are marked on each curve.

BM5) showed increases greater than one, with BP1 exhibiting the largest increase (8.784). For the remaining 23 sensors, the increase was below one. The average RMSE was 1.107, representing a 0.753 increase compared with the training dataset. Similar trends were observed for MAE. On three sensors (BP1, BP7, and BM5), MAE increased by more than one, while on the other 24 sensors, the increase was below one. The average MAE was 0.718, 0.479 higher than in the training dataset. For  $R^2$ , values on eight sensors remained unchanged relative to the training dataset. The largest decrease was observed on BS5, where  $R^2$  dropped from 0.980 to 0.559. The average  $R^2$  was 0.958, which was 0.035 lower than that in the training dataset.

These results indicate that the optimal regression models exhibited reduced performance on the independent test dataset. However, the decrease was moderate, and the models still maintained acceptable accuracy, demonstrating their generalization capability.

### Ablation tests

In the proposed computational method, XGBoost and MLP played central roles: XGBoost was used to assess feature importance, and MLP was used to construct regression models. To confirm that these choices were reasonable, ablation tests were conducted.

First, we replaced XGBoost with two widely used feature selection algorithms, LightGBM<sup>29</sup> and CatBoost<sup>30</sup>. Feature lists generated by these algorithms were processed using the same pipeline as our method, producing 27 optimal regression models for each case. Their average performance on the training dataset is summarized in Table 4, together with the results of our original method for comparison. Using LightGBM yielded average RMSE, MAE, and  $R^2$  values of 0.407, 0.262, and 0.993, respectively, while CatBoost produced 0.406, 0.261, and 0.994. Compared with these, our method achieved an RMSE advantage of approximately 0.05 and an MAE advantage of about 0.02, while  $R^2$  values were similar across all three methods. On the independent test dataset, the differences were more pronounced. Models based on LightGBM showed substantially poorer generalization, with average RMSE and MAE values of 5.993 and 2.092, respectively, and an average  $R^2$  of 0.921. CatBoost performed better, with results closer to those of our method, but was still inferior: the average RMSE and MAE were approximately 0.6 and 0.2 higher, respectively. These findings indicate that using XGBoost provided better overall results than LightGBM or CatBoost, supporting its selection for this study.

Target feature (Pressure sensor)	Number of variables (Related pressure sensors)	RMSE	MAE	R <sup>2</sup>
MD1	25	0.219	0.149	0.998
MD2	19	0.358	0.232	0.999
MD3	14	0.315	0.223	0.998
MD4	14	0.356	0.242	0.999
BM1	15	0.449	0.247	0.940
BM2	25	0.257	0.166	0.984
BM3	6	0.286	0.228	0.987
BM4	22	0.245	0.150	0.996
BM5	23	0.439	0.310	0.999
BP1	19	0.347	0.246	0.999
BP2	7	0.280	0.196	0.992
BP3	26	0.659	0.399	0.996
BP4	17	1.395	0.787	0.999
BP5	18	0.289	0.187	0.999
BP6	8	0.264	0.205	0.999
BP7	23	0.480	0.355	0.999
BP8	6	0.288	0.186	0.999
BP9	20	0.320	0.216	0.999
BS1	10	0.289	0.212	0.974
BS2	24	0.270	0.200	0.994
BS3	8	0.251	0.176	0.983
BS4	13	0.198	0.138	0.997
BS5	18	0.210	0.162	0.980
BS6	7	0.219	0.180	0.999
BS7	23	0.327	0.230	0.999
BS8	14	0.250	0.203	0.998
BS9	22	0.287	0.218	0.999
Mean		0.354	0.239	0.993

**Table 2.** Performance of the optimal regression model for each target feature (pressure sensor).

Second, we replaced MLP with linear regression and least absolute shrinkage and selection operator (Lasso)<sup>31</sup> as the regression algorithm. The corresponding computational methods were evaluated on both training and independent datasets, with results reported in Table 5. Compared with MLP, both linear regression and Lasso performed significantly worse. On average, RMSE increased by more than six and MAE by more than three. R<sup>2</sup> was about 0.05 lower on the training dataset and about 0.1 lower on the independent dataset. These results demonstrate that MLP is more suitable for this task than linear regression or Lasso. The latter two algorithms can only capture linear relationships, whereas the stress values recorded by pressure sensors at different monitoring points are likely governed by nonlinear dependencies. Because MLP incorporates activation functions, it can model nonlinear relationships and therefore achieves superior performance.

### Intersection analysis

As described in section "Results of the IFS method", 27 optimal MLP regression models were obtained, with their performance and numbers of selected features listed in Table 2. Some of these models required nearly all available features. For example, the optimal model for BM2 used the top 25 features out of 26, making it difficult to isolate the most influential features. To address this, further selection was carried out for the 20 optimal models that used more than ten features. By examining the IFS results in Table S2, we constructed 20 feasible MLP regression models that used fewer features than the corresponding optimal models, at the cost of only slightly higher RMSE values. The RMSE values and the number of selected features for these feasible models are listed in Table 6. For example, while the optimal model for BM2 required 25 features, the feasible model relied on only 8 features, a reduction of 17. The RMSE increased only modestly, from 0.257 to 0.349 ( $\Delta = 0.092$ ). This finding indicates that the 8 selected features were substantially more informative than the remaining 17 for predicting BM2 stress values. Similar observations were made for other feasible models. For the seven models that initially required ten or fewer features, no changes were made; they were retained as feasible models for consistency.

In total, 27 feasible models were obtained. Each model was capable of inferring the stress value at a specific monitoring point from a subset of related pressure sensors. At the system level, however, some sensors contributed disproportionately, being repeatedly selected across many feasible models. Such sensors should be considered essential for the SHM system, as they serve as key predictors of stresses at multiple monitoring points. To identify them, we counted the frequency with which each of the 27 pressure sensors appeared in feasible models (Fig. 5). The analysis revealed that BP8, BM2, BP3, BS6, and BM4 occupied the top five ranks,

Pressure sensor	RMSE	MAE	R <sup>2</sup>
MD1	0.543 (↑0.324)	0.469 (↑0.320)	0.987 (↓0.011)
MD2	1.015 (↑0.657)	0.721 (↑0.489)	0.997 (↓0.002)
MD3	0.535 (↑0.220)	0.416 (↑0.193)	0.997 (↓0.001)
MD4	1.516 (↑1.160)	1.138 (↑0.896)	0.992 (↓0.007)
BM1	0.612 (↑0.163)	0.430 (↑0.183)	0.862 (↓0.078)
BM2	0.507 (↑0.250)	0.368 (↑0.202)	0.906 (↓0.078)
BM3	0.669 (↑0.383)	0.478 (↑0.250)	0.907 (↓0.080)
BM4	0.512 (↑0.267)	0.422 (↑0.272)	0.983 (↓0.013)
BM5	1.597 (↑1.158)	1.335 (↑1.025)	0.999 (→)
BP1	9.131 (↑8.784)	3.240 (↑2.994)	0.997 (↓0.002)
BP2	0.504 (↑0.224)	0.412 (↑0.216)	0.958 (↓0.034)
BP3	1.174 (↑0.515)	0.875 (↑0.476)	0.988 (↓0.008)
BP4	2.192 (↑0.797)	1.598 (↑0.811)	0.999 (→)
BP5	0.523 (↑0.234)	0.469 (↑0.282)	0.999 (→)
BP6	0.608 (↑0.344)	0.365 (↑0.160)	0.949 (↓0.050)
BP7	2.330 (↑1.850)	1.781 (↑1.426)	0.999 (→)
BP8	0.700 (↑0.412)	0.529 (↑0.343)	0.999 (→)
BP9	1.195 (↑0.875)	0.902 (↑0.686)	0.999 (→)
BS1	0.565 (↑0.276)	0.547 (↑0.335)	0.891 (↓0.083)
BS2	0.343 (↑0.073)	0.286 (↑0.086)	0.988 (↓0.006)
BS3	0.496 (↑0.245)	0.413 (↑0.237)	0.937 (↓0.046)
BS4	0.308 (↑0.110)	0.287 (↑0.149)	0.993 (↓0.004)
BS5	0.564 (↑0.354)	0.398 (↑0.236)	0.559 (↓0.421)
BS6	0.356 (↑0.137)	0.292 (↑0.112)	0.997 (↓0.002)
BS7	0.407 (↑0.080)	0.337 (↑0.107)	0.999 (→)
BS8	0.435 (↑0.185)	0.387 (↑0.184)	0.994 (↓0.004)
BS9	0.562 (↑0.275)	0.496 (↑0.278)	0.999 (→)
Mean	1.107 (↑0.753)	0.718 (↑0.479)	0.958 (↓0.035)

**Table 3.** Performance of the optimal regression models on the independent test dataset<sup>§</sup>. §: symbols “↑”, “↓”, “→” indicate the increase, decrease, and equal compared with the performance on the training dataset.

Feature selection algorithm	Training dataset			Independent test dataset		
	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
XGBoost	0.354	0.239	0.993	1.107	0.718	0.958
LightGBM	0.407	0.262	0.993	5.993	2.092	0.921
CatBoost	0.406	0.261	0.994	1.691	0.955	0.968

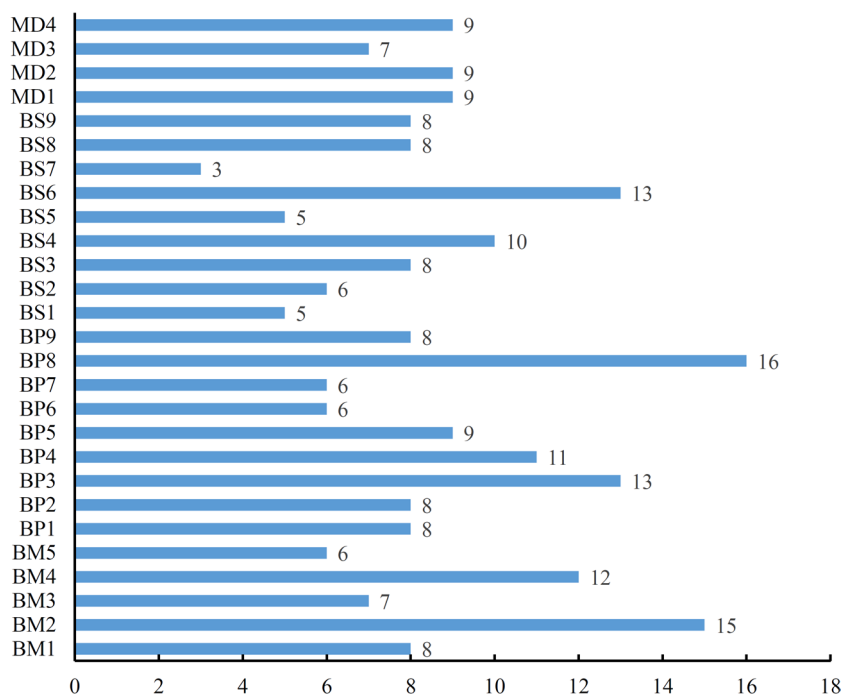
**Table 4.** Average performance of the optimal models on the training and independent test datasets with different feature selection algorithms.

Regression algorithm	Training dataset			Independent test dataset		
	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
MLP	0.354	0.239	0.993	1.107	0.718	0.958
Linear regression	6.603	3.713	0.946	8.159	4.831	0.850
Lasso	6.860	3.761	0.941	8.274	4.888	0.851

**Table 5.** Average performance of the optimal models on the training and independent test datasets with different regression algorithms.

Target feature (Pressure sensor)	Number of other features (Pressure sensors)	RMSE	Optimal RMSE	Difference
MD1	6	0.278	0.219	0.059
MD2	12	0.908	0.358	0.550
MD3	5	0.515	0.315	0.200
MD4	6	0.616	0.356	0.260
BM1	10	0.577	0.449	0.128
BM2	8	0.349	0.257	0.092
BM4	7	0.409	0.245	0.164
BM5	10	0.583	0.439	0.144
BP1	10	0.468	0.347	0.121
BP3	19	0.785	0.659	0.126
BP4	11	1.943	1.395	0.548
BP5	8	0.754	0.289	0.465
BP7	18	0.783	0.480	0.303
BP9	3	0.444	0.320	0.124
BS2	9	0.327	0.270	0.057
BS4	4	0.213	0.198	0.015
BS5	7	0.228	0.210	0.018
BS7	6	0.367	0.327	0.040
BS8	7	0.259	0.250	0.009
BS9	15	0.327	0.287	0.040

**Table 6.** Performance of the feasible regression models for some target features (pressure sensors).



**Fig. 5.** Bar chart to show the frequencies of 27 pressures for constructing feasible MLP regression models.

each being used to infer stresses at more than 11 monitoring points. BM2 and BM4, located on the bottom centerline (midship), likely owe their importance to their proximity to many other sensors (e.g., those at the port bottom). BP8, situated on the port side near L/4, is located in a region of high local stress. BP3 is positioned at the termination of a transverse framing member, a well-known site for stress concentration. BS6, located at frame 60, coincides with the on-bottom operating condition, where the bottom structure rests on seabed sediments and stress concentration occurs.

The locations of these five critical sensors (BP8, BM2, BP3, BS6, and BM4) are illustrated in Fig. 6. Their importance can be explained in terms of classic ship structural mechanics. For a longitudinally framed, double-bottom ship, the normal bending stress in hull plating can generally be expressed as:

$$\sigma = \pm \frac{6M}{t^2} \quad (10)$$

where  $M$  is the bending moment and  $t$  is the plate thickness.

According to this relation, larger bending moments  $M$  produces a larger stress  $\sigma$ . For rectangular thin plates used in hull structures, regardless of aspect ratio, the bending moment and stress attains its maximum at the midpoint of the long edge. From the standpoint of global longitudinal strength, the transverse section at the midpoint along the long-edge direction may therefore experience the peak stress. Therefore, on plating subjected to significant external loads, such as the upper deck and bottom shell, relatively large stresses tend to occur at the mid-length on the upper and lower surfaces of structure in way of the centerline, as well as at locations such as longitudinal primary members and structural discontinuities.

In addition, when the hull is in still water or waves, local hydrostatic and hydrodynamic pressures introduce transverse loads that may cause local deformation or even failure. Under China Classification Society (CCS) Rules, high local stresses are typically expected near one-quarter of the ship length from the bow and stern ( $L/4$  from amidship).

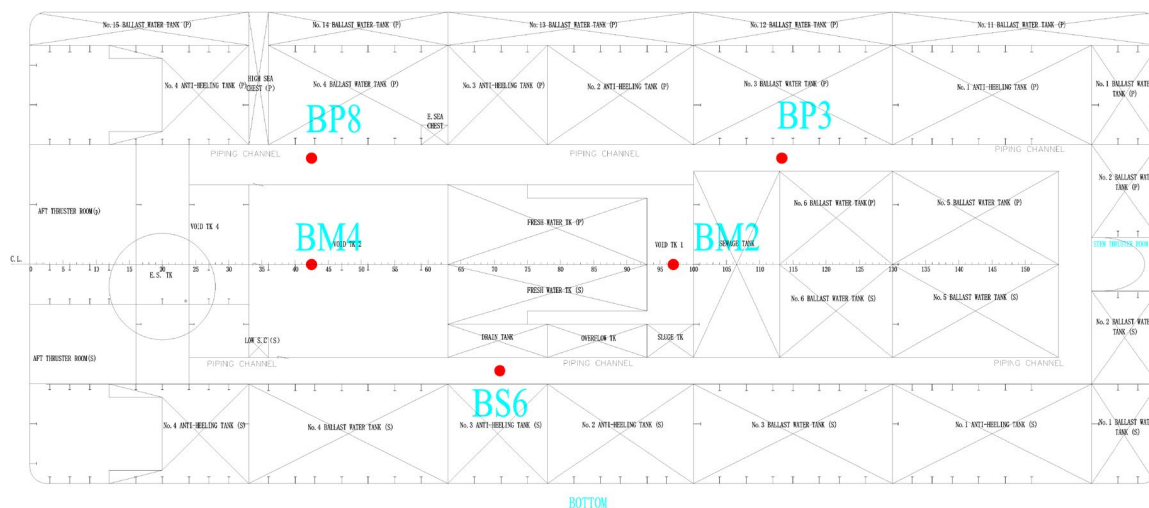
BM2 and BM4, located on the centerline at mid-length ( $L/2$ ) and  $L/4$ , respectively, are therefore prone to elevated stresses and were reinforced by the higher density of nearby sensors. BP8, positioned on the port side near  $L/4$ , is also in a region of substantial local stress. BP3 is located at the termination of a transverse framing member, a typical stress concentration site. BS6, located at frame 60, aligns with the stress concentration region during on-bottom operations, where the bottom structure rests on seabed sediments. Finally, due to the longitudinal symmetry of the hull, stresses at port or starboard locations can often be inferred from symmetric counterparts (e.g., BP8, BP3, and BS6).

Taken together, the importance of BP8, BM2, BP3, BS6, and BM4 is consistent with structural mechanics and operational considerations. These findings confirm both the effectiveness and the interpretability of the proposed method, demonstrating its ability to uncover critical information underlying the monitoring system.

In this study, stress values at four monitoring areas were analyzed: main deck, ship bottom (midship), ship bottom (port), and ship bottom (starboard). For each area, we examined the features used to construct the feasible models. Upset graphs were plotted for each monitoring area (Fig. 7), showing that certain features were consistently related to multiple target pressure sensors. Features associated with more target sensors were evidently more important. To quantify this, we also calculated the frequency of each feature within each area (Fig. 8).

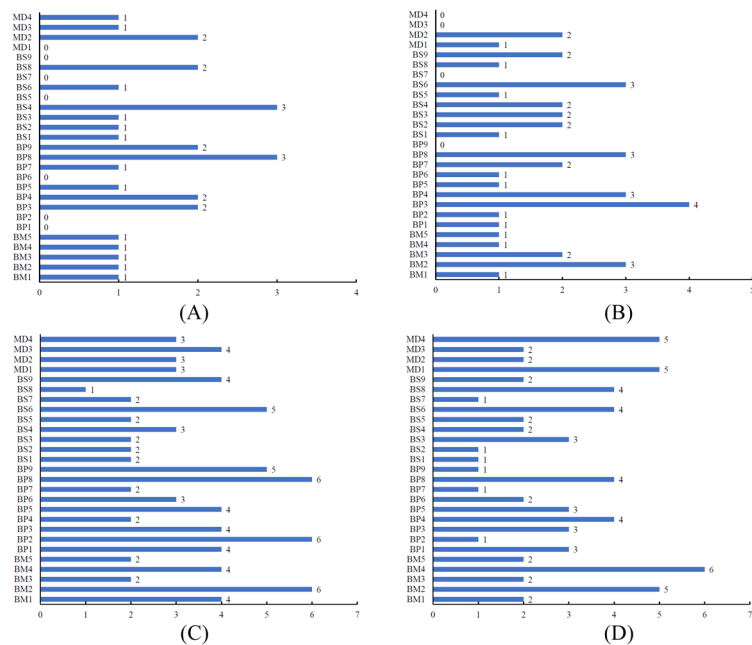
From Fig. 7(A), two features, BS4 and BS8, were highly related to three target sensors beneath the main deck, as illustrated in Fig. 8(A). Because of structural constraints, relatively few sensors could be installed on the main deck, making stress recovery in this region dependent on sensors elsewhere. Structurally, the transverse bulkhead near BP8 and BS4 extends through multiple deck levels and ties the main deck to the bottom shell, thereby providing a load path that transmits stresses beneath the deck.

From Fig. 7(B), one feature (BP3, illustrated in Fig. 8(B)) was highly related to four target sensors at the midship bottom, and four features (BS6, BP8, BP4, and BM2, illustrated in Fig. 8(B)) were related to three target sensors each. Stresses in the bottom shell at midship are strongly coupled to those in the port and starboard side shells, as these regions lie in the same transverse section and are in close proximity. BP3 and BP4 are located near mid-length, where global longitudinal stresses peak, while BP8 and BS6 are positioned near frame 60 and



**Fig. 6.** Monitoring points of five key pressure sensors.





**Fig. 8.** Bar chart to show the frequencies of 27 pressures for constructing feasible MLP regression models at four monitoring areas. **(A)** Bar chart for the monitoring area under the main deck; **(B)** Bar chart for the monitoring area at ship bottom (midship); **(C)** Bar chart for the monitoring area at ship bottom (port); **(D)** Bar chart for the monitoring area at ship bottom (starboard).

## Conclusions

This study developed a ML-based computational method to analyze stress distribution in a ship. Stress data collected from a crane ship were used to evaluate the utility of the approach. For each pressure sensor, an optimal MLP regression model was constructed to infer its stress values from those recorded by other related sensors. All models demonstrated strong fitting performance. In addition, the proposed method identified several key sensors that play central roles within the monitoring system. These findings provide new insights into the mechanisms underlying stress monitoring and offer an interpretable framework for sensor prioritization. Overall, the proposed method presents a novel approach for analyzing stress data and has the potential to enhance the performance and reliability of ship SHM systems.

## Data availability

The data is available upon request. Please contact Dr. Ji Zeng (zengji@shmtu.edu.cn).

Received: 17 August 2025; Accepted: 6 October 2025

Published online: 12 November 2025

## References

- Jeon, M. et al. Estimation of local ice load by analyzing shear strain data from the IBRV aron's 2016 Arctic voyage. *Int. J. Naval Archit. Ocean. Eng.* **10** (3), 421–425 (2018).
- Lee, T. K. et al. Field measurement of local ice pressures on the ARAON in the Beaufort sea. *Int. J. Naval Archit. Ocean. Eng.* **6** (4), 788–799 (2014).
- Foss, G. & Haugse, E. *Using modal test results to develop strain to displacement transformations*. in *Proceedings of the 13th international modal analysis conference*. (1995).
- Ko, W. L., Richards, W. L. & Tran, V. T. *Displacement theories for in-flight deformed shape predictions of aerospace structures*. (2007).
- Li, M. et al. Structural health monitoring of an offshore wind turbine tower using iFEM methodology. *Ocean Eng.* **204**, 107291 (2020).
- Shkarayev, S., Krashantisa, R. & Tessler, A. *An inverse interpolation method utilizing in-flight strain measurements for determining loads and structural response of aerospace vehicles*. (2004).
- Kefal, A. & Oterkus, E. Displacement and stress monitoring of a chemical tanker based on inverse finite element method. *Ocean Eng.* **112**, 33–46 (2016).
- Oboe, D., Sbarufatti, C. & Giglio, M. Physics-based strain pre-extrapolation technique for inverse finite element method. *Mech. Syst. Signal Process.* **177**, 109167 (2022).
- Kefal, A. et al. Three dimensional shape and stress monitoring of bulk carriers based on iFEM methodology. *Ocean Eng.* **147**, 256–267 (2018).
- Wei, P. et al. Real-time digital twin of ship structure deformation field based on the inverse finite element method. *J. Mar. Sci. Eng.* **12** (2), 257 (2024).
- Li, S., Coraddu, A. & Brennan, F. A framework for optimal sensor placement to support structural health monitoring. *J. Mar. Sci. Eng.* **10** (12), 1819 (2022).
- Hassani, S. & Dackermann, U. A systematic review of optimization algorithms for structural health monitoring and optimal sensor placement. *Sensors* **23** (6), 3293 (2023).

13. Karvelis, P. et al. Deep machine learning for structural health monitoring on ship hulls using acoustic emission method. *Ships Offshore Struct.* **16** (4), 440–448 (2021).
14. Sun, C. et al. A data-driven approach to full-field stress reconstruction of ship hull structure using deep learning. *Eng. Appl. Artif. Intell.* **133**, 108414 (2024).
15. Sun, C. & Chen, Z. End-to-end deep learning method to reconstruct full-field stress distribution for ship hull structure with stress concentrations. *Ocean Eng.* **313**, 119431 (2024).
16. Zhang, T. et al. Reconstructing the global stress of marine structures based on artificial-intelligence-generated content. *Appl. Sci.* **13** (14), 8196 (2023).
17. Lu, W. et al. Structural response Estimation method based on particle swarm optimisation/support vector machine and response correlation characteristics. *Measurement* **160**, 107810 (2020).
18. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System*. in *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery. (2016).
19. Liu, H. A. & Setiono, R. Incremental feature selection. *Appl. Intell.* **9** (3), 217–230 (1998).
20. Bao, Y. et al. Recognizing SARS-CoV-2 infection of nasopharyngeal tissue at the single-cell level by machine learning method. *Mol. Immunol.* **177**, 44–61 (2025).
21. Liao, H. et al. Machine learning analysis of CD4+ T cell gene expression in diverse diseases: insights from Cancer, Metabolic, Respiratory, and digestive disorders. *Cancer Genet.* **290–291**, 56–60 (2025).
22. Ma, Q. et al. Machine Learning-driven discovery of essential binding preference in Anti-CRISPR proteins. *Proteom. – Clin. Appl.* **19** (4), e70013 (2025).
23. Wu, W. et al. Machine learning based method for analyzing vibration and noise in large cruise ships. *Plos One.* **19** (7), e0307835 (2024).
24. Niazkari, M. et al. *Applications of XGBoost in Water Resources Engineering: A Systematic Literature Review (Dec 2018–May 2023)* Vol. 174, p. 105971 (Environmental Modelling & Software, 2024).
25. Chen, F. et al. Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. *Sci. Rep.* **15** (1), 25729 (2025).
26. Yao, S. et al. An interpretable XGBoost-based approach for Arctic navigation risk assessment. *Risk Anal.* **44** (2), 459–476 (2024).
27. Huang, M., Zhao, H. & Chen, Y. Research on SAR image quality evaluation method based on improved Harris Hawk optimization algorithm and XGBoost. *Sci. Rep.* **14** (1), 28364 (2024).
28. Kingma, D. P. & Ba, J. *Adam: A method for stochastic optimization*, in *3rd International Conference on Learning Representations*. Louisiana, USA. (2019).
29. Ke, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural. Inf. Process. Syst.* **30**, 3146–3154 (2017).
30. Prokhorenkova, L. et al. *CatBoost: unbiased boosting with categorical features*. *Adv. Neural. Inf. Process. Syst.* **31** (2018).
31. Ransam, J. & Cook, J. LASSO regression. *J. Br. Surg.* **105** (10), 1348–1348 (2018).

### Author contributions

J.Z. designed the research; B.J. and L.X. conducted the experiments; B.J., L.C. and Y.Z. analyzed the results. All authors have read and approved the manuscript.

### Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23356-6>.

**Correspondence** and requests for materials should be addressed to J.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025