MARITIME TRANSPORTATION RESEARCH AND EDUCATION CENTER
TIER 1 UNIVERSITY TRANSPORTATION CENTER
U.S. DEPARTMENT OF TRANSPORTATION

In-Situ Monitoring and Assessment of Post Barge-Bridge Collision Damage for Minimizing Traffic Delay and Detour

08/01/2014-06/30/2016

Principal Investigator(s) –Wei ZHeng
e-mail: wei.zheng@jsums.edu
phone number: 608-332-4298
Jackson State University
Other Author Name – Feng Qian
Final Report Date: 07/31/2016

FINAL RESEARCH REPORT
Prepared for:
Maritime Transportation Research and Education Center

University of Arkansas
4190 Bell Engineering Center
Fayetteville, AR 72701
479-575-6021

ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Transportation under Grant Award Number DTRT13-G-UTC50. The work was conducted through the Maritime Transportation Research and Education Center at the University of Arkansas.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Abstract

This report presents a novel framework for promptly assessing the probability of barge-bridge collision damage of piers based on probabilistic-based classification through machine learning. The main idea of the presented framework is to divide the potential damage region of a pier into multiple discrete sub-regions, and define a classifier for each sub-region based on a probabilistic logistic regression. Several different classification models are considered for each classifier due to the uncertainties associated with the logistic regression model, and trained through Bayesian inference by using simulation data before damage events occur. Bayesian model selection is adopted to select the best classification model and its corresponding effective features extracted from measurements for enhancing the prediction accuracy. The best trained classification models then are used to expeditiously predict the probabilities of damage occurring in all sub-regions. The presented framework can be implemented in a recursive manner by dividing a structure into several hierarchical levels of different divisions of sub-regions, and can be extended to identify other types of structural damage. The effectiveness and applicability of the presented framework are demonstrated through the numerical simulation of identifying barge-bridge collision damage locations of one pier of a prototype bridge. Finally, limitations and future research directions are also discussed.

Keywords Barge-Bridge Collision, Bayesian Inference, Probability, Feature Selection, Logistic Classification Model, Transitional Markov Chain Monte Carlo Method

1 Project description

Piers of bridges across major navigation waterways frequently suffer from barge collisions, resulting in the closure of both bridges and waterways to traffic for assessing the potential damage. Promptly and accurately locating potential collision damage locations provides the basis for further quantifying the damage extent and facilitating the informative decision-making on the operation of highway and navigation channels, thus can significantly reduce the economic losses resulted from unnecessary closure.

Barge-Bridge collision damage may be parameterized by unknown parameters associated with structure physical properties, e.g., the reduced stiffness of damaged regions. Thus, numerical models representing potential damage structures can be established and then damage-related parameters can be determined through the model calibration by matching the model predictions with the field vibration measurements collected from the damaged structures. This model-based damage identification algorithm has been extensively studied in the past [1-3]. The damage parameters are usually obtained through optimization by minimizing the error between the field measurements and predictions of numerical models [4-6]. However, optimization algorithms always give a single best solution to structural damage identification problem, and cannot address the uncertainties associated with field measurements and numerical models.

To quantify uncertainties associated with the structural model and measurement errors, Bayesian inference framework has been proposed [7] and adopted for the vibration-based damage identification approach, including its application for assessing the potential damage of piers due to barge collisions [8]. In such application, uncertain parameters are represented by their probability distribution functions (PDFs) and updated by using the available measurements with possible numerical structural models. In implementation of the existing probabilistic-based damage identification, however, the structural model

needs to be executed for each sample of unknown structural parameters to determine the likelihoods of those parameters for depicting their probabilistic distribution. This demands significant computational time of hours or days for complex structures [9-10]. Since this intensive computation has to be conducted with measured responses from structures after potential damage occurs, it is almost impossible to apply the probabilistic-based damage identification approach to determine the existence of the potential damage for prompt decision-making after barge collision incidents.

To promptly identify the potential damage and its location in bridge piers after barge-bridge collisions, this report proposes and explores a novel probabilistic-based classification framework for locating collision damage based on machine learning through Bayesian inference. The novelty of the proposed framework lies in its ability to identify probabilities of damage occurrence in specific locations through the probabilistic classification and shift intensive computational burden from the assessment process phase after barge-bridge collision events to the prior preprocessing phase in advance of any incident occurrence, resulting in a significant reduction of computing time in comparing with traditional probabilistic-based inference for damage identification. To authors' best knowledge, this shift of the computational burden has not been explored for probabilistic-based assessment on potential bridge pier damage due to the barge-bridge collision and may initiate a new perspective for probabilistic-based approach to identifying other types of damage by using vibration measurements.

Within the proposed framework, the potentially damaged region of a structure is artificially divided into several discrete sub-regions, and a two-class classifier for each sub-region then is defined based on the probabilistic-based logistic regression model. Different classification models are considered for each classifier due to the uncertainties associated with the logistic regression model, and trained through Bayesian inference with effective feature vectors along with their corresponding location labels that describe whether collision damage occurs in the sub-region or not. The feature vectors capturing

effective information of damage occurrence in different sub-regions are extracted from simulated dynamic responses of structural systems by using Principle Component Analysis (PCA) [11]. By using Bayesian model selection, different type of effective features and their corresponding different types of classification models for the classifier of each sub-region can be explored and evaluated in terms of their model evidences. The most effective feature vector corresponding to the best model can be identified. The optimal threshold for distinguishing the damaged and undamaged scenarios from classification labels is determined by maximizing the rate of correct classification of all probabilistic samples of prediction labels obtained from Bayesian inference for all events in the training dataset.

In application phase, the best trained classification models can be used to expeditiously classify each sub-region in terms of the probability of damage based on the input feature vectors obtained from field measurements after barge-bridge collisions. Uncertainties associated with the classifiers can be included in the obtained probability of damage. The proposed framework and its applicability are illustrated and examined through numerical simulation of damage identification of barge-bridge collision based on the finite element model of a prototype bridge. The limitations of the presented framework and feature research directions are also discussed.

2 . Methodological Approach

2.1 Feature Extraction and Label Defining

Let $\{\hat{\omega}_r , \hat{\varphi}_{rj} , r=1, \cdots, m, \ j=1, \cdots, N_s\}$ denotes the modal propriety data simulated from a undamaged structure, where $\hat{\omega}_r$ is the r-th modal frequency, $\hat{\varphi}_{rj}$ is the r-th mode shapes at different measurement locations, and m and N_s are the numbers of observed modes and measured degrees of freedoms (DOFs) respectively. Let $\{\omega_r , \varphi_{rj} , r=1, \cdots, m, \ j=1, \cdots, N_s\}$ be the modal frequencies and mode shape components at N_s measured DOFs obtained from a collision damage scenario simulated through the

calibrated numerical structural model of a bridge structure. Thus, the vector of the *r*-th modal property change $\boldsymbol{\varepsilon}_r = [\varepsilon_r^{\omega} \quad \varepsilon_{rj}^{\varphi}]$ before and after collision damage can be obtained separately for modal frequencies and mode shapes as followings:

$$\varepsilon_r^{\omega} = \omega_r - \hat{\omega}_r \tag{1}$$

$$\varepsilon_{rj}^{\varphi} = \varphi_{rj} - \hat{\varphi}_{rj} \tag{2}$$

where ε_r^{ω} and $\varepsilon_{rj}^{\varphi}$ are the changes of modal frequency and mode shape components of the *r*-th mode respectively. Assembling those modal change vectors yields the matrix of modal property changes below:

$$\boldsymbol{\varepsilon}_{m\times(Ns+1)}=[\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2,\cdots\boldsymbol{\varepsilon}_m]^T \tag{3}$$

This matrix can be interpreted as changes of structural characters caused by the occurrence of damage.

Although the above matrix $\boldsymbol{\varepsilon}$ may be directly used in the proposed framework to identify collision damage locations, manipulating all the measurements such a matrix within an identification algorithm may be too redundant and time-consuming. Thus, the PCA [11] is applied to extract representative features of matrix $\boldsymbol{\varepsilon}$ in Eq. (3) to capture effective information on the possible damages occurring in different sub-regions of a structure. The effectiveness of the PCA method for feature extraction from measurements has been verified by Magalhaes et al.[12] and Huang et al. [13]. In the presented framework, only the first principal component of the modal change matrix of a damage accident may be considered by projecting all the changes of frequencies and mode shapes into the first principal component through PCA [11]. In other word, the N_s+1 column vectors in matrix $\boldsymbol{\varepsilon}$ are compressed into a unidimensional feature vector $X=[x_1, x_2,\ldots, x_m]^T$.

For a specific sub-region of a bridge pier, two classes of its condition are defined: one for damage locating in this sub-region and the other for damage occurring out of this sub-region. Let $\mathbf{D}=\{(\mathbf{X},\mathbf{Y})\}=\{(X_1, Y_1), \ldots, (X_i, Y_i), \ldots, (X_N, Y_N); i=1, 2\ldots N\}$ denotes the events dataset, where $X_i=[x_{i1}, x_{i2},$

…, $x_{im}]^T$ and $Y_i \in \{0,1\}$ are the feature vector and label for the *i*-th event, and Y_i=1 claims that damage locates in the sub-region and Y_i=0 means that damage is out of this sub-region.

2.2 Logistic Classification Model with Linear Discriminant Function

The presented framework is actually to establish a binary classifier for each specified sub-region and implemented through a set of classifiers for all divided sub-regions. As an illustration, Fig.1 presents the sub-region divisions of the potential damage area of a bridge pier and their corresponding classifiers, which are defined based on the logistic regression model of a linear discriminant function. Generally, each sub-region has its own binary classifier as shown in Fig.1 based on the following logistic regression model [14]:

$$\hat{Y}_i = \varphi(f(X_i|\boldsymbol{\alpha})) = \frac{e^{f(X_i|\boldsymbol{\alpha})}}{1+e^{f(X_i|\boldsymbol{\alpha})}} \tag{4}$$

where $\hat{Y}_i$ is the output prediction label of the logic regression model for the input feature vector X_i of observed modal change matrix in Eq. (3), and $\boldsymbol{\alpha}$ is unknown model parameter vector to be determined for the function $f(X_i|\boldsymbol{\alpha})$ in the logistic regression model as defined in the following.

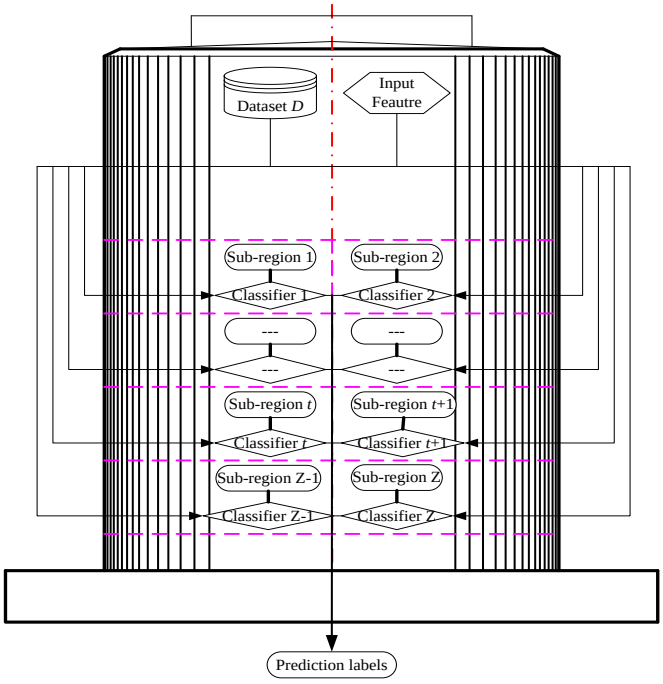


Fig.1 Sub-region division of a bridge pier

In the presented framework, the function $f(X_i|\boldsymbol{\alpha})$ in Eq. (4) is a linear discriminant function of feature vector X_i, and defined by unknown coefficient vector $\boldsymbol{\alpha}=[\alpha_0, \alpha_1, \ldots, \alpha_m]^T$ in the following way:

$$f(X_i|\boldsymbol{\alpha})=\sum_{k=1}^{m}\alpha_k x_{ik}+\alpha_0 =[\alpha_0, \alpha_1, \cdots, \alpha_m]\bullet[1\ x_{i1}, x_{i2}, \ldots, x_{im}]^T=\boldsymbol{\alpha}[1\ X_i^T]^T \quad (5)$$

Theoretically, different dimensions and types of feature vectors can be extracted from the modal change matrix as defined in Eq. (3) for a given damage event by including different number of modal modes. There is uncertainty about how many modal modes should be selected for effective classification. This uncertainty about selecting feature vectors leads to the uncertainty of the logistic regression model in Eq. (4). Thus, different classification models for the classifier of a specific sub-region are considered in terms of different dimensions of the linear discrimination functions, leading to different dimensions of model parameter vector α, which correspond to the different selection of extracted feature vectors.

With given training data D obtained from known damage events, the unknown coefficients α in Eq. (5) for the classifier can be determined through matching the prediction labels of the regression model (4) with the actual labels of known events in the training dataset, i.e., minimizing the square sum of errors $\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2$.

With the trained logistic classification model for each sub-region, the label of a new barge-bridge collision event can be obtained to categorize this new event into one of two classes: damage occurring in this sub-region, i.e., the first class, if $f(X_i|\boldsymbol{\alpha}) > 0$, or no damage locating in this sub-region, i.e., the second class, if $f(X_i|\boldsymbol{\alpha}) < 0$. In practical application for the input feature vector X_i of a specific damage event, the larger the discriminant function $f(X_i|\boldsymbol{\alpha})$ is, the more possible the event is categorized into the first class; and the smaller $f(X_i|\boldsymbol{\alpha})$ is the more likely the event is categorized into the second class. Thus, it can be reasonable to determine that $f(X_i|\boldsymbol{\alpha})=0$, resulting in $Y_i=0.5$, is a theoretical division between the above two classes.

2.3 Training Logistic Model through Bayesian Probabilistic Inference

By fitting the prediction labels obtained from the classifier to its true labels of all known events in training dataset, the logistic model can be trained by determining its model parameter vector $\boldsymbol{\alpha}$ through several existing optimization approaches, such as least squares, Fisher's linear discriminant, and perceptron algorithm [15]. However, those approaches to training the logistic model are deterministic methods that can only determine one optimal parameter vector $\boldsymbol{\alpha}$ and cannot quantify uncertainties associated with classification model and measurement errors.

As an alternative method, Bayesian probabilistic inference can be used to estimate unknown parameter vector $\boldsymbol{\alpha}$ with prior knowledge and quantify the aforementioned uncertainties. In Bayesian probabilistic inference, uncertainty parameters are described through their probability density function

(PDF) [16]. The presented framework applies Bayesian probabilistic inference to update the posterior PDF $p(\boldsymbol{\alpha}|\mathbf{D}, M_j)$ of unknown parameter vector $\boldsymbol{\alpha}$ for the classification model class M_j with given training data **D**. To consider uncertainties associated with the classifier, a set of candidate classification model classes $\mathbf{M}=\{M_j, j=1, 2, \cdots, N_M\}$ can be considered in terms of different selections of feature vectors as described previously in the section 2.2.

To apply Bayesian theorem, the likelihood function has to be established based on the assumption that the error between the prediction label $\hat{Y}_i$ obtained from the logistic model and the true label Y_i of known event in training data **D** is a zero-mean stationary normally-distributed stochastic process with a shared standard deviation σ. This assumption is demonstrated by the maximum information entropy principle [16-17] to be able to incorporate the maximum uncertainties into the unknown parameters. Thus, the likelihood function can be formulated as the PDF of multiple independent and normally distributed errors:

$$p(\mathbf{D}\,|\,\boldsymbol{\alpha}, M_j) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2\right\} \tag{6}$$

It should be noted that the standard deviation σ is also unknown and needed to be inferred through Bayesian inference. Thus, the uncertain parameter vector in Bayesian inference becomes $\mathbf{x}= \{\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_m, \sigma\}^{\mathrm{T}} = \{\boldsymbol{\alpha}^{\mathrm{T}}, \sigma\}^{\mathrm{T}}$.

According to Bayesian theorem, the posterior PDF of uncertainty model parameters **x** can be expressed as:

$$p(\mathbf{x}\,|\,\mathbf{D}, M_j) = \frac{p(\mathbf{D}\,|\,\mathbf{x}, M_j)\cdot p(\mathbf{x}\,|\,M_j)}{p(\mathbf{D}\,|\,M_j)} \tag{7}$$

where $p(\mathbf{D}|\mathbf{x}, M_j)$ is the likelihood function as defined in Eq. (6); $p(\mathbf{x}|M_j)$ is the prior PDF of uncertain parameter vector $\mathbf{x}$ and usually is assumed based on prior knowledge or judgment; and the denominator $p(\mathbf{D}|M_j)$ is referred to as model evidence of model class M_j , which reflects the quality of model class M_j, and is actually a normalization constant that ensures the integral of posterior PDF over parameter space equal to unity. Thus, it can be expressed as:

$$\begin{aligned} p(\mathbf{D}|M_j) &= \int p(\mathbf{D}|\mathbf{x}, M_j) \cdot p(\mathbf{x}|M_j) d\mathbf{x} \\ &= \int \left(\frac{1}{(\sqrt{2\pi}\sigma)^N} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2\right\}\right) \cdot p(\mathbf{x}|M_j) d\mathbf{x} \end{aligned} \tag{8}$$

Typically, directly integrating the multidimensional exponential function in Eq. (8) and obtaining the posterior PDF of uncertain parameter $\mathbf{x}$ in Eq. (7) is impractical. Thus, stochastic simulation methods have been proposed as an alternative to numerically estimate the posterior PDF in Eq. (7) and model evidence in Eq. (8) by generating statistic samples drawn from the posterior PDF of uncertain parameter $\mathbf{x}$. In the presented framework, the Transitional Markov Chain Monte Carlo (TMCMC) method proposed by Ching and Chen [18] is adopted for conducting stochastic simulation. Interested readers are referred to [18-19] for more detailed information and implementation of TMCMC sampling algorithm.

2.4 Feature Selection Based on Bayesian Model Evaluation

Bayesian theorem also includes the model class section, which can be also applied at model class level to evaluate different model classes. In the presented framework, different classification models can be defined from different linear discrimination functions in the logistic regression model as defined in Eq. (4) and Eq. (5). Those different classification models actually correspond to the selection of different features extracted from the modal change matrix as defined in Eq. (3). Thus, Bayesian model selection can provide a new perspective for the selection of effective feature vectors for defining and training classifiers in the presented framework.

The posterior model probability for classification model class M_j among the classification model classes $\mathbf{M}=\{M_j, j=1, 2, \cdots, N_M\}$ can be obtained with the prior model probability $p(M_j|\mathbf{M})$ and given training data **D** as follows:

$$p(M_j|\mathbf{D},M)=\frac{p(\mathbf{D}\,|\,M_j)\times P(M_j\,|\,\mathbf{M})}{\sum_{j=1}^{N_M} p(\mathbf{D}\,|\,M_j)\times P(M_j\,|\,\mathbf{M})} \tag{9}$$

Usually, identical prior probabilities can be assumed for all candidate model classes, i.e., $p(M_j|\mathbf{M})=1/N_M$, if there is no any prior knowledge. The most plausible model can be usually identified in terms of its model evidence $p(\mathbf{D}|\,M_j)$, which can be obtained as a by-product from the TMCMC [18-19].

2.5 Probability of Damage Predicted by Classification Model

Through implementation of Bayesian inference with stochastic simulation through the TMCMC [18-19], total K samples $\{ x_k, k=1, 2, \cdots, K\}$ of the classification model parameter **x** can be obtained to depict the posterior probability distribution of the unknown parameter **x**. With those samples of the classification model parameter **x**, total K prediction labels can be obtained from the logistic regression model for the given input feature vector X of a new barge-bridge collision event or damage event. Those prediction labels describe the probability distribution of the predicted labels of the considered event represented by feature vector X.

For the classification model based on the logistic regression, each individual prediction label $\hat{Y}^k$ usually can be interpreted as the probability of damage occurring in the considered sub-region. With the probability distribution of the prediction label Y for a given event obtained from Bayesian inference, the probability that the prediction label is equal to Y can be estimated based on the PDF of prediction labels as $P(Y \mid \mathbf{x}, \mathbf{D}, M_j)\bullet dY$. Thus, the probability that damage occurring in the considered sub-region can be obtained by $Y\bullet P(Y \mid \mathbf{x}, \mathbf{D}, M_j)\bullet dY$. According to the total probability law [20] and Monte Carlo integration

[21], the prediction probability that the damage occurring in the considered sub-region can be estimated based on all the prediction labels of the classification model for the given event:

$$\hat{Y} = \int Y P(Y \mid \mathbf{x}, \mathbf{D}, M_j) dY = \frac{1}{K} \sum_{k=1}^{K} \hat{Y}^k \tag{10}$$

where $\hat{Y}^k$ is the prediction label of the regression model (4) with the discriminant function defined by the k-th sample x_k of the classification model parameter $\mathbf{x}$.

From Eq. (10), it can be clearly noted that the predicted probability of the damage occurring in the considered sub-region is actually the mean value of all the prediction labels of the classification model for the given event. Compared to existing deterministic methods, however, the advantage of the presented probabilistic-based framework can not only provide a best prediction label, but also can characterize the probability distribution of the prediction labels to quantify uncertainties or confidence level of the prediction results, which can be illustrated in the probabilistic distribution of the prediction labels as shown in the next section and the subsequent numerical simulation.

2.6 Optimal Threshold of Prediction Labels and Modified Probability of Potential Damage

Theoretically, based on the classifier defined in the section 2.2, the prediction label is expected to be 1 for damage occurring in the considered sub-region or 0 for non-damage locating in this sub-region. However, the actual prediction label provided by a specific trained classification model of the considered sub-region for a given event often ranges from zero to one due to the impact of the inaccurate model and other uncertainties associated with the imperfect devices, environmental noises, and computational numerical errors. Those uncertainties may lead to misclassification of the trained classifier, which includes two cases: (1) the event of the damage actually occurring in the current sub-region may be erroneously classified as no damage event with a prediction label less than 0.5 and approaching zero, and

(2) the event of damage actually occurring out of the current sub-region may be erroneously categorized as damage event with a prediction label larger than 0.5 and close to one.

Thus, even though the theoretical threshold of the prediction labels, which divides between the damaged and undamaged scenarios for a specific sub-region can be taken as 0.5 as described in the section 2.2, this theoretical threshold may cause some misclassification of sub-regions for all events in the training dataset due to the aforementioned uncertainties. To mitigate such a misclassification problem, an optimal threshold of the prediction labels is proposed in the presented framework to minimize the rate of misclassification or maximize the rate of correct classification for all samples of prediction labels of all known events in the training dataset as followings.

The rate of correction classification of a probabilistic-based classifier based on a given thresholds δ can be defined below. Fig.2 illustrates the histogram of prediction labels of two events A and B that correspond to damage occurring out of and in a given sub-region respectively. For the event A of damage occurring out of the considered sub-region, the prediction label is expected to be less than the defined threshold δ. Thus, any samples of the prediction label less than δ are regarded as the correct classification, while those exceeding δ are considered as the misclassification. This leads to the formulation of the rate of the correct classification r_A of the classification model for the event A as:

$$r_A = \frac{\int_{Y \le \delta} P(Y_A \mid \mathbf{x}, \mathbf{D}, M_j) dY_A}{\int P(Y_A \mid \mathbf{x}, \mathbf{D}, M_j) dY_A} = \frac{n_A}{K} \tag{11}$$

where $P(Y_A \mid \mathbf{x}, \mathbf{D}, M_j)$ is the PDF of the prediction labels of the event A, n_A is the number of the samples of the prediction label less than the defined threshold δ, and K is the number of the total samples.

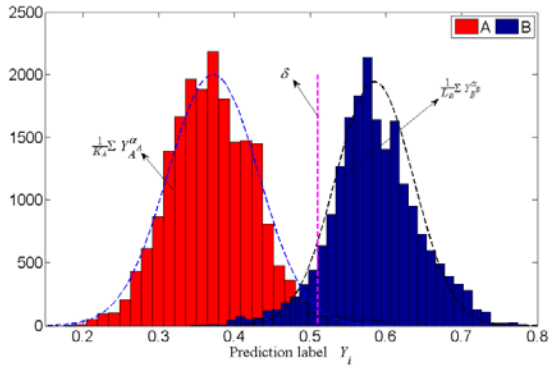


Fig.2 Schematic diagram of threshold and prediction labels

For event B of damage occurring in the considered region, the prediction label is expected to be greater than δ. Thus, any samples of the prediction label larger than the defined threshold δ are regarded as the correct classification, while those less than the defined threshold δ are considered as the misclassification. Thus, the rate of the correct classification r_B for the event B of the damage within the sub-region can be calculated as followings:

$$r_B = \frac{\int_{Y \geq \delta} P(Y_B \mid \mathbf{x}, \mathbf{D}, M_j) dY_B}{\int P(Y_B \mid \mathbf{x}, \mathbf{D}, M_j) dY_B} = \frac{n_B}{K} \tag{12}$$

where $P(Y_B \mid \mathbf{x}, \mathbf{D}, M_j)$ is the PDF of the prediction labels of the event B, n_B denotes the number of prediction labels exceeding δ, and K denotes the number of total prediction labels.

The impact of the defined thresholds on the rate of correct classification of a probabilistic-based classifier can be illustrated by the following example. Fig.3 depicts the prediction labels of seven exemplary events. The first four events in Fig.3 are those with damage occurring within the considered sub-region and the other three are those with damage out of the sub-region. Three defined thresholds δ_1=0.4, δ_2=0.5, and δ_3=0.6 are also plotted in Fig.3. If the defined threshold is selected as δ_1=0.4 , all the first four events can be correctly classified because all their prediction labels exceed δ_1. However, the last event is misclassified as the event of damage occurring in the considered sub-region for its prediction

labels greater than δ_1. As δ_1 further decreases, this situation is gradually exacerbated because more prediction labels of events of damage actually out of the sub-region will exceed the decreased threshold, leading to more misclassification. On the contrary, if the defined threshold is chosen as δ_3=0.6, all the last three events can be correctly classified as the event of damage occurring outside of the sub-region. Nevertheless, the third event of damage occurring in the sub-region is obviously misclassified because its prediction labels are less than the threshold. As δ_3 increases, this misclassification can further deteriorate since more events of damages truly occurring in the sub-region are misclassified as those outside the sub-region because their prediction labels are less than the increased threshold. When the defined threshold is selected as δ_2=0.5, all the seven exemplary events can be correctly classified, leading to the maximum rate of classification. Thus, the defined threshold of δ_2=0.5 can be regarded as the optimal threshold for the classification model based on the available known events.

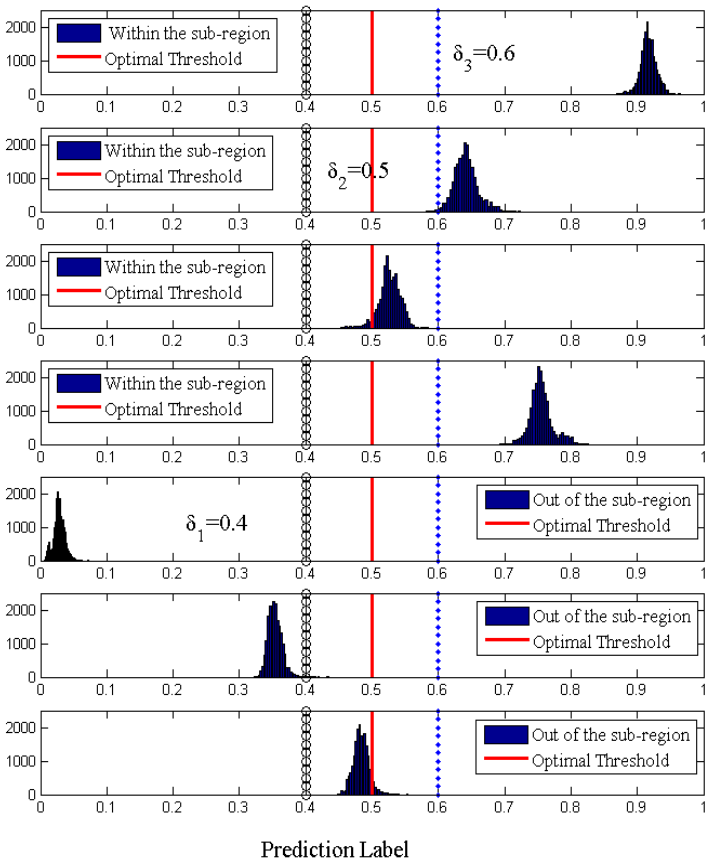


Fig.3 Illustration of optimal threshold definition

If the training dataset **D** contains *w* events of damage occurring out of the considered sub-region and g events of damage occurring in the sub-region, the rate of correct classification of the classifier model for the considered sub-region for all events in the training dataset based on a defined threshold δ can be formulated as:

$$R = \frac{\sum_{\lambda=1}^{w} \int_{Y \le \delta} P(Y_\lambda \mid \mathbf{x}, \mathbf{D}, M_j) dY_\lambda + \sum_{\tau=1}^{g} \int_{Y \ge \delta} P(Y_\tau \mid \mathbf{x}, \mathbf{D}, M_j) dY_\tau}{\sum_{i=1}^{N} \int P(Y_i \mid \mathbf{x}, \mathbf{D}, M_j) dY_i} = \frac{\sum_{\lambda=1}^{w} n_\lambda + \sum_{\tau=1}^{g} n_\tau}{NK} \tag{13}$$

where n_λ is the number of prediction labels of the λ-th event of damage occurring out of the sub-region that are less than δ, n_τ is the number of prediction labels of the τ-th event of damage occurring in the sub-region that are greater than δ, K is the number of total samples of prediction labels of each event, and N is the number of the total number of considered events.

It is worth noting that the rate of correct classification R in Eq. (13) completely depends on the selection of the defined threshold δ. By maximizing the rate of correct classification R for a specific sub-region for all events in training dataset **D,** an optimal threshold δ can be determined. It is reasonable to assume that the optimal threshold obtained in such a way can result in the largest rate of the correct classification for the considered sub-region for any new barge-bridge collision or other damage.

With the aforementioned rate of correct classification of a classifier, the probability of no damage locating in the considered sub-region determined by a classifier for a newly observed event can be considered to be composed of two aspects: one is the percentage of its prediction labels less than the optimal threshold as defined in Eq. (11), which can be interpreted as the probability of damage occurrence in the considered region for an observed event if the classifier is correct ; and the other is the maximum

rate of correct classification R of the classifier as defined in Eq. (13), which can be interpreted as the probability of correctness of the classifier based on the historical or past known events. Thus, the probability of no damage occurring in this considered sub-region can be modified as the following:

$$P(Y=0) = R \times \int_{Y \le \delta} YP(Y \mid \mathbf{x}, D, M_j)dY \times 100\% = R \times \frac{1}{K}\sum_{\alpha=1}^{L} Y_\alpha \times 100\% \quad (14)$$

where Y_t is the t-th prediction label less than the optimal threshold δ; L is the number of the labels that are less than δ among the total K prediction labels. The probability of on damage is considered here for modification due to inaccuracy of the classifier in terms of the rate of correct classification as determined in the above. This modification can lead to the reduced probability of no damage on the conservative side. Therefore, the modified probability of damage occurring in the considered sub-region for a newly observed event can be obtained as

$$P(Y=1) = 1 - P(Y=0) \quad (15)$$

where *P(Y=1)* is the probability of damage occurring in the considered sub-region, and P(Y=0) is probability of no damage as determined in Eq. (14). The above modification can be interpreted as that the probability of damage can be increased in considering the inaccuracy of the classifier in terms of the rate of correct classification of the classifier.

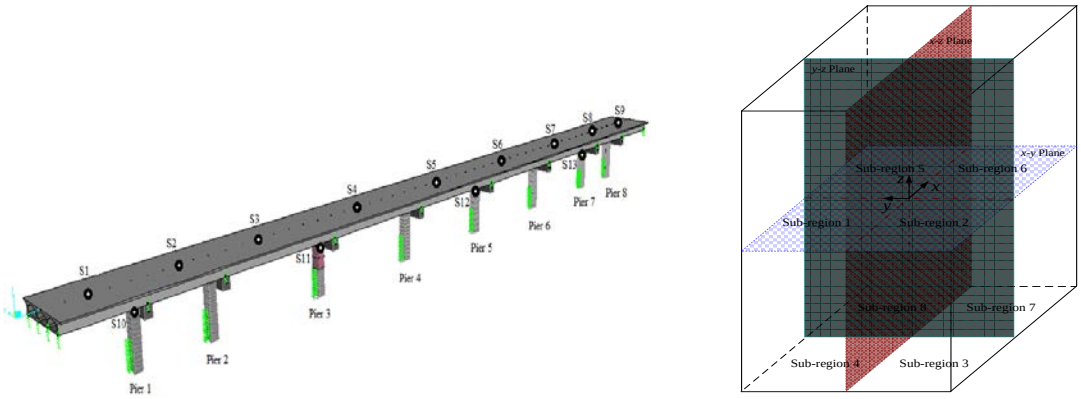
3. Results/Findings

3.1 Simulation of Damage Identification of Barge-Bridge Collision

3.1.1 Finite Element Model of the Prototype Bridge

To illustrate the proposed framework and verify its applicability, training data and test data of different damage scenarios of barge-bridge collision were simulated from the finite element (FE) model of a prototype bridge. The prototype bridge is 155m (510ft) in length with nine spans and eight piers. The FE model of the selected bridge is established in SAP 2000/Bridge 14 [22] under the linear elastic assumption as illustrate in Fig.4 (a). The cross sections of bridge decks, girders, and piers are modeled as

the same as their real properties of the prototype bridge. For simplicity, an equivalent pile is used to replace the group of piles under each bridge pier. The section properties of the equivalent piles were determined based on the group equivalent pile model proposed by Mokwa and Duncan [23]. The interaction between soil and pile foundation is modeled by a series of discrete springs distributed along the equivalent pile (see Fig.4 (a)) based on the selected soil's *p-y* relation. The potential damage regions of bridge piers that are mostly probable to be collided by barges are modeled using solid elements provided in SAP 2000/Bridge 14 [22].



(a) (b)

Fig.4 (a) FE model of the prototype bridge (b) The division of the potential damage area

In this illustration example, the third pier near the middle span of the prototype bridge, namely Pier3 located in the main water channel as shown in Fig.4 (a), is considered vulnerable to barge-bridge collision. Therefore, the potential damage region is assumed to be the middle portion of the Pier3 and was further divided into eight different sub-regions by three planes of *x-y*, *x-z*, and *y-z* as shown in Fig.4 (b). Thus, total 8 two-class classifiers ranging from Classifier 1 to Classifier 8 are required to be established or trained to locate the potential damage in one of the eight sub-regions once vibration measurements are obtained after barge-bridge collision occurs in this pier.

All simulated modal property data of the bridge are assumed to be the modal properties obtained by processing vibration measurements captured after barge-bridge collision incidents from the distributed sensors at the thirteen selected observation points S1-S13 as shown in Fig.4 (a). The sensors S1-S9 were assumed to be placed at the middle points of the nine spans for capturing the dynamic response along *z* direction, and S10-S13 were assumed to be installed at the junctions of piers and girders for collecting the dynamic response in *x* direction. In this simulation, the modal analysis was performed on the structural model of the damaged and undamaged bridge to generate those modal property data. The change of the bridge modal properties was obtained by subtracting the modal measurement of undamaged bridge prior to the damage from the modal responses of the damaged bridge.

3.1.2 Training Data and Evaluation Data

For each sub-region, ten collision accidents with different degrees of damage were simulated by assigning random reduction factors in the range of (0 1) to the elastic modulus of the sub-region respectively, meanwhile keeping the elastic modulus of other sub-regions unchanged. Thus, for each sub-region, ten modal response matrixes composed of natural frequencies and modal shapes can be obtained from 13 measured DOFs (or N_s=13) corresponding to the distributed sensors in different damage extent. By subtracting the measured modal data $\{\hat{\omega}_r, \hat{\varphi}_{rj}, r=1, \cdots, m, \ j=1, \cdots, N_s\}$ of the undamaged structure from those modal response matrixes of ten collision (or damage) events, ten modal change matrixes can be obtained based on Eq. (3). From those matrixes, ten feature vectors can thereby be extracted to represent those events of collision damage occurring in each considered sub-region respectively. The first five feature vectors of the simulated collision damages in each sub-region are selected to constitute training dataset **D**, resulting in 40 collision events in the training dataset for each classifier of eight sub-regions, while the last five feature vectors in each sub-region make up the test (or evaluation) dataset for each classifier, resulting in the other 40 collision events in the test dataset.

To consider the impact of field environment on modal measurements, three level noises of 0%, 5% and 10% are added to the simulated modal responses. Thus, the contaminated modal responses can be expressed as $\overline{\omega}_r = \omega_r(1+\eta \cdot rand)$ and $\overline{\varphi}_{rj} = \varphi_{rj}(1+\eta \cdot rand)$, where $\overline{\omega}_r$ and $\overline{\varphi}_{rj}$ are the noisy natural frequency and modal shape, η=noise level of 0%, 5% or 10% and *rand* is a random number within [-1 1].

3.2 Implementation of Presented Framework to Establish Classifiers

3.2.1 Extracting Features from Modal Changes and Establishing Classification Model Classes

For a specific collision or damage event, five different sets of modal properties were considered by selecting the first 5, 7, 10, 12, 15 mode of frequencies and modal shapes respectively. By compressing the modal change matrixes of those modal properties into a unidimensional vector through PCA, five dimensionally different feature vectors can be derived for each event. Thus, five different classification models can be established for the classifier of each sub-region.

As an example, Table 1 lists the five possible classification models for Classifier1 corresponding Sub-region1 in Fig.4 (b). The dimensions of the discriminant functions and unknown parameters are also tabulated in Table 1. It is necessary to state here that the possible classification models of other seven classifiers (Classifier 2-8) were also established in the same way.

Table 1 Model Classes of Classifier1

Model class	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$
Considered numbers of modal modes	5	7	10	12	15
Dimension of discriminant function	6	8	11	13	16
Unknown parameters	α_0~α_5, σ	α_0~α_7, σ	α_0~α_{10}, σ	α_0~α_{12}, σ	α_0~α_{15}, σ

3.2.2 Training Classification Models through Bayesian Inference with TMCMC Simulation

Bayesian inference with the TMCMC sampling [18-19] was adopted to train classification models for determining the unknown parameters of these models. The prior distributions of unknown model

parameters were proximately determined based on engineering judgment for implementing Bayesian inference. To obtain more accurate estimation of the prior distribution of unknown parameters, the Linear Discriminant Analysis (LDA) proposed by Fisher (1936) [24] for two-class problems was first applied by using training dataset to predetermine unknown model parameters or coefficients of the discriminant function in Eq. (5).

The second column of Table 2 lists those model parameters or coefficients of discriminant function, i.e., $\hat{\boldsymbol{\alpha}}$, predetermined by using the LDA method with noiseless modal data for classification model M_5 of Classifier 1. Then, those identified parameters $\hat{\boldsymbol{\alpha}}$ were taken as the mean of the prior distribution of those unknown parameters $\boldsymbol{\alpha}$ for Bayesian inference. In the first attempt in this simulation, the prior distributions of the unknown parameters in the TMCMC simulation were drawn from the uniform distribution over the range of [-3 $\hat{\boldsymbol{\alpha}}$, 5 $\hat{\boldsymbol{\alpha}}$]. Besides, the prior distribution of the shared error deviation σ in Eq. (6) was taken as a uniform distribution over the range of [0 2]. 10000 samples of potential parameter vector **x** for the classification model class M_5 of Classifier 1 were generated and updated to represent the prior or posterior PDF of **x**.

Table 2 Model Parameters of Model M_5 Identified by LDA and Bayesian Inference

Parameters	0% noise		5% noise		10% noise	
	LDA	**TMCMC**	**LDA**	**TMCMC**	**LDA**	**TMCMC**
α_0	-0.797	-6.986	-0.321	-1.561	-0.0789	-0.380
α_1	13.745	-0.133	0.179	0.503	-0.003	-0.006
α_2	0.055	0.229	0.100	0.345	0.025	0.010
α_3	3.533	32.496	-0.162	-0.790	0.051	-0.084
α_4	-16.966	-30.720	-0.037	-0.078	-0.015	-0.041
α_5	0.245	2.203	0.006	0.008	-0.008	-0.031
α_6	-0.066	-0.369	-0.019	0.013	0.019	0.013
α_7	0.012	0.079	-0.057	-0.179	-0.000	-0.000
α_8	-0.075	-0.066	0.006	0.027	-0.016	-0.070
α_9	-0.012	-0.019	0.067	0.322	0.030	-0.056
α_{10}	0.010	-0.008	0.116	0.552	-0.010	-0.046
α_{11}	-0.031	-0.210	0.027	0.104	0.071	0.349
α_{12}	-0.063	-0.122	0.024	-0.029	-0.059	-0.281
α_{13}	-0.378	-3.117	-0.006	-0.025	-0.034	-0.163
α_{14}	-0.097	-0.281	0.046	-0.100	-0.012	-0.042
α_{15}	-0.068	-0.273	0.028	0.132	0.005	-0.012
σ	-	1.685	-	1.998	-	1.986

Note: "-" means no deviation available. Results from TMCMC is the mean values of derived samples of each model parameters.

However, the histograms of the posterior PDF of parameters α_3 and α_5 exhibit that their samples heavily concentrate in the boundary of the sample space, as illustrated in Fig.5, indicating that the ranges of the prior parameters were not properly provided. Thus, the prior samples of all unknown parameters were eventually taken as uniform distribution from the expended range of [-8$\hat{\boldsymbol{\alpha}}$, 10$\hat{\boldsymbol{\alpha}}$] for each model class in this simulation.

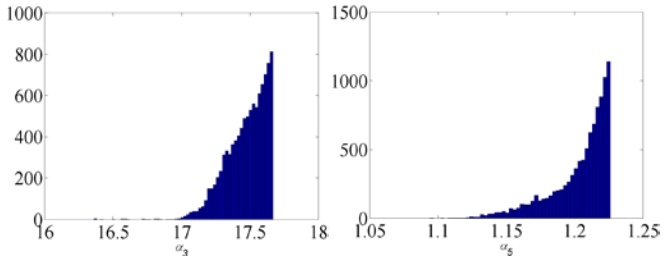


Fig.5 Histograms of the posterior PDF of unknown parameters α_3 and α_5

As an illustration, the results from the implementation of both LDA and TMCMC for the classification model M_5 of Classifier 1 with the training dataset are presented. In such implementation, only five feature vectors corresponded to the events of the simulated collision damage in Sub-region1, while the other thirty-five feature vectors were associated with the events of the simulated damage in other sub-regions. Thus, in the training dataset, only the first five labels were defined to be one and the other labels were assigned to be zero, leading to the two classes among 40 events.

Fig.6 presents the histograms of the samples of unknown parameters obtained from the TMCMC simulation for the M_5 of Classifier1. The mean values of samples inferred from Bayesian method using noise free data are tabulated in the column under TMCMC in Table 2. The identified results of the two methods for the M_5 of Classifier1 using different level noise measurements are also presented in Table 2. Significant differences can be found between the parameters identified from the LAD method and the

sample mean values obtained from the TMCMC simulation, even though the samples of the prior PDF of unknown parameters were drawn from the uniform distribution centered at the results from the LDA method.

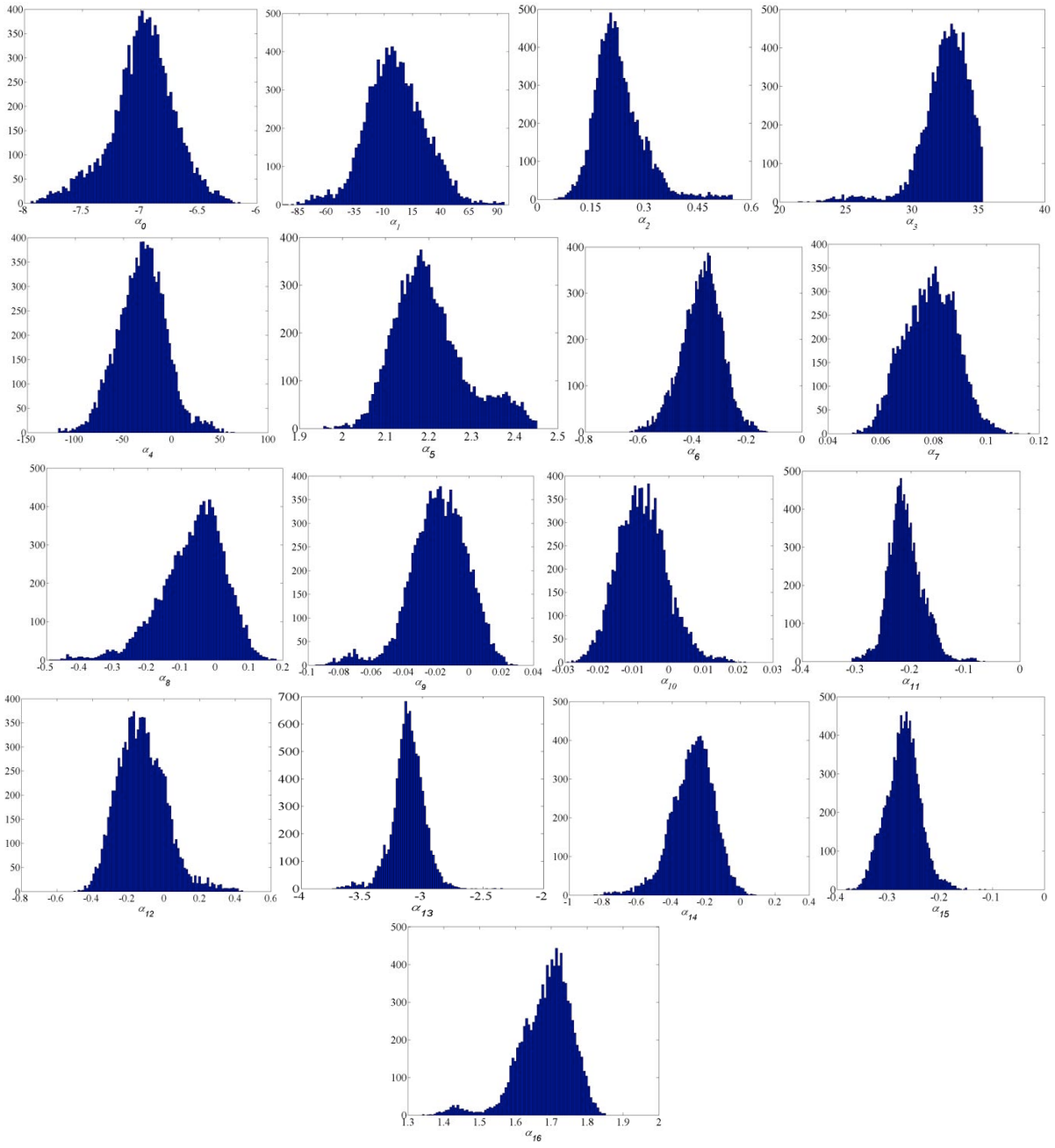


Fig.6 Histograms of the posterior PDF of each complements unknown model parameter vector $\boldsymbol{\alpha}$

3.2.3 Evaluation of Trained Classifiers

Once the eight classifiers corresponding respectively to the eight sub-regions were trained, they can be evaluated by using test data through comparing the predicted probabilities with the known events. In this study, total 10000 statistical samples of prior PDF of the unknown parameters were taken and updated to samples of posterior PDF of the model parameters in the Bayesian inference, resulting in 10000 prediction labels for a specific input feature vector.

As an example, Fig.7 illustrates the histograms of the prediction labels of model class M_5 of Classifier1 for two input feature vectors X_A and X_B, which respectively represent two accidents, i.e. damage in and outside of the sub-region1 in the noiseless test dataset. It can be found that all prediction labels of M_5 for feature vector X_A were greater than 0.5, indicating the damage is correctly located in Sub-region1. The damage scenario represented by feature vector X_B was also correctly identified outside of Sub-region1 because all its prediction labels less than 0.5. It is noteworthy that the trained classifier only took approximately a quarter second to provide samples of the prediction label for the given input feature vector of a specific event, indicating a very high efficiency over traditional damage detection algorithms. Those results illustrate that the presented method for locating potential damage of collision or other types of damage has acceptable accuracy and efficiency.

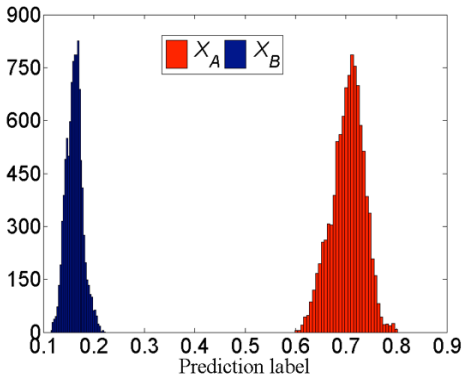


Fig.7 Predictions of the classification model class M_5 of Classifier1

3.2.4 Selection of Effective Feature and Classification Model Based on Bayesian Model Evaluation

Five dimensionally different feature vectors extracted from modal properties as shown in Table 1 lead to five different classification models for each sub-region. Those different classification models may be directly evaluated and compared in terms of their rates of correct classification of events in both training and test dataset. With prediction labels obtained from the trained classification model by using samples of model parameters, the rate of correct classification of a classification model class can be determined based on Eq. (13) by comparing its prediction labels with the defined threshold as defined in the Section 2.6.

The rates of correct classification of the five models of each classifier for both noiseless training dataset and test dataset are shown in Table 3. Results show that the model M_5 outperforms the other four model classes for both training and test dataset for Classifier 1 and Classifier 6. For Classifier 4, the model M_2 can 100% correctly identify the locations of damage events in both datasets. The model M_3 has the best performance for Classifier 8. It is noteworthy that there are more than one model classes in Classifier 2, 3, 5, 7 that can give 100% correct predictions for both datasets.

On the other hand, Bayesian model selection can provide an alternative way to evaluate model classes based on model evidences estimated from the TMCMC simulation. The log-evidences of the five model classes for each classifier are presented in Table 3. Results show that the log-evidence of the model M_5 is the largest among the five model classes for Classifier 1, 4, 6, and 8. For Classifier 2, 3, and 7, the model M_1 has the largest log-evidence than others. For Classifier 5, the model M_2 has the largest log-evidence. The probabilities of all model classes were computed from Eq. (9) with their model evidences and are also presented in Table 3. The results suggest that the best classification model classes for all classifiers ranked from Classifier1 to Classifier 8 are the model M_5, M_1, M_1, M_5, M_2, M_5, M_1, and M_5 respectively.

Table 3 Evaluations of Model Classes (0% noise)

Model			M_1	M_2	M_3	M_4	M_5
Classifier1	Rates (%)	Train dataset	87.50	90.00	87.50	97.50	100.00
		Test dataset	87.50	95.00	87.50	90.00	100.00
	Model evidence	log[f(D|M)]	-1.85×10^3	-1.50×10^3	-1.73×10^3	-830.30	-287.05
	Model	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier2	Rates (%)	Train dataset	100.00	95.00	100.00	100.00	100.00
		Test dataset	100.00	92.50	100.00	100.00	100.00
	Model evidence	log[f(D|M)]	18.89	-704.50	-54.09	-37.35	17.54
	Model probability	$p(M_j$|D,M)	0.79	0.00	0.00	0.00	0.21
Classifier3	Rates (%)	Train dataset	100.00	100.00	97.50	92.50	100.00
		Test dataset	100.00	100.00	92.50	92.50	97.50
	Model evidence	log[f(D|M)]	14.20	-384.86	-843.00	-1.31×10^3	-402.28
	Model probability	$p(M_j$|D,M)	1.00	0.00	0.00	0.00	0.00
Classifier4	Rates (%)	Train dataset	87.50	100.00	95.00	92.50	100.00
		Test dataset	87.50	100.00	95.00	90.00	97.50
	Model evidence	log[f(D|M)]	-1.70×10^3	-477.97	-789.82	-768.69	-45.55
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier5	Rates (%)	Train dataset	87.50	100.00	95.00	100.00	100.00
		Test dataset	87.50	97.50	92.50	100.00	100.00
	Model evidence	log[f(D|M)]	-1.89×10^3	-130.22	-1.62×10^3	-269.56	-391.83
	Model probability	$p(M_j$|D,M)	0.00	1.00	0.00	0.00	0.00
Classifier6	Rates (%)	Train dataset	87.50	92.50	87.50	90.00	100.00
		Test dataset	87.50	82.50	87.50	87.50	97.50
	Model evidence	log[f(D|M)]	-1.98×10^3	-1.07×10^3	-1.23×10^3	-1.40×10^3	-569.25
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier7	Rates (%)	Train dataset	100.00	92.50	100.00	95.00	100
		Test dataset	100.00	97.50	100.00	90.00	95.00
	Model evidence	log[f(D|M)]	17.39	-1.07×10^3	-326.16	-1.36×10^3	-496.89
	Model probability	$p(M_j$|D,M)	1.00	0.00	0.00	0.00	0.00
Classifier8	Rates (%)	Train dataset	97.50	90.00	100.00	97.50	100
		Test dataset	97.50	87.50	97.50	82.50	95.00
	Model evidence	log[f(D|M)]	-2.09×10^3	-1.40×10^3	-537.75	-617.90	-380.58
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00

Those results also show that both model evidence and model probability have a positive correlation with the rate of correct classification of the classifiers, i.e., the larger the model evidence or model probability, the higher the rate of correct classification of the corresponding model class. It is reasonable to conclude that the feature vector corresponding to the best model class is the most effective one for locating damage position. Thus, Bayesian model evaluation may be used as an alternative and effective tool for selecting feature vectors.

Table 4 and 5 present the rates of correct classification, as well as log-evidences of all model classes of each classifier for both training and test datasets under 5% and 10% noise environments. Those results show that noise has considerable impact on the performance of classification models, resulting in

that the best model of each classifier varies in different noise levels. The best models that have larger log-evidences for classifiers ranked from Classifier 1 to Classifier 8 are respectively M_5, M_5, M_1, M_4, M_5, M_1, M_3, and M_1 in 5% noise environment, while they are M_4, M_5, M_3, M_3, M_2, M_4, M_4, and M_5 in 10% noise situation. Although noise has evident effect on the performance of classification models, the rates of correct classification of the best models in different noise levels are still desirable. The lowest rate of correct classification for training data comes from the best model M_4 for Classifier 7 in the eight classifiers, which is still up to 90.00%, while the lowest rate of correct classification for test data is produced from the model M_3 for Classifier 4 and is still up to 85.00%.

3.2.5 Optimal Threshold and Predicted Probability of Damage

The rate of correct classification of a trained classifier can be calculated using Eq. (13) with the optimal threshold that is determined by maximizing the correct classification rate for all known events in the training dataset as described in the section 2.6. As an example, Fig.8 presents the change of the rate of correct classification R along with different thresholds for the model M_5 of Classifier1 in training dataset with 5% noise level. It explicitly shows that the optimal threshold of the model M_5 of Classifier1 is δ=0.46 and the corresponding maximum rate of correct classification is R=99.97%.

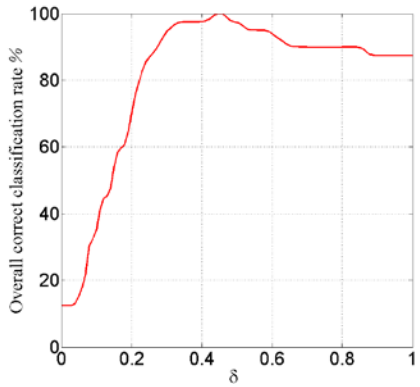


Fig.8 Variation of rate of correct classification R along with different threshold δ

Table 4 Evaluations of Model Classes (5% Noise)

Model			$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$
Classifier1	Rates (%)	Train dataset	82.50	90.00	90.00	85.00	97.50
		Test dataset	87.50	80.00	80.00	75.00	95.00
	Model evidence	log[f(D|M)]	-1.78×10^{3}	-1.15×10^{3}	-1.14×10^{3}	-2.04×10^{3}	-582.87
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier2	Rates (%)	Train dataset	82.50	92.50	100.00	87.50	100.00
		Test dataset	75.00	87.50	95.00	87.50	97.50
	Model evidence	log[f(D|M)]	-1.25×10^{3}	-1.44×10^{3}	-278.78	-1.69×10^{3}	-272.92
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier3	Rates (%)	Train dataset	92.50	87.50	87.50	90.00	85.00
		Test dataset	85.00	85.00	87.50	90.00	80.00
	Model evidence	log[f(D|M)]	-592.12	-1.44×10^{3}	-1.29×10^{3}	-1.12×10^{3}	-1.25×10^{3}
	Model probability	$p(M_j$|D,M)	1.00	0.00	0.00	0.00	0.00
Classifier4	Rates (%)	Train dataset	97.50	87.50	87.50	100.00	92.50
		Test dataset	85.00	85.00	87.50	100.00	85.00
	Model evidence	log[f(D|M)]	-624.41	-1.40×10^{3}	-1.16×10^{3}	-231.83	-937.93
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	1.00	0.00
Classifier5	Rates (%)	Train dataset	85.50	87.50	87.50	100.00	100.00
		Test dataset	87.50	87.50	77.50	92.50	100.00
	Model evidence	log[f(D|M)]	-1.29×10^{3}	-1.32×10^{3}	-1.43×10^{3}	-554.14	-87.67
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier6	Rates (%)	Train dataset	100.00	100.00	100.00	100.00	100.00
		Test dataset	97.50	100.00	100.00	95.00	100.00
	Model evidence	log[f(D|M)]	-34.47	-152.01	-247.20	-363.53	-279.12
	Model probability	$p(M_j$|D,M)	1.00	0.00	0.00	0.00	0.00
Classifier7	Rates (%)	Train dataset	82.50	87.50	100.00	97.50	87.50
		Test dataset	82.50	87.50	100.00	90.00	87.50
	Model evidence	log[f(D|M)]	-1.27×10^{3}	-1.36×10^{3}	-252.89	-863.97	-1.05×10^{3}
	Model probability	$p(M_j$|D,M)	0.00	0.00	1.00	0.00	0.00
Classifier8	Rates (%)	Train dataset	100.00	87.50	95.00	97.50	95.00
		Test dataset	100.00	80.00	92.50	85.00	87.50
	Model evidence	log[f(D|M)]	-215.42	-2.06×10^{3}	-509.33	-559.70	-601.74
	Model probability	$p(M_j$|D,M)	1.00	0.00	0.00	0.00	0.00

Once the optimal threshold and the rate of correct classification R of one classification model are determined, the probabilities of damage of a specific sub-region of the bridge pier after real barge-bridge collision can be determined by using Eq. (15). Table 6 presents the probabilities of damage predicted by the best model of each sub-region in noiseless test dataset, where C_l-M_j refers to as the model M_j of Classifier l. The true locations of the damage events in test dataset are also listed in the last column of Table 6. Besides, the optimal threshold of each classification model is given in the last row of Table 6. Those results show that most of the damage locations can be correctly identified in terms of the probability of damage in the sub-region. Among all events in test dataset, there is only one misclassification of damage location for the forty events, i.e. the 19th case marked with an asterisk. Thus, the probability of correct classification of the presented framework is up to 97.50% for the given noiseless test data.

Table 5 Evaluations of Model Classes (10% Noise)

Model			M_1	M_2	M_3	M_4	M_5
Classifier1	Rates (%)	Train	87.50	90.00	90.00	100.00	95.00
		Test dataset	87.50	87.50	85.00	97.50	90.00
	Model evidence	log[f(D|M)]	-1.53×10^3	-1.45×10^3	-1.21×10^3	-534.84	-553.33
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	1.00	0.00
Classifier2	Rates (%)	Train	87.50	90.00	92.50	87.50	97.50
		Test dataset	87.50	87.50	95.00	87.50	87.50
	Model evidence	log[f(D|M)]	-1.75×10^3	-1.57×10^3	-947.98	-1.14×10^3	-566.28
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	0.00	1.00
Classifier3	Rates (%)	Train	87.50	87.50	100.00	90.00	90.00
		Test dataset	87.50	87.50	90.00	85.00	90.00
	Model evidence	log[f(D|M)]	-726.43	-864.54	-720.32	-1.29×10^3	-1.09×10^3
	Model probability	$p(M_j$|D,M)	0.002	0.00	0.998	0.00	0.00
Classifier4	Rates (%)	Train	87.50	87.50	92.50	87.50	85.00
		Test dataset	87.50	87.50	85.00	85.00	82.50
	Model evidence	log[f(D|M)]	-1.46×10^3	-1.72×10^3	-856.01	-1.73×10^3	-1.52×10^3
	Model probability	$p(M_j$|D,M)	0.00	0.00	1.00	0.00	0.00
Classifier5	Rates (%)	Train	87.50	100.00	87.50	85.00	87.50
		Test dataset	87.50	97.50	90.00	887.50	87.50
	Model evidence	log[f(D|M)]	-1.03×10^3	-527.88	-1.27×10^3	-1.69×10^3	-1.53×10^3
	Model probability	$p(M_j$|D,M)	0.00	1.00	0.00	0.00	0.00
Classifier6	Rates (%)	Train	100.00	87.50	95.00	100.00	100.00
		Test dataset	100.00	87.50	95.00	100.00	100.00
	Model evidence	log[f(D|M)]	-157.80	-1.34×10^3	-669.49	-63.64	-379.05
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	1.00	0.00
Classifier7	Rates (%)	Train	87.50	87.50	85.00	90.00	90.00
		Test dataset	87.50	87.50	90.00	90.00	85.00
	Model evidence	log[f(D|M)]	-1.34×10^3	-1.28×10^3	-1.55×10^3	-1.14×10^3	-1.16×10^3
	Model probability	$p(M_j$|D,M)	0.00	0.00	0.00	1.00	0.00
Classifier8	Rates (%)	Train	87.50	87.50	95.00	97.50	92.50
		Test dataset	87.50	82.50	82.50	77.50	87.50
	Model evidence	log[f(D|M)]	-1.10×10^3	-1.98×10^3	-877.63	-897.11	-842.57
	Model probability	$p(M_j$|D,M)		0.00	0.00	0.00	1.00

The prediction probabilities of damage for noise level of 5% and 10% are presented in Table 7 and 8, respectively. It can be noted that with the increase of noise level, the rate of correct classification is slightly reduced and the optimal threshold also decreases. This decreasing of optimal threshold makes prediction results more conservative due to increased uncertainties. The misclassification events are marked with superscripted asterisks at the upper right of the numbers of the events. The rate of correct classification of the presented framework is still up to 97.50% for the test dataset with noise level of 5% and 72.50% for the test dataset with the noise level of 10%.

Table 6 Predicted Probabilities of Damage Locations (Noise Level 0%)

Event No.	C_1- M_5	C_2- M_1	C_3- M_1	C_4- M_5	C_5- M_2	C_6-M_5	C_7- M_1	C_8- M_5	**True Location**

1	**100.00**	0.00	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region1
2	**100.00**	0.00	0.00	3.85	0.00	100.0	0.00	**85.55**	Sub-region1
3	**100.00**	0.00	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region1
4	**100.00**	0.00	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region1
5	**100.00**	0.00	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region1
6	0.00	**100.00**	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region2
7	0.00	**100.00**	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region2
8	0.00	**100.00**	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region2
9	0.00	**100.00**	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region2
10	0.00	**100.00**	0.00	0.00	0.00	0.23	0.00	0.00	Sub-region2
11	0.00	0.00	**100.00**	0.00	0.00	0.23	0.00	0.00	Sub-region3
12	0.00	0.00	**100.00**	0.00	0.00	0.23	0.00	0.00	Sub-region3
13	0.00	0.00	**100.00**	0.00	**100.00**	0.23	0.00	0.00	Sub-region3
14	0.00	0.00	**100.00**	0.00	0.00	0.23	0.00	0.00	Sub-region3
15	0.00	0.00	**100.00**	0.00	0.00	0.23	0.00	0.00	Sub-region3
16	0.00	0.00	0.00	**100.00**	0.00	0.23	0.00	0.00	Sub-region4
17	0.00	0.00	0.00	**100.00**	0.00	0.23	0.00	0.00	Sub-region4
18	0.00	0.00	0.00	**100.00**	0.00	0.23	0.00	0.00	Sub-region4
19*	0.00	0.00	0.00	**0.00**	0.00	0.23	0.00	**100.00**	Sub-region4
20	0.00	0.00	0.00	**100.00**	0.00	0.23	0.00	0.00	Sub-region4
21	0.00	0.00	0.00	0.00	**100.00**	0.23	0.00	0.00	Sub-region5
22	0.00	0.00	0.00	0.00	**95.79**	0.23	0.00	0.00	Sub-region5
23	0.00	0.00	0.00	0.00	**100.00**	0.23	0.00	0.00	Sub-region5
24	0.00	0.00	0.00	0.00	**88.11**	0.23	0.00	0.00	Sub-region5
25	0.00	0.00	0.00	0.00	**100.00**	0.23	0.00	0.00	Sub-region5
26	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	Sub-region6
27	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	Sub-region6
28	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	Sub-region6
29	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	Sub-region6
30	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	Sub-region6
31	0.00	0.00	0.00	0.00	0.00	0.23	**100.00**	0.00	Sub-region7
32	0.00	0.00	0.00	0.00	0.00	0.23	**100.00**	0.00	Sub-region7
33	0.00	0.00	0.00	0.00	0.00	0.23	**100.00**	0.00	Sub-region7
34	0.00	0.00	0.00	0.00	0.00	0.23	**100.00**	0.00	Sub-region7
35	0.00	0.00	0.00	0.00	0.00	0.23	**100.00**	5.69	Sub-region7
36	0.00	0.00	0.00	0.00	0.00	0.23	0.00	**100.00**	Sub-region8
37	0.00	0.00	0.00	0.00	0.00	0.23	0.00	**100.00**	Sub-region8
38	0.00	0.00	0.00	0.00	0.00	0.23	0.00	**100.00**	Sub-region8
39	0.00	0.00	0.00	0.00	0.00	0.23	0.00	**100.00**	Sub-region8
40	0.00	0.00	0.00	0.00	0.00	0.23	0.00	**99.30**	Sub-region8
R (%)	100.00	100.00	100.00	100.00	100.00	99.77	100.00	100.00	
δ	0.5	0.5	0.5	0.5	0.5	0.52	0.50	0.52	

Note: Ci refers to the i-th divided sub-region and Mj refers to the j-th classification model that is selected for the i-th sub-region.
R is the overall correct classification rate calculated by Eq. (13) with the prediction labels of training data.
The superscript * denotes the incorrectly located damage according to the prediction probabilities.

Table 7 Predicted Probabilities of Damage Locations (Noise Level 5%)

Event No.	**C_1- M_5**	**C_2- M_5**	**C_3- M_1**	**C_4- M_4**	**C_5- M_5**	**C_6- M_1**	**C_7- M_3**	**C_8- M_1**	**True location**
1	**100.00**	0.00	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region1
2	**100.00**	0.00	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region1
3*	**2.62**	0.00	**100.0**	0.00	0.00	0.00	0.00	0.00	Sub-region1
4	**100.0**	**100.00**	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region1
5	**100.0**	0.00	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region1
6	0.03	**100.00**	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region2
7	0.03	**100.00**	**100.0**	0.00	0.00	0.00	0.00	0.00	Sub-region2
8	0.03	**66.77**	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region2
9	0.03	**100.00**	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region2
10	0.03	**97.66**	2.73	0.00	0.00	0.00	0.00	0.00	Sub-region2
11	0.03	0.00	**100.0**	0.00	0.00	0.00	0.00	0.00	Sub-region3
12	0.03	0.00	**100.0**	0.00	0.00	0.00	0.00	0.00	Sub-region3
13	0.03	0.00	**83.07**	0.00	0.00	0.00	0.00	0.00	Sub-region3

14	0.03	0.00	**100.0**	0.00	0.00	0.00	0.00	0.00	Sub-region3
15	0.03	0.00	**44.21**	0.00	0.00	0.00	0.00	0.00	Sub-region3
16	0.03	0.00	2.73	**100.00**	0.00	0.00	0.00	0.00	Sub-region4
17	0.03	0.00	2.73	**100.00**	0.00	0.00	0.00	0.00	Sub-region4
18	0.03	0.00	2.73	**100.00**	0.00	0.00	0.00	0.00	Sub-region4
19	0.03	0.00	2.73	**100.00**	0.00	0.00	0.00	0.00	Sub-region4
20	0.03	0.00	2.73	**100.00**	0.00	0.00	0.00	0.00	Sub-region4
21	**100.0**	0.00	2.73	0.00	**100.00**	0.00	0.00	0.00	Sub-region5
22	0.03	3.99	2.73	0.00	**100.00**	0.00	0.00	0.00	Sub-region5
23	0.03	0.00	**100.0**	0.00	**100.00**	**100.00**	0.00	0.00	Sub-region5
24	0.03	0.00	**100.0**	0.00	**100.00**	0.00	0.00	0.00	Sub-region5
25	0.03	0.00	**100.00**	0.00	**99.91**	**47.91**	0.00	0.00	Sub-region5
26	0.03	0.00	2.73	0.00	0.00	**100.00**	0.00	0.00	Sub-region6
27	0.03	0.00	2.73	0.00	0.00	**100.00**	0.00	0.00	Sub-region6
28	0.03	0.00	2.73	0.00	0.00	**100.00**	0.00	0.00	Sub-region6
29	0.03	0.00	2.73	0.00	0.00	**100.00**	0.00	0.00	Sub-region6
30	0.03	0.00	2.73	0.00	0.00	**100.00**	0.00	0.00	Sub-region6
31	0.03	0.00	2.73	0.00	0.00	0.00	**100.00**	0.00	Sub-region7
32	0.03	0.00	2.73	0.00	0.00	0.00	**100.00**	0.00	Sub-region7
33	0.03	0.00	2.73	0.00	0.00	0.00	**100.00**	0.00	Sub-region7
34	0.03	0.00	2.73	0.00	0.00	0.00	**100.00**	0.00	Sub-region7
35	36.51	10.32	2.73	0.00	0.00	0.00	**100.00**	0.00	Sub-region7
36	5.66	0.00	2.73	0.00	0.00	0.00	0.00	**100.00**	Sub-region8
37	0.03	0.00	2.73	29.23	0.00	0.00	0.00	**100.00**	Sub-region8
38	0.03	0.00	2.73	0.00	0.00	0.00	0.00	**100.00**	Sub-region8
39	0.03	0.00	2.73	0.00	0.00	0.00	0.00	**100.00**	Sub-region8
40	0.03	0.00	2.73	0.00	0.00	0.00	0.00	**100.00**	Sub-region8
R (%)	99.97	100.00	97.27	100.00	100.00	100.00	100.00	100.00	
δ	0.46	0.48	0.37	0.50	0.40	0.45	0.50	0.29	

Note: R is the overall correct classification rate calculated by Eq. (13) with the prediction labels of training data.
The superscript * denotes the incorrectly located damage according to the prediction probabilities.

Table 8 Predicted Probabilities of Damage Locations (Noise Level 10%)

Event No.	**C_1- M_4**	**C_2- M_5**	**C_3- M_3**	**C_4- M_3**	**C_5- M_2**	**C_6- M_4**	**C_7- M_4**	**C_8- M_5**	**True location**
1	**97.49**	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region1
2	**100.00**	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region1
3	**100.00**	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region1
4	**100.00**	0.14	0.00	3.07	4.38	0.00	4.58	5.0	Sub-region1
5	**100.00**	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region1
6	0.00	99.99	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region2
7	0.00	100.0	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region2
8*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region2
9	0.00	100.0	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region2
10*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region2
11	0.00	0.14	**100.00**	3.07	0.00	0.00	20.83	5.0	Sub-region3
12*	0.00	100.0	**0.00**	3.07	0.00	0.00	4.58	5.0	Sub-region3
13	0.00	0.14	**100.00**	3.07	0.00	0.00	100.0	5.0	Sub-region3
14	0.00	0.14	**100.00**	3.07	0.00	0.00	4.58	5.0	Sub-region3

15	0.00	0.14	**100.00**	3.07	0.00	0.00	4.58	5.0	Sub-region3
16	0.00	0.14	0.00	100.0	0.00	0.00	4.58	5.0	Sub-region4
17*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	100.0	Sub-region4
18	0.00	0.14	0.00	100.0	0.00	0.00	4.58	5.0	Sub-region4
19*	0.00	99.93	**100.00**	3.07	0.00	0.00	4.58	5.0	Sub-region4
20*	0.00	100.0	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region4
21	0.00	0.14	0.00	3.07	**100.00**	0.00	4.58	5.0	Sub-region5
22	0.00	0.14	0.00	3.07	**100.00**	0.00	4.58	5.0	Sub-region5
23	0.00	0.14	0.00	3.07	**100.00**	0.00	4.58	5.0	Sub-region5
24	16.01	0.14	0.00	3.07	**100.00**	0.00	4.58	5.0	Sub-region5
25	0.00	0.14	**100.00**	4.4	**99.94**	0.00	4.58	5.0	Sub-region5
26	19.31	0.14	0.00	3.07	0.00	**100.00**	4.58	5.0	Sub-region6
27	0.00	0.14	0.00	3.07	0.00	**100.00**	4.58	5.0	Sub-region6
28	0.00	0.14	0.00	3.07	0.00	**100.00**	4.58	5.0	Sub-region6
29	0.00	0.14	0.00	99.99	0.00	**100.00**	4.58	5.0	Sub-region6
30	0.00	0.14	0.00	100.0	0.00	**100.00**	4.58	5.0	Sub-region6
31*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region7
32*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region7
33	0.00	0.14	0.00	3.07	0.00	0.00	10.04	5.0	Sub-region7
34	0.00	0.14	0.00	3.07	0.00	0.00	100.0	5.0	Sub-region7
35	0.00	0.14	0.00	3.07	0.00	0.00	100.0	93.46	Sub-region7
36*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region8
37	0.00	27.97	0.00	100.0	0.00	0.00	4.58	100.0	Sub-region8
38*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region8
39	0.00	0.14	**100.00**	3.07	0.00	0.00	4.58	100.0	Sub-region8
40*	0.00	0.14	0.00	3.07	0.00	0.00	4.58	5.0	Sub-region8
R (%)	100.00	99.86	100.00	96.93	100.00	100.00	95.42	95.00	
δ	0.44	0.45	0.49	0.35	0.41	0.36	0.45	0.49	

Note: *R* is the overall correct classification rate calculated by Eq. (13) with the prediction labels of training data.
The superscript * denotes the incorrectly located damage according to the prediction probabilities.

4. Impacts/Benefits of Implementation

4.1 Practical Application

Once a barge-bridge collision event happens, field dynamic measurements can be collected from the collided bridge structure with the sensor network. The best feature vectors are then extracted and input into the best classification models of each of the trained classifiers. With the identified threshold of each classifier, the prediction probability of the damage locating in each of the sub-regions can be determined according to Eq. (15). In this numerical simulation, measurements from the prototype of bridge after a real barge-bridge collision incident can only be simulated by using the bridge structural model, and can be taken from any one event in the test dataset presented in the proceeding. Thus, predicted probabilities of damage occurring in each sub-region from implementing the presented framework can be referred to results tabulated in Table 7 or 8. The evaluation of those results through comparison with the damage scenarios can also be found in those tables as discussed in the above sections.

4.2 Discussion and Limitations and Future Research Directions

The presented framework is based on the elastic structural model to generate the modal properties of the considered structure at different damage extent for training probabilistic classification models. In the structural model, the damage extent is represented by different reduction of elastic stiffness of a specific sub-region. In real world, damage may cause local nonlinear force-deformation relation in the damaged sub-regions. However, the amplitude of real measured structure vibrations is small particularly under normal ambient excitations. Thus, it can be reasonably assumed that the force-deformation relationship of the damaged sub-region can be approximately represented by the linear elastic behavior with reduced elastic stiffness.

Even though training data are generated from numerical simulation by using a structural model, rather than field measurements of historically known real damaged events from the prototype structure in the past, application of trained classifier for identifying probability of damage is not essentially different from other model-based damage identification algorithms. All those algorithms identify damage through determining damage parameters of a structure model by matching simulation results with field measurements. Thus, their identified results are impacted by accuracy of the structural model and field measurements. Within the presented probabilistic-based framework, uncertainties associated with the structural model and field measurements can be better quantified in the results in terms of probability of damage. However, it should be noted that in the illustration and examination of the presented framework, the field measurements from the structure after damage event are simulated based on the structural model with added noise representing measurement errors. Thus, the future research may further examine and verify the presented framework by using real measurements from lab experiment or prototype structure site.

With focusing on illustration of the presented framework, the optimal sensor deployment for collecting effective vibration measurements was not considered in the study. The selected sensor locations and its measured directions, which were represented by selected DOFs in the numerical simulation, may not capture the most effective features of damage locations after barge-bridge collisions. Thus, the optimization of sensor deployment has to be particularly studied when the presented framework is applied for damage location of a specific structure.

Even though the numerical simulation only illustrates identification of potential damage locations in one pier of a bridge due to barge-bridge collision, the proposed framework actually can be extended to locate other types of structural damage in other regions of civil engineering structural members. Besides, the numerical simulation only demonstrates the scenario of damage occurrence in a single sub-region. However, actual damage may synchronously appear in several adjacent sub-regions or in the junction of several sub-regions in a given division of sub-regions. For dealing with such situations in practical application, multi-layer and hierarchical divisions of a specific region at different grid levels and shifted locations can be established. As a result, damage occurring in several adjacent sub-regions or the junction of several sub-regions in a given division sub-regions could be located within a specific sub-region in the other set of divisions of sub-regions. Then, the framework can be implemented for all sub-regions of all considered different sets of division of sub-regions. The damage location can be identified at the sub-region with the largest probability of damage among all considered sub-regions.

The numerical simulation further indicates that more attention should be paid to determine the range or space of unknown parameters of classification models for their prior PDF samples. Although results obtained from the LDA method [24] can be taken as the mean values for samples of parameters of classification model derived from their prior PDF for Bayesian inference with the TMCMC sampling [18-19], appropriate range or space of samples of the prior PDF of those unknown parameters has to be determined by judgment. Inappropriately estimated range of those samples of the prior PDF of those

unknown parameters may result in samples of the posterior PDF of those parameters outside of their space, or concentrated at the boundary of their space (as illustrated in Fig.5). The larger estimated space of samples of the prior PDF of unknown parameters may make adequate exploration of the space of the possible regions of high likelihood of those unknown parameters and avoid the aforementioned problems. However, it would inversely increase computational burdens in the process of TMCMC simulation. Therefore, a more appropriate method to determine the range of prior space is still needed to be further explored in future research.

5. Recommendations and Conclusions

This report presents a novel framework for rapidly identifying potential damage locations using vibration measurements from a structure after damage potentially occurs. With the framework, the probabilities of damage occurrence in different sub-regions of a structural system can be determined for providing the guide for further assessment and inspection. The implementation and applicability of the presented framework are illustrated and examined through simulation of identifying damage locations of a bridge pier after barge-bridge collision events. The novel characteristics of the presented framework lie in that intensively computational simulations of dynamic responses of a structure at different damage levels are conducted prior to damage events to generate data for training the probabilistic classifier based on the binary logistic model. Thus, the trained classifier can be used to quickly identify the probabilities of damage occurrence in specific regions based on vibration measurements, leading to significant reduction of computing time in damage identification processes in comparison with traditional probabilistic-based damage identification algorithms.

Within the presented framework, feature vectors containing sensitive information on possible damage and its locations are extracted from the change of simulated modal properties caused by damage using the Principle Component Analysis. The probabilistic-based logistic model is trained as a classifier

with the extracted feature vectors and labels of known events through Bayesian inference with Transitional Markov Chain Monte Carlo simulation. The most effective features can be selected in terms of model evidence through Bayesian model selection for effective classification. The optimal threshold for distinguishing damaged and undamaged scenarios from the prediction label is proposed through maximizing the rate of correct classification of the classifier for all known damage events in training dataset. Uncertainties associated with available measurements and classifier models are quantified in terms of the probability of damage occurrence in a specified sub-region.

Results from numerical simulation indicate that the presented framework can rapidly determine the probabilities of structural damage locations of a measured event in less than one second. Even though levels of measurement errors or noise have significant impacts on the performances of classification models, the rate of correct classification is larger than 97.5% for all events in test dataset with measurement errors up to 5%, and larger than 72.50% with measurement errors up to 10%. The model evidences obtained from Bayesian inference positively correlate with model performances in terms of the rate of correct classification and can be used to determine effective features for a classifier. The optimal threshold for distinguishing between damaged and undamaged scenarios from classification labels has a decreasing tendency as the noise level increasing, which makes the classification results more conservative. The future research may focus on using real measurements from lab experiment or prototype structure site to further examine the presented framework.

Acknowledge

The authors gratefully acknowledge the support of the Maritime Transportation Research and Education Center through Institute for Multimodal Transportation and the support of the National Science Foundation under the grant NSF/HRD # 1036328. Any opinions, findings,

conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Reference

1. Doebling S W, Farrar C R, Prime M B (1998) A summary review of vibration-based damage identification methods. Shock and vibration digest 30(2): 91-105.
2. Jaishi B, Ren W X (2005) Structural finite element model updating using ambient vibration test results. Journal of Structural Engineering 131(4): 617-628.
3. Simoen E, De Roeck G, Lombaert G (2015) Dealing with uncertainty in model updating for damage assessment: A review. Mechanical Systems and Signal Processing 56: 123-149.
4. Mares C, Surace C (1996) An application of genetic algorithms to identify damage in elastic structures. Journal of sound and vibration 195(2): 195-215.
5. Kouchmeshky B, Aquino W, Bongard J C, Lipson H (2007) Co-evolutionary algorithm for structural damage identification using minimal physical testing. International Journal for Numerical Methods in Engineering 69(5): 1085-1107.
6. Begambre O, Laier J E (2009) A hybrid Particle Swarm Optimization–Simplex algorithm (PSOS) for structural damage identification. Advances in Engineering Software 40(9): 883-891.
7. Beck J L, Katafygiotis L S (1998) Updating models and their uncertainties. I: Bayesian statistical framework. Journal of Engineering Mechanics 124(4): 455-461.
8. Zheng Wei and Yung-Tsang Chen (2014) Novel probabilistic approach to assessing barge–bridge collision damage based on vibration measurements through transitional Markov chain Monte Carlo sampling. Journal of Civil Structural Health Monitoring 4(2): 119-131.
9. Papadimitriou C, Papadioti D C (2013) Fast Computing Techniques for Bayesian Uncertainty Quantification in Structural Dynamics. Topics in Model Validation and Uncertainty Quantification 5: 25-31, Springer New York.
10. Hadjidoukas P E, Angelikopoulos P, Papadimitriou C, Koumoutsakos P (2015) Π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. Journal of Computational Physics 284(2015):1-21.

11. Johnson R A, Wichern D W (1992) Applied multivariate statistical analysis (Vol. 4). Englewood Cliffs, Prentice hall.
12. Magalhães F, Cunha A, Caetano E (2012) Vibration based structural health monitoring of an arch bridge: from automated OMA to damage detection. Mechanical Systems and Signal Processing 28: 212-228.
13. Bin H, Bong-Hwan K and Heung S K (2014) PCA-based damage classification of delaminated smart composite structures using improved layerwise theory. Computers and Structures 141: 26-35.
14. Hosmer Jr D W, Lemeshow S (2004) Applied logistic regression. John Wiley & Sons, Inc., New York.
15. Bishop C M (2006) Pattern recognition and machine learning. Springer, New York.
16. Zheng W, Yu W (2014) Probabilistic Approach to Assessing Scoured Bridge Performance and Associated Uncertainties Based on Vibration Measurements. Journal of Bridge Engineering.
17. Jaynes E T (1983) Papers on probability, statistics and statistical physics, R. D. Rosenkrantz, ed., Kluwer Academic Publishers, Boston.
18. Ching J, Chen Y C (2007) Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. Journal of engineering mechanics 133(7):816-832.
19. Zheng W, Yu Y (2013) Bayesian Probabilistic Framework for Damage Identification of Steel Truss Bridges under Joint Uncertainties. Advances in Civil Engineering 2013.
20. Cheung S H, Beck J L (2009) Bayesian Model Updating Using Hybrid Monte Carlo Simulation With Application to Structural Dynamic Models With Many Uncertain Parameters. J. Eng. Mech. 135(4):243–255
21. Andrieu C, de Freitas N, Doucet A, Jordan M I (2003) An introduction to MCMC for machine learning. Machine Learning 50(1-2): 5–43.
22. CSi Bridge SAP2000 14 [Computer software]. Walnut Creek, CA, Computers and Structures.
23. Mokwa R L, Duncan J M (2002) Simplified Method of Analysis for Laterally Loaded Pile Groups. Pro., 37th Engineering Geology and Geotechnical Engineering Symposium, Boise State University, Boise, Idaho March, 133-146.
24. Fisher R A (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2): 179-788.