WILEY | Hindawi

Research Article

Condition Assessment of Highway Bridges Using Textual Data and Natural Language Processing- (NLP-) Based Machine Learning Models

De-Cheng Feng [ID],[1,2] **Wen-Jie Wang**,[2] **Sujith Mangalathu**,[3] **and Zhen Sun** [ID][1,2,4]

[1]*Key Laboratory of Concrete and Prestressed Concrete Structures of the Ministry of Education, Southeast University, Nanjing 211189, China*
[2]*School of Civil Engineering, Southeast University, Nanjing 211189, China*
[3]*Data Analytics Division, Mangalathu, Mylamkulam, Kottarakara, Kollam 691507, India*
[4]*Construct-ViBest, Faculty of Engineering (FEUP), University of Porto, Porto 4200-465, Portugal*

Correspondence should be addressed to Zhen Sun; sunzhen08@gmail.com

Condition rating of bridges is specified in many countries since it provides a basis for the decision-making of maintenance actions such as repair, strengthening, or limitation of passing vehicle weight. In practice, professional engineers check the textual description of damages to bridge members, such as girders, bearings, expansion joints, and piers that are acquired from periodic inspections, and then make a rating of the bridge condition. The task is time-consuming and labor-intensive due to the large amount of detailed data buried in the inspection reports. In this paper, a natural language processing- (NLP-) based machine learning (ML) approach is proposed for automated and fast bridge condition rating, which can efficiently extract the information of deficiencies in bridge members. The proposed approach involves three major steps, say, data repository establishment, NLP-based textual data processing, and ML-based bridge condition rating prediction. The data repository is established with the inspection reports of 263 concrete bridges, and in total there, are four condition levels for the bridges. Then, the NLP-based textual data processing approach is implemented to calculate the word frequency and the word clouds to visualize the characteristics of bridges in different condition levels. Finally, four typical ML techniques are adopted to generate the predictive model of the bridge condition rating. The results indicate that the NLP-based ML prediction model has an accuracy of 89% and is very efficient so that it can be used for large-scale applications such as condition rating for regional-level bridges.

1. Introduction

The safety of bridges is significant to ensure the functioning of the transportation network and economic development in our society. During their service life, bridges are subjected to various loads such as temperature [1], wind [2], traffic [3–5], and earthquakes [6], which aggravate the defects of bridges. These defects could cause bridge failures if they are not detected and repaired. Wardhana and Hadipriono [7] studied more than 500 bridge failures in the United States between 1989 and 2000, revealing that 76% of them occurred during operation. Xu et al. [8] analyzed 302 catastrophic highway bridge collapses in China between 2000 and 2014, out of which 171 were in-service bridges and had an average service life of only 18.7 years.

Generally, periodic inspection bridge by bridge with visual and nondestructive techniques is widely adopted around the world, such as in the United States [9, 10], Japan [11, 12], Korea [13], China [14, 15], Australia [16], and European countries [17]. In practice, conditions of bridges are rated into different levels [18], and the deficient bridges are given more attention for load-carrying capacity evaluation [19] or the implementation of a vehicle weight limit [4]. Condition rating is mainly performed based on the

textual description of damages to bridge members such as girders, bearings, expansion joints, and piers, and then the results are combined to obtain the overall rating of the bridge. The engineers for condition rating must be familiar with relevant condition rating standards and be experienced with bridge inspection practice. Considering that a condition rating is required for each bridge with a large amount of textual data buried in the inspection reports, it is a time-consuming and labor-intensive task. Therefore, it is paramount to propose methodologies for efficient and accurate condition rating of bridges.

To achieve fast and automated bridge condition rating, in the past few years, machine learning (ML) methods have enjoyed fast development based on relevant information from bridge database or inspection [20–22]. State-of-the-art algorithms include artificial neural networks (ANN) [23–25], clustering [26, 27], decision trees (DT) [28, 29], support vector machines [30–33], ensemble learning methods [34–38], and unsupervised learning methods [39], which have indicated both high accuracy and efficiency. Using the relevant historical inspection and inventory data and records of maintenance, Huang [40] developed an artificial neural network (ANN) model to predict bridge deck deterioration with 11 features such as bridge age, deck length, number of lanes, number of spans, and design load. The predicted condition rating showed an accuracy of around 75%. Li and Burgueno [41] adopted several ANN methods to evaluate bridge abutment conditions in the state of Michigan with 8 input variables of bridge length, bridge width, skew angle, age, annual temperature difference, average daily truck traffic, approach surface type, and structural type. Liu and Zhang [42] presented a deep learning-based highway bridge components condition rating prediction approach with 24 input features including geographic region, structural configuration, and other bridge attributes. A case study on the condition rating of deck, superstructure, and substructure demonstrated an accuracy of over 85%. Xia et al. [43–45] developed a condition assessment approach for network-level bridges using machine learning-based techniques. The maintenance scheme is also optimized based on a deterioration model [46].

These studies indicate that ML-based methods have the potential to achieve efficient and reliable condition ratings for bridges. However, they mainly utilized the structural parameters and geographic or environmental features as input variables instead of directly utilizing the detailed textual inspection data. In fact, detailed bridge inspection reports contain rich information on bridge deficiencies, which is also usually required by bridge operators. However, it is challenging to be directly used in the ML methods because the textual descriptions in bridge inspection reports are usually nonstandard and follow various writing patterns, which hinders effective information extraction. Moreover, the textual description in the inspection reports is actually at the component level, while the final condition rating is at the structure level. Hence, a highly nonlinear and complicated

relationship exists between the input variables (textual description) and out variables (bridge condition rating). These two major issues make it difficult to build up simple ML models for the prediction of the bridge rating.

To address similar challenges where huge amount of textual data are presented, natural language processing (NLP) technology has been adopted recently to convert text data into digital vectors for better data mining. For instance, Zhang and El-Gohary [47] utilized semantic NLP and machine learning techniques to extract concepts from documents such as building codes and match them to concepts in the industry foundation classes for building information modelling. Le and Jeong [48] implemented NLP to detect data elements from text documents and used machine learning to extract and classify roadway data items. A human-encoded test showed that the precision and recall rate are 92.76% and 81.02%, respectively. Liu and El-Gohary [49] used semisupervised conditional random fields for extracting information on deficiencies and maintenance actions from bridge inspection reports, which achieved a precision of 94.1%. Mangalathu and Burton [50] applied an NLP-based postdisaster damage evaluation approach using the long short-term memory (LSTM) deep learning algorithm to classify building damages after an earthquake.

Based on the above-mentioned background, the present paper aims to develop a bridge condition rating approach with direct textual data from inspection reports by combining the NLP and ML methods. In total, 263 bridge inspection reports are collected, and the NLP is utilized to process the textual data in the inspection reports to convert them into vectors. Afterwards, they can be fed to the state-of-the-art ML algorithms to train the condition rating predictive model. The novelty of the present paper can be explained in three-fold as follows:

(1) With the NLP, the textual description in the inspection reports can be directly used as input, and the term frequency (TF), term frequency-inverse document frequency (TF-IDF), and word clouds can be obtained to illustrate the extracted information of the bridges and show a direct visualization of the reports

(2) The ML methods can achieve high accuracy and automated prediction of the condition rating of the bridges, avoiding the time-consuming and labor-intensive works by experienced engineers

(3) Since the approach is accurate, efficient, and automatic, it can be extended in a large-scale application, such as condition rating of the regional bridge network

A novel approach for estimating bridge condition rating is introduced, utilizing NLP and ML techniques. Findings in this study suggest that the NLP-based ML approach can capture the information buried in the inspection reports and provide a promising tool for efficient and accurate bridge condition rating.

2. The Proposed Method for Textual Data-Based Bridge Condition Assessment

With tens of thousands of in-service bridges being inspected every year, it is a challenging task to implement the condition rating. This paper proposes an approach that combines NLP with ML techniques for rapid, automatic, and reliable condition rating. As illustrated in Figure 1, the approach is composed of three steps, i.e., data repository establishment, NLP-based textual data processing, and an ML-based bridge condition rating prediction model. Herein, the datasets are established with inspection reports of 263 highway bridges in China, which are classified into four condition ratings according to the condition levels, say, A, B, C, and D, with the damage level increasing from light to severe. The three steps are described as follows.

In step 1, the data repository is established with inspection reports of a regional network of 263 bridges. Each bridge consists of different components, such as the deck, girder, bearing, expansion joints, pier, and foundation. The defects in each component of the bridge are recorded in the report after inspection by engineers using visual or nondestructive techniques. For example, textual descriptions of concrete cracks could be expressed as "A longitudinal crack is observed in the deck bottom plate with a width of 0.15 mm and a length of 2.2 m." The condition rating of each bridge is determined as the weighted sum of the damage extent of all the components that are rated by professional engineers, and it is extracted as the output for prediction in the ML model. Note that the guidelines to rate the damage extent of the components are defined by national standards.

In step 2, NLP-based textual data processing is performed on the inspection reports to convert the text into vectors, which can be directly used as the input of the ML model. The text in reports is preprocessed to remove redundant information such as punctuation marks, excessive space, and stop words. Two word frequency analysis parameters, i.e., TF and TF-IDF, will illustrate the extracted information in the inspection reports. Word clouds with TF or TF-IDF can visualize the condition-level information in the inspection reports of bridges in different condition ratings.

In step 3, the textual data processed in step 2 is fed into the ML model to build up the predictive model of the bridge condition ratings. The datasets are divided into training and testing sets as 70% and 30%, respectively. In the training process, the grid search method is adopted to find the hyperparameters with optimal performance for the ML models, while the 10-fold cross-validation (CV) is conducted to avoid any bias due to random sampling of the training dataset. Then the state-of-the-art ML models are evaluated on the testing dataset, and four typical measures, i.e., precision, recall, $F1$-score, and accuracy, are used to estimate the model performance.

3. Data Repository for the Bridge Condition Assessment

3.1. Introduction on the Regional Network of Bridges with Inspection Reports. As mentioned above, the data repository is established with inspection reports of a regional network of highway bridges located in Jiangsu Province, China. A total of 263 bridges in different health conditions are collected, which are reinforced concrete or prestressed concrete bridges. The construction of these bridges ranges from the 1980s to the 2000s, and they are mostly in satisfactory condition. The spatial distributions of the inspected bridges on the highway network are illustrated in Figure 2, in which we can find that they are located on 8 different highways, say, highways G2, G15, G25, G2501, G36, G40, G205, and S102, where G and S represent national and provincial highway, respectively. They almost cover the whole area of Jiangsu Province, and are representative of regional bridges.

The variations of bridge types and condition ratings are further illustrated in Figure 3. Figure 3(a) shows the distribution of the bridges with different span lengths, i.e., super long-span bridges (>150 m), long-span bridges (40–150 m), medium-span bridges (20–40 m), small-span bridges (5–20 m), and culverts (<5 m). Among medium-span bridges and culverts take the most and fewest percentage of 50.57% and 6.46%, respectively. The distribution of bridges with condition ratings is displayed in Figure 3(b) according to the results of the inspection reports. Among the 263 bridges, 164, 82, 13, and 4 bridges were labelled as Levels A, B, C, and D, respectively. Note that damage of bridges becomes severe from Levels A to D, which requires different maintenance actions. The dataset covers the major bridge types and condition ratings, which can potentially contribute to predicting the condition rating for a new bridge.

For an illustration of the bridge inspection reports, an example is shown herein, and the textual descriptions are used as input in the machine learning model. The contents of the inspection report are shown in Figure 4 which is translated into English. Chapter 1 usually describes the overview of the inspected bridges, such as the location and type. It is followed by two chapters showing the purpose and basis of the inspection and inspection instruments and components, respectively. Chapter 4 introduces the inspection content, method, and condition rating procedures. Chapter 5 provides the statistical analysis of the inspection results, while Chapter 6 is focused on analysis and maintenance suggestions about the main typical defects. The conclusion is given, and the annexe shows the summary of information on the bridges and defects. Figure 5 illustrates an example of the translated description of defects in the expansion joints extracted from Chapter 6 of the inspection report. The main defects are described, and the causes are then explained. Suggestions are given finally for treatment of the defects.
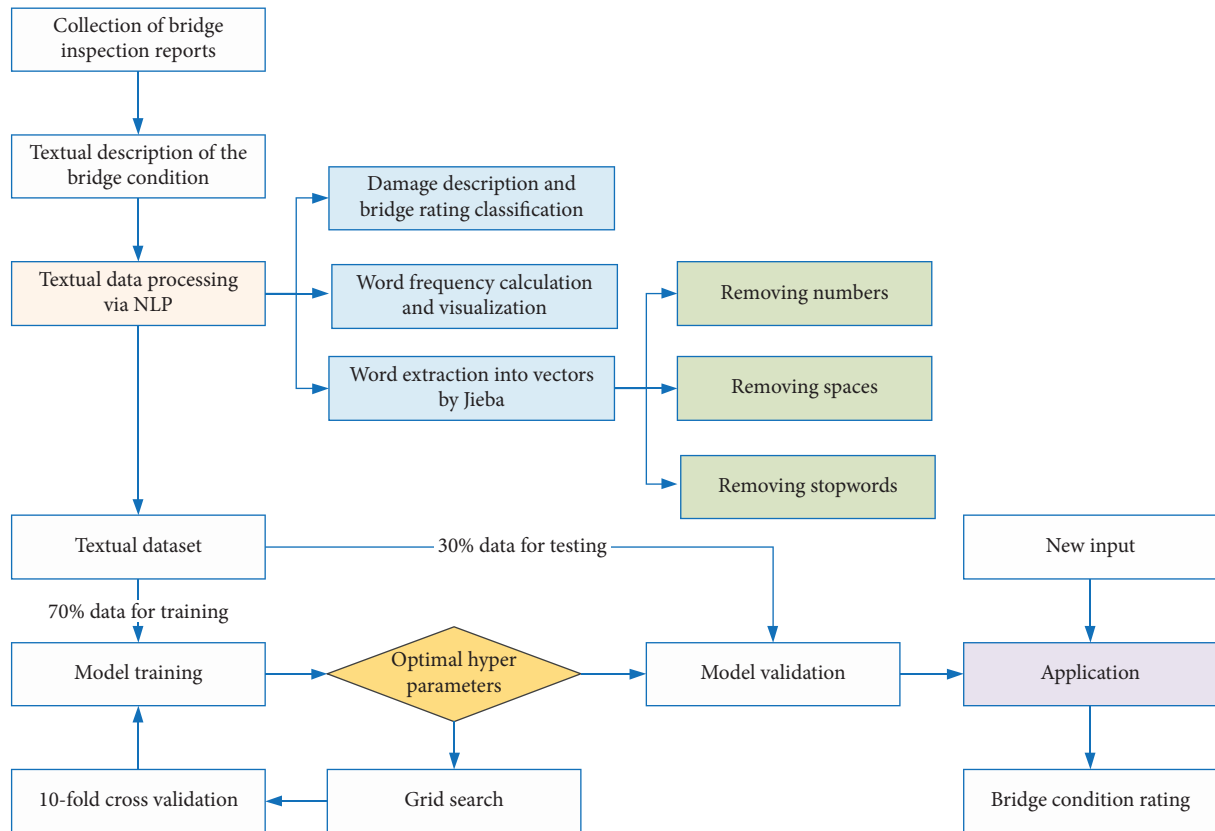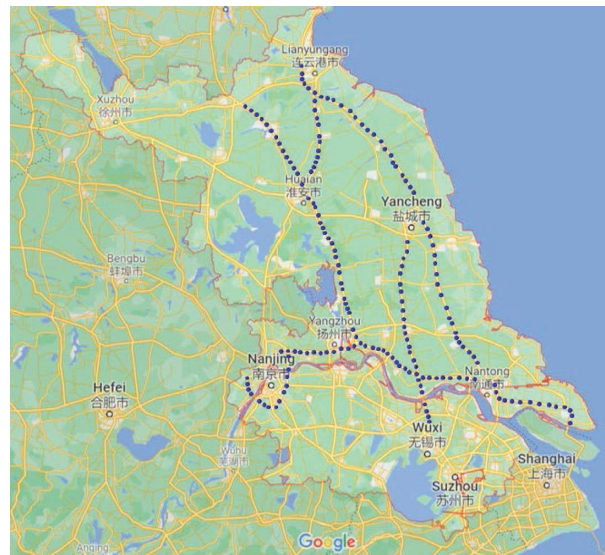
FIGURE 1: Framework of the proposed method.

FIGURE 2: Distribution of the collected bridges in Jiangsu Province, China.

3.2. Condition Rating Rules of the Bridges. The inspection reports were completed after the field inspection of the bridges visually or with nondestructive devices, and then the condition rating of bridges is conducted by experienced engineers through information processing of the textual descriptions of bridge member damages. The engineers should have professional knowledge about the bridges and be familiar with relevant inspection and maintenance standards. Note that the bridge condition rating contains different levels in different counties; for example, 10, 4, and 5 rating levels are adopted in the United States, Japan, and China, respectively. In this paper, the condition rating was conducted following the Chinese code for maintenance of highway bridges [51], where the bridge health condition is
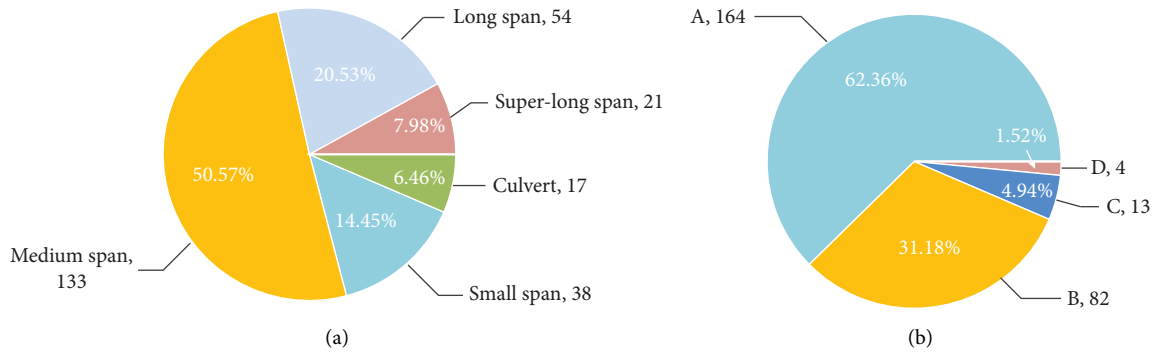
FIGURE 3: Information about the bridges in the data repository. (a) Distribution of the bridge types. (b) Distribution of the bridge condition ratings.

Contents

FIGURE 4: Contents of the inspection report.

6.4.1 The expansion joint rubber strip damage and section steel rupture

During the inspection, it was found that the rubber strips of the expansion joints of the xxxx Bridge were damaged, and the section steel was ruptured.

Cause analysis: 1) The rubber strips between the expansion joints are aging, deformed, and defective; 2) The expansion joints have abnormal expansion and contraction under vehicle and environmental loads, causing the rubber strips to fall off and damage; 3) The anchoring connection of the expansion joints is damaged under the repetitive impact of the vehicles.

Suggestions: 1) Replace the rubber strip; 2) Replace the expansion joint section steel, and re-weld.

FIGURE 5: Example of description of the damage in expansion joints.

divided into five levels, i.e., Levels 1 to 5. The bridges of Level 1 are in good condition, bridges of Level 2 require only minor repairs or daily maintenance to ensure their safety, bridges of Level 3 are in a relatively poor state with many defects, and a small number of functional diseases have taken place, which need to be repaired accordingly. In bridges of Levels 4 and 5, a lot of severe material defects have appeared and the structural functionality greatly deteriorated, thus the safety of these bridges is questionable and the traffic should be closed for special inspection and major repairs. Considering that only a limited number of inspected bridges are rated as Levels 4 and 5 in the dataset so they are combined together and finally four levels are designed as the condition ratings of the bridges, say, Levels A, B, C, and D, representing the damage level of the bridge, increasing from light to severe. A detailed description of the condition rating of bridges in four condition levels is presented in Table 1.

According to the current maintenance standard of highway bridges [51], the detailed condition rating rule is shortly described below. Note that no cable-supported bridges are included in the 263 bridges, and the rating method is applicable. The condition rating of a bridge is defined as follows:

$$D_r = 100 - \sum_{i=1}^{n} \frac{R_i W_i}{5}, \qquad (1)$$

where D_r is the final rating score of the bridge, ranging from 0 to 100; R_i is the rating of each member ranging from 0–5; W_i is the weighting coefficient of each member, and satisfies $\sum_{i}^{n} W_i = 100$; and n is the total number of the components. According to the rating score, the condition is classified into the above-mentioned four levels as follows:

(1) Level A: $D_r \geq 88$

(2) Level B: $88 > D_r \geq 60$

(3) Level C: $60 > D_r \geq 40$

(4) Level D: $D_r < 40$

The weighting coefficient W_i of each member is shown in Table 2. It can be observed that different members in the bridge have different weights due to the load-carrying mechanisms. The substructures such as abutment, pier, and foundation have the highest weighting coefficient, followed by the main and ordinary load-bearing members in superstructures. In addition, the rating criterion of each member R_i is also given in Table 3, which is determined according to the defect condition that is evaluated by experienced engineers. The specific rule actually involves three steps. First, the rating of the member is determined according to the observed defects and is rated as 0, 1, and 2 based on the degree of the defects (from small to big and/or few to many and/or light to severe); then the rating is further modified considering the effect of the member defects on the structural function, which is divided into three levels, say, no, or unimportant effect, small or secondary effect, and big or important effect. After the combination of the two steps, the rating is again modified to reflect the time-dependent

evolution and development of the defects. If the defects are stable, the rating score will be minus 1; if the defects develop fast, then the score will be increased by 1; otherwise, if the defects develop slowly, the score keeps the same. Finally, the score of the member ranges from 0 to 5, representing the member in perfect, good, fairly, poor, and dangerous conditions, respectively. In summary, this rating procedure is very complicated and requires lots of prior professional knowledge to complete this task, thus, it will be time-consuming and labor-intensive.

4. NLP-Based Textual Data Processing and Information Extraction

4.1. NLP-Based Processing of the Textual Data. The bridge reports in the dataset are written in human language, so they are usually very long to include as much information as possible. Only experienced engineers can extract key information from the report. In order to develop fast and automatic predictive models for condition rating, a reliable information extraction method is required to process a large amount of textual description data in the reports. Herein, we adopted NLP to conduct this task, in which NLP refers to transforming human language into machine language that computers can process.

The NLP process involves tasks such as automatic information retrieval, text and speech processing, and language translation. NLP methods fall mainly into three categories, which are rule-based, statistics-based, and machine learning-based. NLP in early times are usually based on hand-crafted rules in information extraction from a large amount of textual data [52]. The common rules are usually related to syntax or semantics, which refer to grammar and meaning, respectively. Statistics-based NLP methods utilize probability on sequences of words, which usually uses TF and TF-IDF as measures in the framework. ML-based NLP makes use of algorithms that learn from input textual data and focus on the most common patterns in the data.

In this paper, NLP is used to analyze the lexical and grammatical structure on the basis of text in the reports and then convert the whole text into discretized word vectors. To this end, the following preprocessing procedures are conducted on the text data in the reports: (1) removing punctuation marks such as periods, commas, and brackets; (2) removing excessive space; and (3) removing stop words such as articles, conjunctions, pronouns, and prepositions. By removal of these terms, little information is lost, and more focus can be given to the important words. In addition, the size of the dataset can be reduced which saves much training time in the next step. A few typical examples are demonstrated in Table 4.

It is observed that the stop words such as "the," "is," "and," and "in" are removed with the processing procedure, and it can avoid the mistaken separation of meaningful descriptions. For example, the word vectors extracted by NLP are "superstructure," "box girder," "diaphragm," "vertical crack," "diagonal crack," "wet joints," "transversely cracked, and "efflorescence" in condition Level B, which can

TABLE 1: The condition of bridges in the four rating levels.

Level	Overall condition	Key members condition	Secondary members condition
A	Perfect	Good functionality and good material property	Good functionality, minor damage in material (less than 3%)
B	Good	Good functionality, minor damage in material (less than 3%), crack less than threshold	Some members with moderate damage in material (less than 10%)
C	Fairly poor	Some members with moderate damage in material (less than 10%), crack greater than threshold, minor functional damage	Many members with severe damage (10 to 20%), degradation in function
D	Poor	Many members with severe damage in material (10 to 20%), crack greater than threshold, severe weathering, peeling, spalling and corrosion, moderate functional damage	More than 20% members with severe damage, loss of function

TABLE 2: Weighting coefficient W_i of each member.

Unit	Member	Weighting coefficient
Superstructure	Main load-bearing member	20
	Ordinary load-bearing member	5
	Bridge deck system	11
	Bearing	3
Substructure	Abutment and foundation	23
	Pier and foundation	24
	Foundation scour	8
	Regulating structures	3
	Wing wall	1
	Slope	1
Others	Others	1

correctly represent the information in the original textual descriptions.

4.2. Word Vectorization and Frequency Analysis. After preprocessing of the text datasets, the textual descriptions are converted to word vectors that can be directly fed into the ML algorithms, and also, the word frequency analysis can be conducted to obtain the frequently used words. Here, two measures are adopted to reflect the importance of words in the textual data of inspection reports, i.e., TF and TF-IDF. Term frequency (TF) is defined as the frequency of a word that occurs in the text, which reflects its importance and is expressed as

$$TF(w, d) = \frac{count(w, d)}{size(d)}, \quad (2)$$

where $count(w, d)$ represents the times of a word w occurs in a text d; and $size(d)$ represents the total number of words in text d.

It can be seen from equation (2) that a word will have a large TF value if it often occurs in the report. However, some words that are not so relevant to the defects of the member will also have a large TF value, e.g., the words about the member types will also have a high TF value. To solve this problem, a new index, inverse document frequency (IDF), is proposed to measure the informativeness of a word w, i.e.,

$$IDF(w, D) = \log \frac{N}{df(w, D) + 1}, \quad (3)$$

where N represents the total number of documents in the data repository; and df represents the total number of a word in the document set D.

Evidently, the more common a word is in the document, the larger the denominator, the smaller the IDF, and the closer it gets to zero. Then, the index TF-IDF is the multiplicative value of TF and IDF as follows:

$$TF - IDF(w, d, D) = TF(w, d) \cdot IDF(w, D). \quad (4)$$

It can be seen that TF-IDF is proportional to the number of occurrences of a word in the document and inversely proportional to the number of occurrences of the word in the entire language. Therefore, TF-IDF tends to filter out common words and retain important ones. Compared with TF, which sometimes is not enough to measure the importance of a word in the report, TF-IDF has the advantage of being simple, fast, and easy to understand.

After the vectorization of the bridge reports and calculation of the TF and TF-IDF measures, we can obtain the word cloud, where the key words are extracted to demonstrate their importance. The larger the words shown in the word cloud, the higher TF they have. Figure 6 gives the word clouds for the four condition Levels A, B, C, and D to make a direct comparison between the key words in the textual data of the four different levels. It is illustrated that the frequent words are highly related to the health condition of the inspected bridges regarding the existing deficiencies, which means the information on bridge deficiencies can be efficiently captured and visualized with the NLP. Note that though the bridge inspection reports were in Chinese, the proposed approach is also applicable to inspection reports in other languages.

In addition, the first 10 words in the four condition levels ranked according to their TF values and TF-IDF values are also analyzed herein, which are listed in Tables 5–8. It is noticed that in bridges with condition Level A (Table 5), the bridge component names such as "deck," "superstructure" and "substructure" appear the most, with TF values of 1.45, 0.99, and 0.99, respectively. The top TF-IDF values of 0.426 and 0.413 belong to "inspection" and "deck." In comparison, words describing the bridge deficiency appear less frequently, and "crack" has TF and TF-IDF values of 0.98 and 0.244, respectively. However, in bridges with condition Level B (Table 6), words that describe the bridge deficiency become more in the data, e.g., "crack" and "damage", respectively, have TF values of 3.05 and 2.83, which are also the top two most frequent words. The top three words ranking following the TF-IDF values are "rebar rusting," "damage," and "crack," whose TF-IDF are 0.427, 0.413, and 0.409, respectively.

In Table 7, the two most frequent words are "damage" and "crack," whose TF values are 3.08 and 2.85 and TF-IDF values are 0.421 and 0.358, respectively. It is found that the frequency of words describing deficiencies obviously increased in Levels B and C compared with Level A. Furthermore, for bridges with Level D (Table 8), the TF values for "cracking" and "crack" reach 8.75 and 8.00, and the corresponding TF-IDF values are 0.567 and 0.487. The TF and TF-IDF values for other words describing deficiencies are also very high. For example, the TF values are 6.50, 5.75, and 5.00 for "rebar rusting," "damage," "exposed rebar," etc.

In general, it is observed that words of bridge member names are most frequent in Level A, while words describing bridge defects such as "crack" and "damage" appear fewer times. In comparison, the words of bridge deficiency are more frequent in bridges of the other three condition levels. The frequent words describing defects observed in Level D clearly stand out. However, Levels B and C can hardly be separated due to the close frequency of words representing defects with TF or TF-IDF. The tabulated word frequencies

TABLE 3: Rating criterion R_i of each member in the bridge.

Defect status and scale					Combined rating		
Defect degree and scale		Degree			Small ⟶ big Few ⟶ many Light ⟶ severe		
		Scale			0	1	2
Effect on the structural function	No, unimportant	0			0	1	2
	Small, secondary	+1			1	2	3
	Big, important	+2			2	3	4
Combination of the above two evaluations		—	0	1	2	3	4
Modification on the defect development	Stable	−1	—	0	1	2	3
	Develop slowly	0	0	1	2	3	4
	Develop fast	+1	1	2	3	4	5
Final condition rating result			1	2	3	4	5

offer a comprehensive overview of the damages across all bridges at each level. It presents an integrated perspective on the words associated with the condition of the bridges. However, it cannot aid in identifying the damage level of each specific bridge. Therefore, the TF-IDF of the inspection report for each bridge is utilized as the input in the machine learning model for the prediction of the condition rating of the bridges.

5. Adopted Machine Learning Techniques

Four commonly used algorithms, i.e., DT, SVM, RF, and GBRT, are employed to build up accurate and reliable prediction models for the condition rating of bridges. A brief introduction of the backgrounds is first presented herein.

5.1. Decision Tress (DT). DT is a tree-structured model that can be used for both classification and regression problems [53]. A DT is generally made up of tree nodes and directed edges. Generally, there are three kinds of nodes in a DT, i.e., a root node, several internal nodes, and several leaf nodes. The decision process of the model starts from the root node, where a certain feature of the sample will be tested and split to the child internal nodes according to the results. The sample is tested and split recursively until it reaches the leaf node, which is taken as the final decision result. The path from the root node to each leaf node corresponds to a decision path. According to the split criterion, different algorithms for building DT can be used. Here, the classification and regression tree (CART) algorithm which is based on the mean square error is adopted. Supposing the input space can be divided into N subspaces $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N$, the DT model can be expressed as

$$\hat{y} = f_{DT}(\mathbf{x}),$$
$$= \sum_{n=1}^{N} c_n \mathbf{I}(\mathbf{x} \in \mathbf{R}_n), \tag{5}$$

where $\mathbf{x}$ is the input variables; $\mathbf{I}(\cdot)$ is the indicator function; and c_n is the prediction value by the nth leaf node.

Moreover, to avoid the over-fitting issue, tree pruning should be also carried out using some specific loss function to make the tree simple and robust. The major model parameters are the maximum depth of the tree; maximum leaf nodes; minimum samples for split; minimum samples of leaf node; etc.

5.2. Support Vector Machine (SVM). Support vector machine (SVM) is a kind of supervised ML method proposed by Vapnik with colleagues at AT&T Bell Laboratories [54]. In essence, the input variables are mapped to a high-dimensional feature space by nonlinear transformation, and then a hyperplane is found in the space for linear classification or regression. This mapping process generally adopts the kernel function method. The specific learning strategy of SVM is to maximize the left and right data interval of the hyperplane, which can be abstracted into the mathematical problem of solving convex quadratic programming, and the learning algorithm of SVM is the optimization algorithm of solving convex quadratic programming. The main parameters are parameters of kernel function and penalty coefficient in optimization. Taking the binary classification problem as an example, the SVM model can be defined as

$$\hat{y} = f_{SVM}(\mathbf{x}),$$
$$= \beta^T \mathbf{x} - b. \tag{6}$$

where β and b denote the normal vector and the bias constant from zero, respectively.

5.3. Random Forest (RF). The DT and SVM, or some other supervised learning methods, are actually individual-type algorithms, which only generate a single predictive model. In ML, there are also ensemble learning methods that can generate several predictive models and then combine them into one according to some rules, which has been proven to be obviously superior to individual learning methods. RF is one typical ensemble method that belongs to the bagging family [55]. It takes advantage of two powerful techniques,

TABLE 4: Examples of textual data to word vectors by NLP.

Conditions	Textual description	Word vectors by NLP
Level A	The accumulated mud is mainly found in expansion joints in bridge deck systems and auxiliary structures	("Bridge deck system," "auxiliary structures," "expansion joint," "accumulated mud")
Level B	Vertical cracks and diagonal cracks are found on the diaphragm of small box girder in the superstructure. Wet joints were transversely cracked and efflorescence is found	("Superstructure," "box girder," "diaphragm," "vertical crack," "diagonal crack," "wet joints," "transversely cracked," "efflorescence")
Level C	In the superstructure inspection, transverse cracks and efflorescence were found in the bottom plate of the slab. The main reinforcement was exposed and corroded seriously. Large areas of stirrups were exposed and corroded. The reinforcement bars were exposed and corroded seriously, and some stirrups were broken	("Superstructure," "slab," "bottom plate," "transverse crack," "efflorescence," "main reinforcement," "expose," "corroded," "seriously," "large area," "reinforcement bars," "partly," "broken")
Level D	Garbage accumulation at the top of the abutment, vertical cracks, water invasion, vertical cracks of the pier, excessive shear deformation, and cracking of the bearing, poor bolt anchorage of the pot bearing are found in the substructure	("Substructure," "abutment," "top," "garbage," "accumulation," "vertical cracks," "water invasion," "pier," "vertical cracks," "bearing," "shear deformation," "excessive," "crack," "pot bearing," "bolt," "poor anchorage")

Figure 6: Word clouds for the four condition levels of the bridges. (a) Level A. (b) Level B. (c) Level C. (d) Level D.

Table 5: Word frequency and TF-IDF values for the top 10 frequent words in bridges with condition Level A.

Words	Frequency	TF	Words	TF-IDF
Deck	237	1.45	Inspection	0.426
Superstructure	163	0.99	Deck	0.413
Substructure	163	0.99	Found	0.378
Crack	160	0.98	Substructure	0.365
Cracking	158	0.96	Expansion joints	0.307
Damage	148	0.90	Cracking	0.255
Expansion joints	143	0.87	Damage	0.245
Deck bottom plate	113	0.69	Crack	0.244
Longitudinal	111	0.68	Superstructure	0.231
Lateral	108	0.66	Deck bottom plate	0.203

Table 6: Word frequency and TF-IDF values for the top 10 frequent words in bridges with condition Level B.

Words	Frequency	TF	Words	TF-IDF
Crack	250	3.05	Rebar rusting	0.427
Damage	232	2.83	Damage	0.413
Rebar rusting	207	2.52	Crack	0.409
Concrete	167	2.04	Exposed rebar	0.351
Cracking	159	1.94	Cracking	0.276
Exposed rebar	152	1.85	Concrete	0.258
Deck	128	1.56	Deck	0.240
Lateral	111	1.35	Vertical	0.239
Vertical	104	1.27	Found	0.214
Longitudinal	90	1.10	Expansion joints	0.201

Table 7: Word frequency and TF-IDF values for the top 10 frequent words in bridges with condition Level C.

Words	Frequency	TF	Words	TF-IDF
Damage	40	3.08	Damage	0.421
Crack	37	2.85	Crack	0.358
Deck bottom plate	31	2.38	Slab bridge	0.354
Slab bridge	26	2.00	Deck bottom plate	0.354
Cracking	25	1.92	Cast-in-place	0.299
Deck	24	1.85	Exposed rebar	0.272
Lateral	23	1.77	Deck	0.265
Cast-in-place	21	1.62	Cracking	0.256
Concrete	20	1.54	Expansion joints	0.231
Exposed rebar	20	1.54	Lateral	0.212

Table 8: Word frequency and TF-IDF values for the top 10 frequent words in bridges with condition Level D.

Words	Frequency	TF	Words	TF-IDF
Cracking	35	8.75	Cracking	0.567
Crack	32	8.00	Rebar rusting	0.498
Rebar rusting	26	6.50	Crack	0.487
Damage	23	5.75	Exposed rebar	0.428
Exposed rebar	20	5.00	Vertical	0.386
Vertical	18	4.50	Damage	0.381
Efflorescence	16	4.00	Efflorescence	0.342
Lateral	15	3.75	Lateral	0.218
Longitudinal	9	2.25	Box girder	0.171
Spalling	8	2.00	Longitudinal	0.137

i.e., bootstrap and random feature selection. At every step of building a tree, N samples are randomly selected as a subset with bootstrap from the entire database, and K features are randomly selected and used as the split node to build a DT with the subset. After several (or M) steps, one obtains M DTs from the above procedure, and the final output is the average of all the outputs by the M trees as follows:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^{M} f_{\text{DT},i}(\mathbf{x}), \quad (7)$$

where $f_{\text{DT},i}(\cdot)$ is the weak learner by the DT method at step $i \leq M$. In addition to the original parameters of DT, the number of trees is an extra yet important parameter.

5.4. Gradient Boosting Regression Tree (GBRT). GBRT is another family of ensemble learning method [56], which is also developed based on the DT method. Unlike RF which builds single predictive models in parallel using Bootstrap, it is a boosting family which generates a single predictive model in sequence according to the model performance at

every step. The initial database is used to train a single model, and then each sample in the database will get a weight according to the prediction accuracy of this model. The weighted new dataset will be used to train a new model in the next step. After several steps, one attains several single models and each model will have a weight computed based on its loss function. The final result is the weighted sum of these models, i.e.,

$$\widehat{y} = \sum_{i=1}^{M} \alpha_i f_{\mathrm{DT},i}(\mathbf{x})\gamma, \tag{8}$$

where α_i is the weight of the weak learner at step $i \leq M$. Similar to RF, a major parameter for GBRT is the number of trees. In addition, a learning rate $\gamma \in [0, 1]$ is usually employed to overcome the overfitting issue.

6. Implementation and Application of the Proposed Method

To implement the state-of-the-art machine learning models, i.e., the single learner DT, SVM, and the ensemble learner RF, GBRT, the whole database is firstly divided into 70% training and 30% testing sets. Then the hyperparameters are determined for the learners. Finally, the models are adopted on the testing sets to predict the condition rating of the bridges.

6.1. Tuning of Hyperparameters. In the machine learning models, hyperparameters should be specified in advance. In single learners like DT, key parameters include the maximum depth of the tree, the minimum samples of the leaf node, and the minimum samples for the split. In SVM, one important parameter is kernel, and the radial basis function (rbf) is adopted. Another two parameters are C and gamma, which are measures of error and curvature of the decision boundary. Similarly, the hyperparameters are also required in the ensemble learners, i.e., RF and GBRT, such as the number of estimators, the maximum depth of the trees, the minimum samples of leaf node, and the learning rate.

Inappropriate selection of hyperparameters will lead to problems of underfitting or overfitting. Therefore, a strategy is adopted as follows. Firstly, a range of parameter values are selected based on previous experience and the grid search method is then used to determine the value in an iterative manner. 10-fold cross-validation is carried out to overcome the bias induced in random sampling of the training set [57, 58]. With this strategy, the optimal hyperparameters are obtained as shown in Table 9.

Moreover, the learning curves of the four ML models are provided in Figure 7 to consider the effects of the size of the training data. For the training sets, the accuracy is low at the beginning, but increases quickly and becomes stable in DT learner. In comparison, the other three learners share a similar trend. The accuracy starts as high as around 1.0 but decreases gradually before stabilizing. This is because overfitting usually occurs when training sets are very few. Then the accuracy becomes stable with the increase of

TABLE 9: Final hyperparameters for the ML models.

Model	Parameter	Values
DT	Maximum depth of the tree	3
	Minimum samples of leaf node	5
	Minimum samples for split	5
SVM	Kernel	rbf
	C	100
	Gamma	0.01
RF	Number of estimators	110
	Maximum depth of the tree	4
	Minimum samples of leaf node	1
GBRT	Maximum depth of the tree	3
	Learning rate	0.01
	Number of estimators	70

training data. On the other hand, the accuracy for the testing sets all starts at low values, while it improves with the number of training sets. It can be observed that the model accuracy becomes stable when the training samples reach around 180 (around 70% of the whole datasets). This verifies that the split of the whole datasets into 70% and 30% is reasonable.

6.2. Prediction Results and Discussion. After choosing the hyperparameters, the prediction is carried out. The performance of the different machine learning models is evaluated using a confusion matrix. The confusion matrix is a plot of actual versus predicted classes, in which an element C_{ij} is the number of actual samples known to be in class i, but predicted as class j. It is hence clear that the diagonal and off-diagonal elements represent the number of correctly and incorrectly classified samples, respectively.

Four common measures of model performance are used for the evaluation of the prediction results. The accuracy of a machine learning model is defined as the ratio between the number of correctly classified samples and the total number of samples. Precision is defined as the percentage of predicted samples that are correctly classified by the learning model, while recall rate is defined as the percentage of the actual samples that are correctly predicted. The higher precision and recall rate demonstrate the better prediction performance of the machine learning algorithm. When precision and recall indicate the different performance of models, it is difficult to compare, so the $F1$-score is a measure of a model's accuracy calculated with the harmonic mean of precision and recall as follows:

$$F1 = \frac{2 \cdot \mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}. \tag{9}$$

The confusion matrix for training and testing sets with the four machine learning algorithms is shown in Figure 8. It can be seen that the DT, SVM, RF, and GBRT models have an accuracy of 82%, 84%, 81%, and 89% on the testing set, respectively. They all display acceptable prediction accuracy. The model performances on Levels A and B are usually better than Levels C and D, which is due to the class imbalance, i.e., there are fewer data in categories Level C and
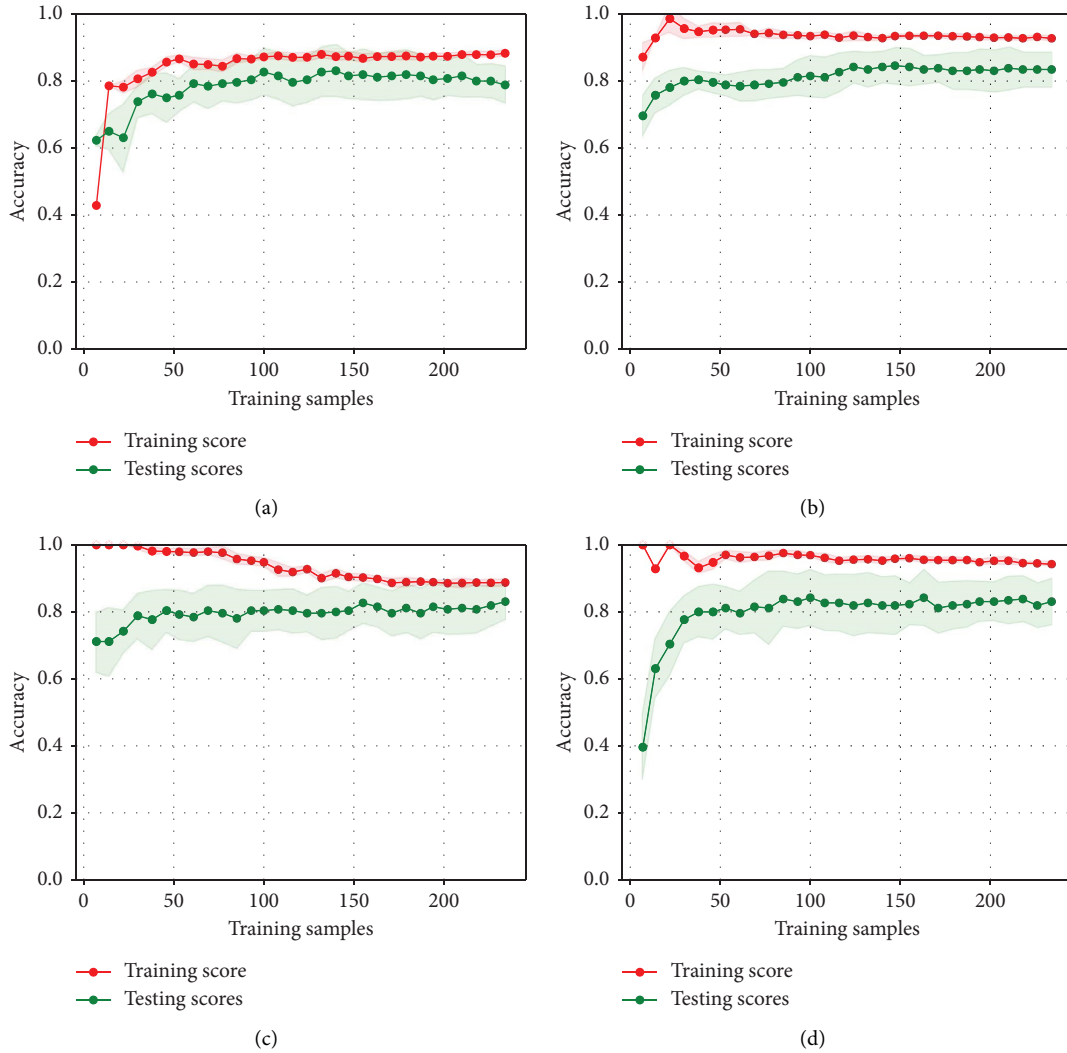
FIGURE 7: Learning curves for different ML methods (a) DT. (b) SVM. (c) RF. (d) GBRT.

D. This issue was shown in previous studies [59] to impair the classification performance. Though approaches were investigated in machine learning fields, the fundamental solution is to incorporate more inspection data on the more deficient (Levels C and D) bridges.

The results of the prediction with the four machine learning models are summarized in Table 10. On condition rating Level A, the $F1$-score of the four models are 0.88, 0.89, 0.88, and 0.93, while on Level B, the results are 0.78, 0.78, 0.76, and 0.88. Similarly, for Level C samples, $F1$-scores are 0.75, 0.67, N/A, and 0.40 for the four models, respectively. The results indicate that all models have the best performance in category Level A, which is followed by Level B and Level C. This is due to the difference in the number of training sets. It is obvious that more training data can generate better prediction performance. There are only 9 and 2 training sets in condition rating Levels C and D,

respectively, and the prediction on the testing sets is poor compared with Levels A and B. Besides, GBRT demonstrates the highest $F1$-score in different rating levels, which clearly outperforms other models. It should be noted that the prediction results are very instructive, especially for future work with an emphasis on the collection of inspection reports of bridges in more deficient levels.

Previous studies were usually devoted to the prediction of the health conditions of bridge components such as concrete slabs [60], concrete decks [40], bridge abutment walls [41], and the superstructure and substructure [42]. While in this study, the condition rating of the bridge structure is carried out to provide comprehensive guidance for stakeholders of bridges. Furthermore, a prediction accuracy of 89% is demonstrated with the case study of 263 highway bridges, showing the applicability to more bridges.
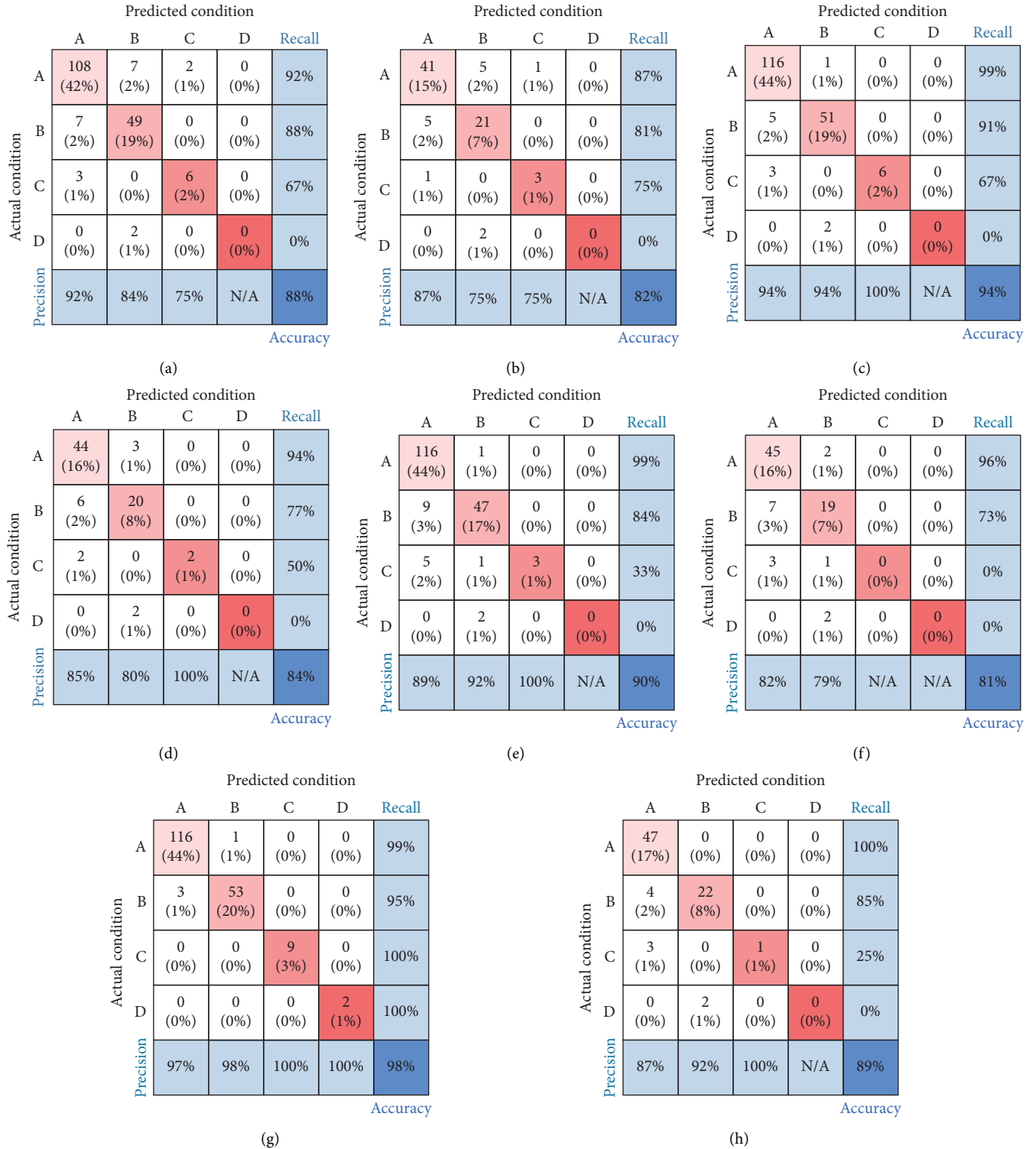
FIGURE 8: Confusion matrix for training and testing sets with different ML methods. (a) Training dataset: DT. (b) Testing dataset: DT. (c) Training dataset: SVM. (d) Testing dataset: SVM. (e) Training dataset: RF. (f) Testing dataset: RF. (g) Training dataset: GBRT. (h) Testing dataset: GBRT.

TABLE 10: Performance measures for condition assessment ML models of the testing set.

Measure		Model			
		DT	SVM	RF	GBRT
Precision	A	0.87	0.85	0.82	0.87
	B	0.75	0.80	0.79	0.92
	C	0.75	1.00	—	1.00
	D	—	—	—	—
	Average	0.79	0.88	0.90	0.93
Recall	A	0.89	0.94	0.96	1.00
	B	0.81	0.77	0.73	0.85
	C	0.75	0.50	0.00	0.25
	D	0.00	0.00	0.00	0.00
	Average	0.61	0.55	0.42	0.53
F1-score	A	0.88	0.89	0.88	0.93
	B	0.78	0.78	0.76	0.88
	C	0.75	0.67	—	0.40
	D	—	—	—	—
	Average	—	—	—	—
Accuracy		0.82	0.84	0.81	0.89

7. Conclusions

Condition rating of existing bridges is an essential indicator of bridge health, which is closely connected with potential maintenance procedures. As professional background in inspection practice and familiarity with the relevant maintenance standards are involved in the condition rating process, the traditional human-based assessment is often time-consuming and laborious.

This paper explores a rapid condition rating method using a machine learning model based on textual inspection data of bridges. The NLP technique is used to extract the information in the text description of the bridge inspection reports and convert them into digital vectors. The methodology is verified with the condition rating of 263 highway bridges of different types and health conditions in Jiangsu, China. The main contributions in this study are summarized as follows:

(1) Automatic information extraction is achieved using an NLP-based technique, i.e., word vectorization, which can capture the most relevant features of the existing bridge deficiencies buried in the enormous and complicated textual data in inspection reports. This has been clearly demonstrated with the word clouds of the bridges in four condition rating levels.

(2) Based on knowledge about bridge deficiencies and inspection practice, TF and TF-IDF can illustrate the hidden influence of important words in the condition rating of bridges. With the measures of the top 10 frequent words, the abundant textual data in inspection reports are adaptively correlated with the condition rating of the bridges. This "explainable" nature of the approach can provide guidance for inspectors to improve efficiency in the inspection practice.

(3) The ML-based approach utilizes the valuable while nonstandard textual information in the bridge inspection report, to achieve efficient and accurate bridge condition rating prediction. This is significant as the details of component-level defects are intelligently exploited to realize bridge-level assessment.

(4) With state-of-the-art machine learning models, a close agreement is obtained between the prediction and the actual condition ratings. The results indicate that the most accurate model in this study, i.e., GBRT, can reach an accuracy of 89% on the testing sets.

The proposed method saves much time and labor in condition rating and can significantly complement the traditional human-based assessment. It is also worth noting that the proposed method uses words and frequencies as input and can be applied to inspection reports in different languages. A limitation is the imbalanced dataset with fewer samples in class such as Level D, which may affect the performance of the trained model. The future study will focus on the collection of bridges with severe damage conditions. With rapid and reliable condition ratings, effective maintenance actions can be more wisely focused on bridges with poor health conditions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The conclusions and opinions in this paper are of the authors and do not necessarily reflect those of the mentioned institution.

Conflicts of Interest

The authors declare that there no conflicts of interest.

Acknowledgments

References

[1] Y. Zhou, Y. Xia, Z. Sun, and Y. Fujino, "Analytical formulation of the temperature-induced deformation of multispan suspension bridges," *Structural Control and Health Monitoring*, vol. 29, no. 6, p. 2937, 2022.

[2] X. W. Ye, Z. Sun, and J. Lu, "Prediction and early warning of wind-induced girder and tower vibration in cable-stayed

bridges with machine learning-based approach," *Engineering Structures*, vol. 275, Article ID 115261, 2023.

[3] G. Miluccio, D. Losanno, F. Parisi, and E. Cosenza, "Traffic-load fragility models for prestressed concrete girder decks of existing Italian highway bridges," *Engineering Structures*, vol. 249, Article ID 113367, 2021.

[4] S. Z. Chen, D. C. Feng, and Z. Sun, "Reliability-based vehicle weight limit determination for urban bridge network subjected to stochastic traffic flow considering vehicle-bridge coupling," *Engineering Structures*, vol. 247, Article ID 113166, 2021.

[5] G. Miluccio, D. Losanno, F. Parisi, and E. Cosenza, "Fragility analysis of existing prestressed concrete bridges under traffic loads according to new Italian guidelines," *Structural Concrete*, vol. 24, no. 1, pp. 1053–1069, 2022.

[6] D. M. Siringoringo and Y. Fujino, "Seismic response of a suspension bridge: insights from long-term full-scale seismic monitoring system," *Structural Control and Health Monitoring*, vol. 25, no. 11, p. 2252, 2018.

[7] K. Wardhana and F. C. Hadipriono, "Analysis of recent bridge failures in the United States," *Journal of Performance of Constructed Facilities*, vol. 17, no. 3, pp. 144–150, 2003.

[8] F. Y. Xu, M. J. Zhang, L. Wang, and J. R. Zhang, "Recent highway bridge collapses in China: review and discussion," *Journal of Performance of Constructed Facilities*, vol. 30, no. 5, Article ID 4016030, 2016.

[9] D. M. Frangopol, J. S. Kong, and E. S. Gharaibeh, "Reliability-based life-cycle management of highway bridges," *Journal of Computing in Civil Engineering*, vol. 15, no. 1, pp. 27–34, 2001.

[10] A. Manafpour, I. Guler, A. Radlińska, F. Rajabipour, and G. Warn, "Stochastic analysis and time-based modeling of concrete bridge deck deterioration," *Journal of Bridge Engineering*, vol. 23, no. 9, Article ID 4018066, 2018.

[11] Y. Fujino and D. M. Siringoringo, "Recent research and development programs for infrastructures maintenance, renovation and management in Japan," *Structure and Infrastructure Engineering*, vol. 16, no. 1, pp. 3–25, 2020.

[12] Z. Sun, T. Nagayama, M. Nishio, and Y. Fujino, "Investigation on a curvature-based damage detection method using displacement under moving vehicle," *Structural Control and Health Monitoring*, vol. 25, no. 1, p. 2044, 2018.

[13] Y. Jeong, W. Kim, I. Lee, and J. Lee, "Bridge service life estimation considering inspection reliability," *KSCE Journal of Civil Engineering*, vol. 21, no. 5, pp. 1882–1893, 2017.

[14] K. Dai, B. H. Smith, S. E. Chen, and L. Sun, "Comparative study of bridge management programmes and practices in the USA and China," *Structure and Infrastructure Engineering*, vol. 10, no. 5, pp. 577–588, 2014.

[15] Z. Sun, Z. Zou, and Y. Zhang, "Utilization of structural health monitoring in long-span bridges: case studies," *Structural Control and Health Monitoring*, vol. 24, no. 10, p. 1979, 2017.

[16] C. Wu, P. Wu, J. Wang, R. Jiang, M. Chen, and X. Wang, "Ontological knowledge base for concrete bridge rehabilitation project management," *Automation in Construction*, vol. 121, Article ID 103428, 2021.

[17] R. Helmerich, E. Niederleithinger, D. Algernon, D. Streicher, and H. Wiggenhauser, "Bridge inspection and condition assessment in europe," *Transportation Research Record*, vol. 2044, no. 1, pp. 31–38, 2008.

[18] J. M. Hooks, D. M. Frangopol, and others, *LTBP Bridge Performance Primer Technical report, United States. Federal Highway Administration*, Office of Infrastructure, Washington, DC, USA, 2013.

[19] Z. Sun, D. M. Siringoringo, and Y. Fujino, "Load-carrying capacity evaluation of girder bridge using moving vehicle," *Engineering Structures*, vol. 229, Article ID 111645, 2021.

[20] Y. Bao, Z. Chen, S. Wei, Y. Xu, Z. Tang, and H. Li, "The state of the art of data science and engineering in structural health monitoring," *Engineering*, vol. 5, no. 2, pp. 234–242, 2019.

[21] X. Ye, T. Jin, and C. Yun, "A review on deep learning-based structural health monitoring of civil infrastructures," *Smart Structures and Systems*, vol. 24, no. 5, pp. 567–585, 2019.

[22] M. Saifullah, C. P. Andriotis, K. G. Papakonstantinou, and S. M. Stoffels, "Deep reinforcement learning-based life-cycle management of deteriorating transportation systems," *Bridge Safety, Maintenance, Management, Life-Cycle, Resilience and Sustainability*, pp. 293–301, 2022.

[23] Z. Zhang, C. Sun, R. Bridgelall, and M. Sun, "Application of a machine learning method to evaluate road roughness from connected vehicles," *Journal of Transportation Engineering, Part B: Pavements*, vol. 144, no. 4, Article ID 4018043, 2018.

[24] L. X. Le, D. M. Siringoringo, H. Katsuchi, and Y. Fujino, "Stay cable tension estimation of cable-stayed bridge under limited information on cable properties using artificial neural networks," *Structural Control and Health Monitoring*, vol. 29, no. 10, p. 3015, 2022.

[25] Q. Tang, J. Dang, Y. Cui, X. Wang, and J. Jia, "Machine learning-based fast seismic risk assessment of building structures," *Journal of Earthquake Engineering*, vol. 26, no. 15, pp. 8041–8062, 2021.

[26] S. Ghahari, F. Abazarsa, and E. Taciroglu, "Blind modal identification of non-classically damped structures under non-stationary excitations," *Structural Control and Health Monitoring*, vol. 24, no. 6, p. 1925, 2017.

[27] Y. Li, Y. Ding, H. Zhao, and Z. Sun, "Data-driven structural condition assessment for high-speed railway bridges using multi-band FIR filtering and clustering," *Structures*, vol. 41, pp. 1546–1558, 2022.

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

[29] D. C. Feng and G. Wu, "Interpretable machine learning-based modeling approach for fundamental properties of concrete structures," *Journal of Building Structures*, vol. 43, no. 4, pp. 228–238, 2022.

[30] Y. Deng, M. Zhang, D. M. Feng, and A. Q. Li, "Predicting fatigue damage of highway suspension bridge hangers using weigh-in-motion data and machine learning," *Structure and Infrastructure Engineering*, vol. 17, no. 2, pp. 233–248, 2021.

[31] Y. Zhang, X. Wang, Z. Ding, Y. Du, and Y. Xia, "Anomaly detection of sensor faults and extreme events based on support vector data description," *Structural Control and Health Monitoring*, vol. 29, no. 10, p. 3047, 2022.

[32] S. Z. Chen and D. C. Feng, "Multifidelity approach for data-driven prediction models of structural behaviors with limited data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 12, pp. 1566–1581, 2022.

[33] Z. Sun, J. Santos, and E. Caetano, "Vision and support vector machine–based train classification using weigh-in-motion data," *Journal of Bridge Engineering*, vol. 27, no. 6, Article ID 6022001, 2022.

[34] D. C. Feng, Z. T. Liu, X. D. Wang et al., "Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach," *Construction and Building Materials*, vol. 230, Article ID 117000, 2020.

[35] X. Lei, D. M. Siringoringo, Y. Dong, and Z. Sun, "Interpretable machine learning methods for clarification of load-displacement effects on cable-stayed bridge," *Measurement*, vol. 220, Article ID 113390, 2023.

[36] J. Li, J. Li, H. Hao, R. Wang, and L. Li, "Structural damage quantification using ensemble-based extremely randomised trees and impulse response functions," *Structural Control and Health Monitoring*, vol. 29, no. 10, p. 3033, 2022.

[37] D. C. Feng, B. Cetiner, M. R. Azadi Kakavand, and E. Taciroglu, "Data-driven approach to predict the plastic hinge length of reinforced concrete columns and its application," *Journal of Structural Engineering*, vol. 147, no. 2, Article ID 4020332, 2021.

[38] Z. Sun, J. Santos, and E. Caetano, "Data-driven prediction and interpretation of fatigue damage in a road-rail suspension bridge considering multiple loads," *Structural Control and Health Monitoring*, vol. 29, no. 9, p. 2997, 2022.

[39] Y. Yang, L. Sanchez, H. Zhang et al., "Estimation of full-field, full-order experimental modal model of cable vibration from digital video measurements with physics-guided unsupervised machine learning and computer vision," *Structural Control and Health Monitoring*, vol. 26, no. 6, p. 2358, 2019.

[40] Y. H. Huang, "Artificial neural network model of bridge deterioration," *Journal of Performance of Constructed Facilities*, vol. 24, no. 6, pp. 597–602, 2010.

[41] Z. Li and R. Burgueño, "Using soft computing to analyze inspection results for bridge evaluation and management," *Journal of Bridge Engineering*, vol. 15, no. 4, pp. 430–438, 2010.

[42] H. Liu and Y. Zhang, "Bridge condition rating data modeling using deep learning algorithm," *Structure and Infrastructure Engineering*, vol. 16, no. 10, pp. 1447–1460, 2020.

[43] Y. Xia, X. Lei, P. Wang, and L. Sun, "Artificial intelligence based structural assessment for regional short-and medium-span concrete beam bridges with inspection information," *Remote Sensing*, vol. 13, no. 18, p. 3687, 2021.

[44] X. Lei, Y. Xia, Y. Dong, and L. Sun, "Multi-level time-variant vulnerability assessment of deteriorating bridge networks with structural condition records," *Engineering Structures*, vol. 266, Article ID 114581, 2022.

[45] X. Lei, Y. Xia, S. Komarizadehasl, and L. Sun, "Condition level deteriorations modeling of rc beam bridges with u-net convolutional neural networks," *Structures*, vol. 42, pp. 333–342, 2022.

[46] Y. Xia, X. Lei, P. Wang, and L. Sun, "A data-driven approach for regional bridge condition assessment using inspection reports," *Structural Control and Health Monitoring*, vol. 29, no. 4, p. 2915, 2022.

[47] J. Zhang and N. M. El-Gohary, "Extending building information models semiautomatically using semantic natural language processing techniques," *Journal of Computing in Civil Engineering*, vol. 30, no. 5, p. 16004, 2016.

[48] T. Le and H. David Jeong, "NLP-based approach to semantic classification of heterogeneous transportation asset data terminology," *Journal of Computing in Civil Engineering*, vol. 31, no. 6, Article ID 4017057, 2017.

[49] K. Liu and N. El-Gohary, "Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports," *Automation in Construction*, vol. 81, pp. 313–327, 2017.

[50] S. Mangalathu and H. V. Burton, "Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions," *International Journal of Disaster Risk Reduction*, vol. 36, Article ID 101111, 2019.

[51] JTG H, *Code For Maintenance Of Highway Bridges And Culvert*, China Communications Press, Beijing, China, 2004.

[52] S. Sarawagi, *Information Extraction*, Now Publishers Inc, Hanover, MD, USA, 2008.

[53] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[54] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer science & business media, Berlin, Germany, 1999.

[55] A. Liaw, M. Wiener, and others, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[56] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[57] X. Ye, Y. Ding, and H. Wan, "Machine learning approaches for wind speed forecasting using long-term monitoring data: a comparative study," *Smart Structures and Systems, An International Journal*, vol. 24, no. 6, pp. 733–744, 2019.

[58] D. C. Feng, W. J. Wang, S. Mangalathu, and E. Taciroglu, "Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls," *Journal of Structural Engineering*, vol. 147, no. 11, Article ID 4021173, 2021.

[59] M. Alipour, D. K. Harris, L. E. Barnes, O. E. Ozbulut, and J. Carroll, "Load-capacity rating of bridge populations through machine learning: application of decision trees and random forests," *Journal of Bridge Engineering*, vol. 22, no. 10, Article ID 4017076, 2017.

[60] K. Kawamura, A. Miyamoto, D. M. Frangopol, and R. Kimura, "Performance evaluation of concrete slabs of existing bridges using neural networks," *Engineering Structures*, vol. 25, no. 12, pp. 1455–1477, 2003.