Local governments manage 70% of all bridges in Japan. One-half of the number of bridges will reach their service lifetime in the near future. However, local governments have insufficient financial resources, specialists, and technologies to maintain these bridges. Therefore, it is necessary to develop a practical method for the evaluation of bridge soundness to assess the degree of the structural health of a bridge’s superstructure. In this paper, the maintenance priority evaluation of bridge integrity of small and medium span bridges is examined using the support vector machines (SVM) of artificial intelligence (AI) techniques. Based on the model, an algorithm is proposed as a useful feature to provide the engineering expertise for the inspection of bridges. This proposed method was able to substitute engineer judgement to distinguish the health rating of I. The input data were the degrees of deterioration of the structural parts and the output data were the soundness of the structure. As inspection example, 971 inspection data on bridges in Gifu prefecture were used. The results showed that adding the inspection item of exposed direction of steel bars gives good assessment on whether bridge maintenance works are needed or not.

**Key Words:** visual inspection, support vector machine (SVM), receiver operating characteristic (ROC) curves

1. INTRODUCTION

(1) Background

Bridges are key structural elements in the transportation network. In Japan, there are over 700,000 road bridges. Those in the jurisdictions of local governments account for approximately 70% of the total number of road bridges. A large number of bridges were constructed during the high economic growth period from the mid-1950s to the early 1970s, most of which will reach 50 years of age between 2010 and 2025. This fact has given rise to the issue of safety in bridge infrastructure. To tackle this problem, the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) has begun to study ways that can be used to extend the life of bridge infrastructure. Local governments have also been exploring measures to maintain their bridge infrastructure 1), 2).

A bridge management system (BMS) consists of three modules: management, assessment, and maintenance. The primary goal of a BMS is to assist bridge managers in determining the best strategy for bridge maintenance, repair, and rehabilitation with respect to current and future bridge conditions. However, because of restricted funds and a shortage of civil engineering technicians, maximizing the effects of the investment in the improvement of the serviceability and safety of the existing highway system is a major challenge for local government agencies. In 1990s-2000s, the importance of preventive maintenance for better bridge management was realized in Japan. MLIT raised a preventive mainte-
nance initiative and ordered National Highway Offices to implement the present once-in-5-years bridge inspection protocol in 2004. Because preventive maintenance is required to reduce future maintenance costs, MLIT started a subsidy program in 2007 for local governments to establish long-term bridge maintenance programs. As a result, prefecture governments now carry out bridge inspection programs that follow the MLIT protocol. For example, bridges on arterial routes that carry a large part of traffic are now inspected periodically. However, some municipalities still do not inspect their bridges. Some do, but inspection protocol and quality are not standardized. Therefore, it is imperative that a methodology be developed to aid bridge management officials, especially at the project level, in selecting the most appropriate alternatives for improving bridge inspection. An essential requirement in this endeavor is the assessment of bridge integrity, which is required for bridge maintenance, to derive the best strategy for bridge maintenance, repair, and rehabilitation.

In recent years, the amount of information at hand has constantly increased and the derivation of meaningful results from this information is seen as a valuable field of study that is described as the age of technology. Data mining, which is very useful for researchers, aims to reveal the information that is hidden in a large amount of data after a series of operations. Over the past few years, several experimental studies on the evaluation of bridge integrity using artificial intelligence (AI) techniques have been carried out by researchers. Some AI techniques of the support vector machine (SVM) were introduced and some useful different combinations of properties were estimated. Oishi et al. analyzed slope data using the SVM of AI technique to evaluate the risk of dangerous spots for sediment-related disaster by learning data on each slope. The results showed that an SVM can evaluate a risk with higher precision than a conventional rating method can. A study by Sugimoto et al. proposed the use of SVM in the estimation of synthetic safety. Basic to structural inspection is a visual examination of the representative parts of the structure. Inspection data on 2,959 bridges in Hokkaido were used as training data and the health ranking of the bridges was estimated. In a study by Chikata et al., the automation of the comprehensive evaluation of the visual inspection results was performed. The applicability of bridge integrity evaluation was examined by comparing SVM and Learning Vector Quantization (LVQ) in recognizing patterns. Though the results were dependent on the sample data, SVM showed high classification ability with complex parameter setting, while LVQ showed high robustness. Yuki et al. analyzed the use of inspection and repair record of the expansion joint of bridges using SVM to evaluate the setting method of the judgment criteria for repair of civil engineering structures. The results showed that the setting method using SVM was effective for the maintenance management plan of civil engineering structures owing to its high integrity upon evaluation by professional engineers. These previous studies have been conducted to evaluate bridge structures, such as the health ranking of the bridges, the bridge integrity, and the judgment criteria for repair of civil structures, using SVM. Most of these studies proposed that SVM could be used to estimate the health rating (I, II, III, and IV). However, approximately 70% of road bridges were managed by local governments. Generally, the inspection items for small and medium span bridges were limited and little information could be obtained from inspection sheets. From this perspective, in applying SVM analyses, this study undertook to contribute to the underlying research gap on the effect of the weight priority of the inspection item on determining bridge health integrity. Based on a suite of machine learning techniques with learning performance and explainability, the SVM gives high accuracy and explainability. Therefore, using SVM to estimate bridge integrity to assist regional governments is a vital task for the maintenance of bridge infrastructure.

(2) Overview and scope of the study

According to the MLIT protocols, the bridges (with span longer than 2 m) are inspected once every five years. Actual visual inspection is required. The condition of each structural member in each span is translated into maintenance urgency ratings. Maintenance urgency ratings are diagnoses given by experienced engineers in a subjective manner, recommending to bridge owners the needs for action until the time of the next inspection. Furthermore, the need for maintenance works before the next inspection is very important for bridge maintenance priority. Consequently, in this study, an efficient bridge soundness evaluation method is proposed by applying SVM analysis. Additionally, this proposed method is expected to determine the need for maintenance works before the next inspection in place of the engineers’ evaluation. In this study, the weight priority of the inspection item was estimated using SVM analysis. Additionally, the need for maintenance works before the next inspection was determined by the health rating classification of “Poor” (health rating: II, III, IV) and “Good” (health rating: I). The SVM was used to classify the health of the bridge structure. The input data were the degrees of deterioration of items in the detailed site inspec-
tion of structural members. The output data were the degrees of the health of the structural members. In the following inspection example, 83 pieces of inspection data on a reinforced concrete slab bridge in Gifu Prefecture were used. Keeping a balance between correctly classified accuracy and Receiver Operating Characteristic (ROC) curves enabled the weight assessment of the bridge structure members’ inspection items. This proposed method found that utilizing AI techniques provided academic foundation to record the additional inspection items.

2. CASE STUDY OF SMALL AND MEDIUM SPAN BRIDGES

In Japan, because there are huge amounts of existing short and medium span bridges in service, it is becoming one of major social concerns how those bridges can be maintained in good condition during their whole lifetime. Therefore, in this analysis, we studied the small and medium span reinforced concrete slab bridges with span of around 2m to 20m.

(1) Visual inspection of bridges

Gifu Prefecture revised the Revision Bridge Inspection Manual [12, 13]. Gifu Prefecture implemented a unique bridge inspection method based on remote viewing with binoculars. It also created a unique bridge inspection manual to ensure appropriate and accurate inspections [14].

In planning the maintenance priority scheme, the bridge’s integrity and degree of deterioration should be known in order to make engineering decisions with their own implicit knowledge. Maintenance urgency ratings are based on subjective diagnosis by experienced engineers. These ratings are used to recommend the need for maintenance by the time of the next inspection. In every structural member in each bridge span, the results of a health diagnosis can be categorized into one of the four ranks listed in Table 1. According to the country’s standards, a bridge’s health diagnosis has four levels: I, II, III, and IV. These levels correspond to the conventional criteria [12]. In Gifu Prefecture, the conventional criteria have five levels, which are denoted in Arabic numerals as 5, 4, 3, 2, and 1. The national standard of health diagnosis of I is divided into Ia and Ib to meet the criteria used in Gifu Prefecture. Ia indicates “healthy”, and Ib indicates “almost healthy”. According to the current standard, if the health diagnosis of a bridge is judged as Ia or Ib, then detailed inspection A (no damage recorded) will be applied. Additionally, if the health diagnosis of the bridge is judged as II, III, or VI, then detailed inspection B (with damage recorded) will be applied [12, 13]. However, if the health diagnosis of a bridge shows a low amount of deterioration, the degree of health is II. Therefore, even if damage occurred, depending on the specifications of the bridge and the surrounding environment, its health degree could be evaluated as Ib. Hence, the Ib reserve forces of II at the time of the next inspection must be distinguished from the health degree of Ia.

(2) Inspection data and feature extraction

We used the bridge inspection data on 971 small and medium span bridges that are managed by the municipal government (K city, Gifu Prefecture); 75% (725 bridges) are reinforced concrete slab bridges. Figure 1 shows the distribution of the numbers and types of bridges. According to the records of periodic road bridge inspection for small and medium span bridges, 13 types of deterioration of bridge superstructure were included in the bridge inspection records. Figure 2 shows the distribution of the number of bridges and the type of deterioration. Figure 2 shows that the deterioration of peeling and exposed steel bars had a high proportion compared to other types of deterioration. In this case study, we examined the deteriorating feature of peeling and exposed steel bars in reinforced concrete slab bridges. Because of the limited amount of bridge inspection data, only three types of bridge inspection data were used, the corresponding health ratings of which were I, II, and III. Of the 725 reinforced concrete slab bridges, 90% had a health rating of I. Furthermore, the deterioration of peeling and exposed steel bars was not recorded in the inspection records when the health rating of bridge member was assessed at Ib (almost healthy) [3, 12, 13, 18]. Therefore, we focused on the health degree of Ib. We also examined the inspection photos that were not included in the records of periodic road bridge inspections of small and medium span bridges. In this study, the 83 pieces of inspection data pertaining to the deterioration of peeling and exposed steel bars on reinforced concrete slab bridges in Gifu Prefecture were used (see Table 3).

In this study, the statistical analyses of two levels of health were conducted. According to our previous study [9], the different experienced inspectors had the same assessment for classifying the health rating of “Good” (health rating: I) and “Poor” (health rating: II, III, IV). Thus, the inspectors’ influences on the inspection data were not considered in this current study. These levels (“Good” and “Poor”) were used to evaluate the reinforced concrete slab bridges based on actual bridge inspection records in Gifu Prefecture. We examined additional features (additional inspection items) to assess the health status of the reinforced concrete slab by using SVM. These analyses were conducted using the SVM analysis so-
Table 1  Diagnosis of soundness in Gifu Prefecture.

| New Criteria | Description | Conventional Criteria Reference |
|--------------|-------------|---------------------------------|
| Ia Good      | No obstacle to the function of the structure. | 5 Healthy |
| Ib Preventive maintenance phase | There is no obstacle to the function of the structure, but it is desirable to take preventive maintenance. | 4 Almost healthy |
| II Early rehabilitation phase | There is a possibility that the function of the structure may be hindered, so a rehabilitation strategy must be taken. | 3 Start repair: Deterioration occurs |
| III Emergency repair phase | Presence of an obstacle to the function of the structure, or a possibility of occurrence is extremely high. An urgent action must be taken. | 2 Intermediate repairs required: Large deterioration |
| IV           |             | 1 Possible structure failure: Critical deterioration |

Table 2  Standard of classifying inspection parts.

| Damage classification | superstructure | substructure | bearing | others |
|-----------------------|----------------|--------------|---------|--------|
| Main girder           |                |              |         |        |
| Cross beam            |                |              |         |        |
| Slab                  |                |              |         |        |

Fig.1 Distribution of bridge numbers and bridge type.

Fig.2 Distribution of bridge numbers and deterioration type.
The software Waikato Environment for Knowledge Analysis (WEKA) version 3.9. We conducted this experiment using the classification solver in a Library for Support Vector Machine (LIBSVM) with the RBF kernel.

As pointed out in the pre-processing of the inspection data, all practical implementations of SVMs adhered to strict requirements for training and testing. The first requirement was that all data were numerical. Therefore, if categorical features were present, they were converted to numerical values using variable transformation techniques. Because the SVM model implementations in WEKA do not support missing values, we needed to either remove data with missing values or use some form of data imputation. Furthermore, SVMs assume that data are in a standard range. To ensure that all feature variables were treated equally, it was best to use the feature “normalization” before training the model. The input data were the degree of deterioration of each reinforced concrete slab’s inspection item, and the output data were the reinforced concrete slab’s degree of health. We converted the health degree to a numerical value as shown in Table 4. The standard used to classify the inspected structural members of the reinforced concrete slab bridges.

In the inspection example, we defined the y class label as “Good” if the health degree was equal to Ib and “Poor” if it was otherwise. In most cases, the training data should be classified manually and the test data should also be classified likewise in order to obtain the error rate. However, manual classification limits the amount of data that can be used for training, verification, and testing, and how to achieve the best performance with this limited data. To solve this problem, the retention procedure is normally used, which is a certain amount (20% - 30%) of the data-set kept for testing, and then the remaining amount is used for training. However, if necessary, a part of the data to be used for training can also be separated as the validation data. In this study, the health integrity of the inspection data was classified into six cases: case 1, case 2, case 3, case 4, case 5, and case 6, according to the retention procedure. We divided our data to training data-set and test data-set, used the first certain amount (20% - 30%) for test and the remaining for training, then used the second certain amount (20% - 30%) for testing and the remaining for training, and then used the third certain amount (20% - 30%) for testing and the remaining for training. Finally, for each above-mentioned three data-sets, we exchanged the test data and the training data. Therefore, the system accuracy is the average rate of six datasets. The inspection data are provided in Table 3. There are many causes of concrete peeling and exposed steel bars. These include the improper placement of concrete, electrochemical reactions between embedded steel bar within the concrete matrix, and corrosion of embedded reinforcing steel due to exposure to salt environment and combination with cracks due to fatigue. According to the aspect of concrete cover peeling mechanism behavior due to local corrosion, the corrosion distribution in the rebar length direction, length, and area, influences the deterioration of reinforced concrete structures. Based on the data collected from the inspection photos of peeling and exposed steel bars, we defined four features \( (x_1, x_2, x_3, x_4) \) from the collected inspection photos. More specifically, engineers carry out health rating assessments by using these features \( (x_1, x_2, x_3, x_4) \) with their implicit knowledge. We determined that the peeling and exposed steel bars occurred in four places: whole part, edge part, center part, and non-exposed (no exposed steel bars) as shown in Figure 3(a). Figure 3(b) shows three types of exposed direction on steel bars: main bar, distributing bar, and non-exposed. Figure 3(c) shows the distribution of the bridge numbers and the exposed length of steel bars. Figure 3(c) shows a high proportion of the value of non-exposed. If we ignored the value of non-exposed, the range from 0.1 to 0.2 had soundness levels of Ib and II. The range from 0.2 to 0.3 had soundness levels of Ib and II. Ranges greater than 1.0 had soundness levels of II and III (see Figure 3(d)). Therefore, we divided the feature of length (exposed steel bars) into five ranges (see Table 4). Figure 3(e) shows the distribution of the bridge numbers and the exposed area of steel bars. As shown in Figure 3(e), there were a high proportion of values of non-exposed. If we ignored the non-exposed, the range from 0 to 0.02 had soundness levels of Ib and II. Ranges greater than 0.2 had soundness levels of II and III (see Figure 3(f)). Therefore, we divided the feature of area (exposed steel bars) into four ranges (see Table 4). In Gifu Prefecture, the standard for the seismic design of bridges was organized systematically for the first time in 1980 (i.e., road bridge specification) when the foundation of the current standard of seismic design was established. Therefore, we divided the years of construction into three ranges: before 1980, after 1980, and unrecorded (bridge construction years have no recorded bridge inspection) as shown in Figure 3(g). For the feature of bridge length, we used the actual bridge length (see Figure 3(h)). The details of the inspection features and class labels are shown in Table 4.

Because our model’s dataset was small, we applied the classifier in a leave-one-out cross-validation, which is quite useful in dealing with small datasets since it utilizes the greatest amount of training data from the dataset. The leave-one-out cross-validation uses an entire model fit for all the data except...
(a) Place where exposed steel bars and peeling occurred.

(b) Direction of exposed steel bars.

(c) Exposed length of steel bars.

(d) Exposed length of steel bars (except non-exposed).

(e) Exposed area of steel bars.

(f) Exposed area of steel bars (except non-exposed).

(g) Construction year of reinforced concrete slab bridge.

(h) Bridge length of reinforced concrete slab bridge.

Fig. 3 Feature extraction of peeling and exposed steel bars.
a single point. It then makes a prediction at the point that can be compared to the actual value, which is iterated over the training dataset to obtain test errors for the labels of health ratings. WEKA supports the leave-one-out cross-validation type of evaluation. If the dataset wants to get a reasonably realistic evaluation, setting the number of folds equal to the number of rows in the dataset will give one leave-one-out cross-validation. Therefore, with the number of rows equal to k, k-fold cross-validation becomes leave-one-out cross-validation in WEKA. We conducted this experiment by using the classification solver in a LIBSVM with grid search and k-fold cross-validation to find the optimal hyper parameters (a soft margin constant C and width of a Gaussian kernel \(\gamma\)). In our approach, the use of k is equal to the number of rows for each of the training data in the k-fold cross-validation. In addition, a grid search was used on a two-dimensional space established by C and \(\gamma\). In this study, the default discrete set of \(\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}\) was used as the searching set of C, and \(\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}\) was used for that of \(\gamma\). In the output file, the classification error of the learning method, and the evaluation results of the LIBSVM classification process, including Kappa statistics, the mean absolute error, and the root mean squared error are reported. The Kappa statistic is a numerical value in which the expected and observed values are compared. Expected and observed accuracy calculation is used simultaneously in Kappa statistics and it can be easily determined through the confusion matrix. In general, the values of Kappa statistics in the range of 0.00-0.20 are considered as low, 0.21-0.40 as notable; 0.41-0.60 as mediocre; 0.61-0.80 as important and 0.81-1.00 as excellent. There are different error criteria in the evaluation of numerical estimation in data mining. Mean absolute error and root mean squared error are the error criteria used to determine how accurate numerical estimation is. The mean absolute error is calculated by taking the absolute value of the differences instead of taking the errors. The root mean squared error is the squared average of the second order loss function. Lastly, the output file is examined in Table 5, which shows that detailed accuracy, Kappa statistic, Mean absolute error, root mean squared error, and ROC curve values are reported for each class (“Good” and “Poor”). The optimal values of C and \(\gamma\) reached the highest classification accuracy. In chapter 2.3, the training data of six cases corresponding to the optimal parameters of three health labels were used.

(3) Testing results using SVM
In machine learning (ML), we tried to train an SVM model using a different set of inspection data. The algorithm used the training data-set to learn the rules of classification. The test data-set was used to test the model on data that was not used for training. Because the rules learned by the training data-set also were applied to the test data, an error rate was computed. In this study, we developed a classification model, and then we used a confusion matrix to estimate error rates.

Therefore, in this study, the class label y of the SVM output would be the result of either “Good” (health rating: Ib) or “Poor” (health rating: II, III). The data contained attributes that were highly correlated. By using one attribute, a weighted value can be obtained between all available predictions. Figure 4 shows the test results of different cases corresponding to the output of class label y. Figure 4(a) shows the correctly classified accuracy test results of class label y, which were “Good” (health rating: Ib) or “Poor” (health rating: II, III) with different cases corresponding to the use of all attributes and the use of only one attribute. The results of the six cases that used only \(x_2\) (exposed direction of steel bars) yielded a high correctly classified accuracy. The results of the six cases that used only \(x_5\) (bridge construction year) gave a low correctly classified accuracy. Figure 4(b) shows the AUC test results of class label y, which were “Good” (health rating: Ib) and “Poor” (health rating: II, III), corresponding to the use of all attributes and the use of only one attribute. The results of six cases, which only used \(x_2\) (exposed direction of steel bars), yielded a high value of AUC. The results of six cases which only used \(x_5\) (bridge construction year) gave a low value of AUC. These results showed that \(x_2\) (exposed direction of steel bars) feature had a good assessment of classifying the health degree of Ib and II; and \(x_5\) (bridge construction year) feature had a low assessment of classifying the health degree of Ib and II.

3. DISCUSSION
As mentioned above, the testing results of the training data using different cases are obtained as probability values, respectively. The features directly influenced the predictive models, and therefore, the results. We need features that describe the structures inherent in our data. The importance and selection of features provides information about the objective utility of features. However, they need to be created manually.

Figure 4 shows the different cases and their corresponding output of correctly classified accuracy and AUCs for bridge health label y. Among the numerical results (see Fig.4(a)), the correctly classified accuracies of \(x_1, x_2, x_3, x_4\) gives high values compared to \(x_5, x_6\). Moreover, the correctly classified
accuracy of $x_2$ (exposed direction of steel bars) feature gives a high value of 83%. However, the correctly classified accuracy of $x_5$ (bridge construction year) feature gives a low value of 36%. On the other hand, among the testing results (see Fig.4(b)), the AUCs of $x_1, x_2, x_3, x_4$ have high values compared to $x_5, x_6$. The AUC of $x_2$ (exposed direction of steel bars) feature has a high value of 0.84. However, the AUC of $x_5$ (bridge construction year) feature has a low value of 0.47. These results showed that the defined added features ($x_1, x_2, x_3, x_4$) had good assessment of classifying health ratings Ib and II. Whereas, the inspection standard used features ($x_5, x_6$) had insignificant meaning in assessment of classifying health ratings Ib and II. The defined added feature $x_2$ (exposed direction of steel bars) had good assessment of classifying health ratings Ib and II. Therefore, the defined added features ($x_1, x_2, x_3, x_4$)

| Case  | Training data (data number) | Test data (data number) | Ratio of number (training data/test data) | Correctly classified accuracy | Area under ROC curve (AUC) |
|-------|----------------------------|-------------------------|------------------------------------------|------------------------------|---------------------------|
| Case 1| 1~30                       | 31~83                   | 29/54                                    | 85%                          | 0.85                      |
| Case 2| 31~60                      | 1~30, 61~83             | 30/53                                    | 87%                          | 0.86                      |
| Case 3| 61~83                      | 1~60                    | 24/59                                    | 80%                          | 0.80                      |
| Case 4| 31~83                      | 1~30                    | 54/29                                    | 83%                          | 0.83                      |
| Case 5| 1~30, 61~83                | 31~60                   | 53/30                                    | 70%                          | 0.71                      |
| Case 6| 1~60                       | 61~83                   | 59/24                                    | 100%                         | 1.00                      |

Table 4 List of features and class label.

Table 5 Grid search and cross validation of optimal parameter.

(a) Result of correctly classified accuracy with used attributes.

(b) Result of AUC with used attributes.

Fig.4 Testing results of correctly classified accuracy and AUC.
which collected information from inspection photos had useful meaning for assessment of the health rating. Meanwhile, the previously mentioned engineers’ implicit knowledge dominated the testing results. From Fig.4, even though the “unrecorded” data were used in feature $x_5$ (bridge construction year), the testing results of feature $x_5$ (bridge construction year) seem to correspond with the results of the other features. This is because the engineers focused on the structures’ damaged conditions when assessing the health integrity. The bridge construction years are considered for the first time when engineers judge the bridge maintenance priority order. Therefore, the results of feature $x_5$ (“unrecorded” data was used) did not influence the following proposed diagnostic optimization method. The priority of the inspection item can be assessed by these case studies. Furthermore, the weight optimization method predicted the health of bridge integrity in these cases. The importance and selection of features can be used to achieve the objective utility of these features. We need to manually create an objective utility feature of inspection items.

To use AI more effectively in bridge inspections, it is important to distinguish soundness rank in the “Good” class (health rating of I) and “Poor” class (health ratings of II, III and IV). The health ratings of III and IV indicate unexpected deterioration and damage. Moreover, sufficient training data of health ratings of III and IV could not be collected. Therefore, the health ratings of III and IV could be assessed without using AI in this stage. Health rating of II has limited criteria for assessment compared with health ratings of III and IV. Because there were many cases, sufficient training data were collected. In the actual records of periodic road bridge inspection of small and medium span bridges, damage of health rating of Ib was not recorded. However, health rating of Ib was included in health rating of I. If it had additional inspection items, it would be enough to determine health ratings of Ib and II. Hence, there is no problem even if there are no extreme training data. In the meanwhile, the testing results showed that the inspection item of $x_2$ (exposed direction of steel bars), defined by inspection photos in the periodic inspection, gave high assessment of the classification “Good” (health rating: Ib) and classification “Poor” (health rating: II, III). Hence, this result provided the academic foundation for adding inspection item to the periodic inspection; also for the health rating of Ib and II.

The concept of the proposed diagnostic optimization method for assessing bridge soundness is shown in Fig.5. This diagram of bridge soundness diagnostic is adopted in a three-stage classification. According to the need for maintenance works before the next inspection, at the classification stage 1, the good condition (health rating of I) is assessed by the proposed diagnostic optimization method. Furthermore, the SVM has superior ability to calculate and quantify the good condition (health rating of I) of the structure. At classification stage 1, the model of the proposed diagnostic optimization method applied the LIBSVM tool to come up with “Good” (health rating: I) and “Poor” (health rating: II, III, IV) classifications. For better AI application to the periodic inspection of road bridges, this proposed method could be given the additional inspection items to evaluate the good condition by comparing results of the correct classified accuracy and those of the AUC.

Fig.5 Concept diagram of soundness diagnosis.
For the health rating (II) and the health rating (III, IV), maintenance works should be done before the next inspection within five years. Furthermore, quality judgments are required. Consequently, the engineers’ judgments are required in the classification stages 2 and 3. With regard to the timing of maintenance works, the health rating (II) and health rating (III, IV) are assessed by the engineers’ judgments through “close visual inspection” of the whole bridge at classification stage 2. In the critical failure performance of health rating IV, the health rating (III) and health rating (IV) are assessed by the engineers’ judgments through “close visual inspection” of the whole bridge at classification stage 3. This scheme also could be technical support for the diagnostic image of the inspection records.

Local governments are constantly distressed by the insufficient maintenance budgets and maintenance experts. The proposed diagnostic optimization method could assist local governments in determining the health and integrity of bridges, which have been assessed based on the subjective experience of a maintenance expert.

In future research work, there is a need for a method to create a model that can achieve multi-class classification. Based on this case study, this proposed bridge soundness evaluation method utilizing SVM will be useful to continue analyzing the multi-class classification.

4. CONCLUSION

In this study, an ML algorithm was developed to obtain a useful feature for inspecting small and medium span reinforced concrete slab bridges. The following recommendations for assessing the health integrity of bridges should be adopted by using the SVM classification:

(1) The assessment weight of inspection items is obtained for classifying the health ratings in the training data of six cases. The effective assessments of different features could assist local governments to determine maintenance priority estimation in their jurisdictions.

(2) The results of the case study showed that the proposed bridge integrity evaluation method provides the academic foundation for adding the inspection item (exposed direction of steel bars) to the periodic road bridge inspections of small and medium span reinforced concrete slab bridges.

(3) The results of the case study showed that this proposed method constitutes technical support for diagnostic image records when determining health rating of Ib and II.

(4) The results of the case study showed that this proposed method for evaluation bridge soundness using SVM is more useful to continue analyzing the multi-class classification in future research works.

(5) The proposed bridge integrity evaluation method could be used as a comprehensive approach to the maintenance management of small and medium span reinforced concrete slab infrastructures.

ACKNOWLEDGMENT: The authors thank Mr. Hosoe, I. of Dainichi Consultant Inc. for his valuable advice. We are grateful for the kind and constant generous help of Dainichi Consultant Inc., Gifu Prefecture.

APPENDIX A: SUPPORT VECTOR MACHINES

An SVM algorithm is used to find the optimal separating hyperplane that effectively separates the data points into the labelled classes. Let us consider a binary classification. The data points are mapped into a high-dimensional feature space (Hilbert space) by a kernel function K (dot products between data points). For input points \( x_i \in \mathbb{R}^p \) and label of the class of data \( y_i (i = 1 \ldots n) \), the decision function in the feature space can be considered as follows

\[
f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b
\]

where \( b \) is the model bias. Note that only those points that lie closest to the hyperplane have \( \alpha_i > 0 \) and consist the support vectors. Let us assume the primal optimization problem in order to obtain the necessary parameters. The soft margin optimization problem can be formulated as

\[
\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i
\]

subject to:

\[
y_i \left(w^T x_i + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0.
\]

where \( w \) is the weight vector normal to the hyperplane; \( \xi_i \) are the slack variables that hold for misclassification examples; and, consequently, the term \( \sum_{i=1}^{n} \xi_i \) can be considered as a measure of the amount of total misclassifications of the model. The trade-off between maximization of the margin and minimization of error is controlled by cost parameter \( C \).

In the non-linear case, a kernel function is introduced to map the low-dimensional input space to the high-dimensional feature space. There are many possible choices of kernel functions, such as the linear, polynomial, radial basis function (RBF), and sigmoid function. In this study, the RBF kernel
(Gaussian kernel) was utilized according to the following function:
\[ k(x, x') = \exp(-\gamma \|x - x'\|^2) \]
where \( \gamma > 0 \) is a parameter that controls the width of Gaussian. It plays a role in controlling the flexibility of the resulting classifier.

**APPENDIX B: RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE**

The ROC curve was introduced by the signal processing community in order to evaluate the capability of a human operator to distinguish an informative radar signal from noise. The ROC curve has been used mainly in the medical field to assess the usefulness of diagnostic tests [28].

Traditionally, the evaluation of a learned model is done by minimizing the estimation of a generalization error or some other related measures. However, the accuracy (the rate of correct classification) of a classifier is the most frequently used performance measurement [28]. Accuracy can be considered as follows:

\[ \text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \]

where a true positive (TP) is a positive record that correctly predicts a positive. A false negative (FN) is a positive record that is incorrectly predicted as negative. A false positive (FP) is a negative record that is incorrectly predicted as positive. A true negative (TN) is a negative record that is correctly predicted as negative. The ability to obtain the optimal balance classification is described in terms of sensitivity (or true positive rate or positive class accuracy) and specificity (or true negative rate or negative class accuracy) as follows:

\[ \text{Sensitivity} = \frac{TP}{TP + FN} \]
\[ \text{Specificity} = \frac{TN}{TN + FP} \]

Specifically, sensitivity measures the proportion of actual positives that are identified correctly. Specificity measures the proportion of actual negatives that are identified correctly. In addition, the trade-off between sensitivity and specificity can be represented graphically as an ROC curve. It can be understood as the plot of the probability of correctly classifying the positive examples against the rate of incorrectly classifying true negative examples. The ROC curve can be constructed by plotting these pairs of values on the graph with 1-specificity on the x axis and sensitivity on the y axis. In fact, when the data are strongly unbalanced, accuracy may be misleading because the all positive classifiers or all negative classifiers may achieve a very good classification rate. Situations in which datasets are unbalanced occur frequently in real-world problems. In these cases, model evaluation is based on criteria other than accuracy. The metrics extracted from ROC curves, such as the area under the ROC curve (AUC), can be a good alternative for model evaluation. The reason is that the ROC curve can determine the difference between errors in positive or negative examples [28].

Ana Maria Simundic’s [29] study on diagnostic accuracy showed the shape of an ROC curve, and the AUC helps us estimate the discriminative power of a test. The area under the curve can have any value between 0 and 1, and it is a good indicator of the overall quality of the test. Generally, the relation between AUC and diagnostic accuracy (as described in Table 6) and the ROC curve [29] can be plotted as shown in Fig.6. In our study, the AUC rating was determined according to Table 6.

**REFERENCES**

1) Emergency Recommendation by Social Infrastructure Maintenance Strategy Subcommittee, Engineering Section, Engineering Committee, Panel on Infrastructure Develop-

| AUC   | Diagnostic Accuracy |
|-------|---------------------|
| 0.9-1.0 | Excellent           |
| 0.8-0.9 | Very good           |
| 0.7-0.8 | Good                |
| 0.6-0.7 | Sufficient          |
| 0.5-0.6 | Bad                 |
| <0.5    | Test not useful      |

**Fig.6** the ROC curve (refer to Simundic et al. 2012).
Social Infrastructure Maintenance Strategy Subcommittee, Engineering Section, Engineering Committee, Panel on Infrastructure Development / Council for Social Infrastructure: Comprehensive Reinforcement of Infrastructure Policy for Full-Scale Maintenance Era - "First Year of Maintenance Policy" Kicks Off (in Japanese). http://www.mlit.go.jp/policy/shingikai/sogo03_sg_000018.html

2) Social Infrastructure Maintenance Strategy Subcommittee, Engineering Section, Engineering Committee, Panel on Infrastructure Development / Council for Social Infrastructure: Ideal of Future Maintenance and Renovation of Social Infrastructure (Intermediate Summary) (in Japanese). http://www.mlit.go.jp/report/press/sogo03_hh_000046.html

3) Ministry of Land, Infrastructure, Transport and Tourism: Bridge Inspection Manual, 2004 (in Japanese).

Masahiro, S. and Takashi, T.: Bridge inspection standards in Japan and US, Proc. of 29th US - Japan Bridge Engineering Workshop, Tsukuba, Japan, 2013.

5) Arong, Yiliguoqi and Murakami, S.: Study on utilization of SVM in soundness evaluation of bridge superstructure, Journal of Japanese Society of Steel Construction (JSSC), Vol. 25, No. 100, pp. 43-56, 2018.

6) Ohishi, H., Kobayashi, H., Yui, Y., Tanaka, H., Nakayama, H. and Furukawa, K.: Evaluation of the danger of slopes in view of the effective value of measure constructions using support vector machine, Journal of Japan Society of Civil Engineers, Ser. F, Vol. 63, pp. 107-118, 2007 (in Japanese).

7) Sugimoto, H., Ichima, K., Abe, J. and Furukawa, K.: On synthetic health evaluation of infrastructure by SVM and its application to ranking of structures, Journal of Japan Society of Civil Engineers, Ser. A, Vol. 65, pp. 658-669, 2009 (in Japanese).

8) Chikata, Y., Aso, K., Kameda, J. and Kido, T.: Applicability of SVM and LVQ to evaluation of bridge integrity, Journal of Applied Computing in Civil Engineering, Vol. 16, pp. 175-184, 2007 (in Japanese).

9) Yuki, K., Kobayashi, H., Ohishi, H., Sugimoto, H., Iida, T. and Furukawa, K.: Evaluation for judgement criteria of repair on civil engineering structure by support vector machine, Journal of Japan Society of Civil Engineers, Ser. F4 (Construction and Management), Vol. 68, pp. 52-61, 2012 (in Japanese).

10) Gunning, D.: Explainable artificial intelligence (XAI), Defense Advanced Research Projects Agency (DARPA). https://www.darpa.mil/program/

11) Fukushima, T., Fujimaki, R., Okanohara, D. and Sugiyama, M.: Outlook for big data and machine learning, Cutting-edge technological challenges and expanding applications, Information Management, Vol. 60 (8), pp. 543-554, 2017 (in Japanese).

12) Road Maintenance Division of Gifu Prefecture Government: Bridge Inspection Manual, 2016 (in Japanese). http://www.pref.gifu.lg.jp/shakai-kiban/doro/doro-iji/11657/index_57545.html

13) Road Maintenance Division of Gifu Prefecture Government: Bridge long-term repair plan, 2018 (in Japanese). http://www.pref.gifu.lg.jp/shakai-kiban/doro/doro-iji/11657/index_7745.data/h29_bridge.pdf

14) Hosoe, I., Kato, N. and Morimoto, H.: Scenario for rational maintenance and management of bridges, Proc. of 10th International Conference on Inspection, Appraisal, Repair and Maintenance of Structures, Hong Kong, pp. 207-214, 2006.

15) Chang, C. C. and Lin, C. J.: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology, Vol. 2 (3), pp. 1-27, 2011.

16) Ben, H. A. and Weston, J.: A user's guide to support vector machines, Data Mining Techniques for the Life Sciences, pp. 223-239, 2010.

17) Mitchell, T. M.: Machine Learning, New York, McGraw-Hill, 1997.

18) Ministry of Land, Infrastructure, Transport and Tourism: Bridge periodic inspection procedure, 2018 (in Japanese). https://www.mlit.go.jp/road/sisaku/yobohozen/tenken/yobo3_1_6.pdf

19) Qiao, D., Nakamura, H., Tran, K. K., Yamamoto, Y. and Miura, T.: Experimental and analytical evaluation of concrete cover spalling behavior due to local corrosion, Journal of Structural Engineering, Vol. 61A, pp. 707-714, 2015.

20) Chapelle, O.: Training a support vector machine in the primal, Neural Computation, Vol. 19 (5), pp. 1155-1178, 2007.

21) Remco, R. B., Eibe, F., Mark, H., Richard, K., Peter, R., Alex, S. and David, S.: WEKA Manual for Version 3-6-5, University of Waikato, Hamilton, New Zealand, 2011.

22) Landis, J. R. and Koch, G. G.: The measurement of observer agreement of categorical data, Biometrics, Vol. 31 (3), pp. 159-174, 1977.

23) Witten, I. H. and Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, San Francisco, Morgan Kaufmann, 2005.

24) Gökhan, A. and Nuri, D.: An analysis program used in data mining: WEKA, Journal of Measurement and Evaluation in Education and Psychology, Vol. 10 (1), pp. 80-95, 2019.

25) Cortes, C. and Vapnik, V.: Support vector networks, Machine Learning, Vol. 20, pp. 273-297, 1995.

26) Schölkopf, B. and Smola, A.: Learning with Kernels, MIT Press, Cambridge, MA, 2002.

27) Cristianini, N. and Shawe, T. J.: An Introduction to Support Vector Machines, Cambridge UP, Cambridge, UK, ISBN 0-521-78019-5, 2000.

28) Rakotomamonjy, A.: Optimizing area under ROC curve with SVMs, Proc. of 1st International Workshop on ROC Analysis in Artificial Intelligence, Spain, pp. 71-80, 2004.

29) Simundic, A.: Diagnostic accuracy - part 1: Basic concepts, Journal of Near-Patient Testing & Technology, Vol. 11, pp. 6-8, 2012.

(Received October 15, 2018)