Deterioration prediction of existing concrete bridges using a LSTM recurrent neural network

Pengyong Miao[1] (corresponding author), Hiroshi Yokota[2], Yafen Zhang[1]
1. Graduate School of Engineering, Hokkaido University, Sapporo, Japan
p.y.miao@outlook.com;
1509117525@qq.com;
2. Faculty of engineering, Hokkaido University, Sapporo, Japan
yokota@eng.hokudai.ac.jp;

Abstract: Bridge censored databases can be used to analyze and assess structural deterioration conditions, but conducting the analysis is difficult. This difficulty occurs because many factors affect deterioration, and the time span of the data for these factors depends on the years in service of the respective bridge. In addition, the values of some factors are not regularly observed. The present study uses the long short-term memory (LSTM) to consider twelve potentially influencing factors to recognize the relationships between these factors and deterioration grades. Testing the model on an inspection database of 3,368 bridges indicates that the LSTM model obtained an accuracy of exceeding 80%, i.e., outperforms the performance of a multilayer perceptron model. For four types of bridges, the LSTM model shows the equivalent performance. In addition, the predictive ability of the model for coastal bridges is slightly superior to non-coastal bridges. No significant differences in accuracy are determined between different deck areas. Practically, the model can predict bridge deterioration paths, and could help decision makers formulate predictive intervention strategies for improving the quality of maintenance management.

Key words: deterioration prediction; inspection database; long short-term memory (LSTM); maintenance schedule formulation; potentially influencing factors; prediction model.

1. Introduction

Predictive maintenance is an increasingly popular method for structural soundness maintenance (Stenström et al., 2016). Progress has been made in taking preventive measures, but predicting structural deterioration remains challenging. This difficulty occurs because a targeted predictive maintenance strategy relies heavily on a reasonable assessment of a structure's past and present conditions and the accurate prediction of deterioration. In other words, predictive maintenance strategies need sufficient data on structural conditions and an appropriate analysis method. Onsite inspection databases for existing bridges collected by engineers during regular inspections constitute abundant information on structural conditions, thus providing the possibility of using the database for formulating predictive maintenance strategies. Potential applications of such databases include evaluating structural conditions, predicting deterioration, and developing maintenance strategies (Chandra & Zhang, 2012).

To select a suitable method, it is necessary to understand the characteristics of the censored database. Since the conditions of an existing bridge are typically dynamically changing, the conditions at a time affect its conditions at the subsequent time (Jeong et al., 2019). In addition, deterioration of existing bridges is associated with many factors, such as traffic volume, chloride, and snowfall. Therefore, the cumulative effects of these factors on deterioration must be evaluated. Furthermore, the time-span of the data for these factors depends on the years in-service of the respective bridge. Some values of many time-dependent factors are irregularly observed. In summary, the method used to analyze the database should be able to: handle problems with many factors, consider the cumulative effects of time-dependent factors, and deal with nonlinear relationships between factors and

deterioration. Conforming to the aforementioned requirements, possible methods are detailed below.

Traditional time series methods using linear models have been widely applied to fit the collected data. In addition, methods such as the autoregressive moving average (ARMA) model and its variants (Dorffner, 1996; Ho et al., 2002) have been shown effective in various practical applications. However, these methods cannot solve nonlinear relationships. To address this issue, different nonlinear autoregressive models have been developed (Connor et al., 1994; Qin et al., 2017). Few of these nonlinear autoregressive models can capture the long-term temporal dependence appropriately. Chandra & Zhang (2012) indicated that neural networks had achieved promising results in the prediction of chaotic time series. Carvalho et al. (2019) presented a systematic literature review of the machine learning methods applied to predictive maintenance. Miao et al. (2019) introduced a multilayer perceptron model considering twelve factors with an accuracy of 65% in predicting bridge deterioration. Recurrent neural network (RNN) is an artificial neural network architecture specialized in learning time-related patterns from time series data and has been employed in many fields (Giles et al., 2001; Choi et al., 2017). Many studies have also applied this method to practical engineering. Jeong et al. (2019) used this method to monitor bridge vibrations caused by loads. Wang et al. (2018) applied this method to predict the quality problems of structural components in construction projects. Razavi et al. (2015) used this model to predict the load-deflection of carbon fiber strengthened reinforced concrete (RC) slab.

Since an RNN can process time-series and consider the time dependence of the respective factor, it can be used to assess the cumulative effect of factors on the deterioration. In

addition, an RNN is able to analyze time-series data with different time spans, enabling it to process the inspection databases composed of bridges with different service years. Another feature of an RNN is that it can assess the combined effects of many factors and establish nonlinear relationships between factors and deterioration (Lipton et al., 2016). These features of the RNN model make it an optimal method for analyzing the inspection database of bridges for predictive maintenance.

This study explores the feasibility of establishing an RNN prediction model for predicting deterioration by correlating the potentially influencing factors and deterioration. Specifically, twelve potentially influencing factors are considered for each bridge. The deterioration conditions of bridges are denoted by grades (Section 3.1.2). In the implementation, the annual average of each factor is treated as an element of a vector (Figure 1). Many vectors are finally obtained for each bridge, and the number of vectors depends on the years in service. Then, all vectors are input into the RNN model in chronological order to calculate the cumulative effects. Verification is conducted using an inspection database with 3,368 bridges. Firstly, the performance of the RNN model is compared with that of a multilayer perceptron (MLP) model in terms of four metrics (Section 4.1). Secondly, the performance of the RNN model is evaluated according to different types of bridges, different environments, and different deck areas. Furthermore, error analysis is performed to detect defects of the prediction model.

2. Methodology

2.1 Problem representation

The inspection database is first converted into vectors in chronological order to represent

potential factors as computable sequences. Figure 1 provides an example of defining the time series of all factors for a bridge as a set of vectors. Suppose N factors influence a bridge that has been in service for T years. The occurrence of all factors in chronological order is illustrated in Figure 1 (a). The values of all the factors at each time step constitute a vector. Since the bridges are inspected every five years or more often as needed, the time step is the same as the inspection period. The value of a factor is zero if its value has not started to be observed. For example, the traffic volume is zero at t_1. Finally, an $N \times T$ matrix is obtained, as shown in Figure 1 (b). The matrix can then be used for further computations.

The duration of each factor depends on the initial time and the inspection time. As illustrated in Figure 1 (a), the durations for the environmental factors and traffic load are $t_3 - t_1$ and $t_3 - t_2$, respectively. In practice, time-series data before t_3 can be exploited to establish an RNN model for predicting the performance at time t_4. Given time-series $x_1^n,\ x_2^n,\ \dots,\ x_T^n\ (n = 1,\ 2, \dots,\ N)$ regarding N potential factors, our purpose is to build a model to predict $\hat{y}$ of the actual deterioration grade y. For a bridge, the input and output are represented as $D = (X;\ y)$. X is the input that consists of N factors, and y is the output indicating the deterioration grades. The input is arranged as:

$$X = \begin{bmatrix} x_1^1 & \cdots & x_t^1 & \cdots & x_T^1 \\ \vdots & \ddots & \vdots & & \vdots \\ x_1^n & \cdots & x_t^n & \cdots & x_T^n \\ \vdots & & \vdots & \ddots & \vdots \\ x_1^N & \cdots & x_t^N & \cdots & x_T^N \end{bmatrix}$$

where T and t denote the years in service of a bridge and the time of one observation, respectively. x_t^n is the observed value at time t for the nth factor.

The matrix X only represents the time series for one bridge. Assuming K bridges are included in the inspection database, a corresponding matrix $X_k(k = 1, 2, \dots,\ K)$ with the

same dimension *N* but a different length T_k will be obtained for each bridge. Since different factors are measured on different ranges, values of these factors are necessary to be adjusted to a common scale to improve the efficiency of the training process. Therefore, each factor's value is normalized within 0–1 using:

$$x_t^n = \frac{x_t^n - \min(|x_t^n|)}{\max(|x_t^n|) - \min(|x_t^n|)} \quad (1)$$

where $\min(|x_t^n|)$ and max ($|x_t^n|$) are, respectively, the minimum and the maximum of the input data x^n among K total target bridges.

Then, all inputs are input into an RNN model in chronological order to consider the cumulative effects, generating outputs only at the final sequence step, as shown in Figure 2. x_t, h_t, and $\hat{y}_T$ represent the inputs, hidden layers, and predictions, respectively. As the initial values of weights are randomly assigned during training, the predicted grades are usually inconsistent with the actual grades. Therefore, cross-entropy defined by Equation (2) is used to evaluate the error between the predicted and the actual grades. The optimal model is determined by iteratively updating the model's parameters to minimize the cross-entropy value (Section 2.2). The optimal model can then be used to predict the deterioration grade provided the corresponding time series data. In addition, the optimal model can be further updated by continuously integrating new inspection data.

$$Loss(\hat{y}_T, y) = \frac{1}{T}\sum_{t=1}^{t=T} -(y_t \cdot \log(\hat{y}_t) + (1 - y_t) \cdot \log(1 - \hat{y}_t)) \quad (2)$$

2.2 Recurrent neural network

An RNN is a type of artificial neural network. Among all recurrent neural networks, long short-term memory (LSTM) (Hochreiter & Urgen Schmidhuber, 1997) and gated recurrent units (GRUs) (Cho et al., 2014) are most commonly used. GRU is structurally similar to

LSTM but simpler than LSTM. The RNN model used for deterioration prediction in this study is an LSTM, as illustrated in Figure 2. The LSTM accepts an input vector x_t at time step t and stores the state of the input in the hidden layer h_t. To classify input data, it is necessary to have a layer for predicting grades, which is usually located at the end of the LSTM architecture. The most prominent method to date is using the softmax activation function. After the entire time series X input into the RNN, the softmax activation function is applied to the final hidden layer h_T and produces a value $\hat{y}_T$. $\hat{y}_T$ is the estimated deterioration grade for a bridge. An LSTM-based RNN provides a neural network with a memory function, enabling the neural network to achieve good modeling ability on time series data (Wang et al., 2018).

The dashed box of Figure 2 corresponds to the LSTM unit used for processing current and previous information. The detailed components of the dashed box in Figure 2 are shown in Figure 3. The update and use of cumulative information are controlled by three gates: the input gate i_t, the forget gate f_t, and the output gate o_t (Wang et al., 2018). s_t, g_t, and h_t are the state memory cell, input candidate memory cell, and hidden layer, respectively (Lipton et al., 2016). The calculation of all gates is affected by both the current input x_t and the output of the previous LSTM cell h_{t-1}. The input gate is applied to process the impact of the current input on the status of the memory cell s_t. The forget gate is used to control the influence of cumulative information on the memory cell s_t. The output gate is applied to control the status value of the memory cell s_t. The mathematical formulation of Figure 3 and the update of the LSTM unit can be divided into the following steps (Hochreiter & Urgen Schmidhuber, 1997):

(1) Calculate the value of the current candidate memory cell g_t.

$$g_t = \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \tag{3}$$

(2) Calculate the value of the input gate i_t.

$$i_t = \text{sigmoid}(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{4}$$

(3) Compute the value of the forget gate f_t.

$$f_t = \text{sigmoid}(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{5}$$

(4) Calculate the state value s_t of the current memory cell.

$$s_t = g_t \odot i_t + s_{t-1} \odot f_t \tag{6}$$

(5) Calculate the value of the output gate o_t.

$$o_t = \text{sigmoid}(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{7}$$

(6) Calculate the value of the current hidden layer h_t.

$$h_t = \tanh(s_t) \odot o_t \tag{8}$$

(7) The output of the last LSTM unit is shown by the dashed part of Figure 3 and is calculated by:

$$\hat{y}_T = \text{softmax}(h_T) \tag{9}$$

In these equations, the tanh, sigmoid, and softmax are three commonly used activation functions in the field of deep learning (Hochreiter & Urgen Schmidhuber, 1997). W and b are the parameters for calculating the three gates. In LSTM, the previous hidden layer h_{t-1} and the current input x_t do not directly affect the value of the current hidden layer h_t. Instead, they change the values for gates i_t, f_t, and o_t and the intermediate memory cell s_t. Then, the current hidden layer h_t is determined by s_t and o_t. In the calculation for s_t and h_t, $\odot$ denotes elementwise multiplication. All three gates have values between zero and

one: zero means the information stored in the gate is ignored, and one means the information is accumulated and passed to the next calculation. The functions of the three gates and separate memory cells allows the LSTM unit to save, read, reset, and update long-distance cumulative information. These characteristics of LSTM are especially useful because LSTM can process non-linear problems with many factors (Choi et al., 2017). In addition, the cumulative effects caused by the time-dependent factors can be appropriately solved using the functions of the three gates.

Following procedures from steps (1) to (7), the error can be calculated according to Equation (2). Then, the parameters in Equations (3) to (9) are updated using the derivative of Equation (2) with respect to each parameter as the step size. The iterations will continue until the minimum error is obtained, as shown in Figure 11 with a practical case.

3. Case study

3.1 Data description

A database collected during bridge inspections regarding 3,386 bridges in Hokkaido, Japan, is employed to verify the feasibility of LSTM. The locations of these bridges are shown in Figure 4. We eliminated incomplete data and unreasonable data from the database and selected 3,368 out of 3,386 bridges. Specifically, the bridges with unknown ages are discarded. In conjunction with the age constrain, the constraint of non-decreasing deterioration grade is introduced. For example, a bridge is assessed grade 2 in an inspection and assessed grade 1 in the next inspection, this kind of bridge is not considered. Furthermore, only bridges with consecutive series data are considered, such as some bridges whose traffic

volume is counted only once are unconsidered. Then, we divided the data for the selected bridges into two categories: potentially influencing factors and deterioration grades. Details on these two categories are described in the following subsections. Since the years in service of the bridge vary from a few years to a few decades, the annual average value of each factor is taken as the value of this factor at a certain moment in the time series. Accordingly, a time series is obtained for each factor of each bridge. The time span of the time series is the bridge age.

3.1.1 Potentially influencing factors

Factors that may affect deterioration are regarded as inputs in the prediction model. The bridge features, such as the bridge length, bridge width, elevation, and years in service, are extracted from the inspection database. In addition, factors of the temperature, rainfall, snowfall, and carbon dioxide concentration are derived from Japan Meteorological Agency (JMA, 2019). The traffic volumes are retrieved from MLIT (2005, 2010, and 2015a). Airborne salt and carbonation are calculated according to previous studies (Kobayashi, 2010; JSCE, 2012; Tamakoshi et al., 2012; NRA, 2016). In summary, the potentially influencing factors considered in this study category as: bridge geometry factors, environmental factors, bridge age, and loading conditions, as shown in Table 1. This study only considers concrete bridges (or components), including PC bridges, RC bridges, PC and RC hybrid bridges, and concrete components of steel bridges. Discussions regarding the performance of the model on different types of concrete bridges are described in Section 4.1. Many other factors such as the deicing salt are not considered due to the inability to collect relevant data.

Factors such as the geometry and elevation of the bridge are constant. More details are

described for the factors of the traffic volume, airborne salt, and carbon dioxide concentration.

(1) Large-sized vehicles

According to the survey of the daily traffic volume (MLIT, 2015b), the proportion of large-sized vehicles (including buses, lorries, construction heavy equipment and other special equipment) R_V can be calculated by:

$$R_V = \frac{Number\ of\ large\ size\ vehicles\ daily}{Daily\ traffic\ volume} \times 100\% \quad (10)$$

(2) Airborne salt

The bridges within 1 km of a coastline are considered influenced by the airborne salt. The airborne salt concentration can be calculated using (Kobayashi, 2010):

$$C_{ab} = C_1 \cdot x^{-b} \quad (11)$$

where C_{ab}: Airborne salt concentration ($\mathrm{mdd \cdot NaCl}$),

C_1: 1 km equivalent airborne salt concentration (mdd·NaCl),

x: Distance from a coastline (km), and

b: Degree of distance attenuation (b = 0.6).

C_1 for areas B and C shown in Figure 5 are 1.174 and 0.072, respectively (Kobayashi, 2010). Then, Equation (12) (JSCE, 2012) can be used to compute the concentration of chloride ions that adheres to the surface of the concrete. Only the situation of $C_{ab} \leq 30$ is considered because all C_{ab} data considered in this study satisfy this constrain.

$$C_0 = -0.016 \times {C_{ab}}^2 + C_{ab} + 1.7 (C_{ab} \leq 30) \quad (12)$$

where C_0: Chloride ions concentration of concrete surface ($\mathrm{kg/m^3}$), and

C_{ab}: Airborne salt concentration at the bridge location ($\mathrm{mdd \cdot NaCl}$).

Figure 6 shows the relationship between the concentration of chloride ions on the concrete surface and the distance from a coastline. For comparison, the chloride ions concentration on the concrete surface obtained according to the Standard Specifications for Concrete Structures (JSCE, 2012) is also displayed. Figure 6 shows that chloride ions concentration decreases as the distance from the coastline increases in both areas B and C.

The estimated chloride ions concentrations in area B are slightly larger than the JSCE standard and those in area C are lower than the standard. One possible reason is that the standard considers the results from all regions in Japan. The reason that the value in area B is greater than that in area C is that the wind is stronger on the side of the Sea of Japan and is relatively weak on the Pacific side (Kobayashi, 2010). Then, Fick's second law is applied to determine the penetration of chloride ions into concrete:

$$C(x,t) = C_0\left(1 - \mathrm{erf}\left(\frac{x}{2\sqrt{D_c \cdot t}}\right)\right) + C(x,0) \quad (13)$$

where $C(x,t)$ is the chloride ions concentration ($\mathrm{kg/m^3}$) at a depth of x after t (year); C_0 is the chloride ions concentration on the concrete surface ($\mathrm{kg/m^3}$) and can be obtained from Equation (12); $C(x,0)$ is the initial chloride ions content ($\mathrm{kg/m^3}$), for which the value is usually 0.3 $\mathrm{kg/m^3}$ (Tamakoshi et al., 2012); and D_c is the diffusion coefficient ($\mathrm{cm^2/year}$).

Deterioration is characterized by corrosion of the rebar or cracks propagation. Therefore, the chloride ions concentration at the surface of the rebar is calculated and used as a criterion to identify deterioration. According to previous experimental studies, the corrosion threshold concentration of chloride ions for ordinary steel in concrete in Japan is 1.2 to 2.5 $\mathrm{kg/m^3}$ (Daisoku et al., 1986). In addition, in previous surveys of existing bridges, almost no steel corrosion was observed when the chloride ions concentration was less than 1.2 $\mathrm{kg/m^3}$

(Tamakoshi, 2012). Therefore, 1.2 kg/m^3 is thought to be the chloride ions concentration for the onset of rebar corrosion.

Furthermore, Equation (14) is used to calculate the diffusion coefficient D_c $(\mathrm{cm}^2/\mathrm{year})$.

$$\log(D_c) = -3.9(\mathrm{W/C})^2 + 7.2(\mathrm{W/C}) - 2.5 \quad (14)$$

where $\mathrm{W/C}$ is the water-to-cement ratio.

In this study, there are few structures with a clear water-to-cement ratio. However, Tamakoshi (2009) suggested that if there are no specific data, material information can be obtained according to Table 2. Therefore, the values of the water-to-cement ratio and thickness of the concrete cover are extracted from Table 2.

For each bridge, the chloride ions concentrations on the surface of all bridge components are calculated. According to the inspection guide (MLIT, 2014), the bridge is considered corroded once corrosion occurs for any component. Figure 7 is an example to show the chloride ions concentrations on the surface of the rebar for a bridge superstructure and pier. The results reveal that the corrosion of the pier occurs slightly earlier than that of the superstructure. Therefore, the corrosion time of the pier is considered to be the initial corrosion of the bridge.

(3) Carbonation

According to the investigation of Kishitani (1991), the carbonation depth of concrete structures in the natural environment can be estimated by:

$$d = A \cdot \sqrt{t} \quad (15)$$

where d is the carbonation depth(mm), A is the deterioration index for carbonation (mm /$\sqrt{\mathrm{year}}$), and t is the years in service.

The deterioration index A is calculated by:

$$A = R(4.6\,\mathrm{W}/C - 1.76)/\sqrt{7.2} \tag{16}$$

where R is related to the type of cement, aggregate, and admixture. In this study, R is 1 according to Kawakami (1995) and NRA (NRA, 2016). W/C can be obtained from Table 2.

Equation (15) is obtained assuming that the CO_2 concentration is constant. However, the CO_2 concentration is observed to increase with time in the past decades, as shown in Figure 8. In addition, the measured carbonation depth of a concrete structure is very different from the value estimated by Equation (15) (NRA, 2016). Therefore, the equation is modified as:

$$d = k(t) * A \cdot \sqrt{t} \tag{17}$$

where $k(t)$ is the modified parameter. According to Uomoto (1992), the carbonation depth is related to the square root of the CO2 concentration, and the parameter k is therefore defined by:

$$k(t) = \frac{\sqrt{C(t)}}{\sqrt{C_{t0}}} \tag{18}$$

where C_{t0} is the CO_2 concentration in the study of Kishitani (1991), and the value of $C(t)$ can be obtained from the Japan Meteorological Agency (JMA, 2019). Figure 8 shows the CO_2 concentration over time in Hokkaido.

Accordingly, the carbonation depth of a concrete structure that served for T years is modified as:

$$d = \int_0^T \frac{k(t)\cdot A}{2\sqrt{t}}\, dt \tag{19}$$

If $k(t)$ is a constant value, Equation (19) is simplified as Equation (17).

3.1.2 Inspection results

The deterioration grades of bridges are used to understand the conditions of the bridges.

Based on the Inspection Guidelines (MLIT, 2014), bridges are required to be visually inspected every five years. The results of inspections are categorized by the engineer as: grade 1: healthy; grade 2: preventive maintenance required; grade 3: prompt action required; and grade 4: emergency action required. The grades and corresponding descriptions are given in Table 3. Since grade 4 is considered a critical condition, meaning that these bridges have to be subjected to a repair timely, this philosophy results in the infrequent occurrence of grade 4. Therefore, grade 3 is considered to be the upper threshold, because grade 3 indicates the necessity of early actions. This consideration urges us to build a predictive model to predict the situation before an emergency occurs.

3.1.3 Data characteristics analysis

Twelve factors that have the potential to influence deterioration, which are listed in Table 1, are investigated. All factors are scaled to the range of 0-1 according to Equation (1). Analysis of variance (ANOVA) is an analysis tool used in statistics to observe whether the independent variables significantly impact the dependent variable. Therefore, the ANOVA test is used to analyze factors that affect deterioration. The results are listed in Table 4. A *p*-value of 0.05 or less means that a factor is considered to influence deterioration significantly. For time series factors, such as the temperature, snowfall, and rainfall, different samples are tested, and the smallest *p*-values are summarized in Table 4. The calculated *p*-values show that all factors would significantly influence deterioration; therefore, all factors are considered in the deterioration model.

3.2 Model establishment

A further overview of the steps taken to perform deterioration prediction is given in Figure 9. In the training phase, we first converted the influencing factors from the inspection database to vectors and normalized the vectors to the range of 0-1. The input vectors are then applied to train a deterioration prediction model using the LSTM-based RNN. The model is optimized by iteratively updating the parameters in Equations (3) to (8) (Wang et al., 2018). In the prediction phase, the testing data of a bridge are first converted into normalized vectors. The vectors are then introduced into the trained model, which will calculate the grade of deterioration. According to previous studies (Wang et al., 2018; Lipton et al., 2016), the LSTM model is trained on 70% of the bridge data and validated on 15% of the bridge data, and the remaining 15% is used as the testing set. The training algorithm of the adaptive moment estimation (Adam) is adopted. The optimal configuration of the prediction model is determined by trial and error (Section 4.1).

3.3 Model evaluation

In a previous study (Miao et al., 2019), a multilayer perceptron (MLP) neural network model was established. Therefore, the performance of the LSTM-based RNN is compared with that of the MLP model in terms of the recall, precision, accuracy, F1 score, and TNR. The definitions of these metrics are illustrated taking grade 1 as an example, as indicated in Figure 10. True positive (TP) is when both the prediction and truth are grade 1. False negative (FN) is when the model identifies actual grade 1 as another grade. True negative (TN) is when both the prediction and truth are other grades. False positive (FP) is when the model identifies another grade as grade 1. A model is considered to have a robust predictive ability when it has larger values for the recall, precision, accuracy, F1 score, and TNR. Since a perfect prediction

model is usually inaccessible, the predictions for some bridges will inevitably fail. Therefore, error analysis of the LSTM-based RNN will be performed to assess the model's performance (Section 4.2).

4. Results and discussions

Using the inspection database, we explored the ability of the LSTM-based RNN to model the relationships between potentially influencing factors and bridge deterioration. The optimal configuration of the LSTM model is determined by trial and error, as summarized in Table 5. The accuracy of the model is first tested using 50 to 200 hidden units. The results show that 150 hidden layers yield better results for this database. Different batch sizes are then tested with 150 hidden units as a second step to determine the LSTM configuration. The results indicate that the LSTM model performed better when the model had 150 hidden units and a batch size of 1024.

The training and validation iterated 6000 times and cost about 12 min in total. The error between the predicted value and the ground truth is evaluated using the cross-entropy. As shown in Figure 11, the accuracy converges to approximately 80%, and the error converges to approximately 0.35. The accuracy and the error show that the established model is some distance from perfection (100% accuracy and zero error). Future endeavors are necessary to improve the established model by actions described in Section 4.3.2.

As twelve factors are considered to affect the deterioration, the relationship between the affecting factors and the deterioration grades is a 13D model. To intuitively reflect this relationship determined by the LSTM model, Figure 12 shows the mapping between the affecting factors and the deterioration grades. The results in Figure 12 and the performance of

the model in Section 4.1 reveal that the complicated relationships are preliminarily established.

4.1 Performance of the model

The performance of the LSTM model is first compared with that of the MLP model. Figure 11 shows that the accuracy of the LSTM model exceeded 80%. As a comparison, an MLP model obtained an accuracy of 65% on the same database (Miao et al., 2019). Obviously, the prediction accuracy of the LSTM-based RNN model exceeds that of the MLP model by approximately 17%.

In addition, the recall, precision, TNR, and F1 score of the two models are calculated, as shown in Figure 13. These four indexes show that the LSTM model obtained greater values regardless of the grades. Therefore, it can be concluded that the LSTM model outperforms the MLP model. Through comparative research of the LSTM and MLP models, the results show that the LTSM model yields considerable improvement and has a relative superiority compared to the MLP model. However, it should be remembered that the LSTM architecture is much more complicated. In other words, although an LSTM-based RNN prediction model is more robust, the complex architecture and calculations make it difficult for engineers to establish such a model easily.

Furthermore, the model is applied to PC, RC, PC and RC hybrid, and steel bridges (mainly concrete components) to evaluate the predicted performance of the model for corresponding types of bridges. Figure 14 shows that the model has equivalent performance regardless of the type of bridge. Similarly, the performances of the LSTM model for bridges in coastal and non-coastal regions are evaluated, as shown in Figure 15. The results show that the model's

predictive ability for coastal bridges is slightly superior to that for the non-coastal bridges. One possible reason is that de-icing salt is not considered for non-coastal bridges because relevant data cannot be collected. The accuracies of the model versus the deck area are also assessed, as shown in Figure 16. The LSTM model shows no significant difference in accuracy between different deck areas. The reason is that the established model is unbiased for the different deck areas, although the model is not yet perfect.

4.2 Error analysis

4.2.1 Distribution of incorrect predictions

The percentages of incorrect predictions made by the LSTM model in various situations are calculated, as summarized in Table 6. For example, the situation where an actual grade 1 bridge is predicted to be grade 2 accounts for 17.3% of all incorrect predictions. As shown in Table 6, predicting actual grade 2 bridge to be grade 1 and predicting actual grade 1 bridges to be grade 2 accounts for more than half of all incorrect predictions. The confusion between actual grade 2 and predicted grade 3 is another major source of incorrect predictions. The percentages of incorrect predictions in other conditions do not show noticeable differences.

For the dark marked area in Table 6, the predicted grades are usually larger than the actual grades, which means that the predictions are aggressive. Under aggressive conditions, timely maintenance can be conducted to ensure the soundness of the bridge, but this will increase costs. For the non-dark marked areas, the predicted grades are smaller than the actual grades, which means that the predictions are conservative. If suggestions under conservative conditions are followed, the timely maintenance of the bridge may be overlooked. Aggressive

and conservative situations accounted for 45.7% and 55.3% of all incorrect predictions, respectively. It is believed that using more abundant information (such as quantitative deterioration) will assist in overcoming this challenge and yield fewer incorrect predictions.

4.2.2 Related factors

In addition, we also assessed which factors result in incorrect predictions for these bridges. Among all considered factors in this study, other factors are not necessarily related to incorrect predictions except for the factor of years in service. The relationships between the numbers of incorrect predictions versus the years in service are shown in Figure 17. If incorrect predictions are divided into two parts by 30 years in service, the results show that almost 85% of all incorrect predictions are bridges that have been in service for more than 30 years.

Furthermore, the inspection database is divided into two groups. One group is aged from 0 to 30 years, and the other is aged 31 and above. Then, the percentages of incorrect predictions in the corresponding groups are calculated, as shown in Figure 18. Among the first group, incorrect predictions account for 7.01%, while that rate is up to 29.08% for the second group. Interventions are thought of as the main reasons that lead to the differences between the two groups. Because interventions are mainly implemented for bridges that have been in service for more than 30 years, but relatively few interventions are implemented for bridges that are less than 30 years old. In addition, the values of some factors 30 years ago are estimates. The observed values in the past three decades are more convincing given that the improvement of observation methods and techniques. Therefore, caution should be exercised when making predictions for bridges that have been in service for more than 30 years.

4.3 Discussions

The discussions of the model elaborate on the following: (1) how the predictive capacity of the model assists in improving the maintenance of structures and (2) the challenges/insufficiency of the model and future work.

4.3.1 Usefulness

The results indicated that the prediction model could guide the quality inspection of existing structures. Specifically, provided the time series of all factors for a bridge, future deterioration can be estimated. Figure 19 shows the time series diagram of all factors for an RC bridge. The latest inspected deterioration grade is 1.

Using the prediction model, the deterioration conditions of the bridge in the next fifteen years are predicted, as shown in Figure 20. It can be inferred from the predicted results that the condition of the bridge in the next five years will not deteriorate significantly. Therefore, the inspection of the bridge can be postponed. Of course, if the bridge deteriorates significantly in the next five years, the bridge can be inspected in advance, and corresponding intervention measures can be taken. The prediction model makes it possible for the decision-makers to formulate a flexible inspection schedule according to the deterioration situation.

In addition, the initial achievements in this study provided a case to explore the deterioration prediction for other infrastructures. Then, we can integrate all these deterioration prediction models into the infrastructure management system (IMS) to enable managers or local authorities to manage and formulate intervention strategies for

infrastructure. Furthermore, the limitations revealed in this study (Section 4.3.2) will prompt the collection of more detailed data during the inspection process in the future.

4.3.2 Insufficiency

Our proposed model focuses on making accurate and robust predictions relying on time series data of many potential factors. Therefore, the outcome of the model is closely related to the data quality of each factor. If the information of some factors is missed/incomplete, or the inherent relation between factors and deterioration is not clear, our model may achieve limited predictions or even fail. This means that it is necessary to check the quality of the data for each factor. In addition, although our proposed model considered twelve potentially influencing factors, it does not mean these factors are necessarily related to deterioration, or other factors that are not considered in this model may also impact deterioration. One good thing is that we can include more influencing factors or remove some factors by updating the rows of the matrix X in Section 2.1 and by modifying the configuration (size of input x_t) of the LSTM model (Section 2.2).

Although the performance of the model is improved relative to an MLP model, the architecture of the LSTM-RNN model is much more complicated than that of the MLP. In addition, more data are needed to train an LSTM-RNN model. Therefore, it takes a long time to collect data to establish a model such as that in this study considering the availability of the database, especially in practical engineering. In the calculation of chloride ions penetrate into the concrete, the diffusion coefficient D_c is considered a constant. This assumption is inconsistent with the actual situation, which makes the LSTM model deviate from reality to a certain extent. In addition, error analysis found that the model is prone to be confused

between grades 1 and 2 bridges, because the misclassification between these two grades accounts for half of the failed predictions. It should be cautious when applying the model for predicting conditions of bridges that are more than 30 years old. Because incorrect prediction increases greatly when applying the model for predicting conditions of bridges that are more than 30 years old.

Furthermore, the proposed model is evaluated by using a database containing data from past inspections. However, a good prediction model should be generalizable. This means that the model in the study should fit future observations well but cannot be verified now. However, the requirements will motivate us to improve data collection strategies to acquire more valuable data. In addition, other time series approaches such as the Bayesian approach will be verified and compared with the model in this study in the future.

5. Conclusions

In this study, an LSTM model is established using time series data of many factors in the bridge inspection database, enabling earlier maintenance to be conducted by predicting deterioration conditions. From the verifications and comparisons, the following conclusions can be drawn:

(1) Using an inspection database containing time series of various influencing factors, the LSTM model can assess these factors' cumulative effects, and can establish relationships between potential factors and deterioration grades. In practical applications, the model can provide the future deterioration path for a specific bridge, given time series on these factors. Accordingly, a postponed or advanced maintenance schedule can be formulated. This way of making decisions ahead of emergencies will help decision-makers improve the quality of

maintenance management.

(2) The obtained LSTM model has an accuracy of more than 80%. In addition, the performance of the LSTM model outperforms an MLP model in predicting deterioration regarding indexes of the TNR, F1 score, recall and precision. The reason for the better performance of the LSTM model is that the consideration of the time dependence of time series. In addition, the LSTM model shows the equivalent performance for the four types of bridges. The performance of the model for coastal bridges is slightly superior to that for non-coastal bridges. The model show no significant differences in accuracy are determined between different deck areas.

(3) Error analysis found that the confusion of grades 1 and 2 bridges accounts for half of the failed predictions. In addition, the results show that the factor of years in service is closely related to the percentage of incorrect predictions. Bridges that have been in service for more than 30 years are prone to obtain incorrect predictions because interventions are implemented on those bridges but cannot be considered.

Although the LSTM model produces better results than MLP, the LSTM architecture is much more complicated. In addition, a long-term maintenance database on various factors is needed to train an LSTM model. In practice, this is a huge project because the values of some factors are only collected every five years according to the inspection guideline. Furthermore, it must be admitted that our model is not perfect for the reasons described in Section 4.3.2. In the future, the performance of the LSTM model will be improved by updating the model with more high quality data, and quantifying the degradation grade. Other time series modeling approaches will also be tried.

References

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers and Industrial Engineering, 137(September), 106024. doi: 10.1016/j.cie.2019.106024

Chandra, R., & Zhang, M. (2012). Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction. In Neurocomputing (Vol. 86, pp. 116–123). doi: 10.1016/j.neucom.2012.01.014

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. https://arxiv.org/abs/1406.1078

Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361–370. doi:10.1093/jamia/ocw112

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks*, *5*(2), 240-254.

Daisoku, N., Katawaki, K., Miyagawa, T., Kashino, N., Kobayashi, A. (1986). Durability series of concrete structures. Salt Damage (I). Gihodo Publishing.

Dorffner, G. (1996). Neural networks for time series processing. Neural Network World, 6(4), 447–468.

Giles, C. L., Lawrence, S., & Tsoi, A. C. (2001). Noisy time series prediction using recurrent neural networks and grammatical inference. Machine Learning, 44(1–2), 161–183. doi:10.1023/A:1010884214864

Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Computers and Industrial Engineering, 42(2–4), 371–375. doi:

10.1016/S0360-8352(02)00036-0.

Hochreiter, S., & Urgen Schmidhuber, J. (1997). Long Shortterm Memory. Neural Computation, 9(8), 17351780.

Jeong, S., Ferguson, M., Hou, R., Lynch, J. P., Sohn, H., & Law, K. H. (2019). Sensor data reconstruction using bidirectional recurrent neural network with application to bridge monitoring. Advanced Engineering Informatics, 42(August), 100991. doi: 10.1016/j.aei.2019.100991.

JMA. (2019, October). Annual changes of carbon dioxide concentration by locations. Retrieved from https://ds.data.jma.go.jp/ghg/kanshi/co2timeser/co2timeser.html.

JSCE. 2012. Standard Specifications for Concrete Structures (Design). Japan Society of Civil Engineers, JSCE Guidelines for Concrete.

Kawakami, H., Waki, K.,Tada, M.1995. Evaluation of durability of old building concrete-analysis of investigation results. Fukui University,Faculty of Engineering Research Report, 43(1):51-58.

Kishitani,K., Kobayashi, K., Kashino,K. Uno,Y. 1991. Relationship between rebar corrosion and carbonation in concrete containing chloride. Concrete engineering proceedings. 2(1), 77-84. doi:10.3151/crt1990.2.1_77.

Kobayashi, R. (2010). Field survey on salt infiltration of existing bridges along the coast of Oshima Peninsula, Hokkaido. Concrete engineering proceedings. 48(7): 52-55. doi:10.3151/coj.48.7_52

Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2016). Learning to diagnose with LSTM recurrent neural networks. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 1–18.

Miao,P., Yokota, H., Zhang,Y. (2019,September, 10-11). Prediction-based maintenance of concrete structures by using an artificial neural network. 3rd ACF Symposium: Assessment and intervention of existing structures. Sapporo, Japan.

MLIT. (2005). 2005 National road and street traffic situation survey (general traffic volume survey). Retrieved from http://www.mlit.go.jp/road/census/h17/

MLIT. (2010). 2010 National road and street traffic situation survey (general traffic volume survey). Retrieved from http://www.mlit.go.jp/road/census/h22-1/

MLIT. (2014). Guidelines for Regular Inspection of Road Bridges. Retrieved from https://www.mlit.go.jp/road/sisaku/yobohozen/tenken/yobo4_1.pdf.

MLIT. (2015a). 2015 National road and street traffic situation survey (general traffic volume survey). Retrieved from http://www.mlit.go.jp/road/census/h27/index.html

MLIT. (2015b) Guidelines for general traffic survey. www.mlit.go.jp/road/census/h22-1/data/kasyorep.pdf, Access date: October, 2019.

NRA. 2016, March. Investigation and research on long-term soundness evaluation of concrete structures. Retrieved from https://www.nsr.go.jp/data/000186088.pdf.

Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A dual-stage attention-based recurrent neural network for time series prediction. IJCAI International Joint Conference on Artificial Intelligence, 2627–2633. doi: 10.24963/ijcai.2017/366

Razavi, S. V., Jumaat, M. Z., El-Shafie, A. H., & Ronagh, H. R. (2015). Load-deflection analysis prediction of CFRP strengthened RC slab using RNN. Advances in Concrete Construction, 3(2), 91–102. doi:10.12989/acc.2015.3.2.091

Stenström, C., Norrbin, P., Parida, A., & Kumar, U. (2016). Preventive and corrective maintenance–cost comparison and cost–benefit analysis. Structure and Infrastructure Engineering, 12(5), 603-617. doi:10.1080/15732479.2015.1032983

Tamakoshi, T., Kubota,K., Hoshino,M., & Yokoi, Y. 2012. Examination about countermeasure of salt damage to

concrete bridge based on result of close visual inspection. MLIT. ISSN 1346－7328.No.711: 15-21.

Tamakoshi, T., Okubo, M., & Watanabe, Y. 2009. Research on highway bridges management-Bridge Manage System. MLIT, ISSN 1346-7328.No.523:25-27.

Uomoto, K., & Takada, Y.1992. Factors affecting the neutralization rate of concrete. JSCE Proceedings, 1992(451):119-128. doi:10.2208/jscej.1992.451_119

Wang, D., Fan, J., Fu, H., & Zhang, B. (2018). Research on optimization of big data construction engineering quality management based on RnN-LSTM. Complexity, 2018. https://doi.org/10.1155/2018/9691868

Table 1. Potentially influencing factors

	Attributes (units)
Bridge geometry factors	Length (m) width (m)
Environment factors	Elevation (m) Yearly highest and lowest temperatures (°C) Carbon dioxide concentration (ppm) Airborne salt concentration (mdd•NaCl), Yearly average snowfall (cm) Yearly average rainfall (cm)
Loading condition	Daily traffic volume (vehicles/day)* Rate of the large-size vehicles (%/day)
Bridge age	Years in service (years)

* Statistics on traffic volume include large and small vehicles (MLIT, 2015b).

Table 2. Suggested materials information for bridge components (Tamakoshi, 2009)

Built Year	Item	Salt damage countermeasure category	Structure type		Pier	Abutment
			RC	PC		
Before 1983	Protective Cover(cm)	−	3.5	2.5	7	7
	W/C (%)	−	55	35	60	60
After 1983	Protective Cover(cm)	0~100m*	7	7	7	7
		100~300m	7	5	7	7
		300~500m	5	3.5	7	7
		others	3.5	2.5	7	7
	W/C (%)	−	50	35	55	55

* 0~100m denote distance from a coastline.

Table 3. Deterioration grades and corresponding descriptions (MLIT, 2014).

Grade	Conditions	Descriptions
1	Healthy	A state that the function of the structure is not disturbed.
2	Preventive action required	Although the function of the structure is not hindered, it is desirable to take measures from the viewpoint of preventive maintenance.
3	Early action required	The function of the structure has interfered and measures should be taken as soon as possible.
4	Emergency action required	A condition in which the function of the structure has been or is likely to be impaired, and measures should be taken urgently.

Tabel 4. Scaled values of each influencing factor and corresponding p-values

Factor	Original data range	After scaled		*p*-value
		Mean	Standard deviation	
Elevation (m)	-0.3~1,106	0.0909	0.1347	8.2501×10^{-7}
Bridge length (m)	2~1,433	0.0444	0.0844	1.3558×10^{-22}
Bridge width (m)	0.7-50	0.2129	0.1124	6.8245×10^{-8}
Years in service (years)	1~84	0.3833	0.2008	2.5555×10^{-62}
Carbon dioxide (ppm)	348.86~409.87	0.3933	0.2413	6.8759×10^{-40}
Chloride (kg/m^3)	0~11.79	0.0490	0.1238	8.4375×10^{-10}
Traffic volume (vehicles/day)	0~56,874	0.0946	0.1143	7.8764×10^{-14}
Large-size vehicle rate (%/day)	0~62.6	0.3076	0.1786	6.9693×10^{-35}
Rainfall (cm)	776~1,491	0.3617	0.2268	3.6774×10^{-14}
Snowfall (cm)	12.0~1,263.8	0.4493	0.2166	1.8851×10^{-13}
Highest temperature (°C)	16~26	0.7025	0.1737	2.5443×10^{-14}
Lowest temperature (°C)	-11~-3	0.5851	0.2361	4.4760×10^{-13}

Table 5. Configuration test of LSTM model

LSTM type	Accuracy rate (%) (training/testing)	Average accuracy rate (%) (training/testing)
	Number of hidden units	
50	66.2/69.3 67.3/66.9 67.4/69.5	66.97/68.57
100	70.7/72.4 69.1/69 70.4/69.3	70.07/70.23
150	81.19/80.8 80.06/80.5 80.86/81.08	80.70/80.79
200	75.2/72.2 77.3/72.4 73.8/72.7	75.43/72.43
	Batch size	
256	55.2/54.7 55.18/54.69 55.23/53.2	55.2/54.19
512	68.1/64.4 67.8/66.6 70.1/67	68.67/66.0
1024	81.19/80.8 80.06/80.5 80.86/81.08	80.70/80.79

Table 6. Percentages of incorrect predictions.

Actual grade \ Predicted grade	1	2	3
1	-	17.3%	10.4%
2	34.4%	-	17.0%
3	10.7%	10.2%	-

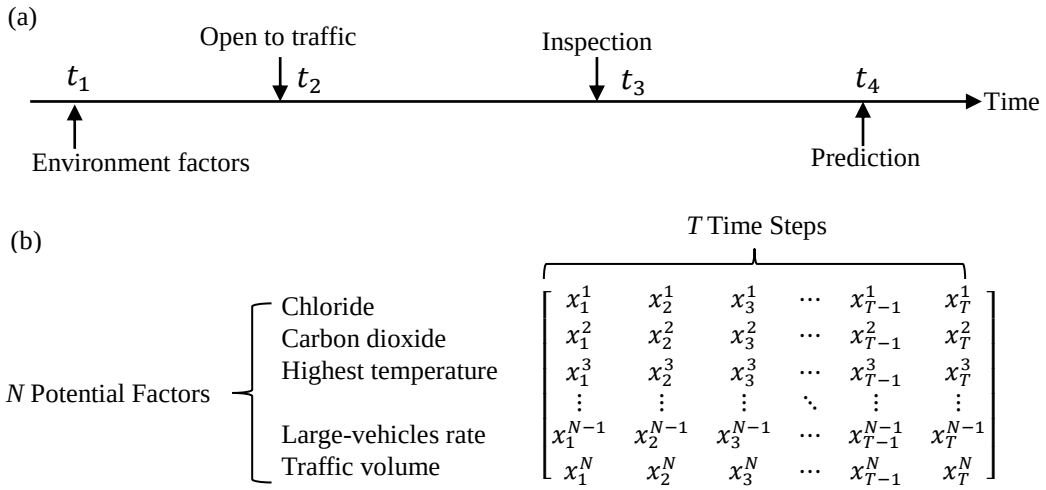

Figure 1. (a) t_i indicates the time at which the event occurs. (b) Data representation of a bridge with *N* potential factors and years in-service *T*.

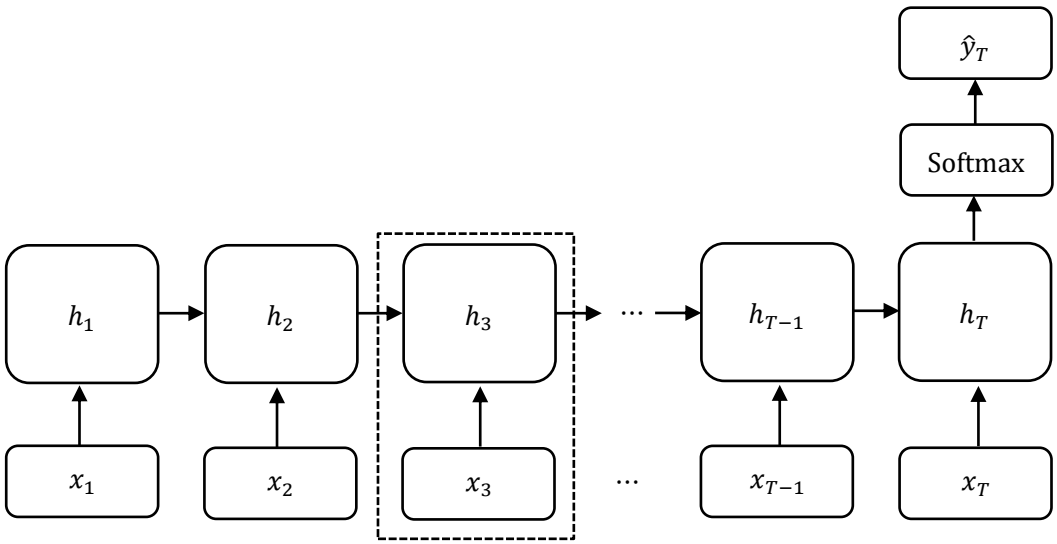

Figure 2. A schematic RNN model for deterioration prediction.

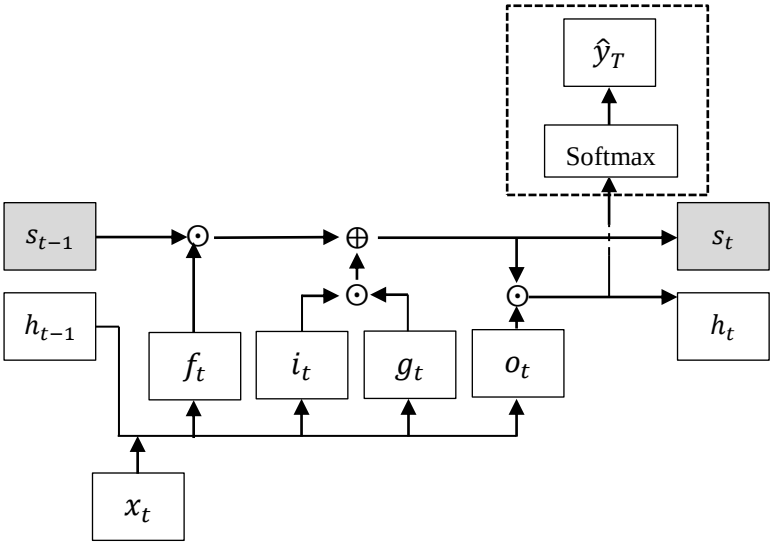

Figure 3. Configuration of the LSTM unit.

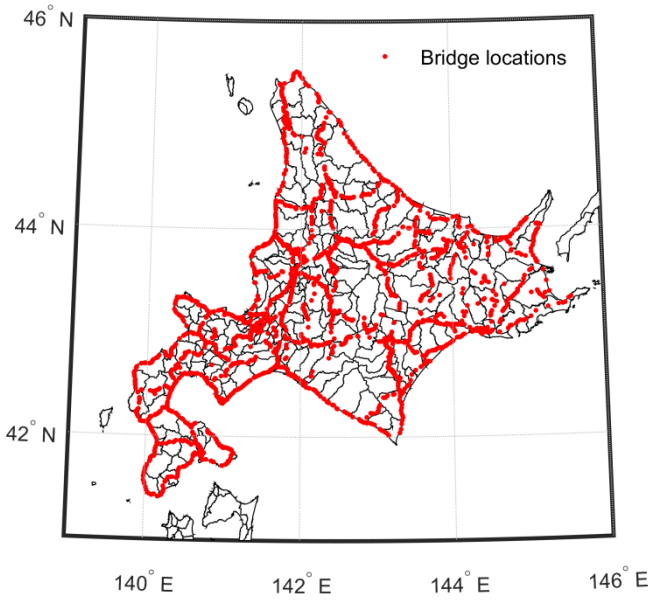

Figure 4. Locations of the bridges.

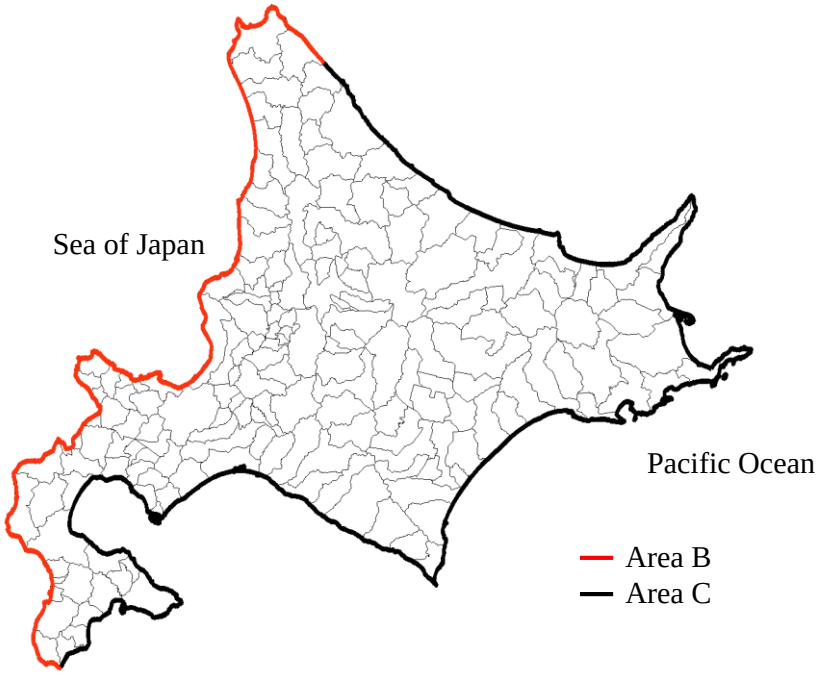

Figure 5. Salt damage area classification in Hokkaido

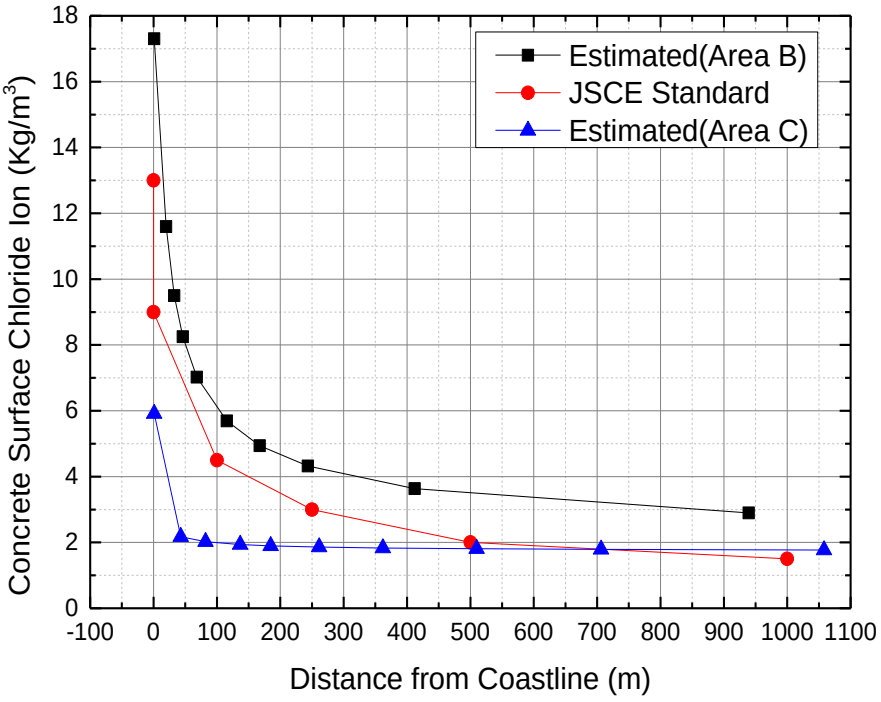

Figure 6. Comparison of estimated chloride ion concentration and the JSCE standard.

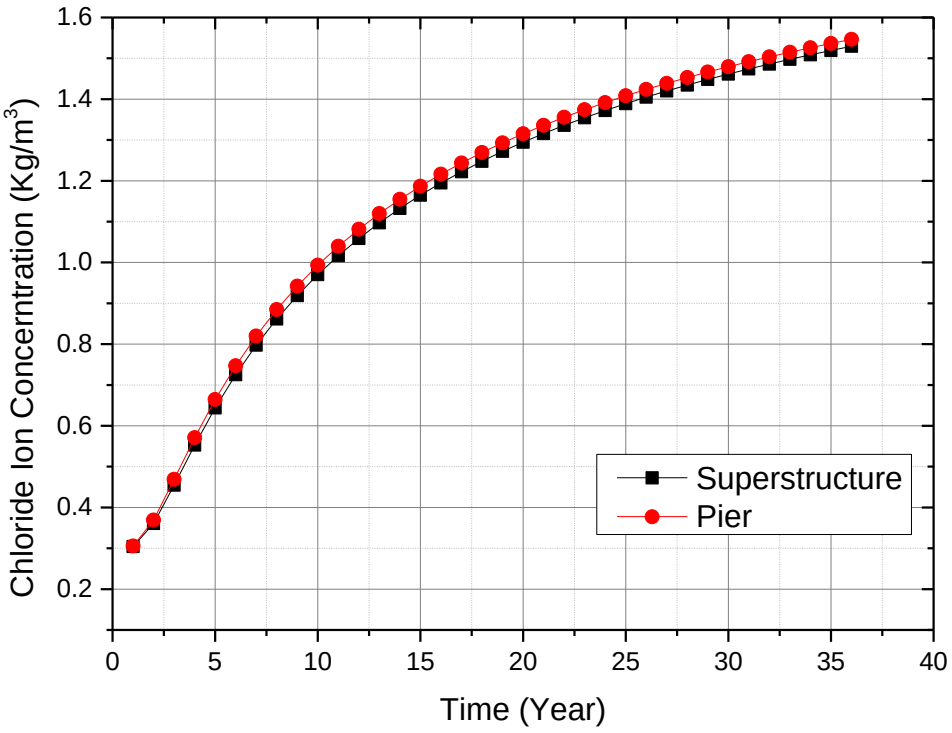

Figure 7. Chloride ions concentration for superstructure and pier

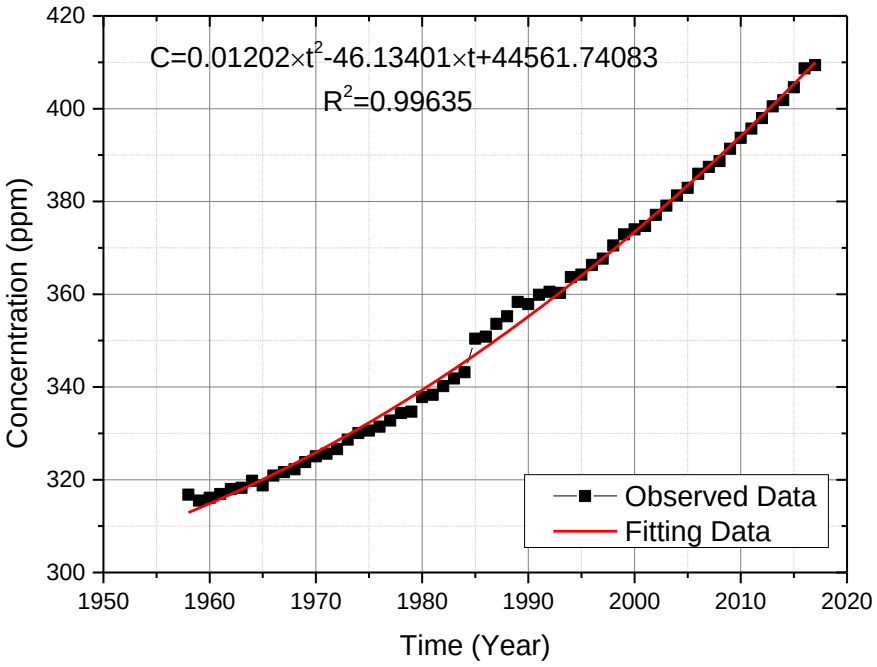

Figure 8. Carbon dioxide concentration versus time

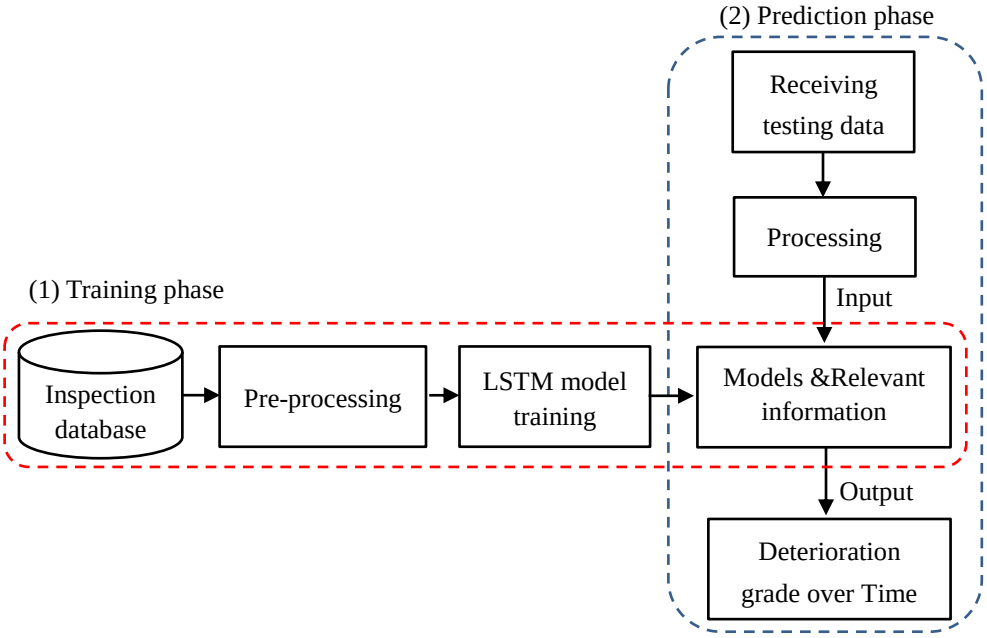

Figure 9. Flowchart of the whole procedures

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$TNR = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{ACC} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{F1} = \frac{2 * Recall * Precision}{(Recall + Precision)}$$

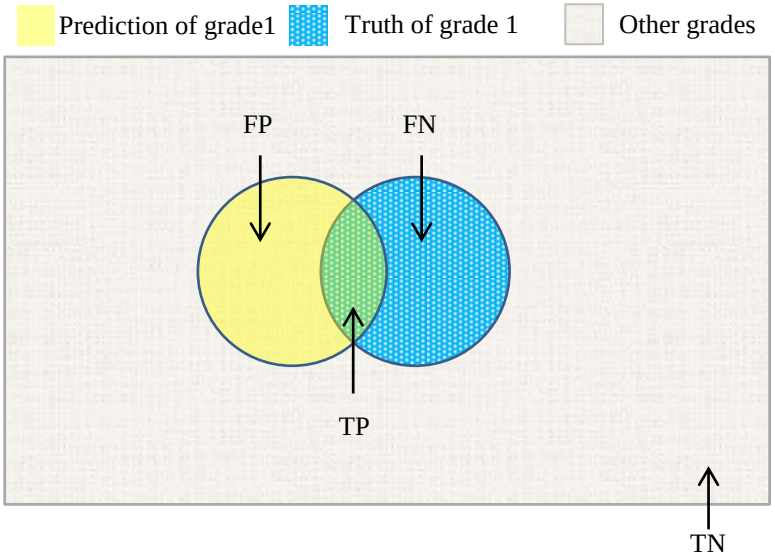

Figure 10. Performance evaluation metrics used in this study (Taking Grade 1 for example)

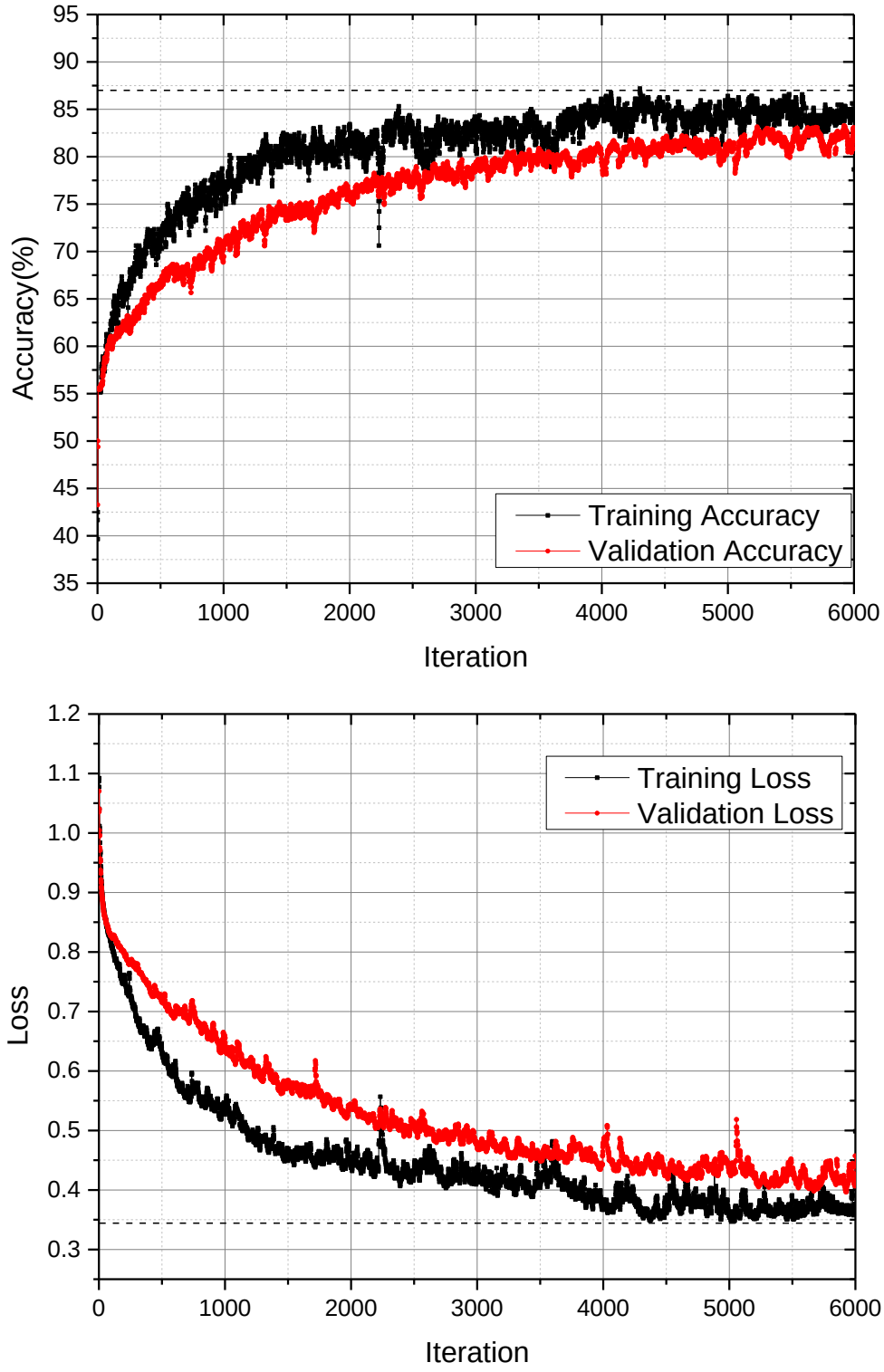

Figure 11. Accuracy and loss with respect to iteration.

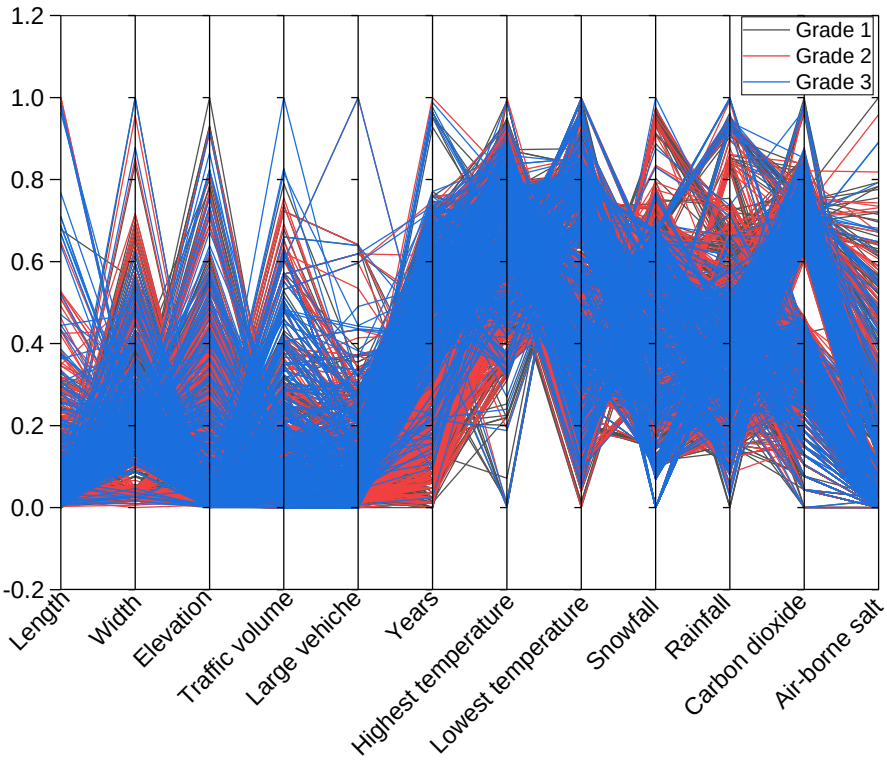

Figure 12. Mapping between affecting factors and deterioration grade

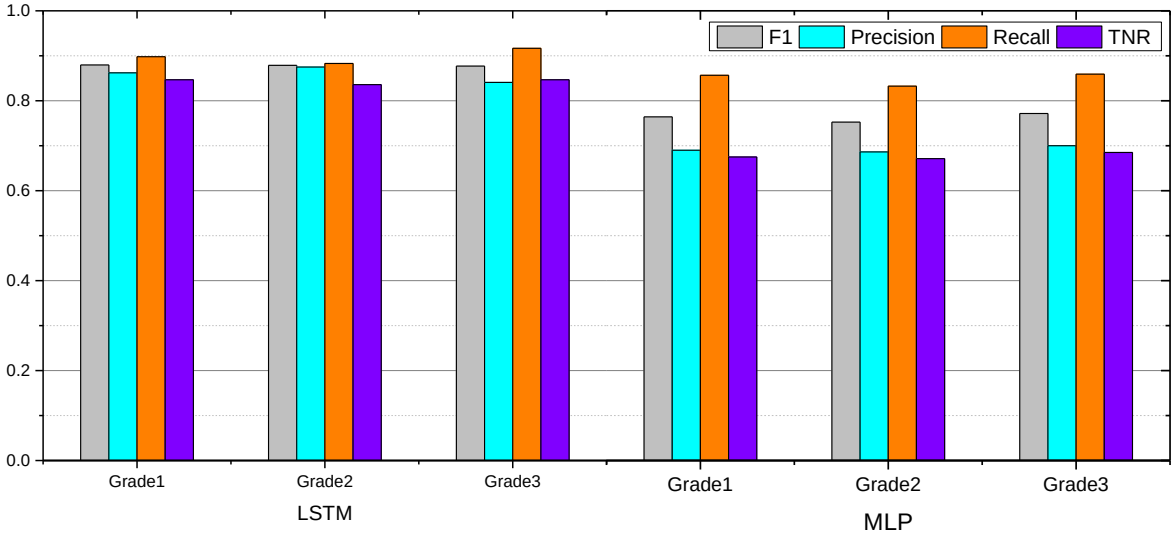

Figure 13. Comparison of the LSTM and MLP models.

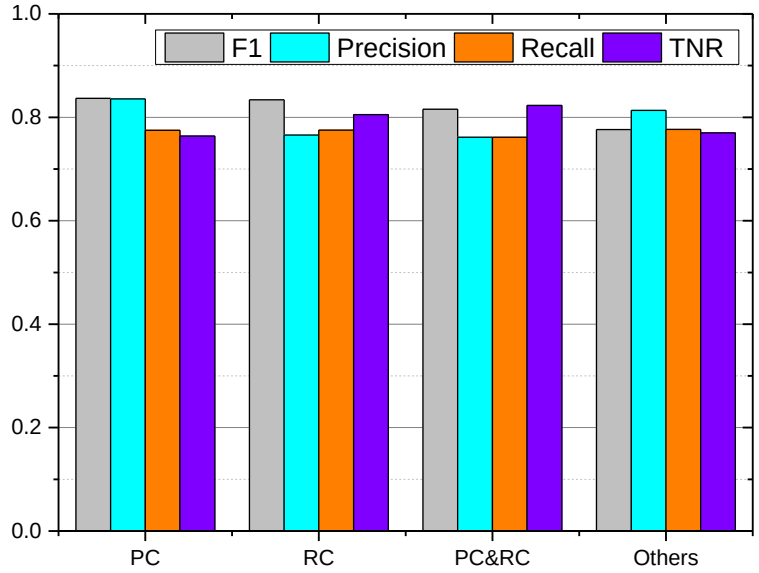

Figure 14 Performance of the LSTM for the PC, RC, PC&RC, and others (steel slab)

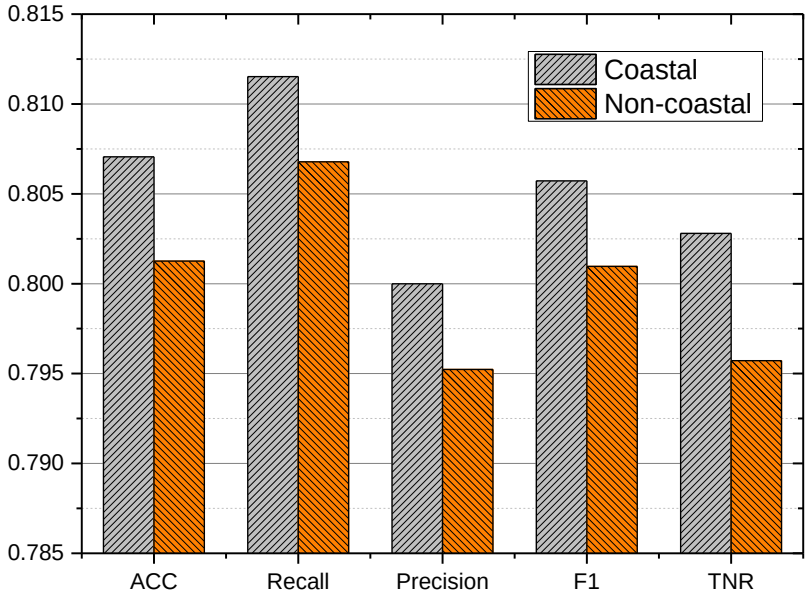

Figure 15. Performance of the LSTM for the costal and non-coastal bridges

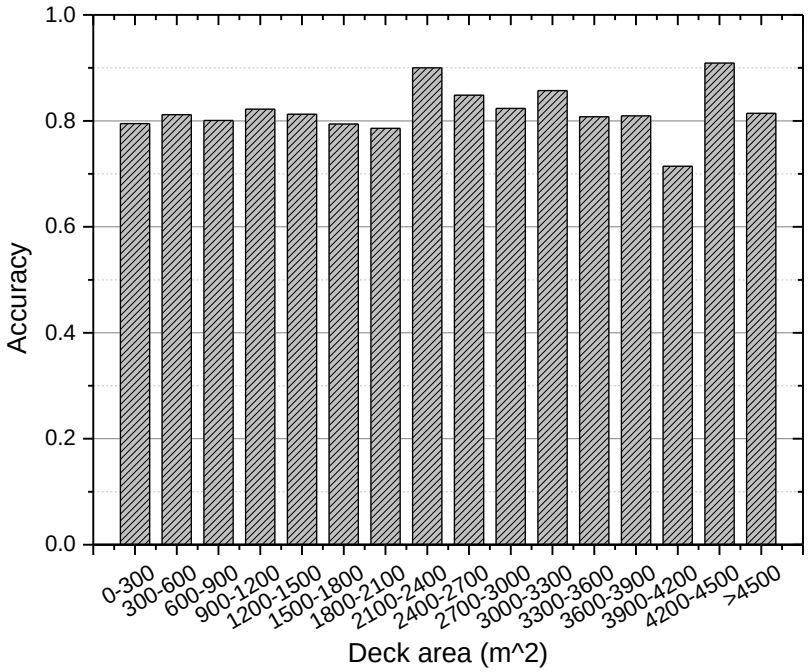

Figure 16. Accuracy of the LSTM versus the deck area

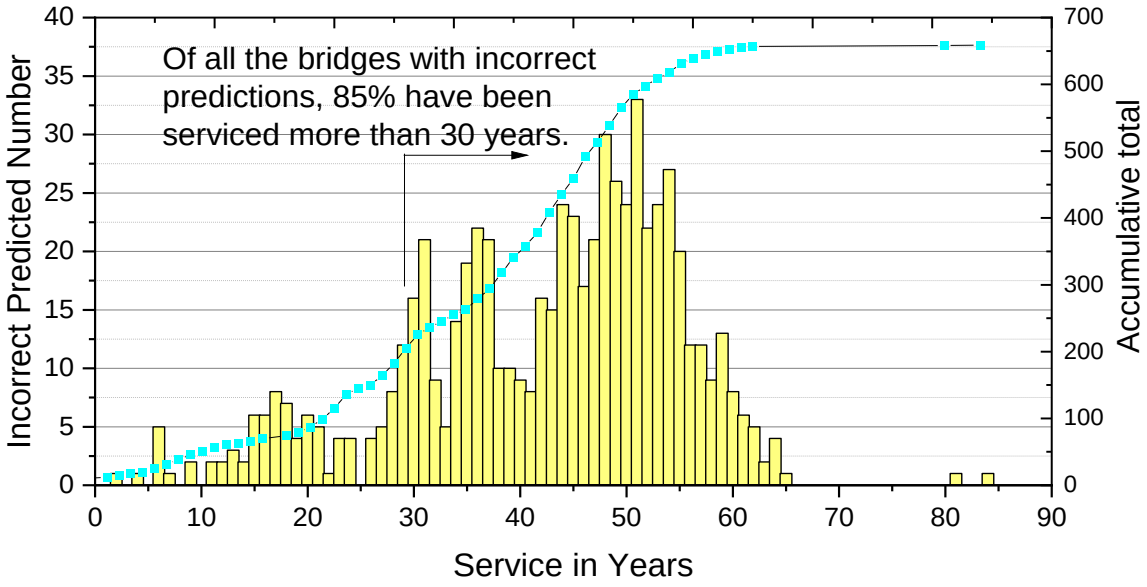

Figure 17. Numbers of incorrect predictions versus years in service

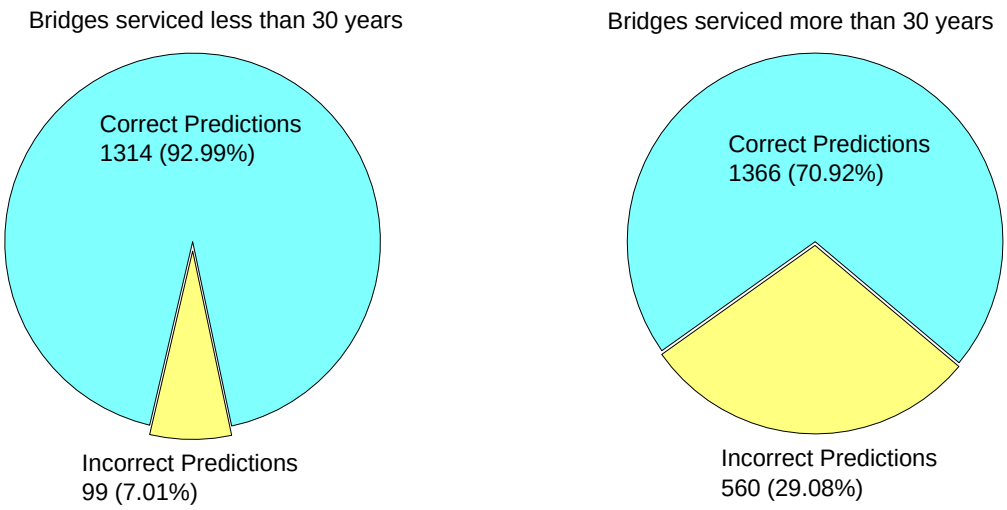

Figure 18. Percentage of incorrectly predictions (Dividing by 30 years)

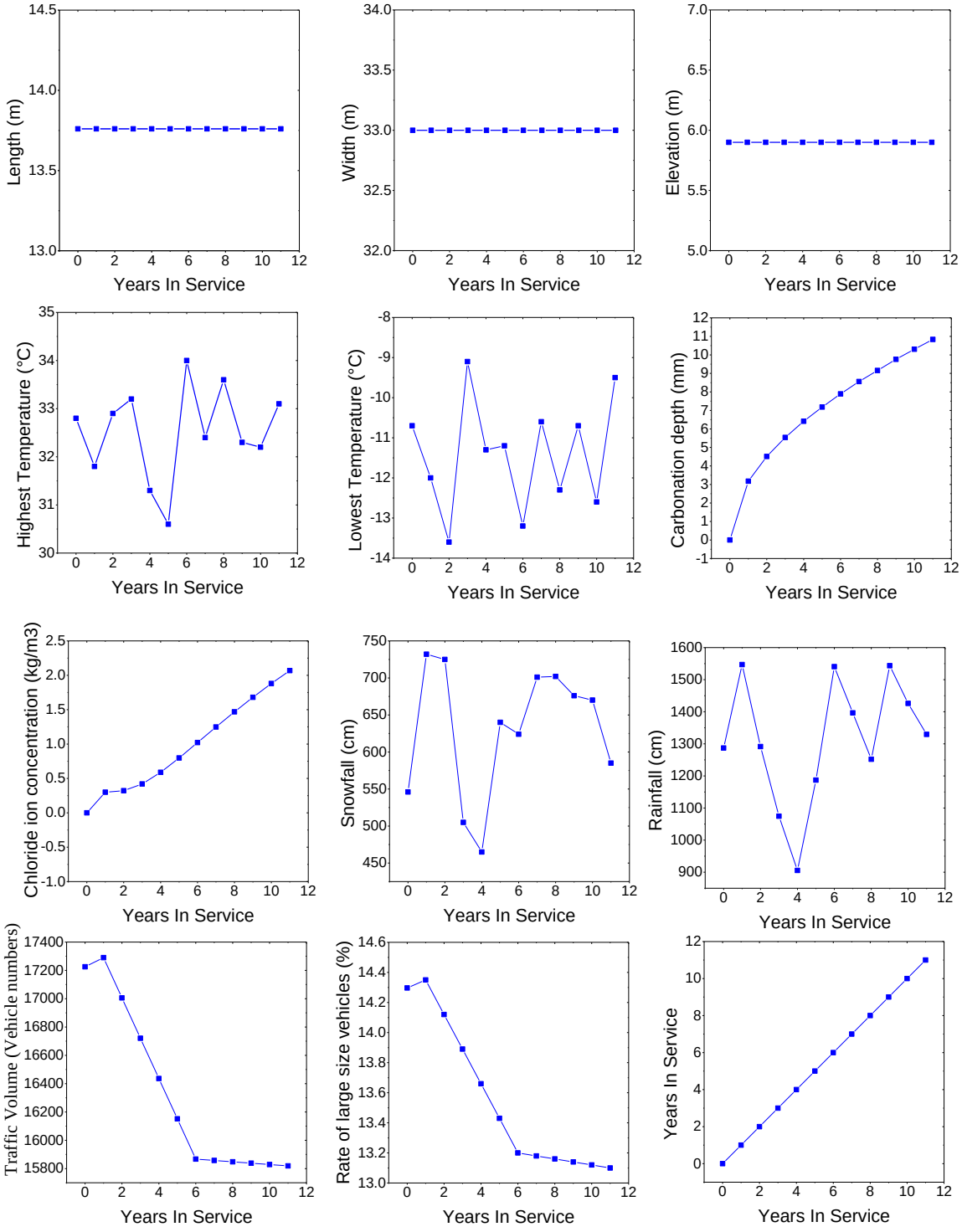

Figure 19. Time series diagram of all factors.

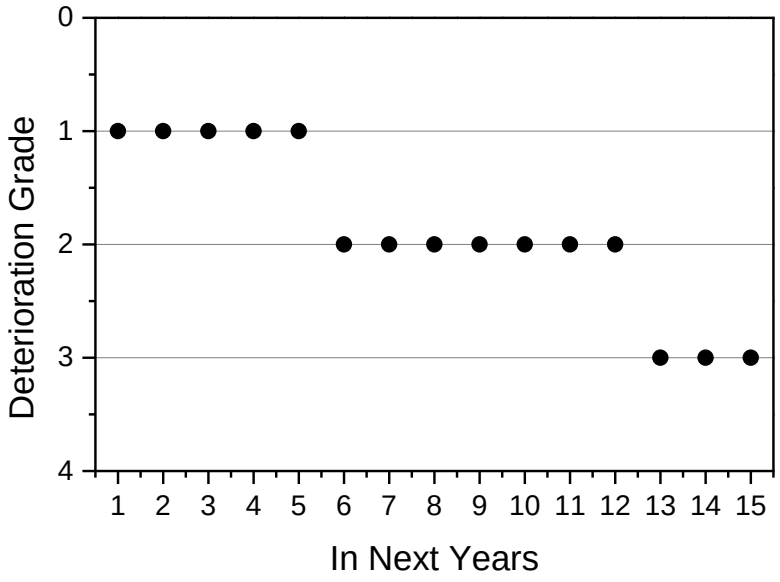

Figure 20. Prediction of deterioration grade in the next fifteen years.