

Article

The Next Failure Time Prediction of Escalators via Deep Neural Network with Dynamic Time Warping Preprocessing

Zitong Zhou ¹ , Yanyang Zi ^{1,*}, Jingsong Xie ², Jinglong Chen ¹ and Tong An ¹

¹ State Key Laboratory for Manufacturing and Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China; tongzhou925@stu.xjtu.edu.cn (Z.Z.); jlstrive2008@163.com (J.C.); antong0531@stu.xjtu.edu.cn (T.A.)

² School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China; jingsongxie@csu.edu.cn

* Correspondence: ziyi@mail.xjtu.edu.cn

Received: 15 July 2020; Accepted: 12 August 2020; Published: 13 August 2020



Abstract: The escalator is one of the most popular travel methods in public places, and the safe working of the escalator is significant. Accurately predicting the escalator failure time can provide scientific guidance for maintenance to avoid accidents. However, failure data have features of short length, non-uniform sampling, and random interference, which makes the data modeling difficult. Therefore, a strategy that combines data quality enhancement with deep neural networks is proposed for escalator failure time prediction in this paper. First, a comprehensive selection indicator (CSI) that can describe the stationarity and complexity of time series is established to select inherently excellent failure sequences. According to the CSI, failure sequences with high stationarity and low complexity are selected as the referenced sequences to enhance the quality of other failure sequences by using dynamic time warping preprocessing. Secondly, a deep neural network combining the advantages of a convolutional neural network and long short-term memory is built to train and predict quality-enhanced failure sequences. Finally, the failure-recall record of six escalators used for 6 years is analyzed by using the proposed method as a case study, and the results show that the proposed method can reduce the average prediction error of failure time to less than one month.

Keywords: failure time prediction; convolutional neural network; long-short term memory; dynamic time warping; escalator

1. Introduction

The escalator is one of the most popular transportation vehicles in densely populated areas. According to the data, almost 2 billion trips are taken using elevators and escalators every day. Once escalator failure occurs, it may lead to a terrible accident [1]. In China, about 40 persons died and thousands of people were injured in escalator accidents every year. According to data statistics from Guangzhou Metro, about 67% of total injuries are escalator-related injuries [2]. Thus, the timely prediction of escalator failure can help the manager and maintenance personnel develop maintenance plans to effectively prevent accidents.

The escalator is a typical human-machine-environment system with many factors involved in safety. One of the links and one personal mistake may cause a serious accident. As a research field related to people's transportation, safety issues have always been the focus of attention [3–5]. An escalator, as a complex system, can cause serious casualties in the event of malfunction. Therefore, the safe operation of escalators deserves in-depth study. Failure and risk prediction are critical to ensuring the safe operation of escalators. When the risk level exceeds a certain safety range, an accident

will occur [6]. Thus, it is necessary to predict the failure time accurately and timely, and arrange maintenance activities in advance, which can effectively reduce the failure rate and reduce accidents. Failure time series modeling analysis can describe the law of failure occurrence by establishing a corresponding mathematical model of related failure data.

Many scholars have done in-depth research on time series analysis in safety research, such as failure maintenance time based on the Weibull distribution [7], shipyard occupational risk assessment based on multivariate regression and genetic algorithm [8], and the Akaike Information Criterion (AIC) based steel mill alarm mechanism [9]. These methods have achieved good effects in the case of large amounts of data. However, for accident-related failure data, it has special features of short samples, non-uniform sampling, and random interference. Traditional data modeling based on statistical laws or deterministic mathematical models requires more labelled data, while data related to safety accidents have typical characteristics as aforementioned, making data modeling problems much more difficult. Neural network technology, as one of the most popular data analysis methods, can build a model containing complex nonlinear relationships without deliberate attention to the mathematical characteristics of the data itself, and has good generalization ability to map the relationship between input and output quickly and effectively, which has been widely used in many fields.

In recent years, with the development of neural network (NN) technology, neural networks have been used in various fields [10–12]. The data modeling by NNs does not deliberately focus on the mathematical characteristics of the data itself but learns to construct the complex relationship between input and output with neurons. Thus, neural networks have good generalization capacity. However, the accuracy of neural networks is sensitive to disturbance in data. In addition, some neural networks only pay attention to the relationship between input and output, ignoring the hidden information of the input themselves, which is unreasonable for time series modeling and prediction. Long short-term memory (LSTM), as one of the recurrent neural networks (RNNs), was proposed to solve the problem of time series prediction [13]. LSTM can analyze and process time series with unknown duration delay. LSTM can improve the memory module of traditional RNNs and avoid the problem where effective historical information cannot be stored for a long time, because of the constant input data [14]. LSTM has also been utilized in time series prediction [15], remaining useful for life prediction [16] and safety analysis [17,18]. However, LSTM does not learn well the characteristics of the data itself for short time series with too few samples. Therefore, this paper considers the short-sequence feature extraction ability of a convolutional neural network (CNN) to extract high-dimensional features, which can help the LSTM unit to understand the data better. In addition, the combination of CNNs and LSTMs in a unified framework has already offered state-of-the-art results in the speech recognition domain [19], health care [20], and power load prediction [21]. In addition, the quality of the data itself makes a big difference to the accuracy of neural network modeling. Meanwhile, considering the short length and random interference in the failure data, it is meaningful to enhance the data quality to improve the prediction accuracy of the neural network before data modeling. Dynamic time warping (DTW) [22], with characteristics of oversampling and similarity matching, is much more suitable for data preprocessing before data modeling, especially for short and interfering failure data. DTW has also been used in many fields [23,24]. For this reason, considering the above-mentioned issues, a new failure time series prediction method for the escalator based on a convolutional long short-term memory neural network combined with dynamic time warping preprocessing (DCLNN) was proposed in this paper.

The remainder of this paper is organized as follows: In Section 2, a time series with high stationarity and low complexity is elected by a comprehensive selection indicator, the main principles of the DCLNN are given, and the diagram of the proposed method for failure time prediction is described. In Section 3, case studies of six escalators are applied with the proposed method and the results are given. Finally, a summary and conclusion are drawn in Section 4.

2. The Proposed DCLNN Method for Failure Time Prediction

The complex working environment and numerous components of the escalator make the failure time data have typical nonlinearity and randomness, which increases the difficulty of data modeling of the escalator failure time series. In this section, an escalator failure time prediction model based on a deep neural network under dynamic time warping preprocessing (DCLNN) will be specifically explained. The DCLNN combined the strategy of DTW pre-processing and CNNLSTM post-learning, and some basic principles about the method are given in the following section.

2.1. The DCLNN Strategy for Failure Time Prediction of Escalator

The accuracy of time series prediction is not only related to the sequence model, but the quality of the time series itself also has a great influence on the prediction result. In this paper, a strategy to combine data quality enhancement with deep neural networks is proposed for time series prediction, which is shown in Figure 1. First, the data sequences are divided into good data and bad data according to a selection indicator; secondly, the bad time series data can be transformed to good data with high quality; finally, the built neural network is used to train and predict the time series.

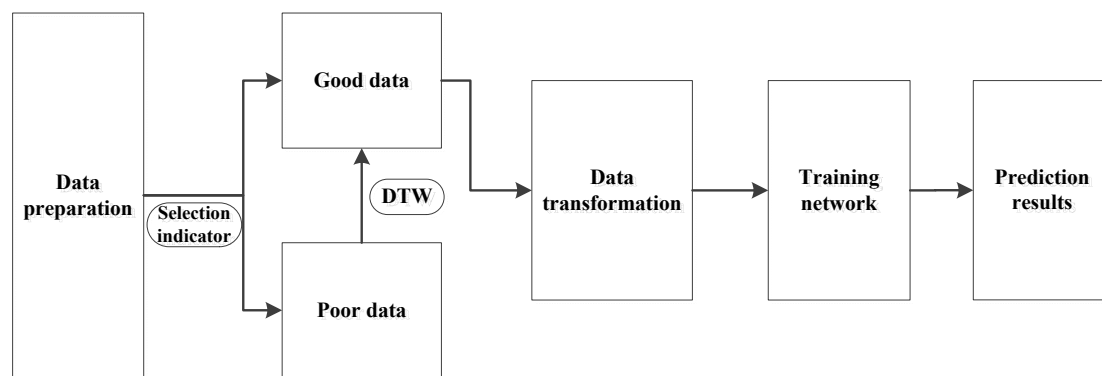


Figure 1. The flowchart of the method proposed in this paper.

Based on the principles of DTW and CNNLSTM, a time series prediction method combining DTW pre-processing and CNNLSTM post-learning is proposed in this paper for escalator failure time prediction, which is called DCLNN.

2.2. Comprehensive Selection Indicator before Data Modeling

In general, data sequences with high stationarity and low complexity can be modeled accurately by the behavior of mathematical modeling. The failure time has certain statistical rules, but with various reasons for failure, different escalators have different failure times. Those regular escalators whose failure time series are stationary and have lower complexity can be selected as a reference. For the same batch of equipment, other escalators have similar failure occurrence rules. In this paper, two stationary indicators (SIs) and two complexity indicators (CIs) are combined to form a comprehensive selection indicator to select excellent referenced time series. Here are details about four indicators:

(1) Stationary indicator

(a) Standard deviation

$$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1)$$

In statistics, the standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the time series tends to be more stationary.

(b) *p*-value

The p -value is the probability value in the statistical significance test, that is, the probability that the hypothesis test occurs. It is generally assumed that if a small probability event occurs, the assumption is true. In this paper, the data sequence is used as the hypothesis test for the stationary data, and the P -value can be calculated as the probability of the unit root test to judge whether the failure time series is stationary or not. A low P -value indicates that the time series tends to be stationary.

$$P_value = P(x \text{ is a non-stationary sequence} | x \text{ is a stationary sequence}) \quad (2)$$

(2) Complexity indicator

(a) Lempel-ziv

The Lempel-ziv indicator is used to characterize the rate at which new patterns appear in a time series. The higher the Lempel-Ziv complexity of the sequence, the greater the degree to which the signal sequence is near-random. The lower the Lempel-ziv indicator, the lower the complexity of the time series.

$$C_{Lempel-ziv} = \frac{C_N(N)}{\lim_{N \rightarrow \infty} C_N(N)} \approx \frac{C_N(N) \times \log_k(N)}{N} \quad (3)$$

$C_N(N)$ is the Lempel-ziv complexity of the time series, and $C_{Lempel-ziv}$ is the normalized complexity. Details about the Lempel-ziv algorithm can be found in Ref. [25].

(b) Sample entropy

The Sample Entropy (SampEn) measures the complexity of the time series by measuring the probability of generating a new pattern in the signal. The greater the probability of the new pattern, the greater the complexity of the sequence. The lower the sample entropy indicator, the lower the complexity of the time series.

$$SampEn = -\ln\{A^k(r)/B^m(r)\} \quad (4)$$

where $A^k(r)$ and $B^m(r)$ are the number of subsets of the original time series that satisfies the requirement within the similar distance r (usually $m = 2$; $k = m + 1$; $r = 0.1 \sim 0.25$).

(3) Comprehensive selection indicator

In order to describe the inherent characteristics of a data sequence, a comprehensive selection indication (CSI) that can describe both stationarity and complexity is constructed in this part to select an inherently excellent data sequence before data modeling. In addition, CSI can be given by the following equation:

$$CSI = STD \times P_value \times C_{Lempel-ziv} \times SampEn \quad (5)$$

CSI can be used to compare the regularity itself between data sequences; the smaller the CSI, the better the data sequence itself and the more accurate the data modeling will be.

2.3. Dynamic Time Warping

The failure time series of escalators are different, and the raw data always have defects (such as disturbances), which are less effective when used directly in the prediction of neural networks. Considering the time delay and local similarity of the failure time series of different escalators, dynamic time warping (DTW) is introduced to pre-process before data modeling [22], and the over-sampling characteristics of data under the space defined by DTW can effectively reduce the impact of the mutated data on the accuracy of the model while increasing the data capacity and finally improve the prediction accuracy of the NN model. An intuitive example is given in Figure 2. The principles of DTW can be further explained as follows:

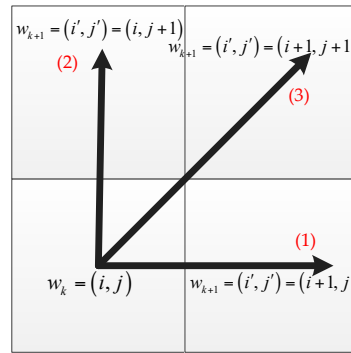


Figure 2. Potential searching directions.

Given two time series X and Y , whose lengths are, respectively, $|X|$ and $|Y|$, the warping path can be given,

$$W = w_1, w_2, \dots, w_k, \text{ where } \max(|X|, |Y|) \leq k \leq |X| + |Y| \quad (6)$$

$$w_k = (i, j), \text{ where } i \text{ is the coordinate in } X \text{ and } j \text{ is the coordinate in } Y \quad (7)$$

The warping path W must start from $w_1 = (1, 1)$ and end with $w_k = (|X|, |Y|)$ to ensure every coordinate in $|X|$ and $|Y|$ would be considered. Besides, the coordinate i and j in $w(i, j)$ of the warping path must be monotonically increasing, which can be constrained by the following equation

$$w_k = (i, j), w_{k+1} = (i', j'), i \leq i' \leq i+1, j \leq j' \leq j+1 \quad (8)$$

The warping path needs to satisfy continuity and monotonicity. According to this, Equation (8) represents the three directions that are in the next optimal path searching. This process can be shown in Figure 3. DTW always searches forward for the closest distance in three directions: Horizontal, vertical, and diagonal upward, to maintain the continuity and monotonicity.

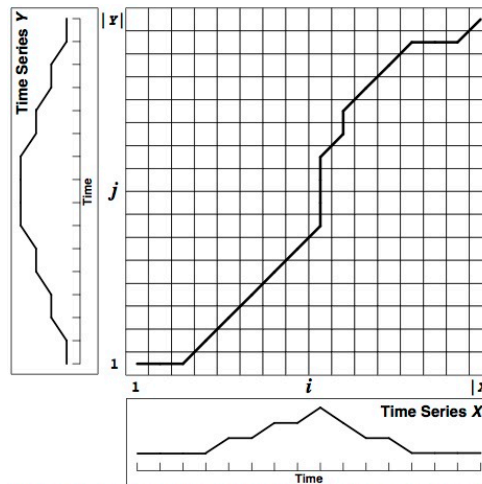


Figure 3. Diagram of dynamic time warping.

For the failure time of the escalator, such a warping path in DTW is physically meaningful. No new values will be added to the original data after DTW with the reference time series.

Finally, the wanted warping path is the shortest path in all possible warping paths

$$D(i, j) = \text{Dist}(i, j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)] \quad (9)$$

DTW is a typical optimization problem. It uses a time warping function that satisfies certain conditions to describe the time correspondence between the test template and the reference template,

and solves the warping function corresponding to the minimum cumulative distance when the two templates match. In addition, the constraint here is the search direction shown in Figure 3 [26]. DTW always searches forward for its nearest distance from three directions: Horizontal direction, vertical direction, and diagonal upward direction. These three directions are consistent with the basic law of the failure time series: (i) The horizontal direction indicates the same failure time; (ii) the vertical direction indicates sudden failure; (iii) the diagonal line indicates the stable development direction of the failure time. It is this constraint that allows the warping time series to learn the inherent failure development information of ‘good data’ while maintaining the original basic physical meaning of itself. Meanwhile, for short time series prediction, DTW can increase the sample capacity for its oversampling characteristics, especially to increase the sample capacity where the data are abrupt, so that the modeling of the disturbance in the neural network can obtain more attention to reduce the influence of disturbance to the whole data model.

2.4. Convolutional LSTM Neural Network

The significant characteristic of the failure time series in this paper is a few samples with non-uniform sampling. When directly using LSTM for data modeling, the features in time series are not obvious, which may lead to poor accuracy of the modeling and prediction. The deep neural network used in this paper is a combination of CNN and LSTM. The main target of this paper is time series prediction. With the help of the CNN’s feature extraction ability, high-dimensional features can be trained to understand the failure time series more comprehensively. Then, these high-dimensional features can be used for time series prediction in LSTM. Based on the advantages of CNN and LSTM, a convolutional LSTM neural network (CLNN) is proposed to achieve the goal of failure time prediction for an escalator.

2.4.1. Convolutional Neural Network

Convolutional neural networks are a type of deep neural network with the ability to act as feature extractors. Such models are able to learn multiple layers of feature hierarchies automatically (also called ‘representation learning’). A typical CNN network structure is shown in Figure 4. This typical CNN includes an input layer, a convolution layer (Conv), a maximum pooling layer (MP), a fully connected layer (FC), and an output layer.

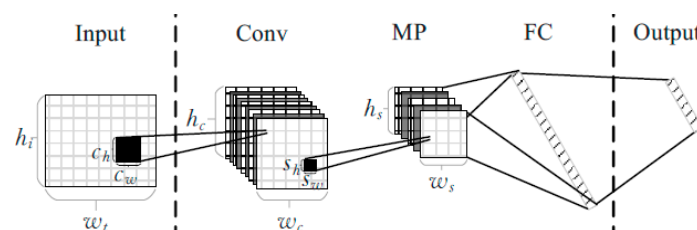


Figure 4. Convolutional neural network (CNN) network structure.

The structure of the CNN network mainly has two characteristics: Sparse connection and shared weights. As shown in Figure 5, unlikely a fully connected neural network, CNNs use the locally connected mode. The neurons in the m layer are connected to the adjacent neurons in the $m-1$ layer. The weights of every neuron in the $m-1$ layer are shared to the neurons in the m layer, which means the weights with the same color in Figure 4 are the same. Due to these advantages, the CNN has the effect of regularization, which can improve the stability and generalization ability of neural networks and avoid over-fitting. At the same time, these two advantages reduce the total amount of weight parameters, which is beneficial to the rapid learning of neural networks. CNNs have good representation learning ability, which can learn excellent features through network training. It is very suitable for a learning process lacking prior knowledge or clear features.

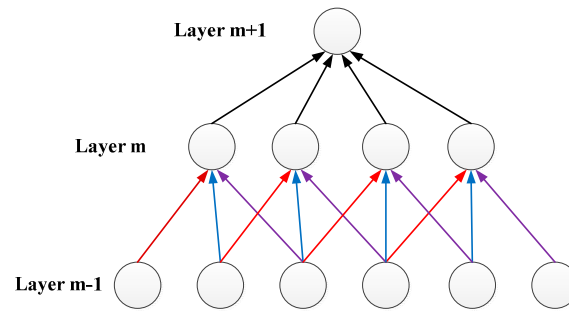


Figure 5. Sparse connection and shared weights in CNN.

In this paper, the failure time in the failure recall data is a typical time series, so a one-dimensional (1D) convolution operation is used to extract the features of failure time sequences.

2.4.2. Long Short-Term Memory Neural Network

After data sequences have been preprocessed, a long short-term memory neural network (LSTM NN) is applied in this paper for data modeling. To overcome the gradient vanish of traditional RNNs, a LSTM NN was proposed. LSTM is one of the special structures of RNNs that can solve the memory problem of neural networks. It can be applied to process and predict important events with long intervals and delays in time series.

With a special gate and cell in the hidden layer, LSTM can effectively update and deliver critical information in time series. Compared to traditional RNNs, LSTM has stronger capacities of information selection and time series learning, which can solve the problem of long-term dependencies by using remote context information for current prediction tasks. An LSTM is composed of one input layer, one hidden layer, and one output layer. Unlike the traditional NN, the basic unit of the hidden layer is a memory block, and LSTM adds a ‘processor’ in the algorithm to judge whether the information is useful or not, which is called a cell [27].

The typical structure of an LSTM cell is shown in Figure 6. An LSTM cell is configured mainly by three gates: Input gate, forget gate, and output gate. The basic idea of LSTM is that when information enters the network, the information that meets the requirements is left by the input gate, the non-compliant information is discarded by the forget gate, and finally new information is generated by the output gate.

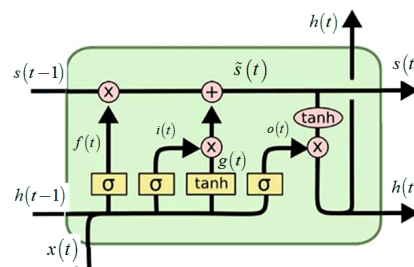


Figure 6. A cell of a long short-term memory neural network (LSTM NN).

Given $x(t)$ is the input time series, the output is denoted as $h(t)$ and the memory of the cell is denoted as $s(t)$. Three gates of the LSTM NN can be given as follows:

(1) Forget gate

This unit is decided by considering the current input, the previous output, and the previous memory. It will produce a new output and change its memory. σ determines how much information can be left.

$$f(t) = \sigma(W_f \cdot [h(t-1), x(t)] + b_f) \quad (10)$$

(2) Input gate

The input of this unit is the same with the forget gate, but the input gate determines the extent to which new memories should affect old memories. Meanwhile, this unit determines how much new information should be delivered to the next cell. Finally, the cell state is updated through discarding the information that needs to be discarded and adding new information.

$$i(t) = \sigma(W_i \cdot [h(t-1), x(t)] + b_i) \quad (11)$$

$$g(t) = \tanh(W_C \cdot [h(t-1), x(t)] + b_C) \quad (12)$$

$$s(t) = f(t) * s(t-1) + i(t) * g(t) \quad (13)$$

(3) Output gate

Based on the new cell state, some information of the new state should be added to the output and the output information of the cell will finally be determined by deterministic information that needs to be output.

$$o(t) = \sigma(W_o[h(t-1), x(t)] + b_o) \quad (14)$$

$$h(t) = o(t) * \tanh(s(t)) \quad (15)$$

where σ denotes the sigmoid function defined in the equation and the tanh function is defined in the equation:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (16)$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

The learning process of LSTM mainly includes the error backpropagation process and optimization algorithm. The backpropagation through time (BPTT) algorithm is applied in the error backpropagation process of LSTM [28].

2.4.3. CLNN: Combination of CNN and LSTM

The architecture of an CLNN is shown in Figure 7. In addition, the network structure of an CLNN consists of a convolutional layer, a pooling layer, LSTM units, and two fully connected layers.

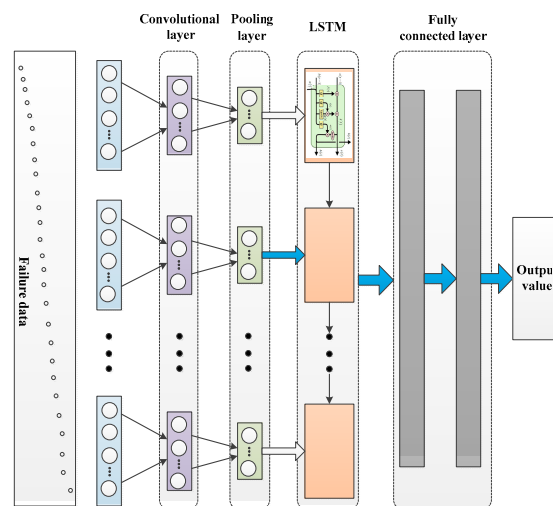


Figure 7. The architecture of a convolutional LSTM neural network (CLNN).

3. Application for Failure-Recall Data with DCLNN Method

3.1. Description of Failure-Recall Data

The current maintenance activities of an escalator are mainly based on two modes: Periodical maintenance and failure-recall maintenance. The maintenance of different components in periodical maintenance is carried out in different cycles (for example, the four types of maintenance items for the escalator with a cycle of 4, 12, 24, and 52 weeks). The failure-recall maintenance aims at the problems discovered by the operator to perform inspection and repair. The failure-recall record can directly reflect the operational safety of the escalator to a certain extent. The failure-recall records of six escalators during 2012.10–2018.10 are investigated. The statistics of failure-recall reasons is shown in Figure 8. There are five main situations in which the failure is recorded, and the cause of component failure accounts for the most.

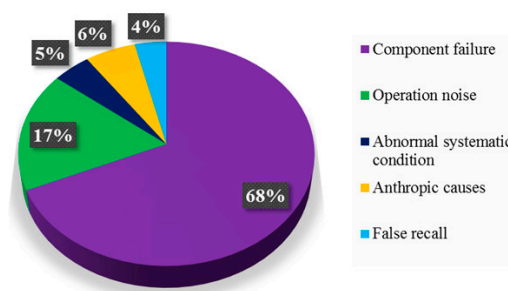


Figure 8. The statistics of failure-recall reasons.

The failure of the escalator is related to the failure of the components, the time of use, and the inherent performance degradation of the system. The collection and analysis of relevant data can provide scientific predictions for escalator failure. In addition, predicting the possible time of escalator failure and implementing maintenance in advance can reduce the chance of escalator failure and reduce casualties. The failure-recall record can reflect the failure occurrence rule and safety condition of an escalator. However, data in the failure-record have the characteristics of short samples, non-uniform samplings, and random interference, which brings difficulty in predicting the failure time of an escalator.

In this paper, failure-recall records of six escalators with the same brand are collected. In the records, some details about every failure time are recorded and shown in Appendix A. Then, the failure time is processed into the corresponding failure interval days, and the failure time curves of the escalators are calculated by the accumulation idea. The processed failure time curve is shown in Figure 9.

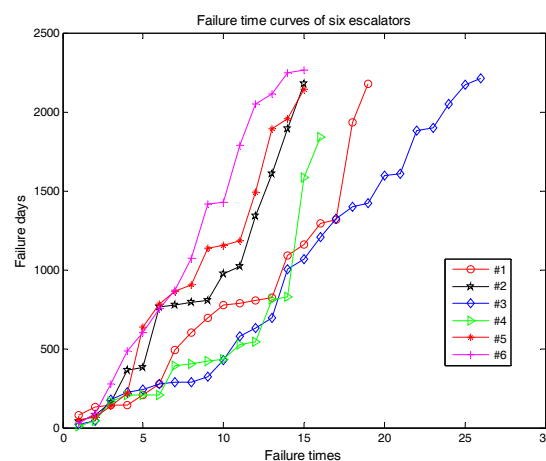


Figure 9. Failure time curve of six escalators.

3.2. Dynamic Time Warping Pre-Processing Before Data Modeling

In this part, failure time datasets of different escalators are distinguished by the abovementioned indicators to find those excellent data sequences with inherent advantages (high stationarity and low complexity) before data modeling. In addition, other data sequences are matched similarly with the selected excellent data using DTW. The above process can be seen as a pre-processing for more accurate data modeling. Table 1 summarizes the stationary indicators and complexity indicators of the original failure time series.

Table 1. Comprehensive selection indicator before dynamic time warping (DTW).

	Indicator	E #1	E #2	E #3	E #4	E #5	E #6
SI	std	144.5452	134.5956	80.0760	192.6437	136.8093	111.0329
	p-value	0.0245	0.1096	0.0046	0.0150	0.0672	0.0921
CI	Lempel-ziv	0.4472	0.5209	0.3616	0.5000	0.5209	0.5209
	SampEn	0.5108	1.3863	0.2076	0.4418	1.0986	0.6931
CSI		0.8090	10.6525	0.0277	0.6383	5.2611	3.6920

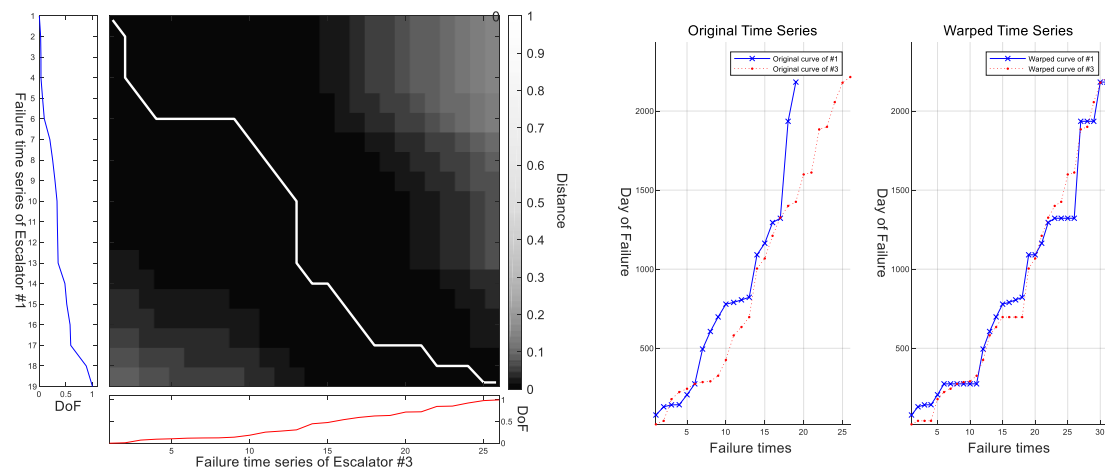
According to the comprehensive judgment from stationarity and complexity, the non-stationarity and complexity of escalator #3 is the smallest. The CSI of Escalator #3 is almost fifty times smaller than those of the others. In order to achieve the purpose of accurate prediction, it is necessary to reduce the influence of the other data disturbances during data modeling. The over-sampling characteristics of DTW can increase the number of time series while weakening the disturbances. Thus, DTW is used to warp other failure time curves with the data of the third escalator as a reference. The specific warping path and warped time series can be found in Figure 10.

The white lines in the left of Figure 10 are the warping path and the grayscale maps show the warping distance between the two time series. Figures on the right show the original time series and warped time series. It is easy to see that the new data after DTW have the same physical meaning as original data for the special warping path. Here, a specific analysis of the warped failure time series of Escalator #1 is given. As can be seen from Figure 10, whether it is the red reference sequence #3 or the blue warped sequence #1, no new failure time is generated on the y -axis, but only extended on the x -axis, where the x -axis represents the order or number of failures. Although some new samples were increased on the x -axis, the occurrence time of failure is the same as before. It can still be regarded as the same failure. This is unambiguous in objective meaning. Meanwhile, it can be seen that due to the oversampling characteristics of DTW, the length of the time series after being warped has been increased. Although the increased samples do not affect the physical meaning of the failure time series, it is possible to increase training samples and increase the network's understanding of the failure time series, which can further improve the training effect and prediction effect of the network.

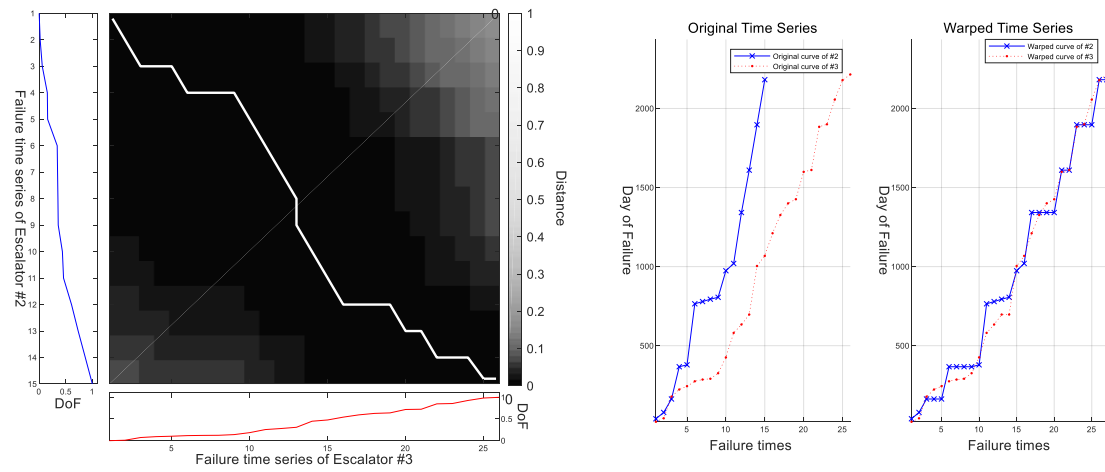
In order to show the function of DTW, above-mentioned stationary indicators and complexity indicators are calculated after DTW with escalator #3 as referenced. Table 2 shows the result after DTW pre-processing.

Table 2. Comprehensive selection indicator after DTW.

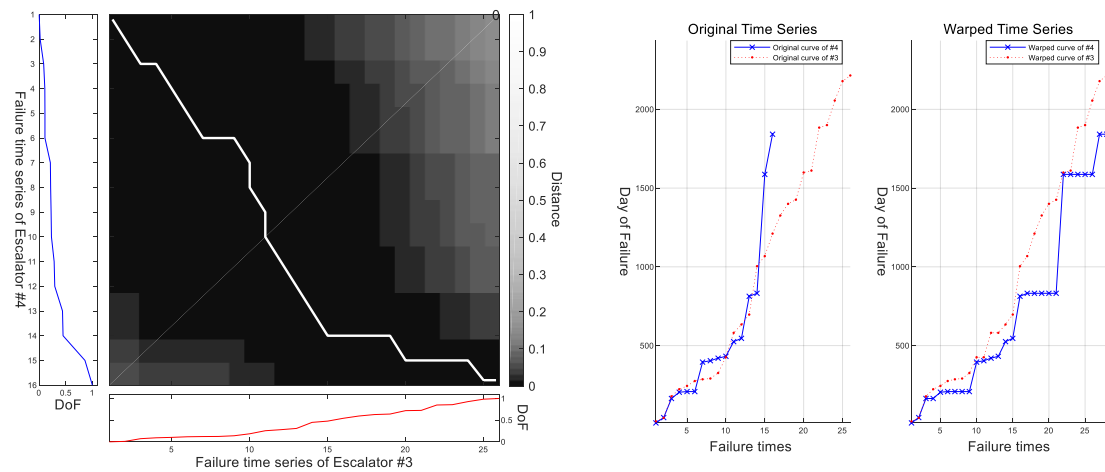
	Indicator	D-M1	D-M2	D-M3	D-M4	D-M5	D-M6
SI	STD	127.73	125.34	80.79	157.41	126.12	114.32
	p-value	0.001	0.001	0.0032	0.001	0.001	0.001
CI	Lempel-ziv	0.3196	0.3522	0.3616	0.3434	0.3522	0.3616
	SampEn	0.3042	0.7419	0.2076	0.4520	0.4187	0.4055
CSI		0.0124	0.0308	0.0194	0.0244	0.0186	0.0168



(a)



(b)



(c)

Figure 10. Cont.

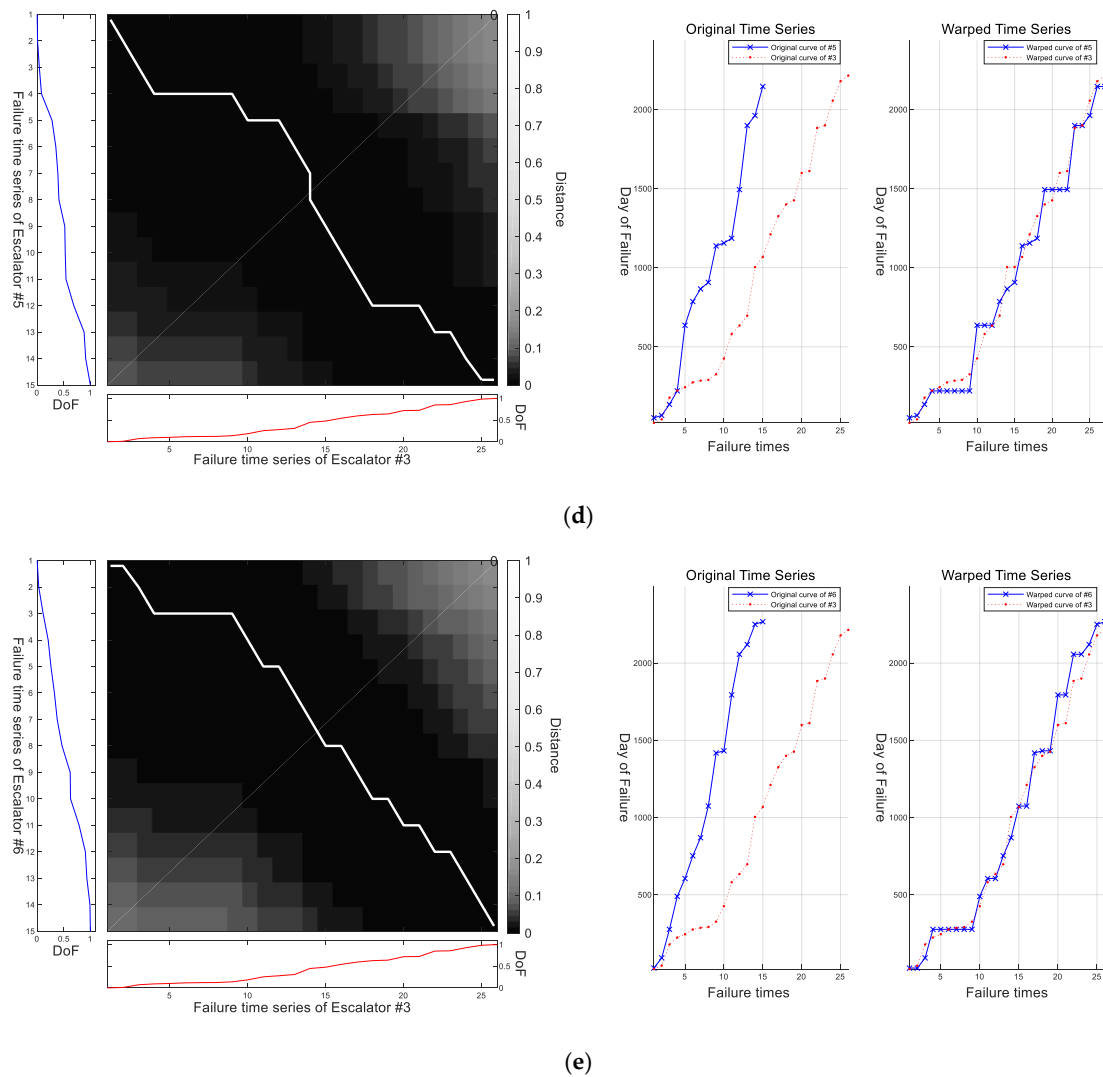


Figure 10. DTW preprocessing results. (a) Warped failure time series of #1. (b) Warped failure time series of #2. (c) Warped failure time series of #4. (d) Warped failure time series of #5. (e) Warped failure time series of #6.

As shown in Tables 2 and 3, the CSI is reduced to below 0.1 after DTW with the failure time series as the referenced sequence. A specific comparison of CSI before and after DTW preprocessing is summarized in Figure 9 for a clear description. As shown in Figure 11, the CSI of the other five escalators are reduced after DTW preprocessing with Escalator #3 referenced. They become more stationary and lower-complexity after matching with the failure time series of Escalator #3.

Table 3. Root-mean-square error (RMSE) in training process.

Training	# Of Escalator		#1	#2	#3	#4	#5	#6
	RMSE	Before DTW	63.64	122.02	73.8	77.69	127.72	93.1
		After DTW	119.75	117.83	73.8	150.8	123.46	107.45

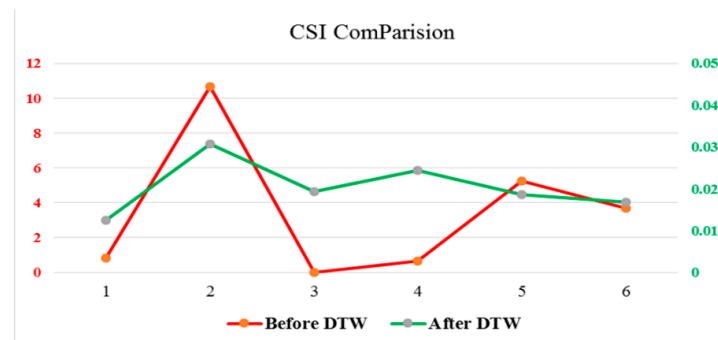


Figure 11. Comparison of comprehensive selection indicator (CSI) before and after DTW pre-processing.

Besides, for the convenience of engineering applications, six sigma was applied to the analysis of the CSI of the failure time series after DTW to try to give a possible threshold of the CSI. In addition, the threshold can be used to distinguish whether the data sequences need to be warped before data modeling. The distribution interval of CSI can be given by $[\mu - 3\sigma, \mu + 3\sigma]$, where μ is the mean value of CSI and σ is the variance. For above-mentioned data sequences after DTW, $CSI_{\text{mean}} = 0.0204$, $\sigma = 0.0064$. Thus, the CSI value can be considered in the interval $[0.0012, 0.0396]$, and the upper limit of the interval is used as the judgment threshold considering the fault tolerance. That is, $CSI_{\text{threshold}} = 0.0396$. For data sequences in this paper, if the CSI is less than $CSI_{\text{threshold}}$, DTW preprocessing is unnecessary before data modeling, otherwise, it is recommended to be warped with a sequence with a CSI less than $CSI_{\text{threshold}}$.

3.3. Results of LSTM and DCLNN

It is easy to see that no straightforward rules can be found directly from the failure time curve in Figure 10. The LSTM is used directly to train and predict the data in Figure 10. During the training process, the prediction goal is the failure time of the last day. Among them, about 90% of the time series data are used to train the LSTM NN, and the last 10% of the data are used to test the model and predict the failure time. The specific architecture and hyperparameters of CLNN can be found in Figure 12. Tensorflow was used to build the above neural networks. A related repository was stored in the link (the same name with paper) on github.

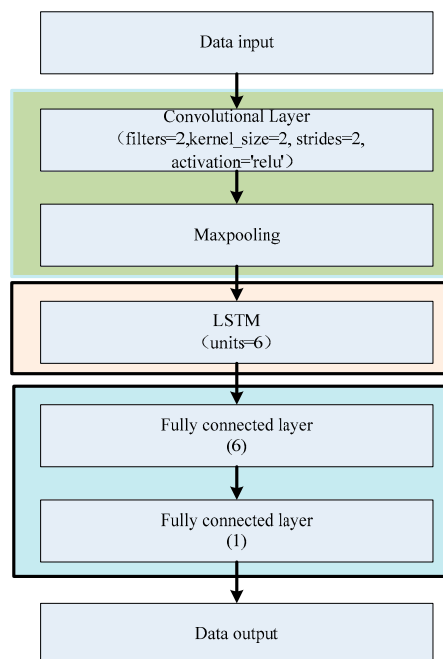


Figure 12. Specific architecture and hyper-parameters.

Taking into account the time cost and prediction accuracy of network training, the recommended training epochs is around 200. Before data training and testing using a neural network, data normalization using Equation (18) is necessary.

$$xn(i) = \frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \quad (18)$$

Then, the original failure time series of six escalators are first trained by the CLNN. The training process and prediction result are shown in Figure 13. Training loss and testing loss are also given in Figure 14.

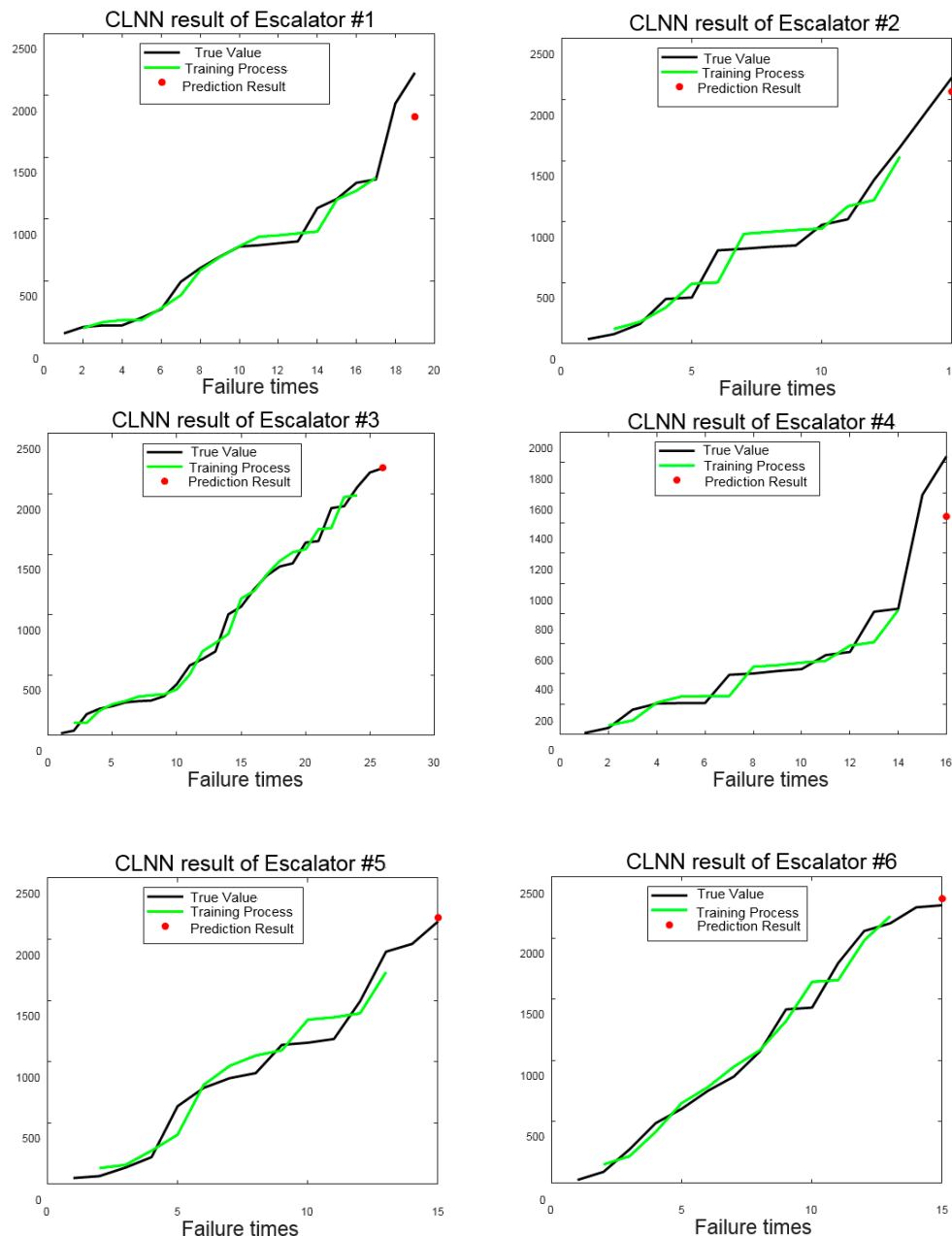


Figure 13. Training process and prediction result using CLNN directly.

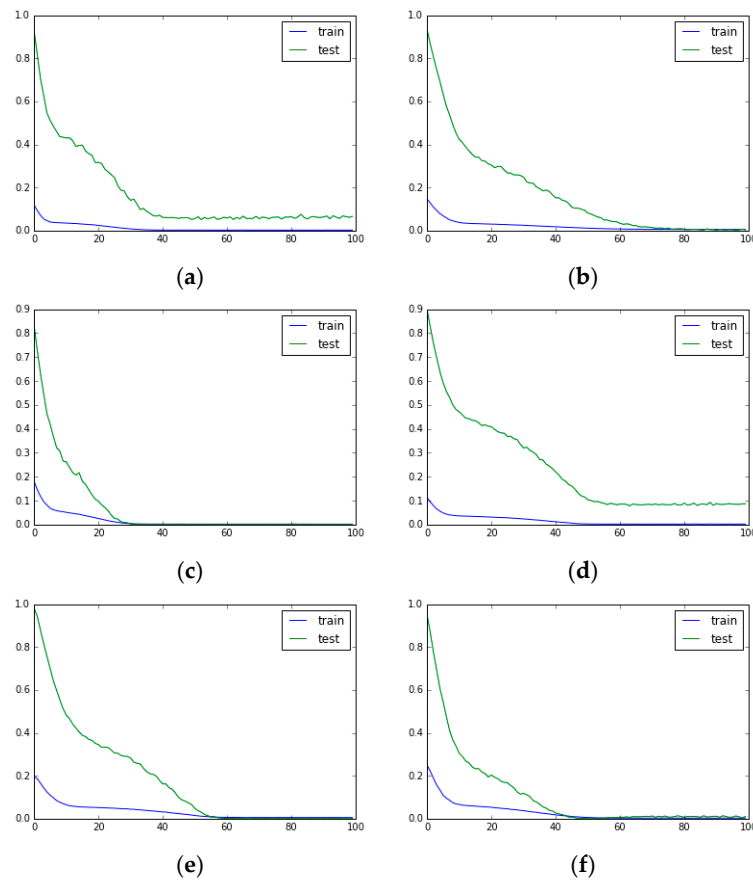


Figure 14. Loss of 6 escalators using CLNN for training and prediction. (a) Escalator #1. (b) Escalator #2. (c) Escalator #3. (d) Escalator #4. (e) Escalator #5. (f) Escalator #6.

Black lines are the true value of the failure time curve, green lines indicate the training process, and the red point shows the prediction result. All escalators except the third one have poor predictions that can be seen from Figure 13 directly. The results are consistent with the result of CSI. The reason for the poor prediction of other escalators are: 1) The length of the data sequence is short, and the number of data used for CLNN model training and testing are also small; 2) the trend of the third escalator is stable with no obvious disturbance, while there are more or less disturbances in the time series data of the other five escalators. In addition, those disturbances make a big difference on modeling accuracy of the CLNN. Then, DTW is used to warp other failure time curves with the data of the third escalator as a reference. Using new data for training and prediction of the DCLNN, the results are shown in Figure 15. Training loss and testing loss are also given in Figure 16.

The prediction accuracy after pre-processing is significantly improved compared to the direct use of the original data. Here, the root-mean-square error (RMSE) is chosen to assess the prediction accuracy of escalators. The RMSE can measure the difference between two time series. The mathematical formula is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y - y')^2} \quad (19)$$

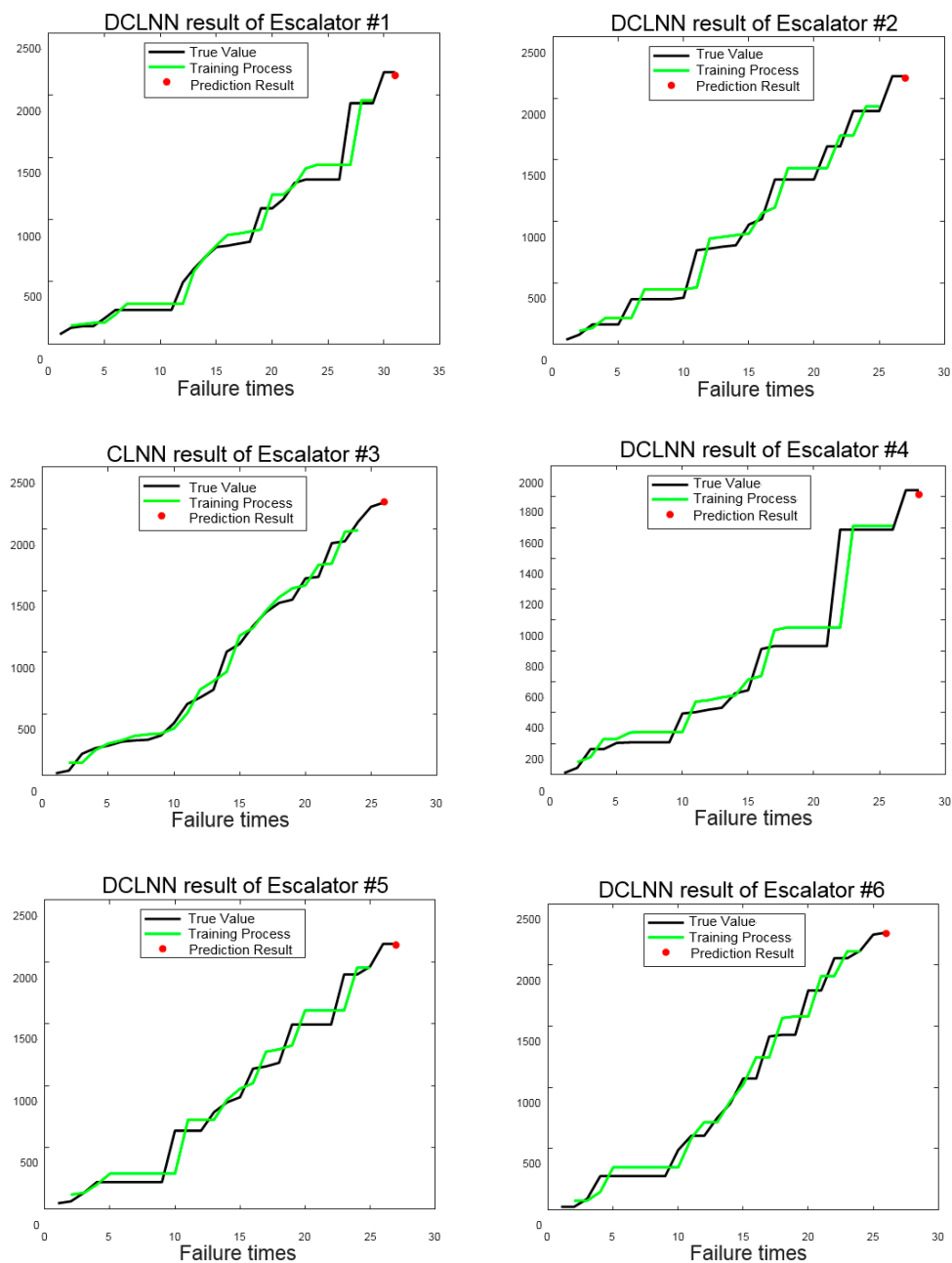


Figure 15. Training process and prediction using DCLNN.

The comparison results are summarized in Tables 3 and 4. Table 3 shows the RMSE of the training process between the original data and warping data. Table 4 shows the RMSE of the prediction process between the original data and warping data. Although the RMSE in the training process changes a little, the RMSE has been reduced a lot in the prediction process. After pre-processing by DTW, the accuracy of the DCLNN is improved.

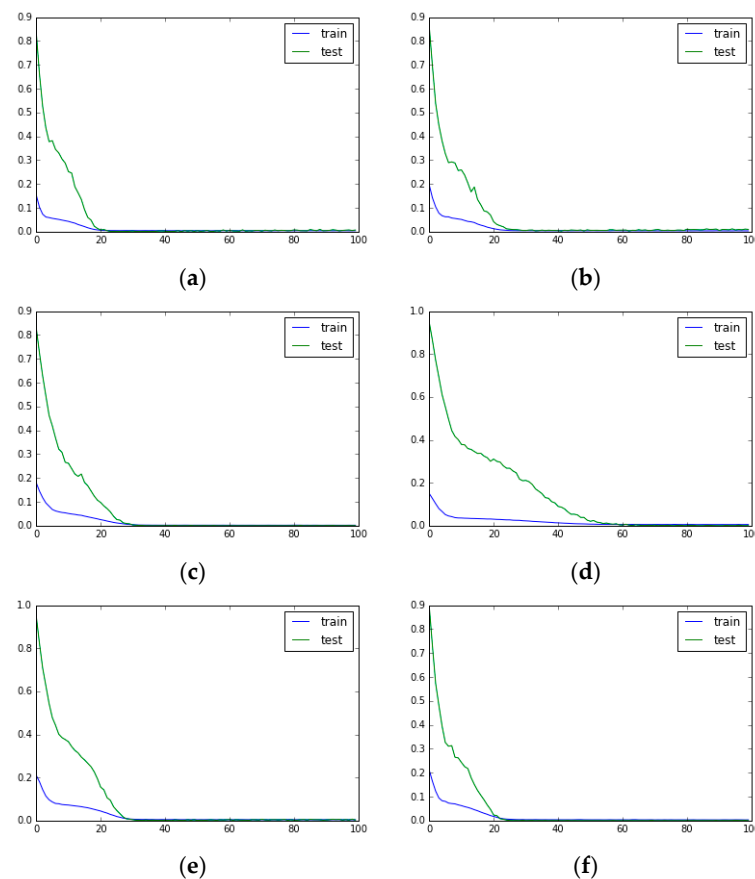


Figure 16. Loss of 6 escalators using DCLNN for training and prediction. (a) Escalator #1. (b) Escalator #2. (c) Escalator #3. (d) Escalator #4. (e) Escalator #5. (f) Escalator #6.

Table 4. RMSE in prediction process.

		# Of Escalator	#1	#2	#3	#4	#5	#6
Prediction	RMSE	Before DTW	384.87	112.78	3.9	397.12	25.3	53.18
		After DTW	25.12	16.81	3.9	30.37	9.3	7.66
	RMSE reduction (%)		93.50	85.09	0	92.35	63.24	85.60
Average MSE reduction percentage: 83.96%								

With DCLNN, the RMSE of the escalator can be reduced by an average of 83.96%. From Figure 15, the accuracy of the prediction result can be reduced from a few hundred days to less than one month, in order to compare the accuracy of the proposed DCLNN model for failure time prediction of escalators. For a better explanation about the effectiveness of the DCLNN, the other three classical time series prediction methods (Kalman Filter, Nonlinear Auto Regressive Neural Network, Elman RNN, LSTM) are applied for data analysis of the third escalator. The results of these methods are shown in Figure 17. The results intuitively show huge errors from the true value compared to results from the DCLNN.

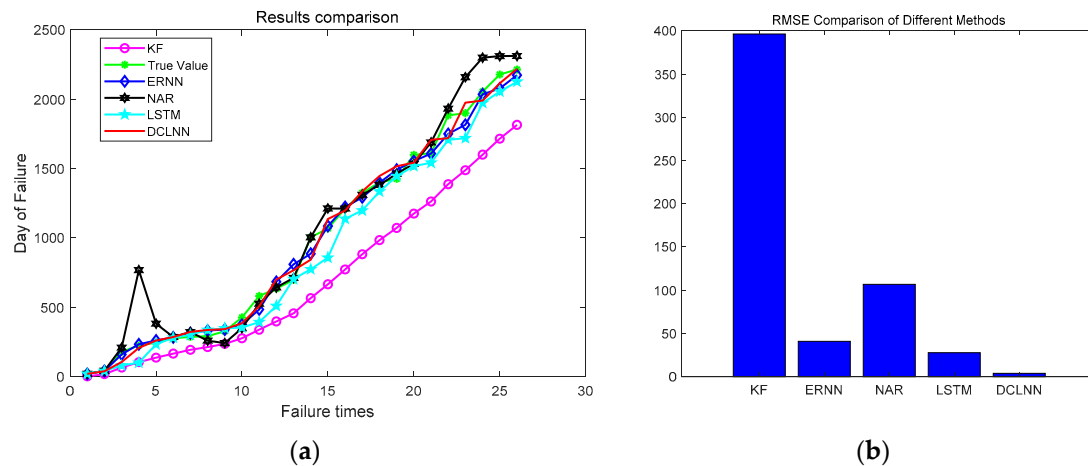


Figure 17. Result comparison. (a) Comparison of different methods. (b) Comparison of RMSE with different methods.

3.4. Function of DTW

The use of DTW to warping time series can improve the prediction accuracy and reduce the mean square error. If the prediction effect of two time series is poor when using the neural network for the first time, the result is not obviously improved after using DTW. However, the result will have a significant improvement after DTW preprocessing if one of the time series is relatively good for prediction. Table 5 concludes all RMSE values when using different time series as referenced data for DTW pre-processing and warping data for prediction using DCLNN.

Table 5. Statistics of RMSE.

	# 1	# 2	# 3	# 4	# 5	# 6
RMSE	384.87	112.78	3.9	397.12	25.3	53.18
#1 referenced	384.87	18.43	6.81	442.35	25.91	52.26
#2 referenced	225.41	112.78	6.85	125.17	12.40	52.48
#3 referenced	25.12	16.81	3.9	30.37	9.3	7.66
#4 referenced	117.56	32.26	10.65	397.12	11.23	43.24
#5 referenced	48.43	13.78	5.86	241.17	25.3	47.77
#6 referenced	119.11	117.85	2.95	61.46	25.44	53.18

Figure 18a is obtained from each column in Table 5. It is clear to see that the MSE is smallest when choosing the third escalator as the reference time series in DTW preprocessing. Figure 18b shows the RMSE improvement of DCLNN prediction after using different escalator data as the reference time series in DTW pre-processing. Similarly, the RMSE improvement of the third escalator is best compared to others. Meanwhile, the result in Figure 18b shows that almost all MSE of the data sequences with different data sequences as referenced can be improved after DTW preprocessing. However, for the third escalator, the RMSE variation is negative. The main reason is that the data quality of the third escalator is best according to the CSI indicator. After DTW is preprocessed with other data as referenced data, the data quality is not as excellent as the third data itself. Therefore, for data with high quality, it is not recommended to use DTW for data preprocessing, otherwise it will reduce the data quality and increase the prediction error.

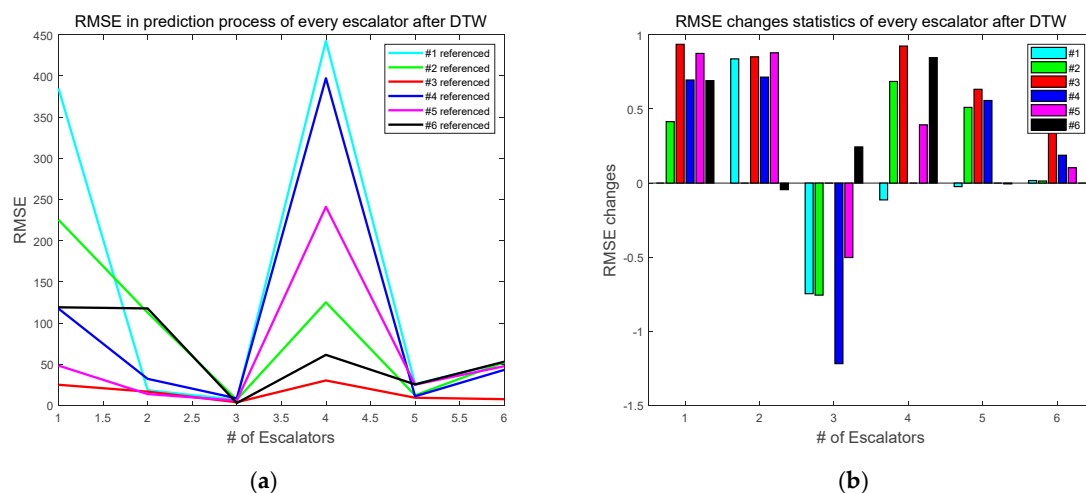


Figure 18. Statistics of RMSE in prediction process. (a) RMSE of every escalator after DTW. (b) RMSE variation statistics after DTW.

4. Conclusions

Failure time, as an important factor for the safe operation of escalators, is used for prediction by the proposed DCLNN neural network in this paper. Due to the short length of the failure time series, and the characteristics of random interference and non-uniform sampling, the failure time prediction is much more difficult. Naturally, higher data quality and a suitable neural network can help solve this problem. Considering the oversampling characteristics and similarity-matching performance of DTW, some inferior data with low stationarity and high complexity can be warped by referring to excellent data while retaining its original physical meaning. Such DTW-based data pre-processing can not only increase the sample numbers of short time series but also reduce the impact of disturbance data on the neural network. After data preprocessing, the combination of CNN and LSTM is used to train and predict the failure time series. With the help of CNN's feature extraction ability, high-dimensional features can be trained to understand the failure time series more comprehensively. Then, these high-dimensional features can be used for time series prediction in LSTM. Based on the advantages of DTW, CNN, and LSTM, a strategy to combine data quality enhancement with deep neural networks is proposed to achieve the goal of failure time prediction for an escalator. Failure-recall data of an escalator are analyzed using the proposed DCLNN method, and the result shows that the proposed method can effectively reduce the root-mean-square error between the predicted value and the real value in the prediction process. At the same time, compared to some classical time series prediction methods such as the Kalman Filter, Nonlinear Auto Regressive Neural Network, and Elman RNN, the prediction accuracy of the proposed method is obviously improved. Besides, the function of the DTW was further analyzed to show that enhancing data quality can effectively improve the accuracy of prediction. The proposed DCLNN method can reduce the prediction time error of escalator failure to less than one month, which can provide scientific guidance for smarter maintenance planning and economic improvement.

However, due to the limited number of samples, the accuracy of the proposed method needs to be improved for better maintenance guiding, and a more reasonable threshold of CSI indicators needs further study before pre-processing in engineering application. Furthermore, the proposed method is currently applicable to the situation of two or more devices to improve the prediction accuracy. In the future, an in-depth study is necessary for the failure time series prediction problem of a single device under a small number of samples.

Author Contributions: Z.Z. proposed the idea of DCLNN, analyzed the failure data, and wrote the paper; Y.Z. proofread the manuscript; J.C. contributed to the conclusion and guided the manuscript; T.A. and J.X. collected the failure data. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0805701 and National Natural Science Foundation of China (No. 51775411).

Acknowledgments: The authors would like to sincerely thank all the anonymous reviewers for the valuable comments that greatly helped to improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The following table is a summary of the failure-record of six escalators in a shopping mall. In this paper, the location and starting date are used for analysis for failure time prediction.

Location	Starting Date	Starting Time	Arrival Date	Arrival Time	Recovery Date	Recovery Time	Maintenance Time (Min)
1U2N	2012/10/23	12:21:00	2012/10/23	12:42:00	2012/10/23	14:00:00	78
1U2N	2013/1/9	13:57:00	2013/1/9	14:20:00	2013/1/9	15:30:00	70
1U2N	2013/3/2	13:47:00	2013/3/2	15:01:00	2013/3/2	19:36:00	275
1U2N	2013/3/14	8:29:00	2013/3/14	8:56:00	2013/3/14	9:43:00	47
1U2N	2013/3/15	8:30:00	2013/3/15	8:50:00	2013/3/15	9:30:00	40
1U2N	2013/5/17	17:43:00	2013/5/17	18:27:00	2013/5/17	19:20:00	53
1U2N	2013/7/25	20:00:00	2013/7/25	20:20:00	2013/7/25	21:00:00	40
1U2N	2014/3/2	14:10:00	2014/3/2	14:30:00	2014/3/2	15:50:00	80
1U2N	2014/6/21	16:35:00	2014/6/21	17:15:00	2014/6/21	18:20:00	65
1U2N	2014/9/21	17:40:01	2014/9/21	18:00:01	2014/9/21	22:15:01	255
1U2N	2014/12/11	16:22:00	2014/12/11	16:37:00	2014/12/11	17:54:00	77
1U2N	2014/12/22	8:36:00	2014/12/22	14:00:00	2014/12/22	17:20:00	200
1U2N	2015/1/7	8:46:00	2015/1/7	9:06:00	2015/1/7	9:30:00	24
1U2N	2015/1/23	8:34:00	2015/1/23	11:40:00	2015/1/23	11:50:00	10
1U2N	2015/10/19	17:00:00	2015/10/19	17:25:00	2015/10/19	20:00:00	155
1U2N	2015/12/31	9:30:00	2015/12/31	9:46:00	2015/12/31	11:10:00	84
1U2N	2016/5/10	9:34:00	2016/5/10	9:50:00	2016/5/10	10:10:00	20
1U2N	2016/6/6	18:27:00	2016/6/6	18:40:00	2016/6/6	19:10:00	30
1U2N	2018/2/9	14:23:00	2018/2/9	14:41:00	2018/2/9	15:10:00	29
1U2N	2018/10/15	9:04:00	2018/10/15	10:00:00	2018/10/15	18:02:49	483
1U2E	2012/9/28	13:11:00	2012/9/28	13:21:00	2012/9/28	15:05:00	104
1U2E	2012/11/5	15:06:00	2012/11/5	15:19:00	2012/11/5	17:20:00	121
1U2E	2012/12/15	13:05:00	2012/12/15	13:23:00	2012/12/15	15:00:00	97
1U2E	2013/3/10	20:10:00	2013/3/10	20:25:00	2013/3/10	22:00:00	95
1U2E	2013/9/30	16:40:00	2013/9/30	17:05:00	2013/9/30	23:30:00	385
1U2E	2013/10/12	13:01:00	2013/10/12	13:24:00	2013/10/12	13:50:00	26
1U2E	2014/11/2	15:58:01	2014/11/2	16:20:01	2014/11/2	17:57:01	97
1U2E	2014/11/16	13:40:00	2014/11/16	14:00:00	2014/11/16	15:18:00	78
1U2E	2014/12/1	9:24:00	2014/12/1	10:00:00	2014/12/1	23:50:00	830
1U2E	2014/12/13	9:39:00	2014/12/13	10:00:00	2014/12/13	10:30:00	30
1U2E	2015/5/30	9:14:00	2015/5/30	9:40:00	2015/5/30	10:20:00	40
1U2E	2015/7/15	21:28:00	2015/7/15	22:05:00	2015/7/15	22:20:00	15
1U2E	2016/6/1	8:43:00	2016/6/1	9:00:00	2016/6/1	9:30:00	30
1U2E	2017/2/24	13:34:00	2017/2/24	14:55:00	2017/2/24	16:30:00	95
1U2E	2017/12/8	9:23:00	2017/12/8	19:42:00	2017/12/8	21:50:00	128
1U2E	2018/9/19	8:53:00	2018/9/19	9:24:00	2018/9/19	11:24:00	120
1U2S	2012/9/25	13:40:00	2012/9/25	14:19:00	2012/9/27	2:30:00	2171
1U2S	2012/10/14	19:07:00	2012/10/14	19:25:00	2012/10/14	20:30:00	65
1U2S	2012/11/5	9:30:00	2012/11/5	10:00:00	2012/11/5	11:20:00	80
1U2S	2013/3/22	15:30:00	2013/3/22	16:10:00	2013/3/22	20:30:00	260
1U2S	2013/5/6	10:00:00	2013/5/6	10:30:00	2013/5/6	12:30:00	120
1U2S	2013/5/27	14:27:00	2013/5/27	14:40:00	2013/5/27	16:00:00	80
1U2S	2013/6/27	9:30:00	2013/6/27	10:00:00	2013/6/27	11:40:00	100

Location	Starting Date	Starting Time	Arrival Date	Arrival Time	Recovery Date	Recovery Time	Maintenance Time (Min)
1U2S	2013/7/8	16:55:01	2013/7/8	17:13:01	2013/7/8	18:20:01	67
1U2S	2013/7/13	11:00:00	2013/7/13	11:30:00	2013/7/13	12:45:00	75
1U2S	2013/8/17	16:23:00	2013/8/17	16:44:00	2013/8/17	17:20:00	36
1U2S	2013/11/25	10:00:00	2013/11/25	10:30:00	2013/11/25	11:30:00	60
1U2S	2014/4/29	11:10:00	2014/4/29	11:30:00	2014/4/29	13:00:00	90
1U2S	2014/6/21	9:10:00	2014/6/21	9:50:00	2014/6/21	11:10:00	80
1U2S	2014/8/23	10:00:00	2014/8/23	10:30:00	2014/8/23	11:30:00	60
1U2S	2015/6/26	11:45:00	2015/6/26	12:20:00	2015/6/26	14:10:00	110
1U2S	2015/8/29	1:22:00	2015/8/29	1:25:00	2015/9/2	4:36:00	5951
1U2S	2016/1/19	13:59:00	2016/1/19	14:17:00	2016/1/19	15:50:00	93
1U2S	2016/5/13	14:22:00	2016/5/13	14:45:00	2016/5/13	16:00:00	75
1U2S	2016/7/26	14:07:00	2016/7/26	14:25:00	2016/7/26	14:30:00	5
1U2S	2016/8/21	15:01:00	2016/8/21	15:23:00	2016/8/21	15:40:00	17
1U2S	2017/2/10	10:12:00	2017/2/10	10:30:00	2017/2/10	10:50:00	20
1U2S	2017/2/22	8:55:00	2017/2/22	9:10:00	2017/2/22	9:30:00	20
1U2S	2017/11/22	9:08:00	2017/11/22	10:02:00	2017/11/22	10:29:46	28
1U2S	2017/12/8	19:25:00	2017/12/8	19:34:00	2017/12/8	19:39:05	5
1U2S	2018/5/13	17:01:00	2018/5/13	17:30:00	2018/5/13	18:03:00	33
1U2S	2018/9/13	15:36:00	2018/9/13	16:00:00	2018/9/13	17:00:00	60
1U2S	2018/10/19	9:40:00	2018/10/19	10:20:00	2018/10/19	10:32:00	12
1U2W	2012/8/7	11:00:00	2012/8/7	11:40:00	2012/8/7	14:30:00	170
1U2W	2012/8/16	14:10:00	2012/8/16	14:25:00	2012/8/16	20:45:00	380
1U2W	2012/9/19	18:15:00	2012/9/19	18:43:00	2012/9/19	20:30:00	107
1U2W	2013/1/18	21:18:00	2013/1/18	21:39:00	2013/1/18	23:57:00	138
1U2W	2013/2/27	12:05:00	2013/2/27	12:20:00	2013/2/27	13:20:00	60
1U2W	2013/3/3	16:03:00	2013/3/3	16:33:00	2013/3/3	22:20:00	347
1U2W	2013/3/3	11:01:00	2013/3/3	11:40:00	2013/3/3	12:00:00	20
1U2W	2013/9/6	18:30:00	2013/9/6	18:50:00	2013/9/6	22:40:00	230
1U2W	2013/9/15	17:28:00	2013/9/15	17:58:00	2013/9/15	18:30:00	32
1U2W	2013/10/1	8:10:00	2013/10/1	8:25:00	2013/10/1	19:10:00	645
1U2W	2013/10/13	16:40:00	2013/10/13	17:07:00	2013/10/13	19:30:00	143
1U2W	2014/1/14	9:33:00	2014/1/14	10:02:00	2014/1/14	10:45:00	43
1U2W	2014/2/4	13:33:00	2014/2/4	13:59:00	2014/2/4	14:30:00	31
1U2W	2014/10/29	8:05:00	2014/10/29	8:50:00	2014/10/29	9:40:00	50
1U2W	2014/11/17	8:10:00	2014/11/17	8:35:00	2014/11/17	10:00:00	85
1U2W	2016/12/11	22:24:00	2016/12/11	22:50:00	2016/12/11	23:30:00	40
1U2W	2017/8/23	9:39:00	2017/8/23	10:00:00	2017/8/23	11:30:00	90
2U3N	2012/8/30	15:10:00	2012/8/30	15:20:00	2012/8/30	18:20:00	180
2U3N	2012/10/20	18:50:00	2012/10/20	19:05:00	2012/10/20	21:10:00	125
2U3N	2012/11/4	16:15:00	2012/11/4	16:15:00	2012/11/4	17:20:00	65
2U3N	2013/1/13	17:45:00	2013/1/13	17:55:00	2013/1/13	19:00:00	65
2U3N	2013/4/8	11:30:00	2013/4/8	12:00:00	2013/4/8	12:40:00	40
2U3N	2014/5/28	15:30:00	2014/5/28	16:00:00	2014/5/28	16:30:00	30
2U3N	2014/10/25	8:45:00	2014/10/25	9:20:00	2014/10/25	12:10:00	170
2U3N	2015/1/13	8:50:01	2015/1/13	9:25:01	2015/1/13	11:40:01	135
2U3N	2015/2/23	13:21:00	2015/2/23	13:54:00	2015/2/23	17:00:00	186
2U3N	2015/10/12	15:00:00	2015/10/12	15:23:00	2015/10/12	17:15:00	112
2U3N	2015/10/30	11:50:00	2015/10/30	12:23:00	2015/10/30	14:20:00	117
2U3N	2015/11/29	17:00:01	2015/11/29	17:20:01	2015/11/29	20:18:01	178
2U3N	2016/10/2	9:09:00	2016/10/2	9:30:00	2016/10/2	10:00:00	30
2U3N	2017/11/11	9:19:00	2017/11/11	9:45:00	2017/11/11	11:30:00	105
2U3N	2018/1/13	9:25:00	2018/1/13	9:53:00	2018/1/13	10:30:00	37
2U3N	2018/7/16	9:15:00	2018/7/16	9:38:00	2018/7/16	11:38:00	120
2U3S	2012/8/3	20:00:00	2012/8/3	20:30:00	2012/8/3	22:30:00	120
2U3S	2012/8/28	9:20:00	2012/8/28	9:40:00	2012/8/28	10:20:00	40

Location	Starting Date	Starting Time	Arrival Date	Arrival Time	Recovery Date	Recovery Time	Maintenance Time (Min)
2U3S	2012/11/2	9:36:00	2012/11/2	10:05:00	2012/11/2	11:30:00	85
2U3S	2013/5/6	10:00:00	2013/5/6	10:30:00	2013/5/6	11:30:00	60
2U3S	2013/12/5	18:30:01	2013/12/5	18:50:01	2013/12/5	19:30:01	40
2U3S	2014/3/31	9:30:00	2014/3/31	10:00:00	2014/3/31	10:30:00	30
2U3S	2014/8/25	11:00:00	2014/8/25	11:30:00	2014/8/25	12:00:00	30
2U3S	2014/12/20	9:08:01	2014/12/20	9:17:01	2014/12/20	9:28:01	11
2U3S	2015/7/13	18:30:00	2015/7/13	19:12:00	2015/7/13	22:30:00	198
2U3S	2016/6/21	9:23:00	2016/6/21	9:40:00	2016/6/21	10:10:00	30
2U3S	2016/7/5	16:50:00	2016/7/5	17:10:00	2016/7/5	17:30:00	20
2U3S	2017/7/2	10:27:00	2017/7/2	10:50:00	2017/7/2	11:00:00	10
2U3S	2018/3/22	14:20:00	2018/3/22	20:55:00	2018/3/22	21:16:00	21
2U3S	2018/5/24	9:20:00	2018/5/24	20:33:00	2018/5/24	21:08:00	35
2U3S	2018/10/2	8:52:00	2018/10/2	10:20:00	2018/10/2	10:38:00	18
2U3S	2018/10/19	10:35:00	2018/10/19	10:36:00	2018/10/19	11:10:00	34

References

- Schminke, L.H.; Jeger, V.; Evangelopoulos, D.S.; Zimmerman, H.; Exadaktylos, A.K. Riding the Escalator: How Dangerous is it Really? *West. J. Emerg. Med.* **2013**, *14*, 141–145. [\[CrossRef\]](#)
- Xing, Y.; Dissanayake, S.; Lu, J.; Long, S.; Lou, Y. An analysis of escalator-related injuries in metro stations in China, 2013–2015. *Accid. Anal. Prev.* **2019**, *122*, 332–341. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chi, C.-F.; Chang, T.-C.; Tsou, C.-L. In-depth investigation of escalator riding accidents in heavy capacity MRT stations. *Accid. Anal. Prev.* **2006**, *38*, 662–670. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, W.; Gong, J.; Yu, P.; Shen, S. Modeling, simulation and analysis of group trampling risks during escalator transfers. *Phys. A Stat. Mech. its Appl.* **2016**, *444*, 970–984. [\[CrossRef\]](#)
- Wang, W.; Li, X.; Pan, Q.-L. Notice of Retraction Risk management based on the escalator overturned accident. In *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*; IEEE: New York, NY, USA, 2013; Volume 2013, pp. 22–27. [\[CrossRef\]](#)
- Saleh, J.H.; Marais, K.B.; Favarò, F.M. System safety principles: A multidisciplinary engineering perspective. *J. Loss Prev. Process. Ind.* **2014**, *29*, 283–294. [\[CrossRef\]](#)
- Corman, F.; Kraijema, S.; Godjevac, M.; Lodewijks, G. Optimizing preventive maintenance policy: A data-driven application for a light rail braking system. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2017**, *231*, 534–545. [\[CrossRef\]](#)
- Tsoukalas, V.; Fragiadakis, N. Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis. *Saf. Sci.* **2016**, *83*, 12–22. [\[CrossRef\]](#)
- Ebrahim, N.; Gholamhossein, H.; Mehdi, J.; Hossein, F.; Morteza, M. Safety performance evaluation in a steel industry: A short-term time series approach. *Saf. Sci.* **2018**, *110*, 285–290.
- Jiang, H.; Li, X.; Shao, H.; Zhao, K. Intelligent failure diagnosis of rolling bearings using an improved deep recurrent neural network. *Meas. Sci. Technol.* **2018**, *29*. [\[CrossRef\]](#)
- Zhang, C.-Y.; Wei, J.; Jing, H.; Fei, C.-W.; Tang, W. Reliability-Based Low Fatigue Life Analysis of Turbine Blisk with Generalized Regression Extreme Neural Network Method. *Materials* **2019**, *12*, 1545. [\[CrossRef\]](#)
- Liu, X.; Liu, Z.; Liang, Z.; Zhu, S.-P.; Correia, J.A.F.O.; De Jesus, A.M.P. PSO-BP Neural Network-Based Strain Prediction of Wind Turbine Blades. *Materials* **2019**, *12*, 1889. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
- Graves, A. Supervised Sequence Labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin, Germany, 2012.
- Yang, B.; Sun, S.; Li, J.; Lin, X.; Tian, Y. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing* **2019**, *332*, 320–327. [\[CrossRef\]](#)
- Guo, L.; Li, N.; Jia, F.; Lei, Y.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109. [\[CrossRef\]](#)
- Yang, J.; Lee, D.; Kim, J. Accident Diagnosis and Autonomous Control of Safety Functions During the Startup Operation of Nuclear Power Plants Using LSTM. *Adv. Intell. Syst. Comput.* **2018**, *787*, 488–499. [\[CrossRef\]](#)

18. Adlen, K.; Abderrezak, M.; Ridha, K.; Mohamed, B. Real-time safety monitoring in the induction motor using deep hierarchic long short-term memory. *Int. J. Adv. Manuf. Technol.* **2018**, *99*, 2245–2255.
19. Sainath, T.N.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia, 19–24 April 2015; pp. 4580–4584.
20. Oh, S.L.; Ng, E.Y.; Tan, R.-S.; Acharya, U.R. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Boil. Med.* **2018**, *102*, 278–287. [[CrossRef](#)] [[PubMed](#)]
21. Tian, C.; Ma, J.; Zhang, C.; Zhan, P. A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies* **2018**, *11*, 3493. [[CrossRef](#)]
22. Stainhaouer, G.; Carayannis, G. New parallel implementations for DTW algorithms. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 705–711. [[CrossRef](#)]
23. Seokgoo, K.; Nam, H.K.; Joo, H.C. Prediction of remaining useful life by data augmentation technique based on dynamic time warping. *Mech. Syst. Signal Process.* **2020**, *136*. [[CrossRef](#)]
24. Sharma, S.K.; Phan, H.; Lee, J. An Application Study on Road Surface Monitoring Using DTW Based Image Processing and Ultrasonic Sensors. *Appl. Sci.* **2020**, *10*, 4490. [[CrossRef](#)]
25. Cui, L.; Li, B.; Ma, J.; Jin, Z. Quantitative trend fault diagnosis of a rolling bearing based on Sparsogram and Lempel-Ziv. *Measurement* **2018**, *128*, 410–418. [[CrossRef](#)]
26. Salvador, S.; Chan, P.K. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **2007**, *11*, 561–580. [[CrossRef](#)]
27. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 326–366.
28. Bodén, M. A Guide to Recurrent Neural Networks and Back Propagation. Ph.D. Thesis, Halmstad University, Halmstad, Sweden, 10 April 2001.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).