This is the Pre-Published Version.

The following publication Q. Wang, S. Bu and Z. He, "Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment With LSTM-RNN," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6509-6517, Oct. 2020 is available at https://doi.org/10.1109/TII.2020.2966033.

Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN

Qi Wang, *Member, IEEE,* Siqi Bu, *Senior Member, IEEE,* and Zhengyou He, *Senior Member, IEEE*

***Abstract*—Current maintenance mode for high-speed railway (HSR) power equipment are so outdated that can hardly adapt to the high-standard modern HSR. Therefore, a new possibility is proposed in this paper to update the obsoleting maintenance mode of the HSR power equipment by adopting both predictive maintenance and proactive maintenance. With the combination of data-driven (predictive) and model-based (proactive) approaches, two principal constituents—the sample generator and the maintenance predictor—are designed. The maintenance predictor which is powered by the long short-term memory recurrent neural network (LSTM-RNN) is developed to realize the goal of predictive maintenance. The sample generator which is formulated by the physical degradation and failure model of HSR power equipment is proposed towards the goal of proactive maintenance. Test results on a gas-insulated switchgear have shown the powerful collaboration between the generator and the predictor, to not only accurately predict future maintenance timing of the switchgear based on historical sample data, but also enrich the data supply proactively to deal with potential data deficiency problems.**

***Index Terms*—High-speed railway (HSR), power equipment, predictive maintenance, proactive maintenance, artificial intelligence (AI), deep learning, recurrent neural network (RNN), long short-term memory (LSTM) network.**

I. Introduction

SINCE firstly completed and put into operation in 2008, high-speed railway (HSR) in China has experienced tremendous and expeditious development over the past decade. With massive existing lines at present as well as an increasing number of new lines entering service in the near future, the HSR operators are being challenged by formidable maintenance tasks and enormous maintenance pressure unprecedentedly [1]. In HSR, the traction power supply system (TPSS) plays a vital role to feed the high-speed trains and constitutes the essential power lifeline. The normal operation of TPSS is guaranteed by the proper functioning of each fundamental component, i.e., the power equipment. Otherwise, any failures or faults in the power equipment will hazard the safety of HSR and even result in strong traffic disruptions or heavy casualties. Since failures and faults are inevitable for the highly intensive and persistent workload of HSR, it is extremely significant to maintain the health, safety and reliability of the power equipment, by appropriate and efficient maintenance.

According to the latest regulation [2], the current maintenance mode for HSR power equipment in China is a combination of corrective maintenance (CM) and preventive maintenance (PM). CM is simply carried out when failures occur, while PM is to repair or replace components at fixed periods before they fail [3], [4]. However, both the existing CM and PM modes are so outdated that they commonly incur either underestimation or overestimation on the maintenance time. An underestimation may lead to serious damage to power equipment and give rise to unaffordable post-contingency loss, while an overestimation can markedly increase maintenance cost and put tons of working pressure on the maintenance personnel. As a result of poor efficiency and high cost, such obsoleting maintenance modes have become a hindrance to modern HSR.

The rapid evolution of maintenance theory and technology in recent years has fostered a series of new maintenance modes, which offers a great opportunity to fully update the obsoleting maintenance modes in the current HSR. One of them is known as predictive maintenance. Referring to the standardized maintenance terminology [4], predictive maintenance is defined as: condition-based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item. Another promising maintenance mode is called proactive maintenance. It has been firstly proposed in [5] and given the definition as: maintenance activity performed to detect and correct root cause aberrations of failures. It is indicated from the above definitions that predictive maintenance is a *data-driven* strategy while proactive maintenance focuses on a *model-based* solution.

Towards the level of predictive maintenance, accurate and convincing prediction on the maintenance timing is the foremost concern, and meanwhile, the biggest obstacle. Fortunately, the recently booming development of deep learning has devoted many powerful prediction models, such as recurrent neural network (RNN) and a variant known as long short-term memory (LSTM) network [6]–[8]. And the LSTM-RNN is proven to have a prominent ability to make predictions on long-run time-sequential data [9], [10]. Nevertheless, the deep learning model is based on supervised learning, which requires a sufficient number of labeled sample data to train. But in the field maintenance practice of HSR, data tend to

This work was supported in part by the National Natural Science Foundation of China under Grant 51807171, in part by the Sichuan Science and Technology Program under Grant 2018RZ0075, in part by the Hong Kong Research Grant Council for the Research Project under Grant 25203917, 15200418, 15219619, in part by the Hong Kong Polytechnic University for the Start-up Fund Research Project under Grant 1-ZE68 and in part by the Open Project of National Rail Transit Electrification and Automation Engineering Technique Research Center in China under Grant NEEC-2019-B01. Paper no. TII-19-4230. *(Corresponding author: Siqi Bu.)*

Q. Wang and S. Bu are with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong (e-mail: clarkstarcraft@gmail.com; siqi.bu@gmail.com).

Z. He is with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, Sichuan 611756, China (e-mail: hezy@home.swjtu.edu.cn).

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

accumulate slowly, especially in those newly opened HSR lines. This creates great contradictions between the *data-driven* demand of predictive maintenance and the deficient data supply. Thus, there is a pressing need to combine the superiority of both *data-driven* (predictive maintenance) and *model-based* (proactive maintenance) modes to address such data deficiency problems.

This paper aims to explore a new possibility to update the obsoleting maintenance mode of the HSR power equipment by a combination of predictive maintenance and proactive maintenance. The contributions of this paper can be summarized as follows.

1) A novel combination of *data-driven* and *model-based* approaches is applied, to maximize their respective superiority by complementary collaboration.
2) As for the data-driven approach, an LSTM-RNN based maintenance predictor is developed, to predict future maintenance timing of HSR power equipment, by giving certain historical sample data.
3) With regard to the model-based approach, a sample generator is proposed based on the physical degradation and failure model of HSR equipment, in order to generate sample data proactively, to deal with the potential data deficiency problems.

II. Related Works

A. LSTM-RNN Prediction Models

As a crucial branch of deep learning models, RNN with its variant LSTM-RNN has proven to be remarkably powerful for various prediction tasks [6]. In the power industry, LSTM-RNN has also shown promising results when tackling a wide range of tough prediction problems, such as power fluctuations identification [11], industrial electrical equipment recognition [12], power markets uncertainty forecasting [13] and short-term residential load prediction [14], etc.

RNN is a specially designed neural network for dealing with time series. In RNN, the interplay between two successive inputs is taken over by a characterized recurrent structure, where the outputs from current time steps feed as inputs to the next time steps [6]. As a result, RNN is endowed with the unique ability to 'memorize' the dependency between each two adjacent time steps [7]. However, RNN has its limitation in long-term dependency memory, because of gradient vanishing (or exploding) problems [8]. Therefore, an evolved form known as LSTM-RNN is explored and polished to handle such long-term dependencies [9]. LSTM-RNN has a totally different but more complicated cell structure which incorporates several gates inside, in order to memorize information selectively [10]. Considering the complicated cell structure, a special algorithm named as backpropagation through time (BPTT) is specially designed for the training of the LSTM-RNN [10], [15].

For the outstanding prediction ability on long time sequences, the LSTM-RNN is fully utilized in this paper, to fulfill the goal of predictive maintenance from the data-driven prospect. Although diverse cell variations of the LSTM-RNN have been invented, this paper uses a typical one proposed in [16].

B. Maintenance of HSR Power Equipment

Comparing to the fast development of HSR, the explorations of maintenance in HSR power equipment has made very limited progress. In this connection, rare studies can be found over the past decade. Early studies have made efforts to optimize the maintenance schedules for power equipment in TPSS [17]–[19], even though these works are still under the outdated PM mode [20]. To go a step further, follow-up studies [21]–[24] have sought help from some newer modes, e.g., the condition-based maintenance (CBM) mode (whose definition can also refer to [4]).

Among these pioneer studies, two innovative papers [23], [24] have made impressively the latest progress. These two papers have conducted deep explorations on the deterioration and failure mechanisms of the power equipment under severe traction load conditions of HSR. On this basis, a hybrid inspection-maintenance strategy has been proposed and optimized to strike the balance of high reliability (availability) and low maintenance cost in the long-run lifetime of the power equipment. However, the maintenance strategy proposed in [23], [24] is still a joint application of PM and CBM. It has no prediction ability so far. This is because the maintenance strategy still requires periodic inspections to reveal non-self-announcing faults inside the power equipment. No matter how far to be optimized, these unavoidable inspections will still

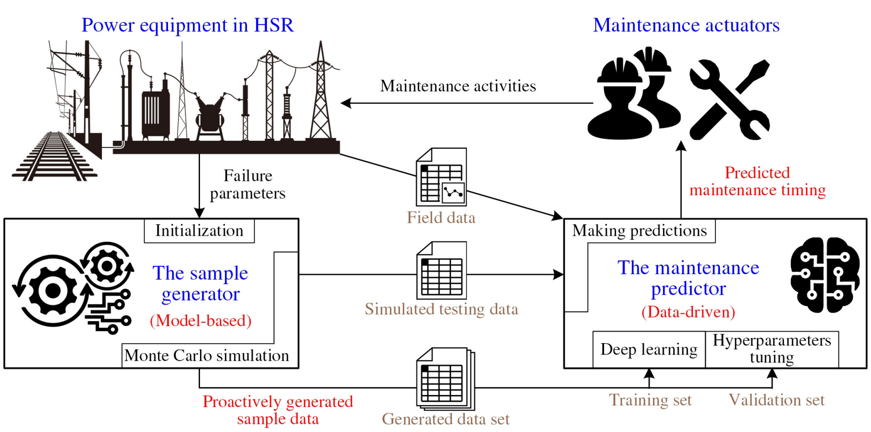

Fig. 1. Framework of the predictive and proactive maintenance for HSR power equipment.

incur additional maintenance (inspection) cost and hold extra maintenance (inspection) time in the limited window period of HSR maintenance.

Inspired by [23], [24], more specifically by the deterioration and failure mechanism revealed in them, another goal of proactive maintenance in this paper is achieved from the model-based (physical degradation and failure model) prospect. Moreover, the limitation of [23], [24] (i.e., lack of prediction ability) is addressed in this paper by the successful application of the LSTM-RNN prediction model.

III. Methodology

A. Conceptual Framework Overview

For a clear overview, the conceptual framework of this paper is exhibited in Fig. 1. As indicated in the framework, two primary components are designed as follows.

1) The sample generator: is basically a physical model which is rigorously derived from the theory of probability and stochastic process. It describes the statistical behavior of the random failure time of the HSR power equipment which undergoes a complicated deterioration and failure mechanism. The generator only requires relevant failure parameters (detailed explanation of which will be given in Section III-B) from the HSR power equipment for initialization. Then it can proactively produce simulated sample data as much as needed, and feed for training, tuning and testing of the maintenance predictor. This model-based sample generator is designed for proactive maintenance.

2) The maintenance predictor: is formulated by LSTM-RNN under the deep learning architecture. After trained, tuned and tested by simulated data from the sample generator, it can make accurate predictions based on historical field data. Then the predictor determines the predicted maintenance timing (i.e., the predicted failure time) of the HSR power equipment, for the maintenance actuators (maintenance personnel) to launch field maintenance activities on it. This data-driven maintenance predictor is the core realization of predictive maintenance.

B. The Sample Generator

As for HSR, one of the most remarkable electrical characteristics is that the load of TPSS (also called the traction load) always performs as frequent and abrupt shocks. It creates a harsh working condition for the power equipment. As a result, the power equipment may not only be prone to instantaneous failure by intense shocks, but also suffer from accelerated degradation under the same shock stresses. The former failure by sudden trauma is defined as the hard failure, while the latter by progressive aging is defined as the soft failure. Accordingly, the HSR power equipment is subject to a competing failure mode by these two failure processes. And the maintenance activities are assumed to restore the equipment to an as-good-as-new state each time.

1) Hard Failure Process: With respect to the hard failure mechanism, considering both temporal and spatial randomness of shocks generated by traction load, a compound Poisson process is introduced to describe the hard failure process of power equipment in TPSS. In the compound Poisson process, the temporally stochastic arrival time of shocks is described by a homogeneous Poisson process while the spatially stochastic magnitudes of shocks are considered to follow a normal distribution. Two thresholds associated with space and time, denoted as H and δ, are introduced respectively to be relevant to the extreme shock model and δ-shock model [25], the detailed meanings of which are explained as follows.

- *Extreme Shock Model*: Extreme shocks are those magnitudes of which exceed the space threshold H. It is one cause of the hard failure. Let Y_i denotes each magnitude of the ith shock which follows a normal distribution with expectation μ_Y and standard deviation σ_Y. The random failure time under extreme shock model is denoted as τ_H. Then the survival function with respect to τ_H can be deformed in regulation of the spatial threshold H by the ith shock as

$$\begin{aligned} F_{\tau_H}(t) &= \Pr\{\tau_H < t\} = \Pr\{\cap\{Y_i < H\}\} \\ &= \left[\int_{-\infty}^{H} \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{(t-\mu_Y)^2}{2\sigma_Y^2}\right) dt\right]^i \end{aligned} \tag{1}$$

- *δ-Shock Model*: δ-shock occurs when the inter-arrival time of any two sequential shocks is less than the time threshold δ. This is another cause of the hard failure. As shocks arrive stochastically according to a homogeneous Poisson process with intensity λ, the power equipment survives when the temporal threshold δ falls within the time interval I_i between the ith and $(i+1)$th shocks. Thus the survival function under δ-shock model, which involves corresponding failure time τ_δ by the ith shock, is formulated as

$$\begin{aligned} F_{\tau_\delta}(t) &= \Pr\{\tau_\delta < t\} = \Pr\{\cap\{I_i > \delta\}\} \\ &= \exp\left(-i\lambda\delta\right) \end{aligned} \tag{2}$$

By merging the extreme shock and δ-shock models, the survival function of the power equipment suffering from hard failures is the product of (1) and (2).

2) Soft Failure Process: On top of instantaneous hard failure by shocks, the power equipment is also subject to progressive degradation. This progressive degradation can be treated as a gradual aging process superposed by several accelerated damages. The respective mechanisms of the gradual aging and the accelerated damage are described by two non-identically distributed stochastic processes as follows.

- *Gradual Aging*: The spontaneous deterioration such as wear, fatigue, erosion and insulation defecting inside the power equipment are principal culprits of the gradual aging process. This monotonically increasing process is assimilated by a gamma process with shape parameter α and scale parameter β. By giving the degradation degree which is denoted as a time-dependent random variable $X(t)$, the characteristic function of $X(t)$ under gradual aging is expressed by

$$\phi_X(t) = \left(\frac{\beta}{\beta - jt}\right)^\alpha \tag{3}$$

where j is the imaginary unit.

- *Accelerated damage*: Load shocks not only contribute directly to hard failures, but also play a significant role in accelerating the progressive degradation. More specifically, shocks will superpose additional accelerated damage onto initial degradation degrees by the basic gradual aging. This accelerating process can be described by another compound Poisson process which combines: a) the same homogeneous Poisson process as in hard failure process with intensity λ and b) a different normal distributed random variable Z_i with expectation μ_Z and standard deviation σ_Z. Z_i is the accelerated damage caused by the ith shock. Let $S_Z(t)$ denote the total accumulated damage till time t, then the characteristic function of $S_Z(t)$ can be derived as

$$\phi_{S_Z(t)}(u) = \exp\left(\lambda t\left[\exp\left(j\mu_Z u - \frac{\sigma_Z^2 u^2}{2}\right) - 1\right]\right) \tag{4}$$

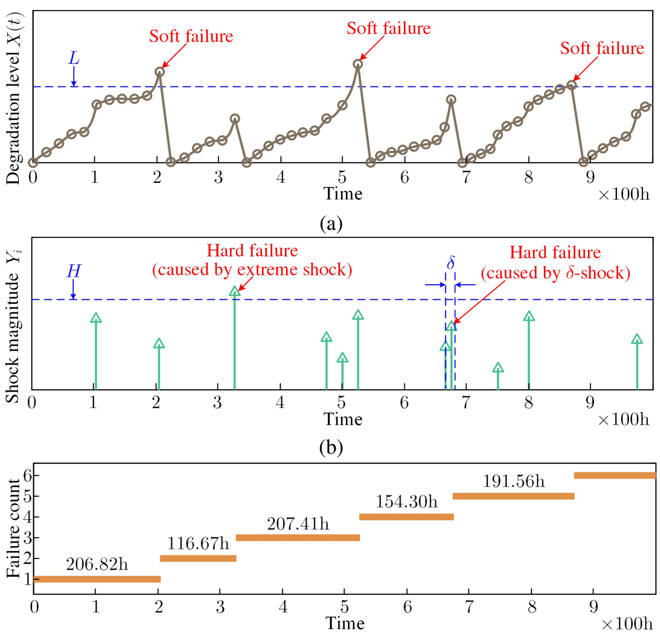

Fig. 2. A numerical example of the sample generator: (a) the soft failure process; (b) the hard failure process; (c) the generated sample data (i.e., the failure time).

Apparently, soft failure occurs when the random failure time τ_L falls within the studied time interval $[0, t]$, which is equivalent to the fact that the overall degradation level reaches a prefixed threshold L. Therefore, the survival function of the power equipment suffering from soft failure can be described by the probability of the event $\{(X(t)+S_Z(t)) < L\}$. And this probability is finally derived into (5) as shown at the bottom of this page. The detailed derivation refers to [24].

In the sample generator model, the failure parameters are basically those variables other than random variables. For a clear explanation, these failure parameters are listed in Table I with their respective significance. Furthermore, a numerical example is demonstrated in Fig. 2 to facilitate an intuitive understanding beyond the complex mathematics of the sample generator.

TABLE I
LSIT OF THE FAILURE PARAMETERS AND THEIR SIGNIFICANCE

Name	Significance
λ	Intensity of the homogeneous Poisson process in the δ-shock model
μ_Y	Expectation of the normal distribution in the extreme shock model
σ_Y	Standard deviation of the normal distribution in the extreme shock model
μ_Z	Expectation of the accelerated damage (normal distribution)
σ_Z	Standard deviation of the accelerated damage (normal distribution)
α	Shape parameter of the gamma process in gradual aging
β	Scale parameter of the gamma process in gradual aging
L	Degradation threshold of the soft failure process
H	Shock magnitude threshold of the hard failure process
δ	Shock inter-arrival time threshold of the hard failure process

C. The Maintenance Predictor

On the basis of the LSTM-RNN, the maintenance predictor is designed as a deep learning model stacked by multiple layers of neural network variants. The overall architecture of the maintenance predictor is sketched in Fig. 3. The architecture shows not only linear stacks of layers, but also layer-sharing (concatenating) structures among them. These concatenating structures are devoted to improving interpretability as well as to enhancing predictability.

Starting with the input layer, the data samples are prepared in the form that each sequence (sample) has a total length (time steps) of 20. The first 19 time steps feed in sequentially as the inputs, while the last one time step is retained as the target to be compared with the final output of the predictor.

Then the input sequences are operated by three successive LSTM-RNN layers with dropout. In each LSTM-RNN layer, a cell variation proposed in [16] is applied, with 20 neurons integrated. And hyperbolic tangent functions are chosen as activation functions of the LSTM cell's input and output units. To further avoid the gradient vanishing and to increase computational efficiency, the activation functions of three gate units in the LSTM cell are configured as rectified linear unit (ReLU) [26]. In addition, dropout is a regularization method to prevent neural networks from overfitting and hence to improve model performance [27]. In the architecture, dropout is treated as a sublayer of the LSTM-RNN layer and the dropout rate is set to 20%.

After merging and outputting from the third LSTM-RNN layer, the sequence is taken over by a fully connected layer,

$$\begin{aligned} F_{\tau_L}(t) &= \Pr\{(X(t)+S_Z(t)) < L\} \\ &= \frac{1}{2} - \frac{1}{2\pi}\int_0^{\infty}\left[\frac{\exp(-juL)}{ju}\cdot\left(\frac{\beta}{\beta-ju}\right)^{\alpha}\cdot\exp\left(\lambda t\left[\exp\left(j\mu_Z u-\frac{\sigma_Z^2u^2}{2}\right)-1\right]\right)\right. \\ &\qquad \left. - \frac{\exp(juL)}{ju}\cdot\left(\frac{\beta}{\beta+ju}\right)^{\alpha}\cdot\exp\left(\lambda t\left[\exp\left(-j\mu_Z u-\frac{\sigma_Z^2u^2}{2}\right)-1\right]\right)\right]du \end{aligned} \tag{5}$$

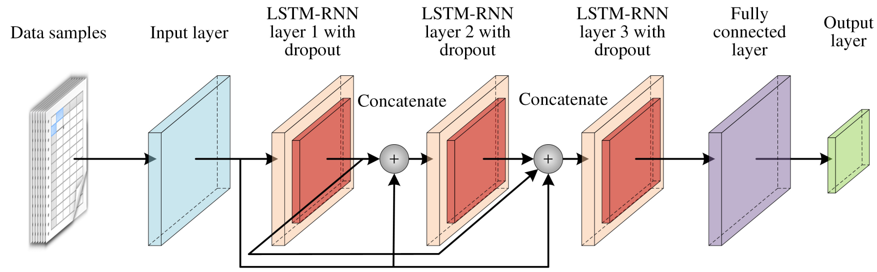

Fig. 3. Architecture schematic of the maintenance predictor.

which consists of 20 neurons with linear activation. Finally, output of the predictor is given by the last output layer, which is indeed another fully connected layer with only one neuron and has sigmoid function as the final activation. By comparing the final output (prediction) with the previous retained target, the loss is calculated, and the predictor can learn to minimize the loss by training and hence present the optimum prediction for the time sequence.

Finally, Fig. 4 summarizes the whole workflow showing the collaboration between the sample generator and the maintenance predictor. By inputting failure parameters, the sample generator can proactively generate sufficient sample data (i.e., recorded failure time). These sample data are then prepared into different data sets which feed into the predictor for training, tuning and testing. When field data (i.e., historical maintenance records) input, the predictor can output predictions of next maintenance timing.

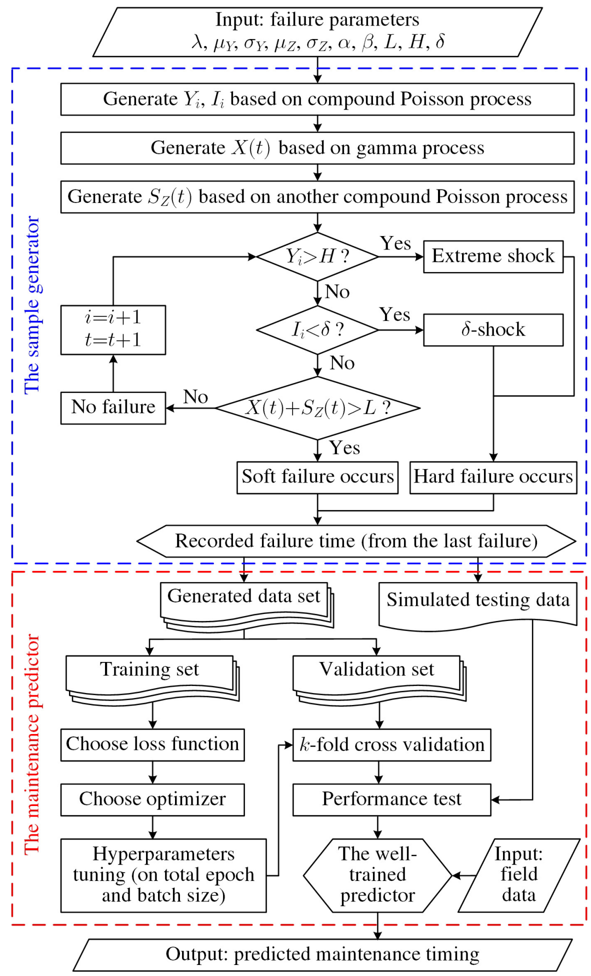

Fig. 4. Workflow of the maintenance predictor with the sample generator.

IV. Case Study

As a frequently switching equipment on the load current, the load-side switchgears are the principal failure-prone components in the TPSS. Usually, a considerable amount of failure records will accumulate in a relatively short period, and hence sufficient samples in size will be collected for training and testing of the predictor. Therefore, in the case study, a gas-insulated switchgear (GIS) installed on the load side of the Fuxin traction substation of Beijing-Shenyang HSR in China is investigated. Values of all failure parameters, involving both hard failure and soft failure, are listed in Table II, with their sources given in footnotes of the table.

TABLE II
FAILURE PARAMETERS OF THE GIS

Paramter	Value	Paramter	Value
λ	8.7245[a]	α	0.0625[b]
μ_Y	2658.0278[a]	β	0.1000[b]
σ_Y	79.4660[a]	L	18.00[b]
μ_Z	6.2590[b]	H	2880.00A[c]
σ_Z	0.2537[b]	δ	28.36h[c]

[a] By maximum likelihood estimate (MLE) method.
[b] By manufacturers.
[c] By the optimization method given in [24].

After the configuration of the failure parameters, the sample generator runs for 10^5 times. And the length of each run is limited to preserve an exact 20 failures count in it. Then a data set that consists of 10^5 samples and 20 observations in each sample is prepared for training of the predictor. In order to obtain relative gradients in the BPTT algorithm, an appropriate loss function should be fixed before the training procedure. In this paper, the log-cosh loss function is chosen and given by

$$C = \frac{1}{N}\sum_{t=0}^{N}\ln(\cosh(\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)})) \tag{5}$$

where $\mathbf{y}^{(t)}$ and $\hat{\mathbf{y}}^{(t)}$ are the outputs vector (i.e., predictions) by the LSTM-RNN cell and the targets vector at time step t.

On top of the loss function, the optimizer is another crucial argument for the training of the maintenance predictor. A

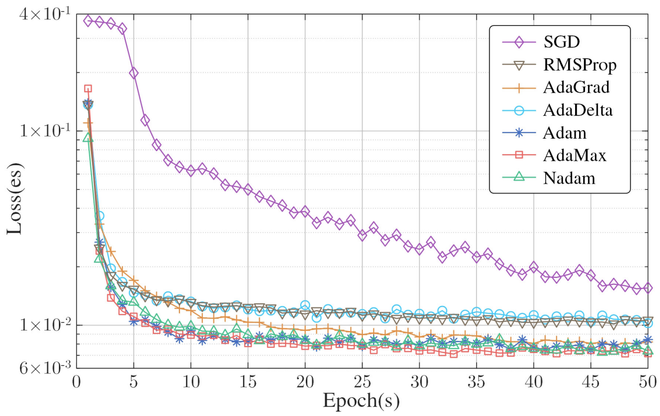

Fig. 5. Comparison of various optimizer candidates in the training procedure.

TABLE III
THE MAINTENANCE PREDICTOR'S PERFORMANCE ON 10-FOLD CROSS VALIDATION

Split	Validation Loss	Split	Validation Loss
Fold #01	6.8357×10^{-2}	Fold #06	1.0114×10^{-2}
Fold #02	9.5907×10^{-3}	Fold #07	7.3478×10^{-2}
Fold #03	9.8628×10^{-3}	Fold #08	7.3287×10^{-2}
Fold #04	6.8639×10^{-2}	Fold #09	6.4272×10^{-2}
Fold #05	4.8911×10^{-2}	Fold #10	2.3303×10^{-2}
Mean	4.4981×10^{-2}	Standard deviation	2.8438×10^{-2}

wisely chosen optimizer is supposed to decrease gradients of the loss function as fast and much as possible. In this case, a group of optimizer candidates is applied to train the predictor. These candidates include stochastic gradient descent (SGD), root mean square propagation (RMSProp), adaptive gradient algorithm (AdaGrad), AdaDelta (an extension of AdaGrad), adaptive moment estimation (Adam), AdaMax (a variant of Adam) and Nesterov-accelerated adaptive moment estimation (Nadam) [28], [29]. The training performance of 50 epochs on each optimizer candidate is illustrated in Fig. 5. It can be observed that AdaMax shows the best performance on both convergence speed and overall losses among all these candidates. Therefore, AdaMax is chosen as the final optimizer to train the predictor in the case study.

For a more robust and efficient training, the mini-batch technique [30] is applied to parallelize gradient descent in the training procedure. As a balance between online learning and the batch method, mini-batch promises a more stable convergence and less computational cost in practice. However, the implementation of mini-batch brings in an additional hyperparameter—the batch size, which requires extra tuning. Furthermore, the total epoch of training is another crucial hyperparameter that can prominently impact the performance of the predictor. Therefore, tuning on these two hyperparameters is fulfilled and the results are shown in Fig. 6.

The tuning is measured by an experiment that runs on a random split of the original data set (holding 80% for training and the remaining 20% for validation). A total number of 100 runs of the experiment are repeated to embrace randomness in

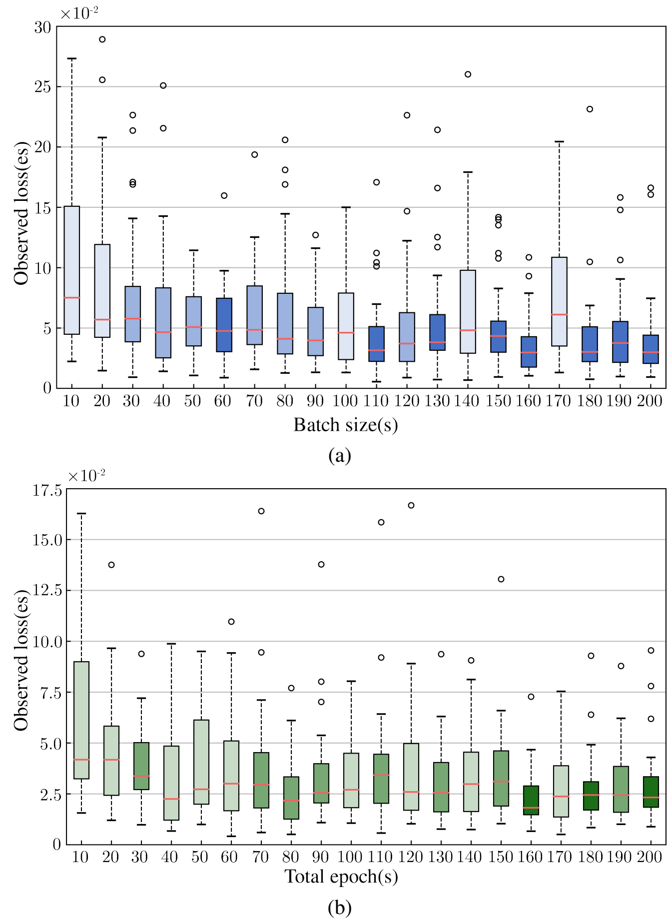

Fig. 6. Results of hyperparameter tuning: (a) variable batch size with a fixed total epoch of 50; (b) variable total epoch with a fixed batch size of 110.

the data resampling and the training procedure. From Fig. 6(a), one can find that batch sizes of 110, 160, 180 and 200 show lower quartiles (Q1, median and Q3) than others. Nevertheless, considering the fact that increasing of the batch size commonly reduces the rate of convergence, a smaller batch size of 110 could be the best choice among them. As for the total epoch, Fig. 6(b) shows that the best skill appears on a choice of 160. Combining with the cognition that less total epoch requires less time to train, an ideal configuration comes to 160 total epoch.

After fixing all these hyperparameters, the predictor should be trained again and the predictor's performance should be evaluated, both based on the original data set. This evaluation quantifies the general prediction performance of the predictor on any unknown future data. A method of k-fold cross validation [31] is applied herein to implement the evaluation. k-fold cross validation is a robust way of evaluation for providing a less biased or less optimistic estimation on the predictor's performance. It operates by randomly dividing the original data set into k subsets, and holding back one subset for validation while training the model on the remaining $k-1$ subsets. By repeating this operation k times, each of the $k-1$ subsets is given one opportunity to be used as validation data in turn. Then the predictor's performance is estimated as the average on all results of these k folds. Since 3, 5 and 10 are commonly empirical choices of k, this case study uses $k = 10$,

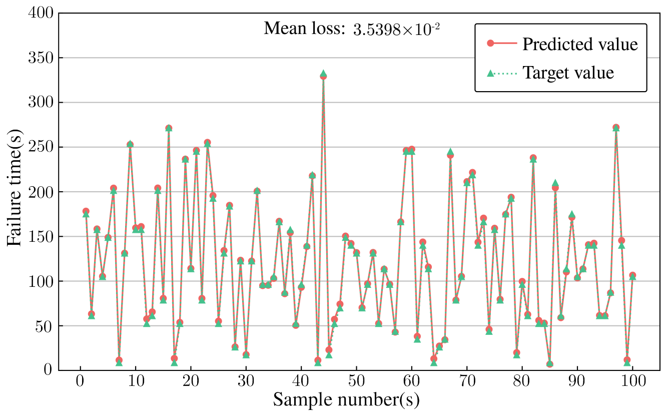

Fig. 7. Prediction performance on simulation data.

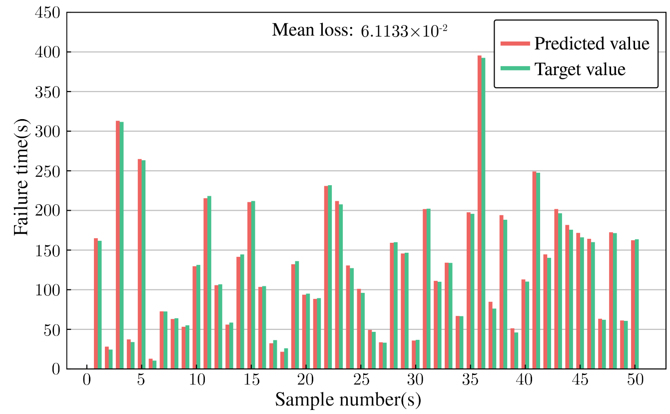

Fig. 8. Prediction performance on field data.

i.e., 10-fold cross validation for an even less bias.

The results of such evaluation are illustrated in Table III, where the predictor shows a mean performance to restrain the prediction loss as low as 4.4981×10^{-2}. Moreover, the fairly low standard deviation of 2.8438×10^{-2} promises slight fluctuations across all folds. This indicates a good generalization ability of the predictor to avoid overfitting problems on future unseen data.

In order to provide a comprehensive and intuitive insight on the predictor's performance, specific tests should be launched on absolutely unseen data which have never been used for training and validation before, and the prediction performance of the predictor on such specific data should be evaluated. The tests fall into two stages. In the first stage, brand new simulation data by the same sample generator are produced to test the predictor. The sample set with a size of 100 is generated in a failure-successive manner within each sample. The test results on this sample set are exhibited in Fig. 7. It is revealed that the predicted values can tightly track the target values of the random failure time. Furthermore, mean loss on this sample set is calculated as 3.5398×10^{-2}, which falls in the range of the general mean performance as $(4.4981 \times 10^{-2} \pm 2.8438 \times 10^{-2})$ in Table III. The results have verified the accurate prediction performance of the predictor on simulated failure time data.

In the second stage, the predictor should manifest its applicable performance when making predictions on actual field data. A block of data from the fault reports and maintenance records of the studied switchgear are collected to constitute the field test data set. The test data set covers a period of 8400.84h (approximate to one year) and contains 69 points in total. To accommodate the input size of the predictor, a sliding window with a length of 20 points is to reconstruct the test data set. As the window slides, the test data set with a total number of 50 samples and a size of 20 points in each sample is obtained. Then the predictor can make predictions on the test data set and present 50 predicted values of the random failure time successively.

The prediction results on the field test data set are demonstrated as the bar chart of Fig. 8. One can find that the predictor still shows its salient ability to present accurate predictions tracking on every target. Although the mean loss calculated as 6.1133×10^{-2} is marginally greater than those on simulation data, it is still restrained at an acceptable low level. In addition, it should be noted that the window length of 20 (nearly 4 months) claims a relatively short period of necessary points to start up the prediction. In other words, all the predictor need is only about 4 months of the history data (i.e., the fault reports and maintenance records), to start with the first prediction. Then the predictor can predict every next failure time of the studied switchgear accurately. Based on these predictions, the maintenance activities can be implemented proactively and precisely ahead of any failure occurrence.

V. Conclusion

The obsoleting maintenance modes have tremendously hindered the efficient maintenance for power equipment in modern HSR. In this paper, a new solution of predictive and proactive maintenance has been proposed for the field maintenance practice of the HSR power equipment. By combining both data-driven (predictive) and model-based (proactive) approaches, two principal constituents, i.e., the sample generator and the maintenance predictor, are designed. Towards the goal of predictive maintenance, the maintenance predictor which is powered by LSTM-RNN, has shown its prominent ability to predict future maintenance timing based on historical sample data. For reaching the goal of proactive maintenance, the sample generator which is developed by the physical degradation and failure model of HSR power equipment, has provided a scientific and rational way to enrich the data supply for the maintenance predictor.

The idea proposed in this paper opens up a new orientation for researchers in the area of maintenance theory. The utilization of deep learning model such as LSTM-RNN is proven to be powerful to realize predictive maintenance, at least on a preliminary stage. Follow-up studies are encouraged to jointly improve the advanced theory of predictive maintenance. Furthermore, the proposed idea also provides a practical paradigm for practitioners working with daily HSR maintenance applications. The combination of model-based and data-driven approaches helps dispel their misgivings versus the uninterpretable pure data-driven approach (e.g., the black box deep learning). And by proactive maintenance, they can produce reasonable sample data actively, rather than wait for the slowly accumulated field data passively.

References

[1] Q. Wang, Z. He, and D. Feng, "A maintenance mode decision method for traction power supply system of high-speed railway," in *2015 IEEE Conference on Prognostics and Health Management (PHM)*, Austin, TX, USA, Jun. 2015, pp. 1–7.

[2] *Maintenance regulations of high-speed railway traction substation*, China State Railway Group Std. TG/GD122-2015, 2015 (in Chinese).

[3] M. Shafiee, "Maintenance strategy selection problem: an MCDM overview," *Journal of Quality in Maintenance Engineering*, vol. 21, no. 4, pp. 378–402, Oct. 2015.

[4] *Maintenance–Maintenance terminology*, The British Standards Institution Std. BS EN 13 306:2010, 2010.

[5] E. C. Fitch, *Proactive maintenance for mechanical systems*, 5th ed. Abingdon, Oxford, UK: Elsevier, 2013.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[7] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. Ibañez, "Detection and classification of continuous volcano-seismic signals with recurrent neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 1936–1948, Apr. 2019.

[8] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International Conference on Machine Learning*, Atlanta, GA, USA, Feb. 2013, pp. 1310–1318.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[10] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[11] S. Wen, Y. Wang, Y. Tang, Y. Xu, P. Li, and Z. Tianyang, "Real-time identification of power fluctuations based on LSTM recurrent neural network: A case study on Singapore power system," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5266–5275, Sep. 2019.

[12] C.-F. Lai, W.-C. Chien, L. T. Yang, and W. Qiang, "LSTM and edge computing for big data feature recognition of industrial electrical equipment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2469–2477, Apr. 2019.

[13] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1203–1215, Mar. 2019.

[14] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[15] A. Graves, "Supervised sequence labelling," in *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2012, pp. 5–13.

[16] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Trans. Neural Netw.*, vol. 12, no. 6, pp. 1333–1340, Nov. 2001.

[17] T. K. Ho, Y. Chi, L. Ferreira, K. Leung, and L. Siu, "Evaluation of maintenance schedules on railway traction power systems," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 220, no. 2, pp. 91–102, Mar. 2006.

[18] S. Chen, T. Ho, and B. Mao, "Maintenance schedule optimisation for a railway power supply system," *International Journal of Production Research*, vol. 51, no. 16, pp. 4896–4910, May 2013.

[19] S.-k. Chen, T.-k. Ho, B.-h. Mao, and Y. Bai, "A bi-objective maintenance scheduling for power feeding substations in electrified railways," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 350–362, Jul. 2014.

[20] L. X. Min, W. J. Yong, Y. Yuan, and X. W. Yan, "Multiobjective optimization of preventive maintenance schedule on traction power system in high-speed railway," in *2009 Annual Reliability and Maintainability Symposium*, Fort Worth, TX, USA, Jan. 2009, pp. 365–370.

[21] H. Liu and Q. Li, "High-speed railway traction power supply equipment condition-based maintenance decision model based on abrasion analysis," *The Open Automation and Control Systems Journal*, vol. 6, no. 1, pp. 1669–1674, Dec. 2014.

[22] S. Lin, A. Zhang, and D. Feng, "Maintenance decision-making model based on POMDP for traction power supply equipment and its application," in *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, Chengdu, China, Oct. 2016, pp. 1–6.

[23] Q. Wang, Z. He, S. Lin, and Z. Li, "Failure modeling and maintenance decision for GIS equipment subject to degradation and shocks," *IEEE Trans. Power Del.*, vol. 32, no. 2, pp. 1079–1088, Apr. 2017.

[24] Q. Wang, Z. He, S. Lin, and Y. Liu, "Availability and maintenance modeling for GIS equipment served in high-speed railway under incomplete maintenance," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2143–2151, Oct. 2018.

[25] K. Rafiee, Q. Feng, and D. W. Coit, "Condition-based maintenance for repairable deteriorating systems subject to a generalized mixed shock model," *IEEE Trans. Rel.*, vol. 64, no. 4, pp. 1164–1174, Dec. 2015.

[26] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (IndRNN): Building a longer and deeper RNN," in *the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 5457–5466.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *the 3rd International Conference for Learning Representations*, San Diego, CA, USA, Jul. 2015, pp. 1–15.

[29] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[30] M. Li, T. Zhang, Y. Chen, and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2014, pp. 661–670.

[31] J. G. Moreno-Torres, J. A. Sáez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, Aug. 2012.

Qi Wang (S'14–M'18) received the B.S. and Ph.D. degrees in Electrical Engineering & Its Automation and in Electrical Engineering, from Southwest Jiaotong University, Chengdu, China, in 2012 and 2018, respectively. From 2016-2017, he worked as a visiting doctoral scholar with University of Tennessee, Knoxville, TN, USA.

He is currently a postdoctoral fellow with The Hong Kong Polytechnic University, Kowloon, Hong Kong. His research interests include data mining, deep learning and artificial intelligence applications in both the electric power system and the high-speed railway traction power supply system.

Siqi Bu (S'11–M'12–SM'17) received the Ph.D. degree from the electric power and energy research cluster, The Queen's University of Belfast, Belfast, U.K., in 2012, where he continued his postdoctoral research work before entering industry. Then he was with National Grid UK as an experienced UK National Transmission System Planner and Operator. He is an Assistant Professor with The Hong Kong Polytechnic University, Kowloon, Hong Kong, and also a Chartered Engineer with UK Royal Engineering Council, London, U.K.. His research interests are power system stability analysis and operation control, including wind power generation, PEV, HVDC, FACTS, ESS and VSG.

He is an Editor of IEEE ACCESS, CSEE J. POWER ENERGY SYST., and PROTECTION AND CONTROL OF MODERN POWER SYSTEMS, and a Guest Editor of IET RENEW. POWER GENER. and ENERGIES. He has received various prizes due to excellent performances and outstanding contributions in operational and commissioning projects during the employment with National Grid UK. He is also the recipient of Outstanding Reviewer Awards from IEEE TRANS. SUSTAIN. ENERGY, IEEE TRANS. POWER SYST., APPL. ENERGY, RENEW. ENERGY and INT. J. ELECTR. POWER ENERGY SYST. respectively.

Zhengyou He (M'10–SM'13) received the B.S. and M.S. degrees in Engineering Mechanics and in Computational Mechanics, in Chongqing University, Chongqing, China, in 1992 and 1995, respectively, and the Ph.D. degree in Power System & Its Automation in Southwest Jiaotong University, Chengdu, China, in 2001.

He is currently a professor with Southwest Jiaotong University, Chengdu, China. His research interests include signal processing, fault diagnosis and reliability analysis in power systems and high-speed railway traction power supply system.