IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 3, 2020, accepted November 9, 2020, date of publication November 16, 2020,
date of current version November 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037935

A CNN-LSTM Model for Tailings Dam Risk Prediction

JUN YANG[1,2], JINGBIN QU[1], QIANG MI[1], AND QING LI[2]
[1]Focused Photonics (Hangzhou) Inc., Hangzhou 310051, China
[2]National and Local Joint Engineering Laboratories for Disaster Monitoring Technologies and Instruments, China Jiliang University, Hangzhou 310018, China

Corresponding authors: Jun Yang (junyang-0309@hotmail.com) and Qing Li (lq13306532957@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0804604, in part by the Zhejiang Key Research and Development Program under Grant 2018C03040, and in part by the National Natural Science Foundation of China under Grant 61701467.

ABSTRACT Tailings ponds are places for storing industrial waste. Once the tailings pond collapses, the villages nearby will be destroyed and the harmful chemicals will cause serious environmental pollution. There is an urgent need for a reliable forecasting model, which could investigate the tendency in saturation line and issue early warnings. In order to fill the gap, this work presents a hybrid network - Long-Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), namely CNN-LSTM network for predicting the tailings pond risk. Firstly, the nonliear data processing method was composed to impute the missing value with the numerical inversion (NI) method, which combines correlation analysis, sensitivity analysis, and Random Forest (RF) algorithms. Secondly, a new forecasting model was proposed to monitor the saturation line, which is the lifeline of the tailings pond and can directly reflect the stability of the tailings pond. The CNN was used to identify and learn the spatial structures in the time series, then followed by LSTM cells for detecting the long-term dependence. Finally, different experiments were conducted to evaluate the effectiveness of the model by comparing it with other state-of-the-art algorithms. The results showed that combing CNN with LSTM layers achieves the best score in mean absolute error (MAE), root-mean-square error (RMSE) and coefficient of determination (R^2).

INDEX TERMS Deep learning, forecasting, LSTM network, real-time warning.

I. INTRODUCTION

At least 84 major tailings dam accidents were reported that caused significant damage from 1960–2020 all over the world [1]. The safety performance of tailings ponds can be obtained by manual observation or measurement analysis from specific sensors. The measurements include the saturation line, displacement, and deformation of the dam body, seepage flow, and dry beach length.

At present, a large number of researchers are devoted on tailings pond monitoring and researchers are mainly focusing on the stability status by monitoring data from sensors and make early-warnings in time. Zhai *et al.* [2] put forward the strategic goals for the development of big data in geology, and discussed the main problems and solutions facing the development of big data early-warning in geology. Huang *et al.* [3] conducted a tailings pond monitoring and early-warning system based on three-dimensional GIS, the response time of the safety monitoring and early warning system is less than 5 seconds. Li *et al.* [5] proposed a displacement analysis method to monitor the displacement of tailings dam online. In this method, the surface displacement and underground displacement of the tailings dam could be determined by the deformation coefficient of the dam bank, and the stability of the tailings dam was determined according to the displacement value. Yang *et al.* [41] proposed a machine learning prediction model to evaluate the saturation line of tailings pond by water level of tailing dam and local rainfall. Hariri-Ardebili and Pourkamali-Anaraki [43] presented a classification method called FEM-SVM to provide the reliable analysis of tailings pond, and the stability category of tailings pond can be evaluated. Gao *et al.* [6] established remote sensing interpretation using high-resolution remote sensing images. As a result, the type, quantity, and geographic location of tailings ponds could be derived from remote sensing images. Necsoiu *et al.* [7] used satellite radar interferometry

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

to monitor the tailings sedimentation by analyzing the images using synthetic aperture radar. Che *et al.* [8] assessed the risk of tailings pond by runoff coefficient, which can simultaneously determine the safety performance of multiple tailings dams. Dong *et al.* [10] set up the alarm system based on the cloud platform, where the phreatic line, rainfall, water level and limit equilibrium state parameters were used to build the prediction function, showing good performance in real-time monitoring. Qiu *et al.* [11] designed a monitoring system of saturation line based on mixed programming. In this study, the saturation line were calculated by flow rate, precipitation transition. Tailings dams are usually located in remote mountainous areas. The structure is very complicated and the dam collapse problems are almost nonlinear. As a result, the stability of the tailings pond cannot be directly observed.

Recently, with the advantages of handling almost any non-linear and linear problems, whatever low- and high-dimensions, neural network and machine learning methods have been effectively composed in real-time risk analysis and evaluation [4], [19]. Zhao [45] conducted a support vector machine model to analyze the slope reliability. Later, Hariri-Ardebili and Pourkamali-Anaraki *et al.* [43] proposed a support vector machine model to predict the tailing pond structural behaviour. Similarly, the machine learning method was also used in the crack damage detection task [12], [46]. However, the role of real-time monitoring cannot be equated with early warning and forecasting. In other words, risk prediction methods could help people perceive risk before it happens. With excellent ability to process time-series, classic prediction model such as Auto-Regressive Integrated Moving Average (ARIMA), neural network, and LSTM have been used in prediction problems [48], [49]. Prochazka [21] detected the different time-series information by the seasonal change using neural network, which clearly indicates that the time series will be affected by time changes. Later, Tseng *et al.* [28] combined the neural network with seasonal time ARIMA model to predict the production value of machinery industry and the soft drink. The researchers analyzed and identified the time series information of training data and gave the prediction value for a few days in advance. Nevertheless, different from LSTM, the ARIMA model only gets a high score at the condition of data with linear correlation or without obvious fluctuation. With the rapid development of deep learning, the CNN and LSTM have been the most popular networks. The CNN can filter out the noise data and extract important features, achieving good performance in images, speech, and time-series recognition [50], [51]. While the LSTM network has the ability to find the linear or non-linear time series information from the shallow and deep network and combine it with current memory [47], [52]. In the study of the prediction of the saturation line, Li *et al.* [53] demonstrated the feasibility of using a single LSTM model and the RMSE was less than 0.3. In the air pollution prediction task, Tao *et al.* [54] combined with the addition of a one-dimensional convolution layers before the GRU model, which obtained good prediction performance. Similarly, Pan *et al.* [55] combined CNN and GRU models in the water level prediction research, which can predict the water level in the next 5 days. Huang *et al.* [32] combined the VMD, CNN, and GRU algorithms to build a hybrid model, which was used to predict the electricity price in different seasons. Huang *et al.* [27] designed a hybrid model that combines deep neural networks with LSTM for predicting the PM2.5, which can play a big role in the prevention and control of PM2.5. These researches provides a good foundation for the establishment of the CNN-LSTM model in the saturation line prediction task. Considering the good performance of CNN and LSTM in the prediction task, a hybrid model CNN-LSTM may achieve better prediction performance to a large extent.

As the most important factor of stability of tailings dams, for every 1-meter drop in saturation line, the safety factor of static stability is increased by 0.05 or more [11]. High saturation line will lead to a decrease of the dam stability and even potentially cause leakage, landslide, and dam break [33], [34]. Therefore, the saturation line is called the lifeline of tailings dams [13]. The stability of tailings dam can be determined by their saturation line position accurately. It is imperative to establish accurate models to predict the height of saturation line and the security situation of tailings ponds. However, the prediction research of tailings pond is almost nonexistent and have poor generalization performance. For this purpose, the goal is to propose a new model that can make full use of the strengths of deep learning. In more detail, utilizing the hidden information of the previous saturation line, the model will predict the value and tendency in the next few days. The proposed model was evaluated by comparing with state-of-the-art models, which shows the two kinds of CNN-LSTM models are the most effective choice, especially the $CNN - LSTM^2$, where convolutional layers play important roles in grabbing more abstract information and pass it on to the LSTM layers. In this work, taking Jiande tailings pond, China, as the study area, three main contributions of the study are presented:

(1) Proposing a NI method using RF algorithm to fill missing values, which saved the time-series information of data as much as possible.

(2) Proposing a new hybrid CNN-LSTM network to solve the tailings pond risk prediction problem, which achieved good performance in MAE, RMSE and R^2.

(3) Comparing the CNN-LSTM model with different hyperparameters and with other state-of-the-art algorithms.

In this work, Pearson correlation coefficient, feature importance of RF model, and sensitivity analysis techniques was employed for the saturation line prediction, especially severed as tools of dimensionality reduction. After the process of dimensionality reduction, only two kinds of monitoring data were needed to restructure the saturation line data. On this basis, the hybrid CNN-LSTM model was established for further tailings pond risk forecast.

II. NUMERICAL INVERSION METHOD

In the monitoring data, due to the problems with sensors and remote data transmission, a small part of the data was missing or abnormal. The missing values may have been generated at any time, and we cannot know when the missing values are generated in advance. Missing values accounted for 1.82% of all collected data. However, because missing values may be generated continuously, for example, no data is generated at 10 consecutive time points. In this case, the data cannot provide valid time series information. Due to the existence of missing data, the accuracy of the model will inevitably be affected. It should be noticed that, for a time-series prediction problem, missing value will cause the loss of time dependence, which will restrict the performance of the prediction model [18], [29], [30], [35]. Hence, we hope to keep the data with good long-term and short-term continue information. Similarly, instead of deleting the abnormal data directly, abnormal saturation line value could be reconstructed by the NI system. The key to the solution is to find the relationship between missing value and other normal values. According to the special relationship, the missing value could be reconstructed by other normal values. However, it is hard to find the precise computing relationship between the saturation line values and other features. The method in this study was to create a direct mapping from the inputs to the outputs, using machine learning, which has the ability of finding the relationship between inputs and outputs [16].

In other words, the NI method was composed to reconstruct the data from building the RF model. By doing so, more data are achievable. In more detail, this NI system includes three steps. First, considering that a large number of parameters may have a strong correlation, it will be difficult to evaluate the importance of a single feature. Taking into account the possibility of missing values for each parameter, we should choose as few parameters as possible as the input of NI method. The study site was Jiande copper mine tailings pond, Hangzhou, Zhejiang Province, China. The monitored data were collected by different sensors installed in Jiande tailings pond, including the saturation line, displacement and deformation of the dam body, seepage flow, and dry beach length sensors. They ensured that we could obtain the desired monitoring data in real time. The research data for this work were collected from the sensors mentioned above from 2018-03-18 to 2019-04-29. For this study, the data were from 5 different positions of tailings dam, specifically the 8, 13, 17, 21, 28, 33 stage of the tailings dam, and the time interval between data was two hours. In total, 8215 data points were collected except the missing values. Notably, the underground displacement and deformation sensors were assembled by multi-section sensors, including a total of 60 single sensors, and each sensor measurement was treated as a feature. The monitoring data had a total of 82 features. The Pearson correlation coefficients [38], [39] were calculated and a heat map was drawn, which helps eliminate the characteristics with a strong correlation (correlation coefficient greater than 0.8). There were 64 left in the final data. The Pearson correlation coefficients is defined as follow:

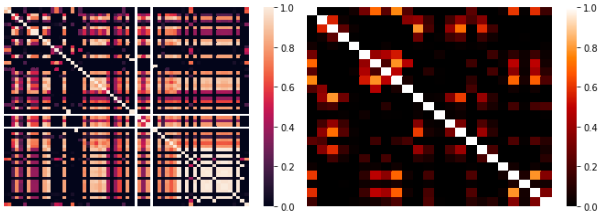

FIGURE 1. The Pearson Correlation heat maps. The left side shows the correlations among original data, the right side shows the correlations among the remaining features.

$$P_{m_i,n_i} = \frac{k\sum m_i n_i - \sum m_i \sum n_i}{[k\sum m_i^2 - (\sum m_i)^2]^{\frac{1}{2}}[k\sum n_i^2 - (\sum n_i)^2]^{\frac{1}{2}}} \quad (1)$$

where m_i, n_i are two different variables, k is the number of variables. From Figure 1, the left side shows the correlations among original data, and the right side shows the correlations among the remaining features calculated by Pearson method. Second, a RF model was composed, where the feature importance ranking generated by RF [22], [31] was composed by sorting the features according to how much accuracy they contributed to the model during building process. Third, posterior judgment is also required. We also interested in which features have great impacts on the output of the trained RF model. Thus, Sobol sensitivity analysis was adopted to explore the contribution of the individual feature and which parameters were influential and drive model outputs more [23]–[25], [39]. After establishing the model through RF, we judged the importance of the features according to whether reducing the specified features will cause the model accuracy to decrease, and ranked the importance of these features. The feature importance ranking according to the RF model and sensitivity analysis results are shown in Figure 2, where the first order represents first-order sensitivity, total order represents global sensitivity. They jointly selected $x3$ (rainfall) and $x4$ (water level) as the most important parameters. Subsequently, RF was used to create a direct mapping finding the linear and nonlinear relationship between inputs ($x3$,$x4$) and outputs (saturation line) to predict the saturation line [16]. Moreover, the abnormal data were deleted and replaced with predicted data by NI method. It should be noticed that rainfall and water level are factors that directly affect the height of the saturation line, and they have a similar time-series information. Therefore, the NI method in this study greatly preserved the time-series information of the saturation line and generate more achievable values for further deep learning prediction.

III. CNN-LSTM PREDICTION MODEL

The study aims to develop the construction of a prediction system for forecasting the saturation line the utilizing state-of-the-art LSTM and CNN networks. What has devoted to the popularity of the convolutional layer is the fact that it

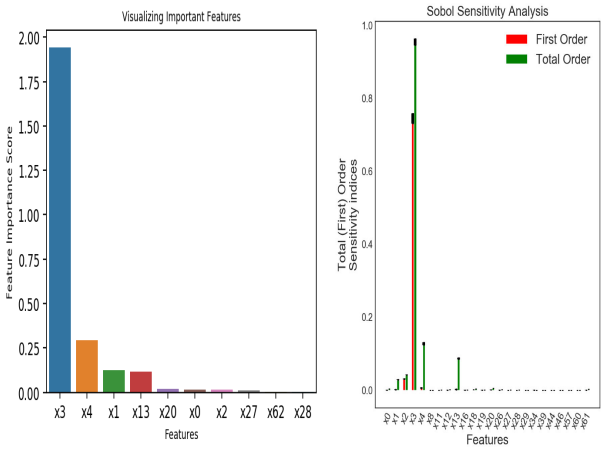

FIGURE 2. The feature importance ranking according to the RF model(left) and sensitivity analysis(right).

is good at extracting and recognizing the spatial structures of the time series in the monitoring data, while the LSTM networks achieve good performance in detecting long-short-term dependence. In light of this, the principal idea of the study is to combine the advantages of CNN and LSTM.

The proposed model in the study is named CNN-LSTM model, including two versions, which include two parts. The first part is convolutional layers and max-pooling layers, while the second part is the LSTM layers. The convolutional layers encode the time-series information, while the LSTM layer decodes the encoded the information from convolutional layers, which later will be flattened and pushed into a fully-connected layer. The CNN-LSTM auto-encoder model is shown in Figure 3.

A. CONVOLUTIONAL AND POOLING LAYERS

The convolutional layers and max-pooling layers detect the spatial structures and features of the saturation line values together reducing the redundant characteristics, respectively. More important, the convolutional layer could extract hidden information in the time dimension, and usually pass higher quality and denser features to the following layers.

More specifically, numerous useful convolved features will be generated by convolution kernels, which are always more important than the original features. As a subsampling method, max-pooling layer saves certain information from the convolved features and reduces the original data dimension. Specifically, the max-pooling layer helps to collect and summarize the features from convolutional layer.

B. LONG SHORT-TERM MEMORY (LSTM)

As a popular type of recurrent neural network(RNN), LSTM achieves good performance in detecting long-term dependencies. The problem named "lack of memory" was solved after LSTM was proposed, which means the time-series information cannot be effectively exhibited. Moreover, "vanishing gradient problem" prevents the RNN for long-time dependencies detecting. The LSTM model is composed of one memory unit and other three interactive gates: memory cell, input gate, forget gate, and output gate. The memory cell memorizes the state from the previous state. The input gate determines how much input data of the network needs to be saved to the unit state at the current moment t. The forget gate controls whether the information will be discarded or enters the input gate as reserved information at time $t-1$. The output gate determines what information will be utilized as the output. Eqs.(1)–(6) briefly describe the update in the LSTM layers.

$$i_t = \sigma(V_i x_t + W_i h_{t-1} + b_i) \tag{2}$$
$$f_t = \sigma(V_f x_t + W_f h_{t-1} + b_f) \tag{3}$$
$$\widetilde{c_t} = tanh(V_c x_t + W_c h_{t-1} + b_c) \tag{4}$$
$$c_t = f_t \otimes c_{t-1} + i_t \otimes \widetilde{c_t} \tag{5}$$
$$o_t = \sigma(V_o x_t + W_o h_{t-1} + b_o) \tag{6}$$
$$h_t = o_t \otimes tanh(c_t) \tag{7}$$

The x_t is the input data at time t, V_* and W_* denote the weight matrices, h_* is the hidden state, b_* is the bias. σ and $tanh$ are the activation function of sigmoid and tanh, respectively. i_t,f_t,c_t and o_t stand for the input gate, forget gate, memory cell and output gate, respectively. The $\otimes$ means the component-wise operation. Finally, output h_t is calculated by output gate and information in memory cell.

C. CNN-LSTM MODEL FOR PREDICTION

In the study, two different CNN-LSTM structures are utilized. The first version named $CNN-LSTM^1$, which consists of two convolutional layers of 16 and 32, a max-pooling layer filters of 2, a LSTM layer of 50, a flatten layer and a fully-connected layer in order. The second version named $CNN-LSTM^2$, which includes one convolutional layer filters of 32, a max-pooling layer filters of 2, a flatten layer, two LSTM layers with unit size of 25, 50, a flatten layer and a fully-connected layer in order. Different parameters are compared for further study. The two kinds of CNN-LSTM structures are shown in Figure 4(a) and Figure 4(b).

IV. DATA PREPARATION

The study site is Jiande copper mine tailings pond, Hangzhou, Zhejiang Province, China, where the amount of mineral copper metal accounts for about 60% of the province's total output. The main mineral products are copper concentrate, zinc concentrate, sulfur concentrate, and by-product gold and silver. The tailings pond level is III. Different geological hazard sensors are installed to monitor the surface displacement, dam body internal displacement, saturation line height, water level, rainfall, and seepage flow [26], [27], [36], [37]. The research data for this work were collected from the sensors mentioned above from 2018-03-18 to 2019-04-29, and the time interval between data was two hours. The saturation line value refers to the distance between the tailings dam and the groundwater, which is measured by liquid level sensors.

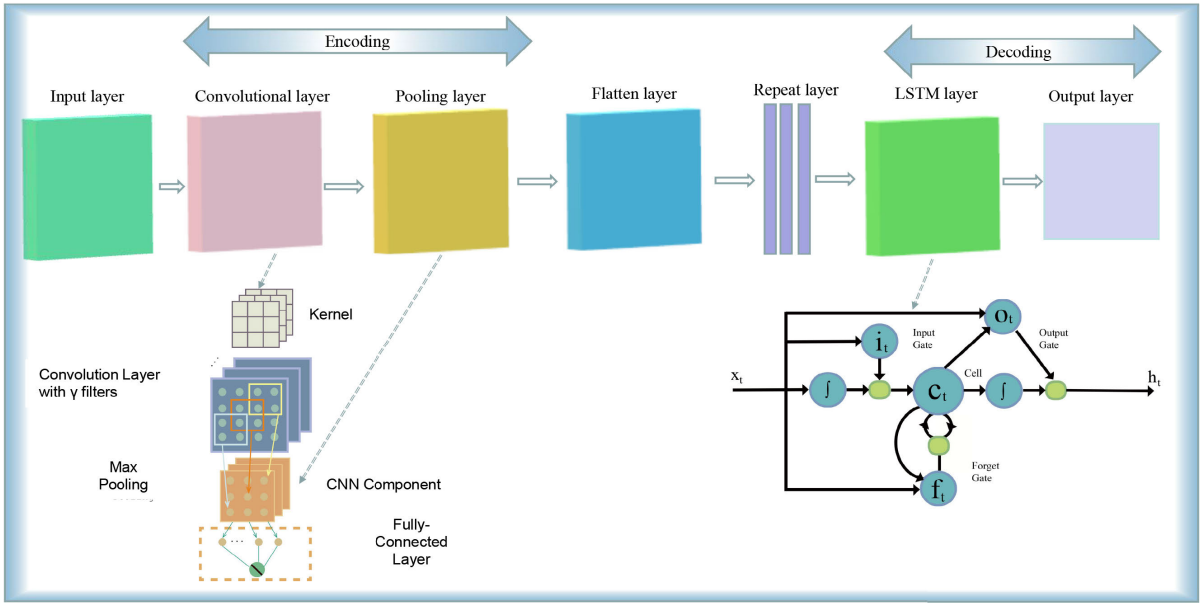

FIGURE 3. The CNN-LSTM auto-encoder model.

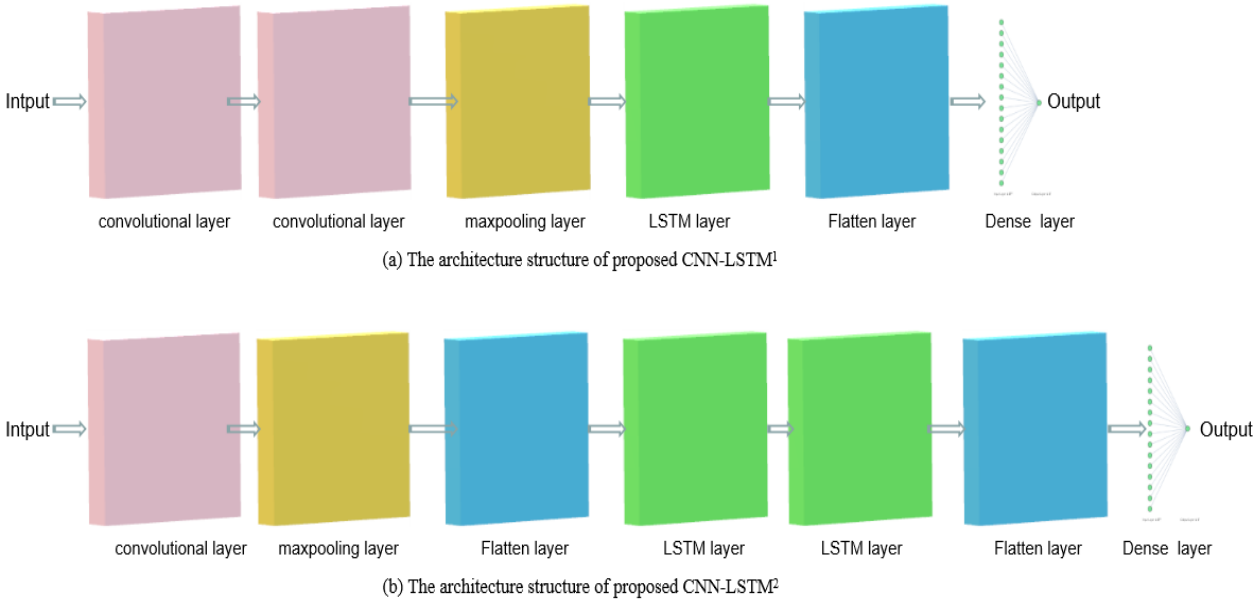

FIGURE 4. The architecture structure of proposed $CNN-LSTM^1$ and $CNN-LSTM^2$.

The tailings dam bank is a slope with steps every 2 meters. Each step has a level sensor for measuring the saturation line, and the monitoring data in the study comes from 5 different positions, specifically the 8, 13, 17, 21, 28, 33 stage of the tailings dam. The 8, 13, 17, 21, 28, 33 only represent the position of the measured saturation line. It should be mentioned that the purpose is to utilize the hidden information of the previous saturation line by the model, by finding out the relationship between data in time series and spatial dimensions, the saturation line value and tendency in the next few days can be predicted.

After collecting the data, the proposed NI system was used to fill the missing value, and the abnormal value were deleted and replaced with predicted value by NI system. Finally, 8365 data point were used for the further study. The continuous monitoring value ensures a wide range of time-series information. It should be noted that the CNN-LSTM model trained and validated on the 8365 data. Among the 8365 data, we randomly choose 70% of the data as the training sets, the 10% as the validation set. The performance of the models was evaluated on the rest 20% data, which is the unseen part during the model building process. For keeping

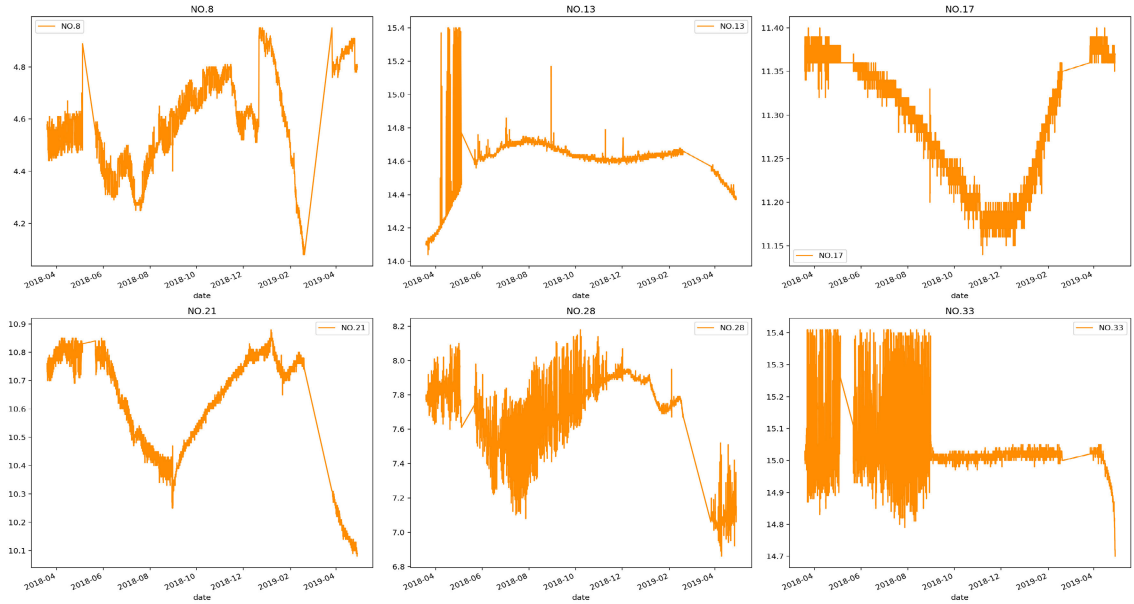

FIGURE 5. The monitored saturation line data at different positions.

the long-short-term dependence in the data, these data cannot be shuffled as usual in traditional deep learning studies. Table 1 shows the describe of the collected data, and the first three rows labeled as 1,2 and 3 are historical monitoring data. The distribution of monitoring data is shown in Figure 5. As is shown in Figure 5, there is a wide range of variation in the monitoring data. These changes are largely affected by tailings pond operations and weather change, such as the discharge of a large amount of wastewater and waste residue on a certain day or the experience of heavy rain.

In order to eliminate the impact of different data dimensions on the calculation, we used $Z-score$ normalization on the data, the formula is as follows:

$$\dot{x} = \frac{x_t - \mu_t}{\sigma_t} \tag{8}$$

where x_t is the input data, μ_t and σ_t are the averages and standard deviation of data.

V. EXPERIMENT AND RESULTS

Two different version $CNN-LSTM^1$ and $CNN-LSTM^2$ were evaluated and compared to show the prediction performance. The simulation hardware environment of this experiment is Intel Core CPU i7-8750. GPU is NVIDIA GTX 1060, and the memory is 6GB. The algorithm is implemented using Python in conjunction with the TensorFlow framework.

The loss value in the training process of the proposed model was calculated by root mean square error (RMSE). In fact, loss calculation during model training and model evaluation are not the same concept. The loss function is only used in the process of model building, and the evaluation function is used to evaluate the completed model. RMSE is very effective in back-propagation calculation of loss values, but not enough to accurately evaluate the performance of the model. RMSE meets an important problem: let us consider that although the model has an error of less than 0.5% in the 98% dataset and 95% in the other 2% dataset, the overall RMSE will be still very high, resulting in this model considered as a poor model. To solve this problem, mean absolute error (MAE) was utilized to evaluate the performance of the established model [56]. What's more, coefficient of determination, denoted as R^2 [57], was also used in the evaluation methodology. It is the proportion of the total variation of the dependent.

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(y_t - \widehat{y_p})^2]^{\frac{1}{2}} \tag{9}$$

$$MAE = \sum_{i=1}^{n}\frac{|y_t - \widehat{y_p}|}{n} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(y_t - \widehat{y_p})^2}{\sum_{i=0}^{n}(y_t - \overline{y_t})^2} \tag{11}$$

where y_t represents the true value, $\widehat{y_p}$ represents predicted saturation line value, $\overline{y_t}$ represents average of true value, and n is the count of data. Figure 6 shows the prediction results of $CNN-LSTM^1$ and $CNN-LSTM^2$ on five different monitoring sites about 1750 test sets.

In this study, we trained the model for 120 epochs with a batch size of 64, RMSE as loss function and Adam for optimizer. The Adam is an improved RMSProp optimizer combining with the moments trick. It is worth noticing that in order to reduce the feature loss during the convolutional layers, same padding operation was conducted during this process. The last but not least, the forecasting sequence length should be set properly to make sure the

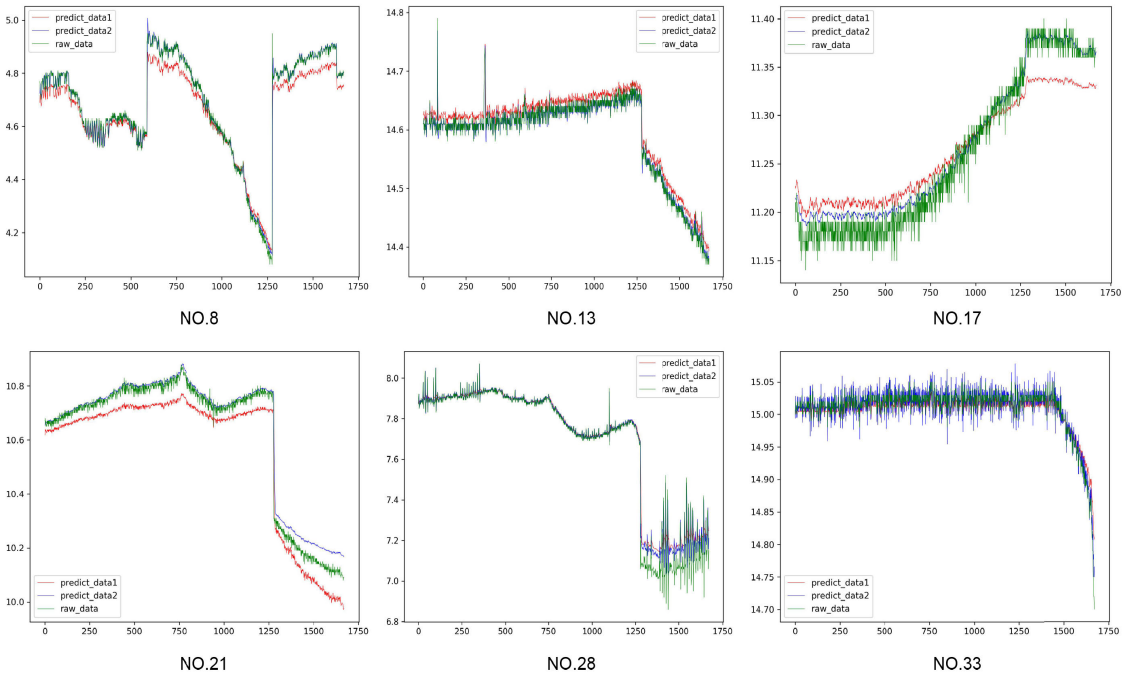

FIGURE 6. The prediction results of saturation line at different positions. The green line, red line, orange line represent the prediction value from $CNN-LSTM^1$, prediction value from $CNN-LSTM^2$ and raw data, respectively. From the prediction result, we can see the $CNN-LSTM^2$ outperform the $CNN-LSTM^1$.

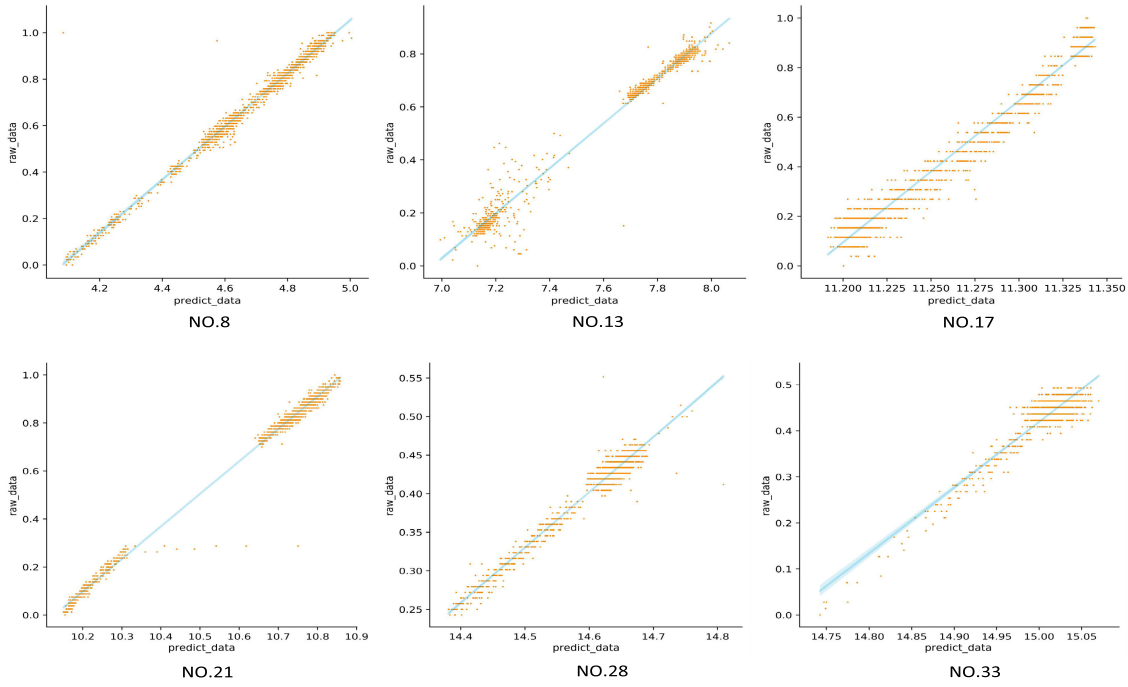

FIGURE 7. The prediction scatters of saturation line at different positions using $CNN-LSTM^2$.

model performance. The sequence length is the forecasting horizon, which specifies how many saturation line monitoring data the model needs at a time to predict the next saturation line value. Specifically, we divided the training data into different sequences of equal length of 10, which means that we use the first 10 data to predict the $11th$ data. On the one hand, considering that a longer sequence length will occupy a huge computer memory, on the other hand, too

TABLE 1. Some datasets used for saturation line prediction.

	NO.8	NO.13	NO.17	NO.21	NO.28	NO.33	Base status
1	4.56	7.78	11.38	10.76	14.10	15.01	Normal
2	4.58	7.79	11.35	10.72	14.16	15.07	Normal
3	4.59	7.81	11.37	10.81	14.18	15.27	Normal
mean	4.57	7.73	11.32	10.67	14.52	15.03	Normal
min	4.08	6.86	11.14	10.08	14.04	14.70	Normal
25%	4.51	7.68	11.28	11.28	14.23	14.96	Normal
50%	4.56	7.80	11.36	10.76	14.61	15.01	Normal
75%	4.61	7.84	11.37	10.80	14.68	15.04	Normal
max	4.95	8.18	11.40	10.88	15.40	15.41	Normal

TABLE 2. Prediction performance of the proposed $CNN-LSTM^1$ model using MAE, RMSE and R^2.

Metrics	NO.8	NO.13	NO.17	NO.21	NO.28	NO.33
RMSE	0.0214	0.0344	0.0491	0.0313	0.0134	0.0342
MAE	0.034	0.012	0.035	0.111	0.141	0.136
R^2	0.885	0.756	0.851	0.702	0.913	0.865

TABLE 3. Prediction performance of the proposed $CNN-LSTM^2$ model using MAE, RMSE and R^2.

Metrics	NO.8	NO.13	NO.17	NO.21	NO.28	NO.33
RMSE	0.0209	0.030	0.0336	0.0170	0.0123	0.0366
MAE	0.0155	0.0096	0.0144	0.0198	0.0299	0.0125
R^2	**0.969**	**0.974**	**0.937**	**0.981**	**0.951**	**0.892**

short time sequence can not detect the long-term and short-term dependencies between data well. We found through experiments that set the sequence length to 10 achieves better performance than, for example, 4, 7, 20. The most important thing for the hyperparameter selection of the model is the learning rate of the network, which has a significant influence on time consumption until convergence [58]. If the learning rate is set too large, the loss function will be difficult to converge, resulting in a lower final detection accuracy; On the contrary, a small learning rate will lead to slow convergence and increase the training time. At first, we chose the learning rate 0.0001, 0.01, 0.001, and then used $cross-val-score$ from $Scikit-learnlib$ for cross-validation, which can help to determine the optimal learning rate for each partial network [40]. The result shows that the minimum MAE, RMSE, R^2 can be obtained when the learning rate is 0.001.

The prediction performance of the proposed $CNN-LSTM^1$ and $CNN-LSTM^2$ are shown in Table 2 and Table 3, respectively. NO.8, NO.13, NO.17, NO.21, NO.28, NO.33 means the different station of saturation line mentioned above. The $CNN-LSTM^1$ consists of two convolutional layers of 16 and 32, a max-pooling layer filters of 2, a LSTM layer of 50, a flatten layer and a fully-connected layer. While the $CNN-LSTM^2$ includes one convolutional layer filters of 32, a max-pooling layer filters of 2, a flatten layer, two LSTM layers of 25, 50, a flatten layer and a fully-connected layer. From Table 2 and Table 3 combing with Figure 6, we can conclude that in terms of RMSE, MAE and R^2, the proposed model $CNN-LSTM^2$ outperform the $CNN-LSTM^1$. More specifically, the model which includes one convolutional layer, one max-pooling layer, a flatten layer, two LSTM layers, a flatten layer and a fully-connected layer is more accurate. In fact, even the convolutional layer is good at extraction and recognition, which could detect the spatial features of the saturation line value well. The deep and abstract features the convolutional layer learned may be different from the ordinary time-series information from the raw data. This is obviously a disadvantage when the monitoring data contains only simple information. Therefore, more convolutional layers and more convolution kernels cannot improve the accuracy of the model. While using one convolutional layer is more suitable and two LSTM layers can capture the long-short-term data dependencies to a significant degree from the result. It is worth noticing that we have tried to add a new LSTM layer with a unit of 25 to the $CNN-LSTM^2$ model, and 83,154 trainable parameters are generated in the network. The parameters that need to be trained is 15 times the total number of data sets. Not only does the model training take twice as much time, but the model is under-fitting and performs poorly. The scatter plots of raw data and predicted saturation line is illustrated in Figure 7, which helps show the prediction performance more intuitively.

TABLE 4. Performance comparison of several machine learning and deep learning models.

Model Type	RMSE	MAE	R^2	Runtime (second)
SVR	0.132	0.101	0.548	–
DTR	0.141	0.073	0.489	–
RFR	0.251	0.193	0.839	–
MLP	0.0504	0.0552	0.798	44.08
RNN	0.0308	0.0402	0.864	47.54
GRU	0.0221	0.0366	0.879	63.55
LSTM	0.0214	0.0358	0.887	77.08
$CNN-LSTM^2$	**0.0209**	**0.0155**	**0.969**	**25.49**

To show the superiority of the proposed model $\&CNN-LSTM^2$, we applied comparative studies with other state-of-the-art machine learning and deep learning models, including the support vector regression (SVR), decision tree regression (DTR), random forest regression (RFR), multi-layer perception (MLP), single gate recurrent unit (GRU), simpleRNN as well as LSTM models. Table 4 presents the RMSE, MAE and R^2 score of these models in the experiments. The SVR, DTR, RFR models are machine

TABLE 5. Prediction cases using different hyperparameters.

	Batch Size	Conv Size	Pooling Size	LSTM Size	RMSE	MAE	R^2	Runtime
Case 1	32	16	2	[25,50]	0.0211	0.0175	0.899	50.87
Case 2	**64**	**32**	**2**	**[50,75]**	**0.0208**	**0.0153**	**0.971**	**49.37**
Case 3	**32**	**16**	**4**	**[25,50]**	**0.0314**	**0.0347**	**0.792**	**21.13**
Case 4	64	32	4	[50,75]	0.0304	0.0332	0.801	21.88
Case 5	**16**	**16**	**2**	**[50,50]**	**0.0297**	**0.0151**	**0.975**	**51.32**
Case 6	128	16	4	[25,50]	0.0321	0.0362	0.784	25.12
Case 7	16	32	2	[25,75]	0.0205	0.0178	0.902	50.52
Case 8	128	32	2	[25,50]	0.0311	0.0163	0.932	37.49
Case 9	**64**	**32**	**2**	**[25,50]**	**0.0209**	**0.0155**	**0.969**	**25.49**

learning models, which are essentially different from deep learning. They do not involve multi-layer networks, so the calculation time is very short, usually less than 0.1 seconds. Compared with deep learning, machine learning cannot effectively learn the information between data, so they can not achieve good performance. In the study, we compared these traditional machine learning with deep learning models to highlight the effectiveness of the models. The results demonstrate that the $\&CNN-LSTM^2$ method significantly outperforms the others in R^2. Besides, the runtime for 120 epochs is much less than other deep learning models. It should be mentioned that over-fitting is one of the trickiest obstacles in applied deep learning. Simply checking the accuracy of the test set cannot effectively prove the accuracy of the model, because the model may be over-fitted. In this task, by observing the loss value, the training set error decreases gradually and tends to be constant. Also, over the test set, the loss tends to be constant. In addition, it is worth noting that there is no obvious pattern in the change of data as can be seen from the original data (Figure 5), and the distribution of training set and test set are quite different. Even if the test set and training set data are very different, the model can still show good predictive performance on the test set. It fully proves that the model has proves that the model has good predictive ability and there is no over-fitting.

In order to build the complete saturation line prediction model and show the reliability of the &CNN-LSTM^2 model together with parameters set, different hyperparameters was compared, such as batch size, filters in of the convolutional layers, max-pooling size, number of LSTM cells in the experiments. Table 5 lists the different situations of combing multiple hyperparameters. In term of the evaluation metrics, although Case 2 and Case 5 achieve a litter bit higher performance than the model using ordinary hyperparameters, the Runtime is almost twice the &CNN-LSTM^2 model, excessive running time will reduce the real-time performance of prediction, especially when the amount of data is very large. The disadvantage is more pronounced for a large amount of data, and this incurs no loss of generality. Case 3 need the least Runtime but achieve low accuracy. As a result, the &CNN-LSTM^2 with one convolutional layer and two LSTM layers become the best performer. This is also in full compliance with deep learning logic. Although the padding method restricts the feature loss of the time-series data to some extent, the pooling layer inevitably loses part of the data information. Considering the accuracy and running time of the model, the model parameters were kept as same as the ordinary model. To be clear, the batch size is equal to 64, one convolutional layer filters of 32, a max-pooling layer filters of 2, two LSTM layers of 25 and 50. When the tailings ponds meet more complex situation, the data will become very complicated and internal time-series information is harder to calculate. At that time, single shallow deep learning layer lacks the capability to capture complex information. Deeper layers of LSTM cells will be more suitable.

VI. DISCUSSION AND CONCLUSION

In this work, a new method was applied to predict the safety of tailings pond according to the saturation line using $CNN-LSTM^2$ model, which is also first used in tailings pond risk prediction. Compared with the traditional methods, the risk evaluation method of tailings ponds has the characteristics of high accuracy and high real-time performance. The contributions of this work is two fold: Firstly, a NI system (including Pearson correlation coefficients, sensitivity analysis and random forest algorithms) was applied for reconstructing missing and abnormal values of saturation line by water level and rainfall. It should be observed that the water level and rainfall have the same time-series information with saturation line. Secondly, two CNN-LSTM models, especially the $CNN-LSTM^2$ model is shown to outperform other state-of-the-art models, such as SVR, DTR, RFR, MLP, RNN, GRU and LSTM. Conclusively, although these models can also achieve good performance, the

$CNN-LSTM^2$ still far ahead in RMSE, MAE, R^2. Moreover, the Runtime of $CNN-LSTM^2$ is another advantage, which is more pronounced in a larger amount of dataset. Thirdly, for a better understanding of the meaning of hyperparameters, more experiments were conducted using different Batch size, convolutional layer filter size, max-pooling size and LSTM cell size.

In tailings pond risk prediction task, these experiments consequently provide applicability of the $CNN - LSTM^2$ model. It is worth mentioning that the $CNN - LSTM^2$ model could also be applied in other time-series predictions including water level prediction, weather prediction and air quality prediction. It is evident that the model can not only to extract and recognize the structures in the time series and spatial features, but also identify long-term and short-term series information of the data.

In the future, we will focus on more factors of the safety monitoring parameters of the tailings pond, such as the underground displacement, ground displacement and dry beach length. Furthermore the risk level corresponding to the monitoring parameters of the tailings pond should be built to more intuitively reflect the safety of the tailings pond in the future work.

NOMENCLATURE

LSTM	Long-Short-Term Memory.
ARIMA	Auto-Regressive Integrated Moving Average.
CNN	Convolutional Neural Network.
NI	Numerical inversion.
RF	Random Fores.
MAE	Mean absolute error.
RMSE	Root-mean-square error.
R^2	Coefficient of determination.
RNN	Recurrent neural network.
NO.8	The 8th stage of the tailings dam.
NO.13	The 13th stage of the tailings dam.
NO.17	The 17th stage of the tailings dam.
NO.21	The 21th stage of the tailings dam.
NO.28	The 28th stage of the tailings dam.
NO.33	The 33th stage of the tailings dam.
SVR	Support vector regression.
DTR	Decision tree regression.
RFR	Random forest regression.
MLP	Multilayer perception.
GRU	Gate recurrent unit.
P_{m_i,n_i}	Pearson correlation coefficient.
t	Time.
i_t	Input gate.
f_t	Forget gate.
c_t	Memory cell.
o_t	Output gate.
h_t	Output.
$\dot{x}$	Input data after using $Z - score$.

ACKNOWLEDGMENT

Jun Yang would like to thank Yixuan Sun, a Ph.D. Student from Purdue University, for giving methodological guidance and reviewing the article in this research.

REFERENCES

[1] World Information Service on Energy Uranium Project. *Chronology of Major Tailings Dam Failures*. Accessed: May 12, 2020. [Online]. Available: https://www.wise-uranium.org/mdaf.html
[2] Z. Qi and L. Xuelong, "Big data epoch: Challenges and opportunities for geology," *Bull. Chin. Acad. Sci.*, vol. 33, no. 8, pp. 825–831, 2018.
[3] L. Huang, F. Miao, and M. X. Wang, "Designing and setting up a monitoring and early warning system for tailings ponds in a region," *China Saf. Sci. J.*, vol. 23, no. 12, pp. 146–152, 2013.
[4] J. Yang, Y. Sun, Q. Li, and Z. Qian, "Measure dry beach length of tailings pond using deep learning algorithm," in *Proc. Int. Conf. Robot., Intell. Control Artif. Intell.*, 2019, pp. 503–508.
[5] Q. Li, W. Tian, and Y. Wang, "Study on displacement analysis methods of tailing ponds on-line monitoring system," *J. Saf. Sci. Technol.*, vol. 8, pp. 47–52, Jan. 2011.
[6] Y. Gao, Y. Chu, and L. Wei, "Remote sensing monitoring and analysis of tailings ponds in the ore concentration area of Heilongjiang province," *Remote Sens. Land Resour.*, vol. 27, no. 1, pp. 160–163, 2014.
[7] M. Necsoiu and G. R. Walter, "Detection of uranium mill tailings settlement using satellite-based radar interferometry," *Eng. Geol.*, vol. 197, pp. 267–277, Oct. 2015.
[8] D. Che, A. Liang, X. Li, and B. Ma, "Remote sensing assessment of safety risk of iron tailings pond based on runoff coefficient," *Sensors*, vol. 18, no. 12, p. 4373, Dec. 2018.
[9] J. Zhang, X. Wu, and G. Zhao, "Study on safety on-line monitoring and warning systems of tailings reservoir," in *Electrical, Information Engineering and Mechatronics*, London, U.K.: Springer, 2012, pp. 589–598.
[10] L. Dong, W. Shu, D. Sun, X. Li, and L. Zhang, "Pre-alarm system based on real-time monitoring and numerical simulation using Internet of Things and cloud computing for tailings dam in mines," *IEEE Access*, vol. 5, pp. 21080–21089, 2017.
[11] P. Qiu *et al.*, "Monitoring system of saturation line based on mixed programming," in *Proc. 2nd Int. Conf. Comput. Eng., Inf. Sci. Appl. Technol. (ICCIA)*. Atlantis Press, 2016.
[12] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun, "Infrared thermal imaging-based crack detection using deep learning," *IEEE Access*, vol. 7, pp. 182060–182077, 2019.
[13] W. Fei-Yue *et al.*, "Stability analysis of tailings dam based on saturation line matrix," *Rock Soil Mech.*, vol. 30, no. 3, pp. 840–844, 2009.
[14] H. E. Chaoyang, J. U. Nengpan, and H. Jian "Automatic integration and analysis of multi-source monitoring data for geo-hazard warning," (In Chinese), *J. Eng. Geol.*, vol. 22, no. 3, pp. 405–411, 2014.
[15] *Code for Design of Tailings Facilities (GB50863-2013)*, 2013.
[16] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2018.
[17] M. Macas, F. Moretti, A. Fonti, A. Giantomassi, G. Comodi, M. Annunziato, S. Pizzuti, and A. Capra, "The role of data sample size and dimensionality in neural network based forecasting of building heating related variables," *Energy Buildings*, vol. 111, pp. 299–310, Jan. 2016.

[18] M. Macas *et al.*, "Sensitivity based feature selection for recurrent neural network applied to forecasting of heating gas consumption," in *Proc. Int. Joint Conf. SOCO-CISIS-ICEUTE*. Cham, Switzerland: Springer, 2014, pp. 259–268.
[19] S. Ganguli and J. Dunnmon, "Machine learning for better models for predicting bond prices," 2017, *arXiv:1705.01142*. [Online]. Available: https://arxiv.org/abs/1705.01142
[20] A. Prochazka and V. Sys, "Time series prediction using genetically trained wavelet networks," in *Proc. IEEE Neural Netw. Signal Process.*, Sep. 1994, pp. 195–203.
[21] A. Prochazka, "Neural networks and seasonal time-series prediction," 1997, pp. 36–41.
[22] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinf.*, vol. 8, no. 1, p. 25, Dec. 2007.
[23] A. Saltelli, T. Ratto, and T. Andres, *Global Sensitivity Analysis: The Primer*. Hoboken, NJ, USA: Wiley, 2008.
[24] A. Mokhtari and H. C. Frey, "Sensitivity analysis of a two-dimensional probabilistic risk assessment model using analysis of variance," *Risk Anal. Int. J.*, vol. 25, no. 6, pp. 1511–1529, Dec. 2005.
[25] A. Kiparissides, S. S. Kucherenko, A. Mantalaris, and E. N. Pistikopoulos, "Global sensitivity analysis challenges in biological systems modeling," *Ind. Eng. Chem. Res.*, vol. 48, no. 15, pp. 7168–7180, Aug. 2009.
[26] N. Shentu, H. Zhang, Q. Li, and H. Zhou, "Research on an electromagnetic induction-based deep displacement sensor," *IEEE Sensors J.*, vol. 11, no. 6, pp. 1504–1515, Jun. 2011.
[27] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.
[28] F.-M. Tseng, H.-C. Yu, and G.-H. Tzeng, "Combining neural network model with seasonal time series ARIMA model," *Technol. Forecasting Social Change*, vol. 69, no. 1, pp. 71–87, Jan. 2002.
[29] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*. Berlin, Germany: Springer, 2003, pp. 107–119.
[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
[31] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010.
[32] C.-J. Huang, Y. Shen, Y.-H. Chen, and H.-C. Chen, "A novel hybrid deep neural network model for short-term electricity price forecasting," *Int. J. Energy Res.*, Sep. 2020, doi: 10.1002/er.5945.
[33] A. Medyńska-Juraszek and L. Kuchar, "Carbon sequestration rates in organic layers of soils under the grey poplar *(Populus x Canescens)* stands impacted by heavy metal pollution," in *Functions of Natural Organic Matter in Changing Environment*. Dordrecht, The Netherlands: Springer, 2013, pp. 365–369.
[34] S. Fiaschi, M. Mantovani, S. Frigerio, A. Pasuto, and M. Floris, "Testing the potential of sentinel-1A TOPS interferometry for the detection and monitoring of landslides at local scale (Veneto region, Italy)," *Environ. Earth Sci.*, vol. 76, no. 14, p. 492, Jul. 2017.
[35] D. Zeng, Z. Wu, C. Ding, Z. Ren, Q. Yang, and S. Xie, "Labeled-robust regression: Simultaneous data recovery and classification," *IEEE Trans. Cybern.*, early access, Nov. 5, 2020, doi: 10.1109/TCYB.2020.3026101.
[36] L. Parra, E. Karampelas, S. Sendra, J. Lloret, and J. J. P. C. Rodrigues, "Design and deployment of a smart system for data gathering in estuaries using wireless sensor networks," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst. (CITS)*, Jul. 2015, pp. 1–5.
[37] L. Parra, S. Sendra, J. Lloret, and J. J. P. C. Rodrigues, "Design and deployment of a smart system for data gathering in aquaculture tanks using wireless sensor networks," *Int. J. Commun. Syst.*, vol. 30, no. 16, p. e3335, Nov. 2017.
[38] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp. 1–4.
[39] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 16, no. 4, pp. 757–765, May 2008.
[40] Z. Wu, S. Liu, C. Ding, Z. Ren, and S. Xie, "Learning graph similarity with large spectral gap," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Mar. 5, 2019, doi: 10.1109/TSMC.2019.2899398.
[41] J. Yang, Y. Sun, Q. Li, and Y. Sun, "Effective risk prediction of tailings ponds using machine learning," in *Proc. 3rd Int. Conf. Adv. Electron. Mater., Comput. Softw. Eng. (AEMCSE)*, Apr. 2020, pp. 234–238.
[42] B. Linda and J. B. P. Soares, "The influence of tailings composition on flocculation," in *The Evolution of Cognition*. Cambridge, MA, USA: MIT Press, 2015.
[43] M. A. Hariri-Ardebili and F. Pourkamali-Anaraki, "Support vector machine based reliability analysis of concrete dams," *Soil Dyn. Earthq. Eng.*, vol. 104, pp. 276–295, Jan. 2018.
[44] V. Ranković, N. Grujović, D. Divac, and N. Milivojević, "Development of support vector regression identification model for prediction of dam structural behaviour," *Struct. Saf.*, vol. 48, pp. 33–39, May 2014.
[45] H.-B. Zhao, "Slope reliability analysis using a support vector machine," *Comput. Geotechnics*, vol. 35, no. 3, pp. 459–467, May 2008.
[46] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017.
[47] J. Li, Q. Dai, and R. Ye, "A novel double incremental learning algorithm for time series prediction," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6055–6077, Oct. 2019.
[48] B. Cortez, B. Carrera, Y.-J. Kim, and J.-Y. Jung, "An architecture for emergency event prediction using LSTM recurrent neural networks," *Expert Syst. Appl.*, vol. 97, pp. 315–324, May 2018.
[49] G. Maragatham and S. Devi, "LSTM model for prediction of heart failure in big data," *J. Med. Syst.*, vol. 43, no. 5, May 2019, Art. no. 111.
[50] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10, 1995.
[51] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[53] J. Li, H. Chen, T. Zhou, and X. Li, "Tailings pond risk prediction using long short-term memory networks," *IEEE Access*, vol. 7, pp. 182527–182537, 2019.
[54] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690–76698, 2019.
[55] M. Pan, H. Zhou, J. Cao, Y. Liu, J. Hao, S. Li, and C.-H. Chen, "Water level prediction model based on GRU and CNN," *IEEE Access*, vol. 8, pp. 60090–60100, 2020.
[56] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, 2005.
[57] A. P. Hutton, A. J. Marcus, and H. Tehranian, "Opaque financial reports, R-square, and crash risk," *SSRN Electron. J.*, vol. 94, no. 1, pp. 67–86, 2008.
[58] Y. Liang, Z. Ren, Z. Wu, D. Zeng, and J. Li, "Scalable spectral ensemble clustering via building representative co-association matrix," *Neurocomputing*, vol. 390, pp. 158–167, May 2020.

JUN YANG received the M.S. degree from the School of Mechanical and Electrical Engineering, China Jiliang University. He was a Visiting Scholar with the Department of Mathematics, Purdue University, West Lafayette, IN, USA. He is currently an Algorithm Engineer at Focused Photonics (Hangzhou), Inc. His research interests include deep learning, image processing, and computer vision.

JINGBIN QU was born in 1979. He received the bachelor's degree in mechatronics from Shanghai Jiao Tong University, Shanghai, China, in 2002. He was engaged in data processing and network with Cisco Systems Inc., for a period of eight years. His current research interests include deep learning, graph convolutional networks, and traffic big data processing and analysis.

QIANG MI received the bachelor's degree in software engineering from Hangzhou Dianzi University, China. He has research experience in big data for a period of ten years. He is currently a Senior Research and Development Engineer. He is also currently a Senior Big Data Algorithm Engineer at Focused Photonics (Hangzhou), Inc. His research interests include financial big data, machine learning, and deep learning.

QING LI was the Dean of the School of Mechanical and Electrical Engineering, China Jiliang University, where he is currently a Professor with China Jiliang University. He is also an Executive Director with the China's Metrological Testing Association. He has presided over completion of more than 60 science and technology projects. His current research interests include dynamic measurement and control and sensing technology. He was selected as a Famous Teacher from the National Ten-Thousand Talents Program in 2017.