



Underflow concentration prediction based on improved dual bidirectional LSTM for hierarchical cone thickener system

Yongxiang Lei¹ · Hamid Reza Karimi¹

Received: 5 January 2023 / Accepted: 28 March 2023 / Published online: 31 May 2023
© The Author(s) 2023

Abstract

In the practical thickener cone systems, the underflow concentration is hard to measure through physical sensors while there exist the high cost and significant measurement delay. This paper presents a novel and deeply efficient long short-time memory (DE-LSTM) method for concentration prediction in the deep cone thickener system. First, the DE-LSTM for thicker systems is developed for feature learning and long temporal preprocessing. Then, the feedforward and reverse LSTM subnetworks are employed to learn the robust information without loss. At last, the experimental verification of an industrial deep cone thicker demonstrates the proposed DE-LSTM's performance outperforms other state-of-the-art methods.

Keywords Underflow concentration prediction · Bidirectional LSTM · Deep learning · Cone thickener system (CTS)

1 Introduction

The paste filling in the mining industry has got high attention in the development of power generation. For sustainable environment protection and the carbon neutral and zero emission target, industrial mining is making a revolution toward smart, intelligent, and clean underground mining [1–5]. Deep cone thickener (DCT) is the critical device to produce underground pasting for mining security. The main principle of the device is that the low-concentration slurry crude from the stope was fed into the DCT, and the gravity of these particles is then dissolved to a larger lump under the effect of flocculant and concentrated at the bottom of the thickener. The high-quality underflow concentration is then produced and the high layer of clean water is recycled by the overflow pipe.

Stable underflow concentration is a critical metric that plays a pivotal role in assessing the performance, security, and stability of industrial production processes. Several quality factors, including the volume of underflow, feed flow,

flocculant dosage, and mass of flocculant, significantly influence the stability of underflow concentration during the production process. Therefore, ensuring high performance and robustness in underflow prediction and control is crucial for efficient CTS production. However, conventional detection techniques fall short in enabling online prediction, while their high computational costs and susceptibility to sampling oscillation are added to this challenge. Through analyzing the production mechanism, it is evident that production variables have a direct correlation with the output underflow concentration, highlighting the need for a cost-effective data-centric approach for underflow concentration prediction [6].

For subsurface paste filling, a deep cone thickener is essential. CTS is a key method for achieving a consistent concentration for underground mining fills. Pipe-blocking mishaps might occur during the thickening process if the subterranean concentration is too high. On the other hand, a low subterranean concentration reduces the quality of the entire backfilled paste and affects the overall safety of the mining operation. Hence, the development of a data-driven model for predicting underflow concentration in the CTS system is of utmost importance. The whole mining paste filling operates on a continuous and hierarchical concept [7]. The CTS was fed with the crude unstable low-concentration slurry flow (almost 20–30%), which was combined with a flocculant to speed up the sinking rate. A mud bed can collect the dissolved particles. The suitable concentration and volume feed flow are created at the bottom of the CTS. The top clean water

Yongxiang Lei and Hamid Reza Karimi contributed equally to this work

✉ Yongxiang Lei
yongxiang.lei@polimi.it

✉ Hamid Reza Karimi
hamidreza.karimi@polimi.it

¹ Department of Mechanical Engineering, Politecnico di Milano, Milan 20156, Italy

from the overflow pipe has also been recycled for future use. The CTS's primary control is to provide a consistent and precise underflow concentration [8]. The concentration of underflow is a key metric for assessing the effectiveness and efficiency of the industrial underground pasting process. It is necessary to identify changes in the different variables due to the inner link for the production quality variable in the deep cone thickening process. Because the response has a long time lag, various traditional methods have also been used as the fundamental unit. This time series architecture may be used to acquire some prior knowledge and historical information, which can subsequently be utilized to anticipate the underflow concentration.

Recently, deep learning has become a hot topic in the modeling, optimization, control and prediction of industrial systems. For instance, Lei et al. proposed a semi-supervised method to tackle the superheat degree identification in the aluminum electrolysis reduction cell, which has a corresponding relation with our industrial case [1]. Yuan et al. investigated a long short-time memory with attention mechanism with the time sequence soft sensor tasks [12]. Some other methods, such as self-supervised learning [2] have also achieved large attention in the industrial process domain [10–18]. While recurrent neural networks can retain historical information using memory and nonlinear functions, they suffer from low accuracy, gradient vanishing, and exploration issues due to their long dependency on historical features. Long short-term memory (LSTM) is an enhanced version of RNN that overcomes these limitations by introducing a new memory unit to store historical information. This makes LSTM ideal for practical applications, as it covers the input gate, cell gate, forget gate, and output gate, retaining only relevant information. In practical applications, the attention network is used for the hierarchical industrial process. Other LSTM-based applications with attention to the mechanism are given in [21–25]; however, the spatial information is largely ignored. Other recent variants of long short-time memory have been presented in [26–30].

In particular, several time sequence models can capture valuable historical information that should be effectively utilized. The prediction of underflow concentration is a crucial challenge in the field of industrial cone thickeners. Recently, some works based on the underflow concentration are reported. For example, Yuan et al. proposed a dual-attention recurrent neural network method for the deep cone thickener, their model can capture the spatial information with the high dimensional series [8]. To our best knowledge, some underflow concentration detection methods have been proposed [29–34]; however, they only consider the static feature prediction, and the adaptive and dynamic mechanism for online learning are largely ignored, so the time series long

prediction for the CTS system still needs to be further investigated.

Although some conventional time series prediction models are widely used in the industrial analysis, such as the principal component analysis (PCA), autoregressive integrated moving average (ARMA), recurrent neural network, and gated recurrent networks [35–40]. All these architectures have been widely used in industrial fields. In the deep cone thickening system, some basic hypotheses are proposed to be considered as follows. (1) The coupled multi-variables are linked with the output underflow concentration variables, which means we can build a data-driven model to learn the complex nonlinear mapping between those variables. (2) The large delay in the industrial case gives the insight that we should propose a time sequence to learn the time series information. (3) The industrial database collected from all the sensors has outliers and noise, which greatly reduces the prediction performance in the applicable CTS. To tackle these problems, a new LSTM method is proposed in this paper. First, the original data from the industrial database is preprocessed by an average moving filter. Second, the encoder transmission channel was leveraged to gain the feature representation of the underflow concentration prediction information. Finally, the attention mechanism with a full-connected layer is used for the final underflow prediction in the CTS system. The main contributions of this paper can be highlighted as follows:

1. A novel DE-LSTM model is proposed to deal with the long-time prediction with robustness and high accuracy for the large time delay CTS pasting system.
2. Compared to the traditional prediction methods such as SVM, LSTM, GRU, and XGBOOST, the performance of the proposed DE-LSTM provides a more accurate and robust result.
3. Stacked long short-time memory with attention mechanism is employed for the long prediction without information loss during the learning process.
4. The industrial application case is proposed to evaluate the high performance of the proposed DE-LSTM method. The experiments verify the prediction MSE has decreased by a large margin with the presented framework.

The remainder of this paper is organized as follows. Section 2 describes the specific problems and gives some fundamental theory background. Section 3 introduces the model formulation process for the underflow prediction in CTS. Section 4 gives the experimental results and the conclusive remarks are given in the final section.

2 Related works

2.1 Problem statement

In the underground paste filling process, the underflow concentration is the top index which can sustain the stable production of the paste, avoid the security hazard and improve the quality of the paste filling. The detailed structure of the deep cone thickener is presented in Fig. 1, and all the processes in the pasting-filling process are included. However, the traditional method for prediction accuracy is limited by the expensive physical device, the large time delay, and the difficulty of online detection. In the long term, the flocculant should react with the feed flow particle for a long time, so the change of the underflow concentration is a long time interval. To the best of our knowledge, the existing production of intelligent DCT systems is scarce, and the tendency of the deep cone thickener is the smart mining industry and paste filling.

Based on the above analysis, the problem can be specified as: Given multi-dimension time input variables (x_1, x_2, \dots, x_t) and temporal underflow concentration (y_1, y_2, \dots, y_t) , we need to learn a distribution under the condition of all the observer of the multi-dimension massive input variables. The specific expression is $p(y_{t+L}, y_{t+L-1}, \dots, y_{t+1}/y_1, y_2, \dots, y_t, x_1, x_2, \dots, x_t)$, T is the sliding moving window size for the underflow future prediction. The aim is to achieve a model with the nonlinear sequential mapping function expressed as follows:

$$\max p(y_{t+L}, y_{t+L-1}, \dots, y_{t+1}/y_1, y_2, \dots, y_t, x_1, x_2, \dots, x_t) = f(\cdot) \tag{1}$$

where the \cdot represents the multi-dimension variables. The main objective needs to obtain high performance in the long interval prediction accuracy and robustness.

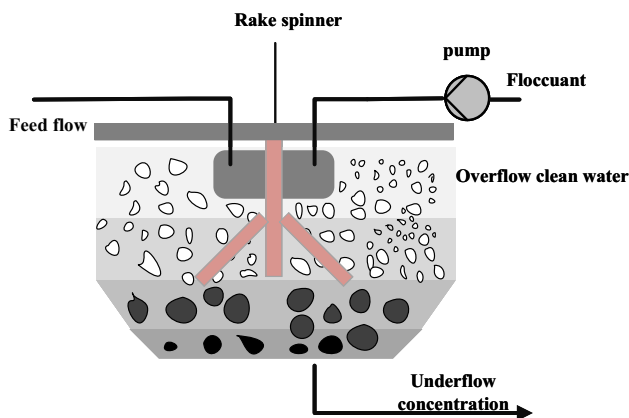


Fig. 1 A schematic of industrial DCT system for the paste filling

2.2 LSTM unit

The next stage for the model consideration should be the option of the time series model. In the underflow concentration prediction task, the primary goal is the time factor. Long short-time memory has the function of preserving historical information. LSTM is proposed by Helrtcher with a four-gate unit which has been proven for superior performance in natural language processing, computation vision, and other domains. The adaptive sequence chunkers are utilized in the model with the capability to bridge arbitrary time lags. Denote $x(t)$ for the input historical vector, $h(t-1)$ represents the precious hidden state, and the external inputs are inherited from the previous cell state $c(t-1)$ (Fig. 2). Then, the forget gate is triggered as:

$$f(t) = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

where W_f and b_f are the weights of the forget gate and the corresponding bias, respectively. A new formation from the input gate and new candidate vectors are calculated as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_{(C)} \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

W_i is the input gate weight and bias is denoted by b_i . The new cell state in LSTM is updated by the following equation:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

The output gate vector can be given by:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t * \tanh(C_t) \tag{7}$$

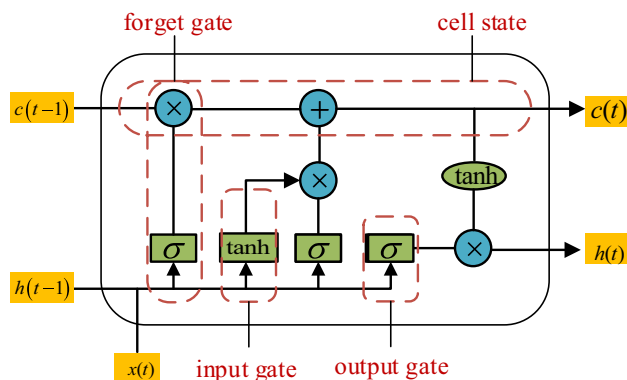


Fig. 2 The schematic diagram of LSTM unit

where W_o is the output weight and b_o is the bias. σ is the sigmoid nonlinear activation function. $*$ is the vector product term.

2.3 Attention mechanism

The weight between the time sequence and the temporal sequence is often in the high-layer future representation in multi-dimensional prediction problems. Attention mechanism-based neural networks have recently shown effectiveness in a variety of activities. In a BiLSTM network with an attention mechanism, the attention technique uses the BiLSTM’s last cell state, or the implicit state of the LSTM, to align with the cell state of the input at the current step. The output state’s correlation with these prospective intermediate stages is then calculated. To improve prediction accuracy and efficiency, related information can be emphasized while irrelevant information is hidden throughout the learning process. In the attentive BiLSTM network, the attention layer’s output A is created using the following Eqs. (8)–(10):

$$M = \tanh(Y) \tag{8}$$

$$\alpha = \frac{(w_a^T M)_i}{\sum_i (w_a^T M)_i} \tag{9}$$

$$A = Y\alpha^T \tag{10}$$

where Y is a matrix that reflects the LSTM model’s captured features, such as the matrix stated above. $Y = [y_1, y_2, \dots, y_t]$. α is a vector that represents feature attention weights. The attention layer’s weight coefficient matrix is Y . A transposition operation is denoted by the symbol T .

3 Method formulation

In this part, the methodology of the proposed DE-LSTM is given. A schematic of the proposed DE-LSTM is given in Fig. 3, The original data collected from the massive sensors are preserved in the industrial database. The main procedures include variable selection, preprocessing, average moving filtering, and DE-LSTM model training. Finally, the trained model is used for the prediction implementation.

3.1 Initial preprocessing

Before the construction of an excellent DE-LSTM model, the preprocessing of the collected data is necessary. Due to the reason that the original data from the massive sensors has the oscillation points, the outlier, and the missing data,

which will greatly influence the accuracy of the proposed DE-LSTM model. Before considering the sliding window to train the data, the standardization and normalization of the training samples and the temporal output are used for improving the quality of the data. The feed flow concentration, the flocculant rate, and the feed amount are the initial input variables.

The second target is the encoder representation of the underflow concentration prediction in the CTS system. The historical information should be preserved in the long temporal term. The LSTM unit is built to learn the temporal information contained in the training samples. The specific representation inherited from the front layer are:

$$\max p(y_{t+L}, y_{t+L-1}, \dots, y_{t+1} / y_1, y_2, \dots, y_t, x_1, x_2, \dots, x_t) = LSTM(\cdot) \tag{11}$$

3.2 Encoder representation

The features from the average moving filter are sent to the deep encoder network to obtain an efficient representation of the underflow concentration. Given the time series of (x_1, x_2, \dots, x_L) , the hidden representation from the multiple stacked LSTM can be represented as the following equation.

$$h_i = LSTM(x_1, x_2, \dots, x_L) \tag{12}$$

the details of the above equation can be extended as:

$$\tilde{\mathbf{i}}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\tilde{\mathbf{h}}_{t-1} + \mathbf{W}_{ci}\tilde{\mathbf{c}}_{t-1} + \mathbf{b}_i) \tag{13}$$

$$\tilde{\mathbf{f}}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\tilde{\mathbf{h}}_{t-1} + \mathbf{W}_{cf}\tilde{\mathbf{c}}_{t-1} + \mathbf{b}_f) \tag{14}$$

$$\tilde{\mathbf{c}}_t = \tilde{\mathbf{f}}_t \odot \tilde{\mathbf{c}}_{t-1} + \tilde{\mathbf{i}}_t \odot \tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\tilde{\mathbf{h}}_{t-1} + \mathbf{b}_c) \tag{15}$$

$$\tilde{\mathbf{o}}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\tilde{\mathbf{h}}_{t-1} + \mathbf{W}_{co}\tilde{\mathbf{c}}_t + \mathbf{b}_o) \tag{16}$$

$$\tilde{\mathbf{h}}_t = \tilde{\mathbf{o}}_t \odot \tanh(\tilde{\mathbf{c}}_t) \tag{17}$$

The encoder representation extracted complex features and learn historical temporal information from the underflow concentration variables. The attention mechanism process from the encoder layer is expressed as follows:

$$M = \sum_{i=1}^N \tanh(h_i) \tag{18}$$

and the transmission of the front layer is written as:

$$= \frac{w_a^T M}{\sum_i (w_a^T M)_i} \tag{19}$$

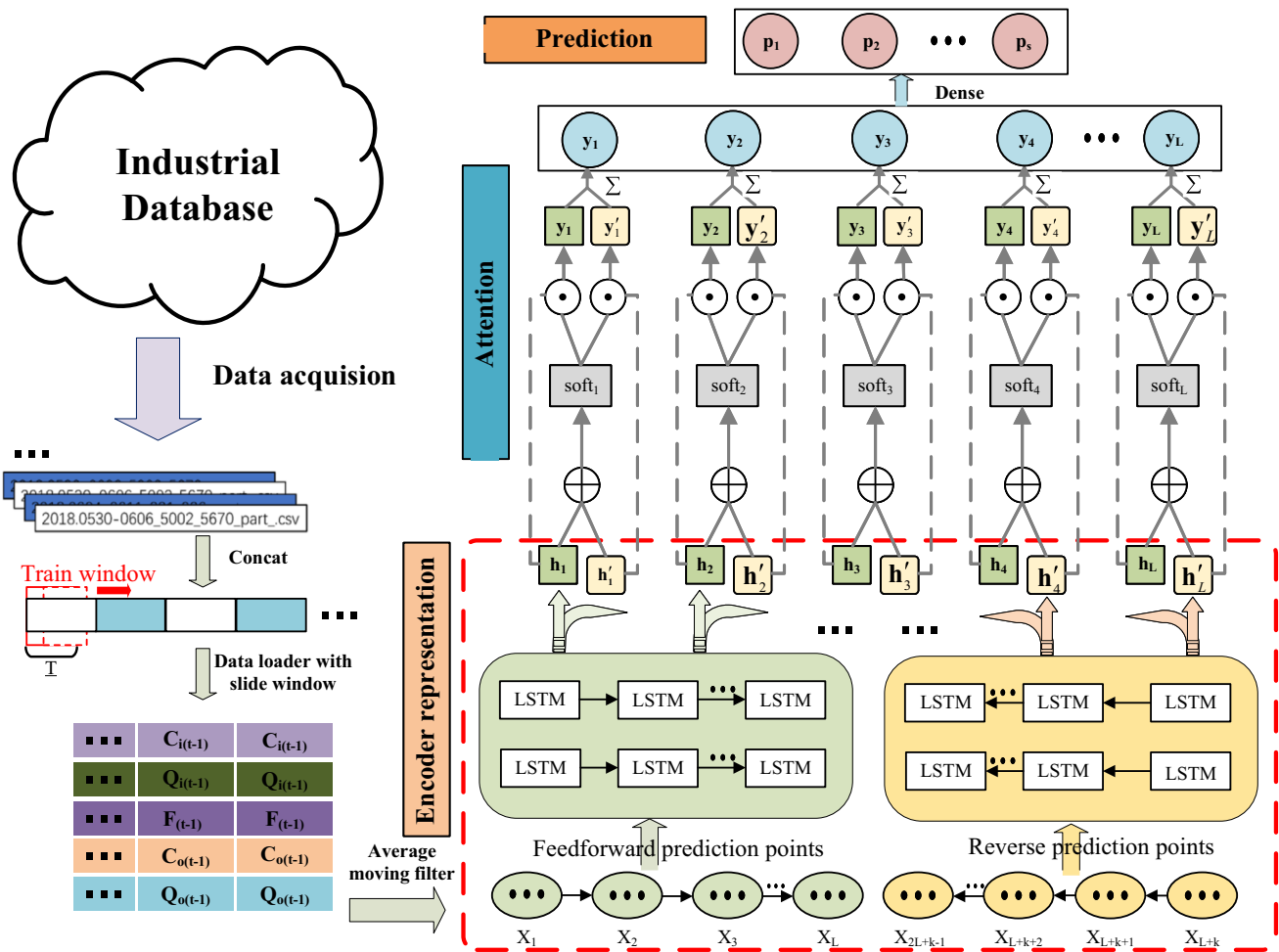


Fig. 3 The whole framework of the proposed DE-LSTM for the underflow concentration prediction

then, the final output of the DE-LSTM is:

$$\hat{y} = \sigma\left(\sum_{i=1}^N h_i\right) \tag{20}$$

The next step is to train the proposed network and gain the weights of the network. The loss function is defined as the least square algorithm, which is specified as follows:

$$Loss = \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_t^i - y_t^i)^2 \tag{21}$$

where N is the total number of samples and T is the size of the sliding moving window. The hidden parameters of the LSTM and the attention learnable parameters can be finetuned by the Adam optimizer, the backpropagation is used for the proposed DE-LSTM training.

3.3 Underflow concentration prediction

The flowchart of the proposed DE-LSTM is given in Fig. 3. The procedure is that the original data collected from the industrial CTS database is reprocessed by the sliding window, and split with the training datasets. Then the training datasets with time series T are transmitted to the average moving filter. The final output of the T steps from the DE-LSTM is the prediction of underflow concentration. Specifically, the dynamic recurrent update mechanism is used in the whole model. The prediction results with the front weights are finetuned by the time interval of 1 h.

The specific algorithm is given in **Algorithm 1**.

3.4 DE-LSTM’s prediction approximation analysis

We state the universal approximation ability of underflow concentration prediction in the cone thickener system in this block. Consider the continuous-time sampling series from the industrial paste filling process, the input series variables

Algorithm 1 The proposed DE-LSTM algorithm.

Require: n samples, learning rate α , initialize the network parameters, L layers, sliding moving windows T , Dropout, batch size $batch_size$, max_epochs .
Ensure: The DE-LSTM prediction model for long-term underflow concentration prediction with (15).
 The hyperparameters of multiple LSTM: W
 Other hidden weights parameters.
while $epoch \neq 0$ **do**
 if $epoch$ is not equal zero **then**
 Training the DE-LSTM with the (16) loss and Adam optimizer.
 Updating the hyper-parameters and weights for each epoch.
 Setting $N \leftarrow N - 1$
 else if $error$ is less than the threshold **then**
 The model is fully trained.
 Output the hyper-parameters and nonlinear projection mapping prediction model.
 end if
end while

are denoted as the compact set $\mathbf{X} = \{\mathbf{C}_i, \mathbf{Q}_i, \mathbf{F}, \mathbf{Q}_o\}_{i=1}^T$. \mathbf{C}_i is the input underflow concentration, \mathbf{Q}_i is the input flow, and \mathbf{F} is the mud pressure level. \mathbf{C}_o is the underflow concentration series. $Pred^{(t)}_{W_{LSTM}, W_{atten}, W_{soft}}$ is the long future prediction steps. The target of our analysis is to guarantee an approximated global prediction ability.

Inspired by the theoretical work in [2], which proves that the single feedforward neural network can approximate arbitrarily well continuous functionals and infinite hidden neurons.

Lemma 1 Given a bounded piecewise nonlinear activation function $\sigma(x)$, where $\sigma(x) \in L^2(x)$ is fully dense in spatial space $x \in \mathfrak{R}^d$, which means the activation of hidden units is from the full-connected units. If the residual error of the nonlinear mapping projection in the prediction function $g: \mathfrak{R}^d \rightarrow \mathfrak{R}^s$ satisfies:

$$\int_R g(x) dx \neq 0 \tag{22}$$

then, the discrete expression can be rewritten as:

$$\sum_{i=1}^N g(x_i) \Delta x \neq 0 \tag{23}$$

which exists for an integer $N > 0, \exists \varepsilon > 0$, we can find:

$$\lim_{n \rightarrow N} \|e_n\| = \lim_{n \rightarrow N} \|g - f\| = \lim_{n \rightarrow N} \left\| \sum_{i=1}^L \beta_i \sigma_i(x) \right\| \leq \varepsilon \tag{24}$$

where e_n is the increment residual error with the number of n hidden neurons.

Theorem 1 Denote the pairwise time series $\mathbf{X} = \{\mathbf{C}_i, \mathbf{Q}_i, \mathbf{F}, \mathbf{Q}_o\}_{i=1}^T$ with T sliding window, the normalized output target (ranged from $[0,1]$) satisfies $\mathbf{C}_o \in \mathfrak{R}^{s \times N}$, $\mathbf{X} \in \mathfrak{R}^{2L \times N}$, $\mathbf{y} \in \mathfrak{R}^{L \times N}$, the dynamical time series prediction can be performed with a projection map from the X to output C , N_{lstm} positive integer for number of LSTM unit, then $\forall \varepsilon > 0, \exists \{W_{LSTM}, W_{atten}, W_{soft}\}_{t=1}^T$, with the infinite memory approximation, such that:

$$\left\| \left(\mathbf{C}_o \right)_j^{(t)} - \underbrace{\sum_{i=1}^N \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y})}_{DE-LSTM} \right\| < \varepsilon, j = 1, 2, \dots, N \tag{25}$$

Proof According to Lemma 1, denote $e_n \in L^2(x)$ as the dense compact subset, $\psi_n^t = \|e_n^t\|^2$ as the Lyapunov function, the whole residual error can be specified as the following equation:

$$\begin{aligned} \Delta &= \psi_{n-1}^t - \psi_n^t = \|e_{n-1}^t - e_n^t\|^2 \\ &= \|e_{n-1}^t\|^2 - 2\langle e_{n-1}^t, e_n^t \rangle + \|e_n^t\|^2 \\ &= \|e_{n-1}^t\|^2 - 2(\langle e_{n-1}^t, e_n^t \rangle - \langle e_{n-1}^t - e_n^t, e_n^t \rangle) + \|e_n^t\|^2 \end{aligned} \tag{26}$$

then, it can be further represented as:

$$\begin{aligned} \Delta &= \left\| \left(\mathbf{C}_o^{n-1} \right)_j^{(t)} - \sum_{i=1}^{n-1} \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right\|^2 \\ &\quad - 2 \left(\left(\mathbf{C}_o^{n-1} \right)_j^{(t)} - \sum_{i=1}^{n-1} \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right) \\ &\quad \left(\mathbf{C}_o^n \right)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \\ &\quad - 2 \left(\left(\mathbf{C}_o^n \right)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right) \\ &\quad \left(\left(\mathbf{C}_o^n \right)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right) \\ &\quad \left(\mathbf{C}_o^n \right)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \\ &\quad + \left\| \left(\mathbf{C}_o^n \right)_j^{(t)} - \sum_{i=1}^{n-1} \sum_{t=1}^T W_{LSTM}, W_{atten}, W_{soft} Pred^{(t)} \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right\|^2 \end{aligned} \tag{27}$$

Let $\rho = (\mathbf{C}_o^{n-1})_j^{(t)} + (\mathbf{C}_o^n)_j^{(t)}$, $v = (\mathbf{C}_o^{n-1})_j^{(t)} - (\mathbf{C}_o^n)_j^{(t)}$, the above equation can be transmitted to:

$$\begin{aligned} \Delta + & \left((\mathbf{C}_o^{n-1})_j^{(t)} - \sum_{i=1}^{n-1} \sum_{t=1}^T \mathbf{W}_{LSTM} \text{Pred}^t \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right. \\ & \left. + (\mathbf{C}_o^n)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T \mathbf{W}_{LSTM} \text{Pred}^t \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right) \\ & \left((\mathbf{C}_o^{n-1})_j^{(t)} - \sum_{i=1}^{n-1} \sum_{t=1}^T \mathbf{W}_{LSTM} \text{Pred}^t \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right. \\ & \left. + (\mathbf{C}_o^n)_j^{(t)} - \sum_{i=1}^n \sum_{t=1}^T \mathbf{W}_{LSTM} \text{Pred}^t \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \right) \\ & = (\rho - DE_LSTM_{(n-1)} - DE_LSTM_{(n)}) \\ & (v - DE_LSTM_{(n-1)} + DE_LSTM_{(n)}) \end{aligned} \quad (28)$$

with the similar condition of (25), (27) with the constraint easily satisfies that:

$$\begin{aligned} \lim_{n \rightarrow N} \|e_{n-1} - e_n\|^2 &= \lim_{n \rightarrow \infty} \|\psi_{n-1}^t - \psi_n^t\| \\ &= \lim_{n \rightarrow N} (\|e_{n-1}\|^2 - \|e_n\|^2) \\ &= \lim_{n \rightarrow N} (\|e_{n-1}\|^2 - \|e_{n-1} - \beta_n \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y})\|^2) \\ &= \lim_{n \rightarrow N} (\rho - DE_LSTM_{(n-1)} - DE_LSTM_{(n)}) \\ &\cdot (v - DE_LSTM_{(n-1)} + DE_LSTM_{(n)}) \end{aligned}$$

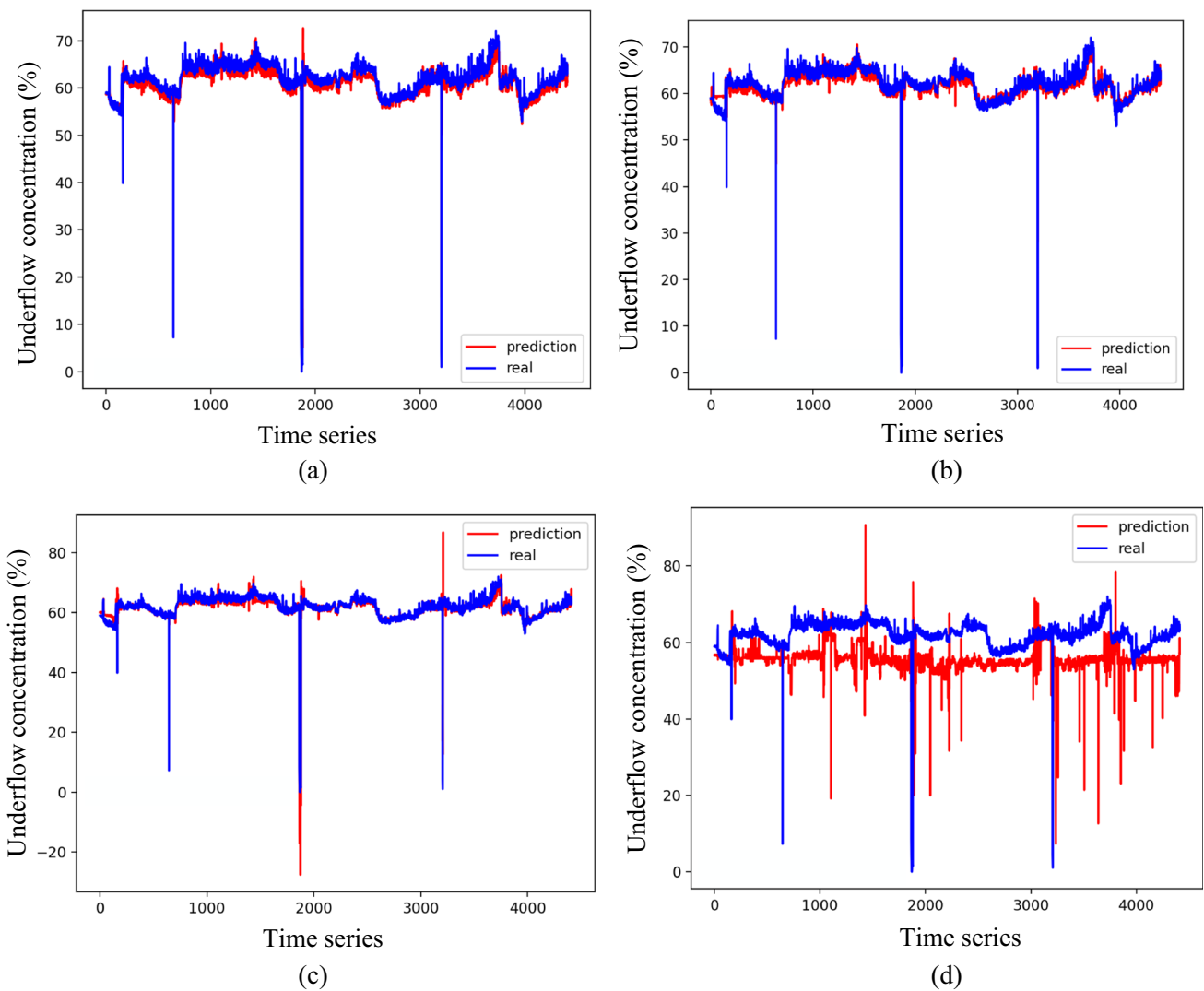


Fig. 4 The proposed DE-LSTM model for industrial underflow prediction results with different batch sizes, (a) batch size = 10; (b) batch size = 15; (c) batch size = 20; (d) batch size = 25

$$\begin{aligned} < \lim_{n \rightarrow N} \left(2\beta_n \langle \rho v, \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y}) \rangle - \beta_n^2 \|\sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y})\|^2 \right) \\ & \lim_{n \rightarrow N} \|\beta_n \sigma_i(\mathbf{C}_o/\mathbf{X}, \mathbf{h}, \mathbf{y})\|^2 \end{aligned} \quad (29)$$

to this end, (22)'s proof is completed. \square

The stated theorem and the explicit expression provided have a very long impact on the foreseeable applications of DE-LSTM to model industrial processes. The long prediction approximation can be guaranteed that inherited the front information. The presented framework has been tested on the realistic underflow concentration prediction tasks for the first time.

4 Experimental case

In this section, we will evaluate the proposed DE-LSTM for the industrial underflow concentration application. We employ our DE-LSTM in the computer device with the i7-core processing with 64GB memory, the Pytorch environment is installed and a GPU of Nvidia 2080Ti is used to accelerate the model training process. The time series was selected with 2.5h for the whole process. Some competitive algorithms such as the recurrent neural network, gated recurrent network, and other attention-based methods are compared in the experiments. The experimental bench originates a 1:10 scaled industrial deep cone thickener with the same functionality. For the experiment, the flow sensors, the mud level sensors, and the rotating speed sensors are extra installed. The capability of this thickener is up to

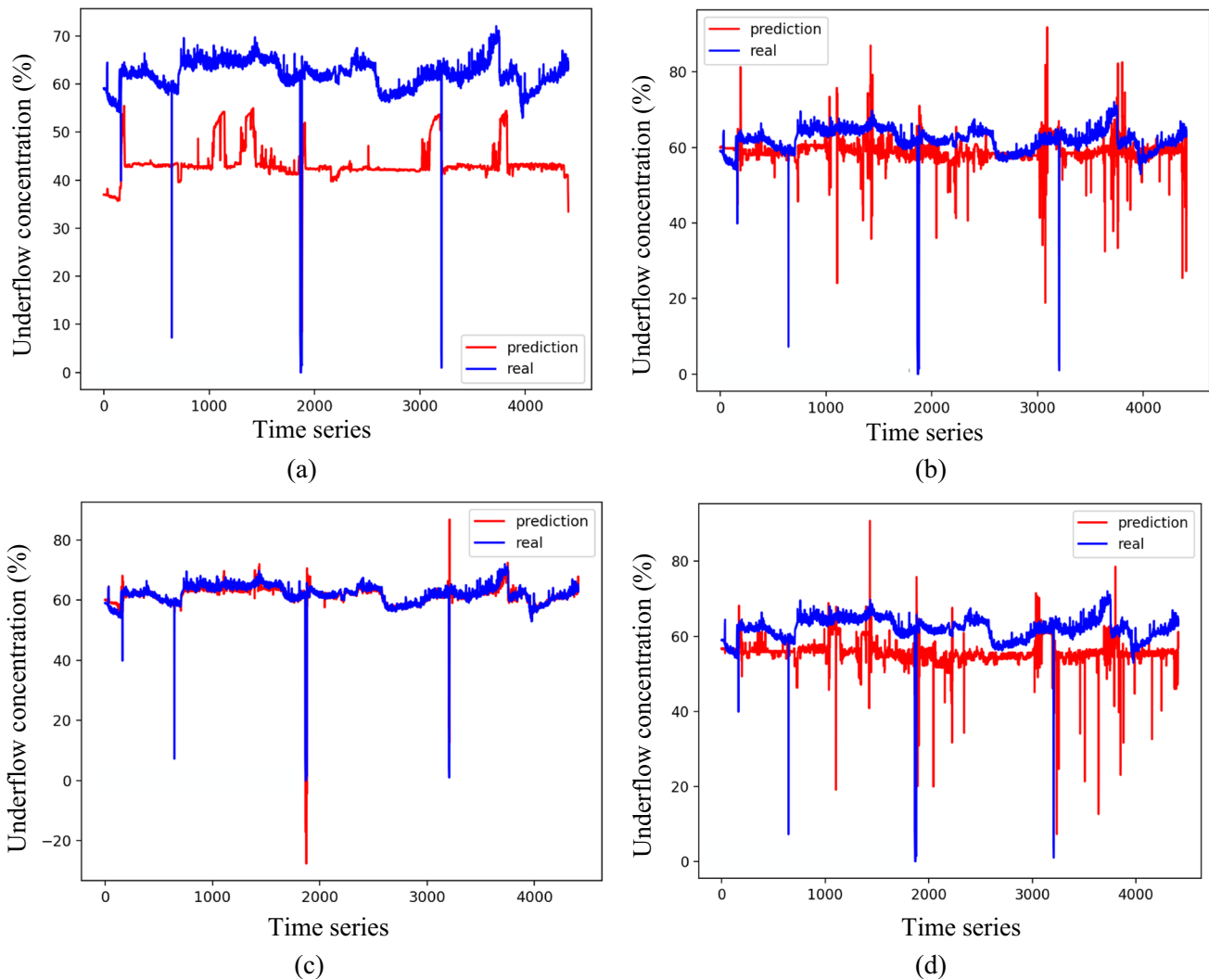


Fig. 5 The proposed DE-LSTM underflow prediction with competitive algorithms comparison; (a) SvR; (b) BiLSTM; (c) GRU; (d) LSTM

600 ~ 1200m³/h with a larger angle cone. Deep cone thickener also called paste thickener is higher than other thickeners which consist of a deep cone, feeding device, stirring device, control box, reagents adding device, and automatic control system. Efficient deep cone thickener is mainly applied in the thickening of minerals or other fine materials. Slender body and flocculants are added in the process of concentration, which accelerates the particle settlement and dewatering process.

Before the training of the process, the sliding window is set to 10, and the decay factor and the learning rate *l_r* are set to 0.01 and 0.001, respectively. The batch size of the experimental setup is with the increase of 5 steps, and the training epoch is set to 30000. The Adam optimizer is selected to fine-tune the whole parameter of the whole network. Finally, the cross-entropy of the prediction value and the actual value is the loss criteria for the iterative training process.

To evaluate the performance of the proposed method, two cases are studied. The proposed DE-LSTM is evaluated by the industrial application of long-time underflow prediction for the CTS system. Some public methods such as support vector regression (SVR), Neural Networks (NNs), Deep belief networks (DBNs), and the temporal recurrent network such as recurrent neural network (RNN), XGBOOST, long short-time memory (LSTM) and gated recurrent unit (GRU) and DARNN are also compared in the experiments throughout. The evaluation indexes are selected from the literature. The RMSE can be given as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_t^i - y_t^i)^2} \tag{30}$$

Also, another index for the evaluation is MAE, it is rewritten as the following equation:

$$MAE = \frac{1}{N_u} \sum_{i=1}^{N_u} |(\hat{y}_t - y_t)| \tag{31}$$

During this period, through training experiments, the influence of training batch size and sliding window length *T* on prediction accuracy was explored. The results are shown in Figs. 3 and 4. On the way, 3000 epochs were trained, the 3 layers of BiLSTM layer that concat the hidden features to the encoder representation, and the hidden layer size is 256. The different batch size is used to testify what's the approximate size to obtain the state-of-the-art prediction performance. In the underflow concentration study, the batch size is set as 10, 15, 20, and 25, respectively, and the corresponding results are reported in Fig. 4. The experimental results demonstrate that under the constraint of batch size equals 15, the algorithm achieves the best prediction performance. While the other three comparisons cause unsatisfactory accuracy, especially

with a batch size equal to 25. The prediction mean square error is big and the prediction curve cannot follow the real underflow concentration timidly, which means the algorithms need to be further fine-tuned.

On the other hand, the existing state-of-the-art competitive algorithms such as the support vector regression, bidirectional long short-time memory, and gate recurrent unit and long short-time memory are used as the benchmark to verify the performance of the proposed algorithm. Figure 5 gives the full intuitive description of these results. Figure 5(a) is the underflow concentration prediction with support vector regression, the prediction value (marked in red) varies from the real value (marked with blue), while the gate recurrent unit's prediction performance (Fig. 5(b)) is better than long short-time memory (Fig. 5(d)) and bidirectional long short-time memory Fig. 5(b). The data collected directly by the actual industrial sensor has disturbances, errors, and occasional sudden outliers, which is the original attribute of the device production. In the proposed DE-LSTM framework, this problem can be alleviated with appropriate filter processing. Experimental results show that using a moving average filter of length [20, 40, 20, 10, 20] can greatly improve the prediction accuracy and reduce the root mean square error (about -96%).

As shown in Table 1, compared to the other competitive methods, DE-LSTM shows the best performance with the lowest RMSE and MAE, the values of 0.2234, and 0.1735, respectively. The SVR with a multi-polynomial activation function also achieves a second performance, compared with the basic LSTM and GRU network. However, if we use other kernels, like linear kernel or Gaussian kernel, the RMSE

Table 1 Evaluation indexes for the different competitive prediction methods

Methods	RMSE	MAE
RNN	0.6712 ± 0.0124	0.3488 ± 0.0082
GRU	0.4634 ± 0078	0.2499 ± 0.0345
LSTM	0.7533 ± 0.0030	0.5119 ± 0.0079
SVR	0.8133 ± /	0.6003 ± /
LGB	0.8091 ± 0.0092	0.6083 ± 0.0118
VAE-WGAN	0.7985 ± 0.0116	0.5847 ± 0.0164
GSTAE	0.7735 ± 0.0068	0.5296 ± 0.0097
SS-DeepFM	0.7650 ± 0.0036	0.5206 ± 0.0073
SSFAN	0.7571 ± 0.0033	0.5235 ± 0.0092
LSTM-DeepFM	0.7497 ± 0.0026	0.5091 ± 0.0071
MPA-RNN	0.7515 ± 0.0034	0.2074 ± 0.0072
DA-RNN	0.851 ± 0.0029	0.2322 ± 0.0052
Attention additive	0.9086 ± 0.0021	0.2445 ± 0.0031
DSTP-RNN	0.8496 ± 0.0051	0.2254 ± 0.0042
Proposed DE-LSTM	0.2234 ± 0.0021	0.1731 ± 0.0013

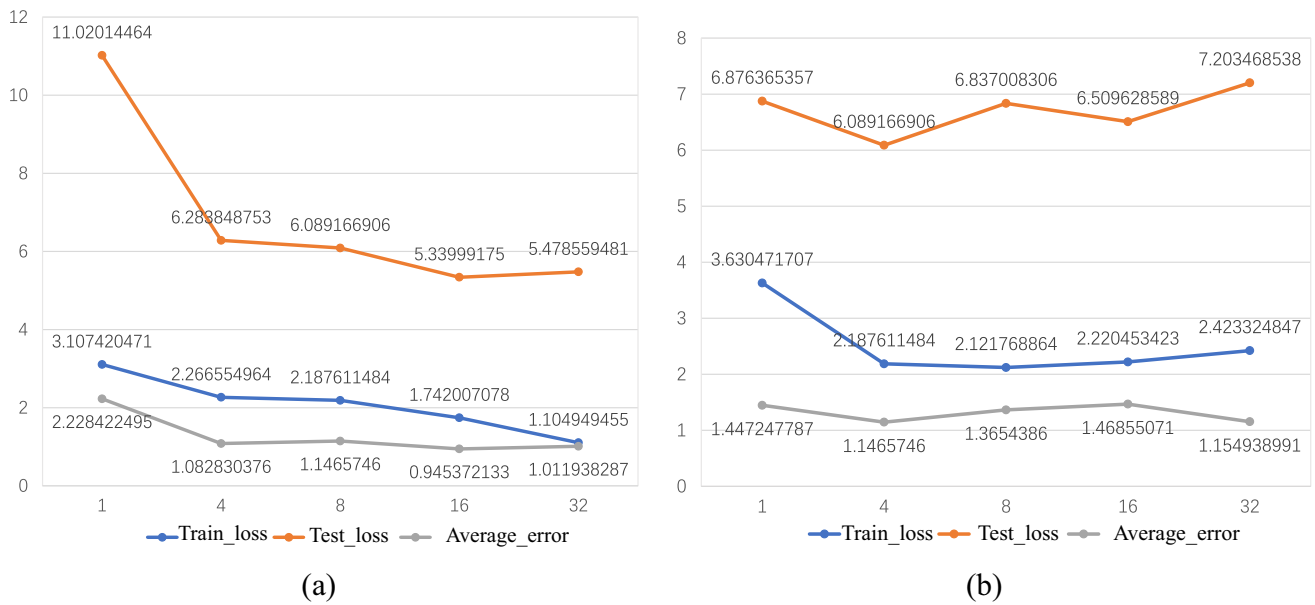


Fig. 6 The measured evaluation indexes comparison for the different algorithms with (a) different batches, (b) different windows and RMSE loss illustration

and MAE are very high (0.9763 in RMSE, 0.7236 in MAE, respectively). In our industrial case, the performance of the LSTM is lower than the GRU network. Compared with the attention mechanism additive, the performance improved a satisfactory amount which means that the attention mechanism learns other complex representations and with the different attention in the underflow concentration process variables. The multi-layer attention in those modules also helped to improve the whole prediction performance, the corresponding error results can be seen in Fig. 6.

On the other side, the addition of the average moving filter in the industrial shows that the prediction accuracy is improved by a large margin (90%). The different sliding time window was used in our experiment and the result shows that the best parameters are [20, 40, 20, 10, 20]. In DE-LSTM, the front of underflow concentration instant is considered to train the whole model with the attention mechanism, the experiment results show that we have leveraged the proposed prediction model in a state-of-the-art situation. The other prediction model is competitively compared in our experiments, the LSTM's performance outperforms the RNN's based model, and the specific RMSE and MAE are 0.7533, and 0.5119, respectively. Our DE-LSTM outperforms the other prediction methods because the hidden information from the hidden variables is fully captured by the proposed average moving filter, encoder, attention, and finally softmax transmission without loss. Besides, the front underflow concentration and reverse input variables are mutually and jointly to be used to train the whole and achieve superior performance and robustness. A remarkable performance has been implemented in this proposed architecture.

5 Conclusion

In this paper, a new data-centric model for the long-time underflow concentration prediction method in a deep cone thickener system, DE-LSTM, is proposed. DE-LSTM learns the historical and spatial information without information loss during the transmission of underflow concentration features for the paste filling process. In our practical case, the use of an average moving filter significantly enhances the prediction performance of the model. The DE-LSTM outperforms other state-of-the-art methods with the lowest RMSE and MAE in underflow concentration prediction while maintaining high robustness. Furthermore, the proposed DE-LSTM can be effortlessly extended to improve its applicability in other industrial domains and processes. For future consideration, we would consider the proposed DE-LSTM as the core bench for the intelligent underground mining system and optimal control platform, which will promote the whole development of the mining industry.

Funding This paper is supported by the Council Scholarship of China (CSC) under grant no. 202006370101. Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

Availability of data Data can be available upon reasonable request.

Declarations

Ethics approval The research does not involve human participants or animals and the authors warrant that the paper fulfills the ethical standards of the journal.

Consent to participate It is confirmed that all the authors are aware and satisfied with the authorship order and correspondence of the paper.

Consent for publication All the authors are satisfied that the last the revised version of the paper is published without any change.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Chen E, Cao H, Li H et al (2021) A big data mining approach for environmental emissions prediction of die casting process. *Int J Adv Manuf Technol* 114(11–12):3779–3791
- Gao X, Guo Y, Hanson DA et al (2021) Thermal error prediction of ball screws based on PSO-LSTM. *Int J Adv Manuf Technol* 116(5–6):1721–1735
- Lei Y, Chen X, Min M et al (2020) A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification. *Neurocomputing* 381:186–195
- Lei Y, Karimi HR, Cen L et al (2021) Processes soft modeling based on stacked autoencoders and wavelet extreme learning machine for aluminum plant-wide application. *Control Eng Prac* 108:104706
- Huan W, Ting L, Yuning C et al (2019) Underflow concentration prediction model of deep-cone thickener based on data-driven. *J China Univ Posts Telecommun* 26(6):63
- Tan CK, Setiawan R, Bao J et al (2015) Studies on parameter estimation and model predictive control of paste thickeners. *J Process Control* 28:1–8
- Fang C, He D, Li K, et al (2021) Image-based thickener mud layer height prediction with attention mechanism-based CNN. *ISA Transactions*
- Takács I, Patry GG, Nolasco D (1991) A dynamic model of the clarification-thickening process. *Water Res* 25(10):1263–1271
- Xiao D, Xie H, Jiang L et al (2020) Research on a method for predicting the underflow concentration of a thickener based on the hybrid model. *Eng Appl Comput Fluid Mech* 14(1):13–26
- Yuan Z, Hu J, Wu D et al (2020) A dual-attention recurrent neural network method for deep cone thickener underflow concentration prediction. *Sensors* 20(5):1260
- Qin Y, Song D, Chen H, et al (2017) A dual-stage attention-based recurrent neural network for time series prediction. [arXiv:1704.02971](https://arxiv.org/abs/1704.02971)
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Li C, Tang G, Xue X et al (2019) Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans Sustain Energy* 11(3):1370–1380
- Jaderberg M, Simonyan K, Zisserman A (2015) Spatial transformer networks. *Adv Neural Inf Process Syst* 28
- Zhang B, Xiong D, Su J (2018) Neural machine translation with deep attention. *IEEE Trans Pattern Anal Mach Intel* 42(1):154–163
- Yuan X, Li L, Shardt YAW et al (2020) Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Trans Ind Electr* 68(5):4404–4414
- Chen J, Li X, Xiao Y, et al (2022) FRA-LSTM: a vessel trajectory prediction method based on fusion of the forward and reverse sub-network. [arXiv:2201.07606](https://arxiv.org/abs/2201.07606)
- Geng J, Yang C, Li Y, et al (2022) MPA-RNN: a novel attention-based recurrent neural networks for total nitrogen prediction. *IEEE Trans Ind Inf*
- Greff K, Srivastava RK, Koutník J, et al (2016) LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232
- Yuan X, Li L, Wang Y (2019) Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans Ind Inf* 16(5):3168–3176
- Xiang S, Qin Y, Zhu C et al (2020) LSTM networks based on attention ordered neurons for gear remaining life prediction. *ISA Trans* 106:343–354
- Bandara K, Bergmeir C, Hewamalage H (2020) LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Trans Neural Netw Learn Syst* 32(4):1586–1599
- Xu X, Yoneda M (2019) Multitask air-quality prediction based on LSTM-autoencoder model. *IEEE Trans Cybernet* 51(5):2577–2586
- Siami-Namini S, Tavakoli N, Namin AS (2019) A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. [arXiv:1911.09512](https://arxiv.org/abs/1911.09512)
- Chen J, Zhang J, Xu X et al (2019) E-LSTM-D: a deep learning framework for dynamic network link prediction. *IEEE Trans Syst Man Cybernet Syst* 51(6):3699–3712
- Ho GTS, Lau HCW, Lee CKM, et al (2006) An intelligent production workflow mining system for continual quality enhancement. *Int J Adv Manuf Technol* 28
- Peng T, Zhang C, Zhou J et al (2021) An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. *Energy* 221:119887
- Li Y, Zhu Z, Kong D et al (2019) EA-LSTM: evolutionary attention-based LSTM for time series prediction. *Knowl-Based Syst* 181:104785
- Bin Y, Yang Y, Shen F et al (2018) Describing video with attention-based bidirectional LSTM. *IEEE Trans Cybernet* 49(7):2631–2641
- Wang Y, Huang M, Zhu X, et al (2016) Attention-based LSTM for aspect-level sentiment classification. In: *Proceedings of the 2016 conference on empirical methods in natural language processing*. pp 606–615
- Ran X, Shan Z, Fang Y et al (2019) An LSTM-based method with attention mechanism for travel time prediction. *Sensors* 19(4):861
- Usher SP, Scales PJ (2005) Steady state thickener modelling from the compressive yield stress and hindered settling function. *Chem Eng J* 111(2–3):253–261
- Koopman B, Cadee K (1983) Prediction of thickening capacity using diluted sludge volume index. *Water Res* 17(10):1427–1431
- Sayah M, Guebli D, Al Masry Z et al (2021) Robustness testing framework for RUL prediction Deep LSTM networks. *ISA Trans* 113:28–38
- Castilho LR, Medronho RA (2000) A simple procedure for design and performance prediction of Bradley and Rietema hydrocyclones. *Minerals Eng* 13(2):183–191
- Wang X, Zhang G, Zhao J, et al (2015) Underflow concentration prediction and external structure parameter optimization of deep cone thickener. *J Chongqing Univ*

37. Tan CK, Setiawan R, Bao J et al (2015) Studies on parameter estimation and model predictive control of paste thickeners. *J Process Control* 28:1–8
38. Yang Y, Xiong Q, Wu C et al (2021) A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ Sci Pollut Res* 28(39):55129–55139
39. Zhang T, Zheng XQ, Liu MX (2021) Multiscale attention-based LSTM for ship motion prediction. *Ocean Eng* 230:109066
40. Xiang S, Qin Y, Zhu C et al (2020) LSTM networks based on attention ordered neurons for gear remaining life prediction. *ISA Trans* 106:343–354
41. Liu Z, Li L, Fang X et al (2021) Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network. *Autom Construction* 125:103647
42. Gui Z, Sun Y, Yang L et al (2021) LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing* 440:72–88
43. He Y L, Chen L, Gao Y, et al (2021) Novel double-layer bidirectional LSTM network with improved attention mechanism for predicting energy consumption. *ISA Transactions*
44. Guo W, Wu C, Ding Z et al (2021) Prediction of surface roughness based on a hybrid feature selection method and long short-term memory network in grinding. *Int J Adv Manuf Technol* 112:2853–2871

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.