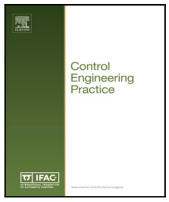
Control Engineering Practice 137 (2023) 105566

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac



DualLSTM: A novel key-quality prediction for a hierarchical cone thickener

Yongxiang Lei [a,b], Hamid Reza Karimi [a,*]

[a] Department of Mechanical Engineering, Politecnico di Milano, Milano, 20156, Italy
[b] School of Automation, Central South University, Changsha, 410083, China

ARTICLE INFO

Keywords:
Underflow concentration prediction
Attention
LSTM
Deep learning
Cone thickener system (CTS)

ABSTRACT

Due to the inaccuracy and significant disturbance of the complex and harsh environment in real industrial processes, the traditional sensor devices cannot meet the high-performance requirement of measuring key quality variables. However, in practical industrial thickener cone systems, the underflow concentration is hard to measure and has a high cost and significant time delay. Furthermore, the higher encoder representation often causes information loss from the process variables. This paper presents a novel efficient dual long short-time memory (LSTM) method for concentration prediction in the deep cone thickener system. To this end, dual feedforward and inverse bidirectional long-short time memory are proposed for feature learning and long temporal prediction. The proposed framework introduces an averaging moving filtering to pass through features, therefore the performance of dual LSTM is increased by a large margin. In addition, the feedforward and reverse bidirectional LSTM are employed to learn the robust information without loss. At last, experimental verification of the performance of an industrial deep cone thickener demonstrates the proposed dual LSTM method outperforms other state-of-the-art methods.

1. Introduction

Resource mining is a critical industry for societal development and progress. Raw mining products have been widely applied to many domains, such as aerospace and medicine. The hierarchical cone thickener is a critical device for paste filling in the mining industry (Fang, He, Li, Liu, & Wang, 2022; Huan, Ting, Yuning, & Aixiang, 2019; Tan, Setiawan, Bao, & Bickert, 2015). The current mining and paste-filling industry still faces some thorny challenges. For example, the underflow concentration of the cone thickener system (CTS) directly determines the quality of the whole pasting process (Takács, Patry, & Nolasco, 1991). The underflow concentration of CTS is a critical key-quality variable that is hard to measure because of the multi-couples, high device cost, and large time delay (Xiao et al., 2020). The accurate and stable underflow concentration can guarantee the stability of the paste-fill process. An oscillating underflow concentration would cause an imbalance and mud oscillation in the thickener. In this large-scale device, the prior task is to control the underflow concentration in the CTS. However, the CTS is a large time-delay process in that the underflow concentration needs to be increased by the paste gravity filling. The filling quality largely relies on the underflow concentration of the CTS. The core requirement of the pump filling is that it should keep the underflow concentration in a small dynamic range. If the underflow concentration is too high, it may cause filling accidents such as blockage and crushing. If the underflow concentration is too low, it cannot meet the requirement of paste filling and results in safety hazards. Therefore, taking into account the mechanical process of CTS and then accurately predicting the underflow concentration are serving as the basis for further control purposes, maintaining the quality of paste mining filling, and avoiding safety hazards (Yuan, Hu, Wu, & Ban, 2020).

The traditional physical device can hardly achieve the online long underflow concentration prediction because the current CTS is a multi-coupled system. The different sampling frequency of pressure and flow leverages the massive integrated sensors to detect and monitor the production process. Furthermore, much historical information and factors among the quality variables should be considered. Based on the industrial case analysis, some solutions to these problems have been proposed. To sum up, the existing methods can mainly be classified into two categories: (1) the mathematical model-centric mechanism (also called the 'White Box') and (2) the model-free data-centric methods ('Black Box'). In the former case, however, the performance of white box-based models is determined by the specific mathematical model of the device, which means the accuracy will be greatly decreased if the model is incorrect. Furthermore, there are a lot of devices that do not have a mechanical mathematical model, so these models lose their significance. As for the latter case, with the development of machine learning, many data-driven self-supervised and semi-supervised methods have been proposed in the literature to improve the modeling of industrial processes (Lei, Karimi, & Chen, 2022; Wang, Yin, Deng,

* Corresponding author.
E-mail addresses: yongxiang.lei@polimi.it (Y. Lei), hamidreza.karimi@polimi.it (H.R. Karimi).

https://doi.org/10.1016/j.conengprac.2023.105566
Received 10 December 2022; Received in revised form 9 May 2023; Accepted 11 May 2023
Available online 24 May 2023

Bo, & Shao, 2022). The main advantage of these methods is that they can learn a nonlinear mapping projection between the input variables and the target outputs. Compared to traditional physical sensor devices (Barua & Sharma, 2022; Chen, Chen, et al., 2022; Chen, Zhang, & Ye, 2022; Goh et al., 2022; Hu, Wang, Li, & Wang, 2022; Liang, Chai, Sun, & Tan, 2022; Lin, Zhang, Li, & Lu, 2022; Shuaiyi, Wang, Zhang, & Wang, 2022; Sun, Sun, Wang, Zhou, & Cai, 2022; Zhang et al., 2022), it is the most economical way to measure key quality, which can save costs and increase profits.

Over the last decades, deep learning technology has become a hot topic in various research fields. For some time-series tasks, many temporal models have emerged. For example, recurrent neural network (RNN) (Qin et al., 2017), long short-time memory (LSTM) (Hochreiter & Schmidhuber, 1997), gated recurrent unit (GRU) (Li, Tang, Xue, Saeed, & Hu, 2019), Transformer (Jaderberg et al., 2015), feature representation (Liu, Wang, Wang, & Yuan, 2021), etc. In Jiang and Yin (2018)'s work, the time series factors have been considered for many plant-wide applications (Zhu et al., 2023). For instance, some improved attention mechanism has been applied to the models (Zhang, Xiong, & Su, 2018). However, in the specific CTS system, some research has been put into the deep cone thickener systems, and the time instant key factor has been ignored. For example, Huan et al. (2019) proposed an underflow concentration prediction method based on the XGBoost model, which achieves satisfactory performance in the underflow concentration of CTS. Nevertheless, the results in Huan et al. (2019) are based on a static prediction method and the historical and temporal information has been lost during the network training stages. The information contained in time-instant factors has been underestimated. Recently, a dual-attention RNN method is used for the deep cone thickener application, their prediction approach considers the temporal pertain to the same quality factors and achieves comparative accuracy and efficiency, while its performance relies on fine-tuning the hyperparameters and the generation ability has been greatly decreased. Yuan, Hu, et al. (2020) introduces a spatiotemporal attention-based LSTM for industrial soft sensor model, which achieves great development in the processes industry. Chen, Li, Xiao, Chen, and Zhao (2022) investigates an automatic identification system (AIS) data-driven LSTM method based on the fusion of the forward sub-network and the reverse sub-network (termed as FRA-LSTM) to predict the vessel trajectory, however, their methods did not consider the noise processing and time-spatial factors. Geng, Yang, Li, Lan, and Luo (2022) proposes a novel attention-based recurrent neural network for nitrogen prediction, where three kinds of attention mechanisms are used in the MPA-RNN model. In Greff, Srivastava, Koutník, Steunebrink, and Schmidhuber (2016), the specific investigation of the forward and backward gradient of the LSTM unit is reported, it covers the implementation of the training step with the backpropagation through time. Some other variants of the recent LSTM-based model can be found in Barua and Sharma (2022), Chen, Chen, et al. (2022), Liu, Wang, Wang, Xie, and Yang (2021) and Yuan, Li, Shardt, Wang, and Yang (2020). Compared to the traditional industry (Lei, Karimi, Cen, Chen, & Xie, 2021; Lei et al., 2022), the data collected directly by the exogenous infrequent sensor in CTS has large disturbances, errors, and occasional sudden outliers. In the proposed framework, this problem can be alleviated with appropriate filter processing.

Based on the aforementioned analysis, the time series prediction of the underflow concentration of CTS has not been yet completely studied. Developed from Yuan, Hu, et al. (2020), a novel underflow concentration method based on deep LSTM and average smooth filtering is proposed. The motivation is that the average smooth filtering is used to prepossess the industrial CTS data, and then a dual bidirectional LSTM (BiLSTM) with that integrated attention mechanism is generated for the underflow concentration prediction. The main solid bullet benefits of this paper are:

• Unlike the traditional RNN, LSTM, and bidirectional LSTM, a new prediction model architecture (DualLSTM) is proposed. The proposed architecture uses the feedforward and reverse BiLSTM as the basic units, which implement the new information learning. Average smoothing filtering is considered in the plant-wide process and thus Gaussian white noise cannot affect the performance of the proposed model. The robustness and generalization ability has been verified. The prediction accuracy has improved by a large margin compared to the existing state-of-the-art methods.

• Distinct from the previous prediction method in Xie et al. (2020), Yuan, Hu, et al. (2020), Yuan, Li, et al. (2020) and Yuan, Qi, Wang, and Xia (2020), a novel back-propagation through time (BPTT) algorithm is designed for the DualLSTM. The proposed solution for the new method is another key contribution of this brief.

• To the best of our knowledge, we first employ the DualLSTM to tackle the underflow concentration prediction in industrial CTS. The proposed model can be a core part of further intelligent control, which is the basis for further optimization control in an integrated platform.

To verify the feasibility and effectiveness of the proposed method, a realistic industrial application in Beijing, China, is covered by the core DualLSTM verification. The following is the structure of the paper. Section 2 contains the foundational efforts, which include the problem description, LSTM unit basics, and average smooth filtering techniques. Section 3 and Section 4 address the prediction methodology and experimental case studies, respectively. Finally, Section 5 concludes the paper.

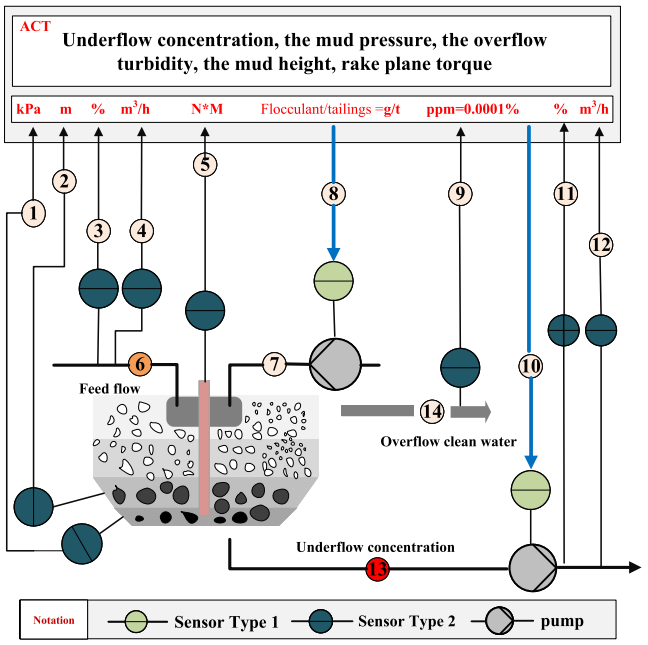


Fig. 1. Test bench of deep cone thickener treatment plant process and the role of our method. All the main process variables have been specified in Table 1. Sensor Type 1 represents the different process sensors, while Sensor Type 2 represents the different speed sensor categories, respectively.

2. Preliminary foundations

This section provides an introduction to the deep thickener system and the paste-filling process problem. It also gives the contents of the LSTM unit and the average smooth filtering. Both aspects are the foundation for the further development of the model framework.

2.1. Problem statement

For subsurface paste filling, a deep cone thickener is essential. The diagram of CTS is shown in Fig. 1. CTS is a key method for achieving a consistent concentration for underground mining fills. Pipe-blocking

Table 1
Variable candidates specification.

Symbol	Quality	Description
①	Mud pressure	KPa
②	Mud height level	m
③	Feed flow concentration	$C\%$
④	Feed flow amount	Q
⑤	The rake torque	N M
⑥	Tailings	t
⑦	The amount of flocculant	g/t
⑧	Flocculant flowrate	m/s
⑨	Overflow turbidity	ppm
⑩	The rotation speed of rake	r/min
⑪	Frontier tth underflow concentration	$C\%$
⑫	Underflow flow amount	Q
⑬	Underflow pressure	KPa
⑭	Overflow channel	$C\%$

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, as a result, the overall safety of the mining operation. Therefore, developing a model to forecast subterranean concentrations in the CTS system is critical. The whole mining paste filling operates on a continuous and hierarchical concept. 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 dissolved particles. The suitable concentration and volume feed flow are created at the bottom of the CTS. The top clean water from the overflow pipe has also been recycled for future use. The CTS's primary control need is to provide a consistent and precise underflow concentration. 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. 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. The specific statement of the problem is

Given: $\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{Y}\right)$

Predict: $\hat{\mathbf{Y}} = \left(y_{T+1}, y_{T+2}, \ldots, y_{T+t}\right)^{\mathrm{T}} \in \mathbb{R}^t$. where $\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{Y}\right)$ is the sequential and historical quality variables series. T denotes as the instant of the underflow concentration. Specifically, we need to learn a nonlinear mapping representation with $\left(\hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_{T+\tau}\right) = F\left(y_1, \ldots, y_T, x_1, \ldots, x_n\right)$.

2.2. Model motivation formulation

The underflow concentration is challenging to estimate using existing methods, so we propose a new data-centric model to address the problem. The underflow concentration is detected using some flow switch put on the unit in the practical plant, which means the online forecast is not possible. Our motivation stems from the fact that the process variables are inextricably linked, allowing us to construct a time sequence model for the key-quality prediction. Furthermore, there is a significant time delay in the processes that deal with the past data.

As shown in Greff et al. (2016), the LSTM unit is a time sequence model that preserves historical data while avoiding recurrent neural network long-dependency. It memorizes the information from the previous circumstance via the memory gate. Three gates protect and control the cell state in the basic LSTM unit, and some repeating modules are added in an LSTM with four interacting layers. To avoid gradient vanishing, the LSTM has the capacity to establish temporal memory through gate switching.

Denote the input historical vector $x_{(t)}$ and the precious hidden state $h_{(t-1)}$, and the external inputs are inherited from the previous cell state $c_{(t-1)}$. Then, the forget gate is triggered as:

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

a new formation from the input gate and new candidate vectors are calculated as:

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \\ \tilde{C}_t &= \tanh\left(W_{(C)} \cdot \left[h_{t-1}, x_t\right] + b_C\right). \end{aligned} \tag{2}$$

The new state of an LSTM cell is updated by

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{3}$$

The output gate vector can be given by

$$\begin{aligned} o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right), \\ h_t &= o_t * \tanh\left(C_t\right). \end{aligned} \tag{4}$$

where σ is the nonlinear activation function, usually, the Sigmoid function, tanh represents the nonlinear tangent activation function, and $*$ represents the point-wise multiplication operation. Parameters such as W_c, W_o, W_i are the related weights and b_i, b_c, b_o are the bias, respectively. The whole weight parameters can be learned with some gradient descent algorithms. So the LSTM can be organized as the basic unit for the network design. However, the existing literature about the prediction with time sequences has been reproduced, their result shows a shortcoming of low prediction accuracy in long-time prediction, which is hard to apply to the industrial CTS for key-quality prediction. Due to the reason that the traditional LSTM can only effectively predict the short time influence between these variables, some historical spatiotemporal sequences are commonly ignored. Besides, in industrial CTS, there are a bunch of outliers, and bidirectional LSTM can give a superior and robust representation of the feature embedding, and the simultaneous feedforward and reverse training can jointly improve the whole accuracy, so we build the BiLSTM as the basic unit. Consider in a far way, the original output of BiLSTM cannot balance the influence of the different weights for the underflow concentration prediction. The frontier encoder representation should be further assigned with different attention. Therefore, in the underflow prediction model design, we employ the blocks of attention mechanisms. Compared to the traditional BiLSTM, in each sliding window, we deploy a dual BiLSTM with an attention mechanism to achieve a robust representation of the entire underflow period. The embedding of the underflow concentration in CTS at each point in time is focused on the proposed DualLSTM method. On the other hand, the underflow concentration is greatly influenced by the frontier time sequences, which also motivates us to build the dual BiLSTM attention method.

Another factor to consider is that our goal is to enhance long-term forecast accuracy by a significant margin, and data is a vital component. Average moving filters are further included in the developed model to smooth and process the training and testing data in order to improve the overall quality of the acquired data and optimize the suggested DualLSTM's performance. As a result, we are able to maximize the proposed DualLSTM performance to an extremely high level. The presented average moving filtering overcomes the outlier's drawbacks and proposes a long-term prediction of underflow concentration in the CTS, which provides us with a strong motivation to develop the DualLSTM method. The entire proposed model framework emerges based on the preceding principles and is likewise formulated in Fig. 4 (see Figs. 2 and 3).

3. Methodology

The core idea of this paper is to propose a prediction method for the underflow concentration of the industrial CTS process. The formulation of the proposed method covers the variables selection, preprocessing and averaging smooth filtering, and whole model training and prediction implementation.

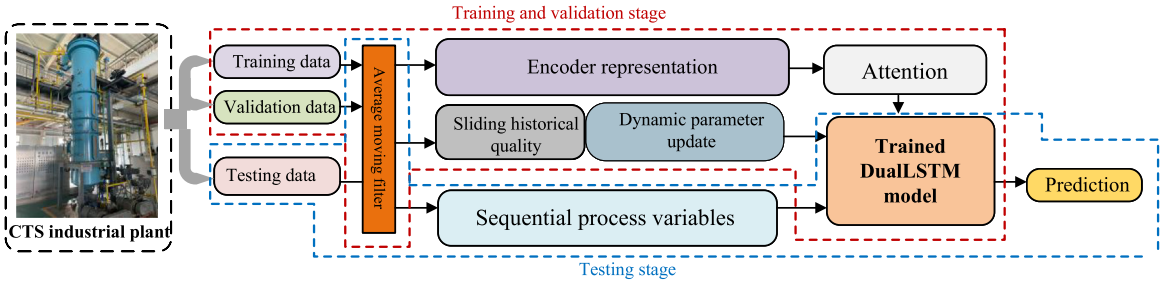


Fig. 2. The model flowchart illustration of DualLSTM for the underflow quality prediction. The original data collected from the industrial plant are transmitted to the preprocessing unit, and then the training process is implemented by the proposed gradient descent algorithm, the testing process is marked in blue, while the training process is marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

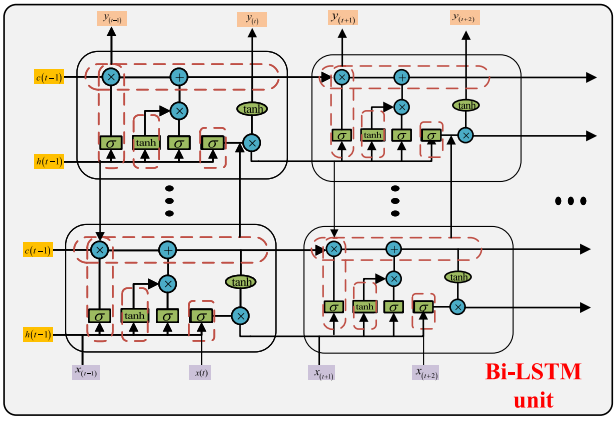


Fig. 3. The basic illustration of BiLSTM unit. The BiLSTM is a bi-directional feature architecture that accepts the feedforward and reverses inputs, the higher feature representations are extracted by the stacked unit.

3.1. Variable analysis

Consider the practical scenario, the number of various variables is consecutively coupled with each other. To achieve an efficient variable quality of the cone thickener variable, a coefficient analysis was performed. The auxiliary variables in the production process should be analyzed for further modeling training. Therefore, the choices of auxiliary variables are the first step in generating the data-driven prediction model. The necessary and vital factors we added to the model are the mud height and pressure. From the basis of mechanical analysis, there is a nonlinear link between these two variables.

The final selection of the dominant variables is the mud pressure, the mud height, the last-instant feed flow, the top volume flow, underflow density, and the front instant underflow concentration.

3.2. Prepossessing policy and average smooth filtering

Because of noisy and non-uniform data in the practical measurement and sampling process with the different sensors, data samples accuracy is limited by the physical device's sampling frequency. Efficient and effective smooth filtering algorithms are utilized in the proposed method. As in Greff et al. (2016), the same N rounds input signals are expectation sampled. In our practical evaluation, the sequence of underflow concentration and feed flow series are tackled with the average filtering algorithm to reduce the Gaussian white noise and smooth the whole signal. The idea of the mean filtering is developed from the convolution kernels, during the training of the whole sequence model, the moving multiple-pass average filters are utilized. The specific moving window is set by ourselves.

3.3. Designed BPTT for DualLSTM

The framework and flowchart of the whole proposed algorithms are given as follows. Two stages of mechanisms are used for the training. The first stage is coordinate training, which means all the historical variable data are used to train the DualLSTM model for acquiring the optimal parameters. The solution process is fine-tuned with a background gradient algorithm. The backpropagation through time (BPTT) (Yuan, Li, et al., 2020) training process is given. With BPTT, the maximum number of instant steps is set along with errors that can be propagated.

In the proposed framework, $h_{(t-1)}$ is the hidden state, $x_{(t)}$ is the input variable vector, and $T_{(t)}$ is the prediction time. Assume the key variables for training and underflow concentration sequences are $\{x_{(1)}, x_{(2)}, \dots x_{(N)}\}$ and $\{T_{(1)}, T_{(2)}, \dots T_{(N)}\}$, where N is the total number of training data in the time sequence. Suppose that the intended behavior of the learning process is to observe the memory gate state at the end of the epoch $[t, t+s]$. The desired underflow concentration or the output loss is:

$$L^{\langle t \rangle}(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_{(t)} - \hat{y}_{(t)}\right)^2 \tag{5}$$

The next target is to train the whole network parameters. Consider the dual Bi-LSTM has two embeddings with feedforward LSTM and the reverse LSTM. Thus, developed from Greff et al. (2016), the DualLSTM fine-tuning policy for the hyper-parameters is designed in the sequel.

Suppose the feedforward and reverse input weights: $\vec{\mathbf{W}}_z$, $\vec{\mathbf{W}}_i$, $\vec{\mathbf{W}}_f$, $\vec{\mathbf{W}}_o$, $\overleftarrow{\mathbf{W}}_z$, $\overleftarrow{\mathbf{W}}_i$, $\overleftarrow{\mathbf{W}}_f$, $\overleftarrow{\mathbf{W}}_0 \in \mathbb{R}^{N\times M}$.

Recurrent feedforward and reverse weights: $\vec{\mathbf{R}}_z$, $\vec{\mathbf{R}}_i$, $\vec{\mathbf{R}}_f$, $\vec{\mathbf{R}}_o$, $\overleftarrow{\mathbf{R}}_z$, $\overleftarrow{\mathbf{R}}_i$, $\overleftarrow{\mathbf{R}}_f$, $\overleftarrow{\mathbf{R}}_o \in \mathbb{R}^{N\times N}$.

Output feedforward and reverse weights: $\vec{\mathbf{p}}_i$, $\vec{\mathbf{p}}_f$, $\vec{\mathbf{p}}_o$, $\vec{\mathbf{p}}_i$, $\overleftarrow{\mathbf{p}}_f$, $\overleftarrow{\mathbf{p}}_o \in \mathbb{R}^N$.

Bias weights: $\vec{\mathbf{b}}_z$, $\vec{\mathbf{b}}_i$, $\vec{\mathbf{b}}_f$, $\vec{\mathbf{b}}_o$, $\overleftarrow{\mathbf{b}}_z$, $\overleftarrow{\mathbf{b}}_i$, $\overleftarrow{\mathbf{b}}_f$, $\overleftarrow{\mathbf{b}}_o \in \mathbb{R}^N$.

The piecewise epoch equations $\mathbf{z}$ are expressed as:

$$\begin{aligned}\vec{\mathbf{z}}^t{}_f &= \overleftarrow{\mathbf{W}}_z\mathbf{x}_f{}^t + \overleftarrow{\mathbf{R}}_z\mathbf{y}_f{}^{t-1} + \overleftarrow{\mathbf{b}}_z,\\ \mathbf{z}_f{}^t &= g(\vec{\mathbf{z}}^t{}_f).\end{aligned} \tag{6}$$

The dual forward and reverse input gates **i**s are:

$$\begin{aligned}\bar{\mathbf{i}}_f{}^t &= \tilde{\mathbf{W}}_i\mathbf{x}_f{}^t + \tilde{\mathbf{R}}_i\mathbf{y}_f{}^{t-1} + \tilde{\mathbf{p}}_i \odot \tilde{\mathbf{c}}^{t-1} + \tilde{\mathbf{b}}_i,\\ \mathbf{i}_f{}^t &= \sigma(\bar{\mathbf{i}}^t_f)\end{aligned} \tag{7}$$

$$\begin{aligned}\bar{\mathbf{i}}_r{}^t &= \overleftarrow{\mathbf{W}}_i\mathbf{x}_r^t + \overleftarrow{\mathbf{R}}_i\mathbf{y}_r^{t-1} + \overleftarrow{\mathbf{p}}_i \odot \overleftarrow{\mathbf{c}}^{t-1} + \overleftarrow{\mathbf{b}}_i,\\ \mathbf{i}_r{}^t &= \sigma(\overleftarrow{\mathbf{i}}_r).\end{aligned} \tag{8}$$

The dual forward and reverse forgot gates **f**s are

$$\begin{aligned}\bar{\mathbf{f}}_f{}^t &= \tilde{\mathbf{W}}_f\mathbf{x}_f{}^t + \tilde{\mathbf{R}}_f\mathbf{y}_f{}^{t-1} + \tilde{\mathbf{p}}_f \odot \tilde{\mathbf{c}}^{t-1} + \tilde{\mathbf{b}}_f,\\ \mathbf{f}_f{}^t &= \sigma(\bar{\mathbf{f}}_f{}^t)\end{aligned} \tag{9}$$

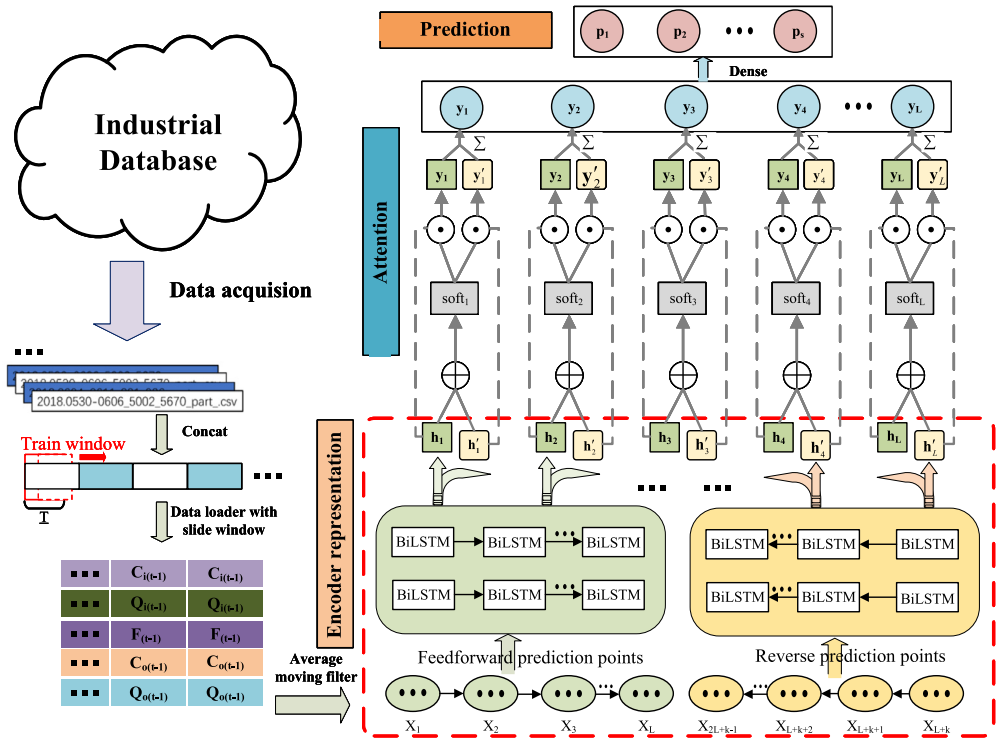


Fig. 4. Flowchart of the proposed underflow concentration prediction method for the deep cone thicker production. In this framework, a feedforward and reverse bi-directional stacked long-short time memory is constructed for obtaining a complex temporal feature extraction, then, the attention mechanism is integrated for the robust feature aggregation, and the output for the prediction is obtained from the high-level abstract features.

$$\begin{aligned}\bar{\mathbf{f}}_r^{\,t} &= \overleftarrow{\mathbf{W}}_f \mathbf{x}_r^{\,t} + \overleftarrow{\mathbf{R}}_f \mathbf{y}_r^{\,t-1} + \overleftarrow{\mathbf{p}}_f \odot \overleftarrow{\mathbf{c}}^{\,t-1} + \overleftarrow{\mathbf{b}}_f,\\ \mathbf{f}_r^{\,t} &= \sigma(\bar{\mathbf{f}}_r^{\,t})\end{aligned} \tag{10}$$

Then the forward and reverse cell states can be represented, respectively, by

$$\begin{aligned}\mathbf{c}_f^{\,t} &= \mathbf{z}_f^{\,t} \odot \mathbf{i}_f^{\,t} + \mathbf{c}_f^{\,t-1} \odot \mathbf{f}_f^{\,t},\\ \mathbf{c}_r^{\,t} &= \mathbf{z}_r^{\,t} \odot \mathbf{i}_r^{\,t} + \mathbf{c}_r^{\,t-1} \odot \mathbf{f}_r^{\,t}.\end{aligned} \tag{11}$$

The forward and reverse outputs **o**s are

$$\begin{aligned}\bar{\mathbf{o}}_f^{\,t} &= \vec{\mathbf{W}}_o \mathbf{x}_f^{\,t} + \vec{\mathbf{R}}_o \mathbf{y}_f^{\,t-1} + \vec{\mathbf{p}}_o \odot \mathbf{c}_f^{\,t} + \vec{\mathbf{b}}_o,\\ \mathbf{o}_f^{\,t} &= \sigma(\bar{\mathbf{o}}_f^{\,t})\end{aligned} \tag{12}$$

$$\begin{aligned}\bar{\mathbf{o}}_r^{\,t} &= \overleftarrow{\mathbf{W}}_o \mathbf{x}_r^{\,t} + \overleftarrow{\mathbf{R}}_o \mathbf{y}_r^{\,t-1} + \overleftarrow{\mathbf{p}}_o \odot \mathbf{c}_r^{\,t} + \overleftarrow{\mathbf{b}}_o,\\ \mathbf{o}_r^{\,t} &= \sigma(\bar{\mathbf{o}}_r^{\,t})\end{aligned} \tag{13}$$

$$\begin{aligned}h_i^{\,t} &= h(\mathbf{c}_f^{\,t}) \odot \mathbf{o}_f^{\,t},\\ (h')_i^{\,t} &= h(\mathbf{c}_r^{\,t}) \odot \mathbf{o}_r^{\,t}.\end{aligned} \tag{14}$$

where σ, g and h are the piecewise nonlinear activation projections, respectively. Since the backward and forward processes are the same in principle but in opposite series order, the equation of the backward process could be derived similarly by replacing $\rightarrow$ with $\leftarrow$. Then we finished the encoder representation of the underflow key quality features, the attention mechanism inherited from the BiLSTM output can be given as the following embedding feedforward process:

$$soft_i = \tanh\left(\vec{W}_h h_i + \vec{b}_h\right) + \tanh\left(\overleftarrow{W}_h h_i' + \overleftarrow{b}_h\right) \tag{15}$$

$$\alpha_i = \frac{\exp\left(soft_i\right)}{\sum_{t=1}^{N} \exp\left(soft_t\right)}, \quad \sum_{i=1}^{N} \alpha_i = 1 \tag{16}$$

$$y = \sum_{i=1}^{N} \alpha_i h_i, \quad y' = \sum_{i=1}^{N} \alpha_i h_i' \tag{17}$$

Then, the final output of DualLSTM would be:

$$p_j = \sum_{j=1}^{s} W_j \cdot y_i \langle 1, \ldots, L\rangle = \sum_{j=1}^{s} \sum_{i=1}^{L} W_j W_y \cdot \left\langle y_i, y_i' \right\rangle \tag{18}$$

where W_j and W_y are the weight parameters that connect the dense layer and attention layer, respectively. Thus, the aim of the Eqs. (6)–(18) is to learn a mapping prediction function:

$$\begin{aligned}&p\left(y_{t+s} | X_1, X_2, \ldots X_L, X_{L+k}, X_{L+k+1}, \ldots, X_{2L+k-1}\right)\\ &= p(y_{t+s} | \langle \mathbf{C}_{i(t)}, \mathbf{Q}_{i(t)}, \mathbf{F}_{(t)}, \mathbf{C}_{o(t)}, \mathbf{Q}_{o(t)} \rangle)\\ &= p\left(y | W \cdot \sum_{i=1}^{L} \left\langle y_i, y_i' \right\rangle\right)\end{aligned} \tag{19}$$

The backpropagation algorithm with Adam optimizer is employed to train the corresponding parameters in the DualLSTM model.

The training algorithm for DualLSTM can be summarized for the real-time propagation to perform the following steps at each time t: (1) Collect the sufficient samples with the input variables matrix [C_i, Q_i, F, C_o, Q_o]; (2) Initialize the network parameters which include the time sliding window T, the prediction steps s, the input time sequence L, the layers of forward and reverse BiLSTM, and the average filter step matrix; (3) Preprocess the whole data with an average moving filter, split all the data into a long sequence, and divide it into the training dataset, validation dataset and the test dataset according to the ratio of 7:2:1, which is decided by the expert experience; (4) Training the whole network according to the designed equations (6)–(18), all the gradients of the partial derivative of the error can be computed and the weights are tuned accordingly; (5) Build a training window with the duration T, slide the data loaded for neural network training, and the loss function is the root mean square error between the predicted value and the true value; (6) After the training, the end-to-end underflow concentration prediction model in industrial CTS is obtained, which can predict the underflow concentration at the next T moment according to the feed

concentration, feed flow rate, underflow flow rate, thickener pressure, and underflow concentration at the previous time of T duration.

Among them, T is an adjustable key hyperparameter, which has a more significant impact on prediction accuracy. In addition, the depth of the feedforward and reverse BiLSTM layer and the size of the hidden layer of the BiLSTM are all key weights that need to be adjusted through training experiments. The attention mechanism captures the attention intensity of a in the sequence at each time point. In the training procedure, we employ the Adam optimizer to acquire the learnable weight parameter with the learning rate l_r.

4. Case study

To evaluate the performance of the proposed method, two cases are studied. The proposed DualLSTM is evaluated by the industrial application of long-time underflow prediction for the CTS system. Some public methods such as support vector regression (SVR), Artificial Neural Networks (ANN), BiLSTM, and the temporal recurrent network such as recurrent neural network (RNN), XGBOOST (Huan et al., 2019), LSTM and gated recurrent unit (GRU) and DARNN (Yuan, Li, et al., 2020) are also compared during the experiments. The evaluation indexes are selected from the literature. The RMSE can be given as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_t^i - y_t^i)^2} \tag{20}$$

Also, another index for the evaluation is MAE, which is rewritten as follows:

$$MAE = \frac{1}{N_u}\sum_{i=1}^{N_u}|\hat{y}_t - y_t| \tag{21}$$

Finally, the prediction uncertainty is quantified by the confidence interval:

$$CI = \bar{y} \pm a\frac{s}{\sqrt{N}} \tag{22}$$

where a is the value from the standard normal distribution for the selected confidence level, and s is the standard deviation. All the evaluations are conducted on the i7-PC Python 3.9 with the package Keras (version 3.8.6) with Tensorflow (version 2.4.1) backend and running on 4 NVIDIA GeForce RTX 2080 Ti GPU and 198 GB RAM. The data is standardized by a standard scaler, which transforms the data by removing its mean value and scaling to unit variance. The default learning rate l_r is set to 0.0001 with an increase of the batch size of 5 for each epoch. To make a complete comparison, the network hyperparameter for the RNN, LSTM, and BiLSTM are both settled in an input layer (constructed by RNN, LSTM, or BiLSTM) with 128 hidden neurons, which are connected by a fully-connected layer with 64 neurons and finally, an output layer with 1 neuron. An ANN model is constructed by $128-64-8-1$ dense architecture. The input size for these models is both normalized as a vector $[batch\_size, time\_seq, num\_feature]$, and the $time\_seq$ is set as the same length with sliding window T, which is 20, $num\_feature$ is set as 12. According to the standard normal distribution coefficient, a is chosen as 1.96 for obtaining the 95% confidence interval.

4.1. Industrial CTS application case

In the practical application, the variable coefficient should be considered. According to the worker's experiences, 12 variables are collected from the database which is sampled from the sensors and can satisfy the same distribution. The main influencing factors of underflow concentration are feed amount, feed concentration, underflow flow rate, mud layer height, rake speed, flocculant ton consumption, etc. Among them, the feed concentration is restricted by the dressing plant's production conditions; the rake's speed is generally determined by the yield stress of the paste and depends on the characteristics of the aggregate; the amount of additive flocculant is determined by laboratory experiments. The presented framework is combined with existing data to build an end-to-end machine learning model which uses feed concentration, feed flow, previously detected underflow flow, and thickener pressure as the original input. Since in the dense process, the current concentration has a strong correlation with the control amount and system state in the past period, therefore, a machine learning model with time series is constructed. The model inputs are feed concentration, feed flow rate, underflow flow rate, thickener pressure, and underflow concentration for a period of time before this time. In the DualLSTM model, to balance the network architecture, the weights of W_j, W_j' are set as 0.5.

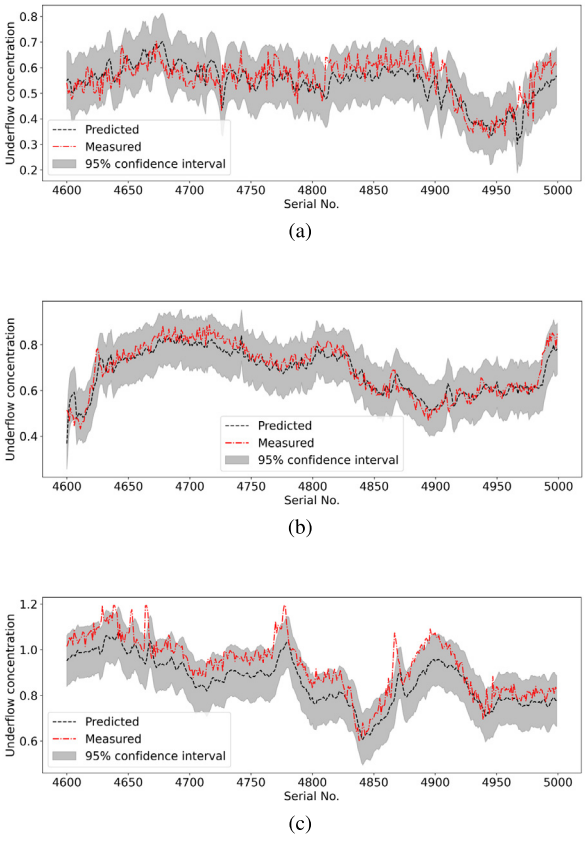


Fig. 5. The proposed ANN model for industrial underflow prediction results. (a) Training dataset; (b) Validation dataset; (c) Testing dataset. All the simulation results are based on the 95% confidence interval.

4.2. Result analysis

During this period, through training experiments, the influence of training batch size and sliding window length T on prediction accuracy was explored. The corresponding comparison results are shown in Figs. 5–7. On the way, 3000 epochs were trained, the 3 layers of BiLSTM layer that concatenates the hidden features to the encoder representation, and the hidden layer size is 256. The sliding window length and the batch size are set as 4 and 8, respectively.

The data collected directly by the actual industrial sensor is subject to certain restrictions such as disturbances, random errors, or occasional sudden outliers, which is the original motivation for developing the proposed model. In the proposed DualLSTM 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 significantly improve the prediction accuracy and reduce

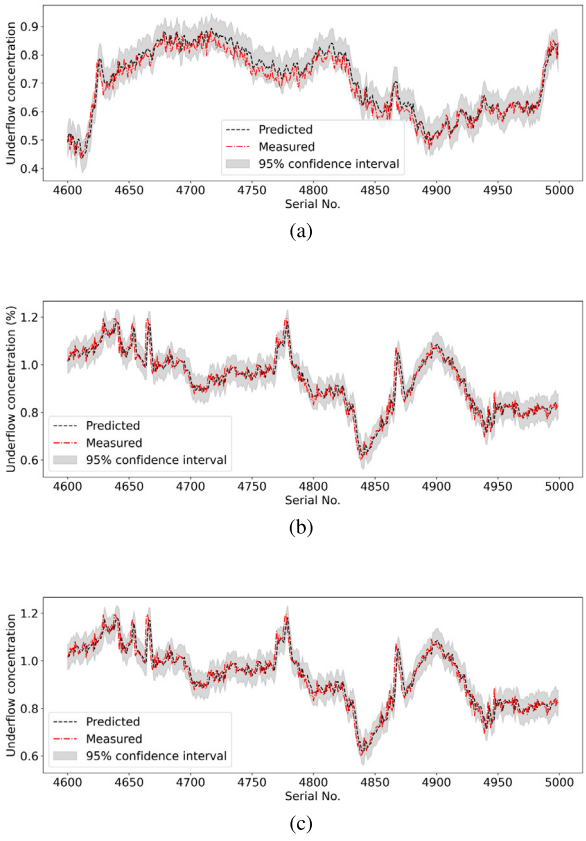


Fig. 6. The proposed GRU model for industrial underflow prediction results. (a) Training dataset; (b) Validation dataset; (c) Testing dataset.

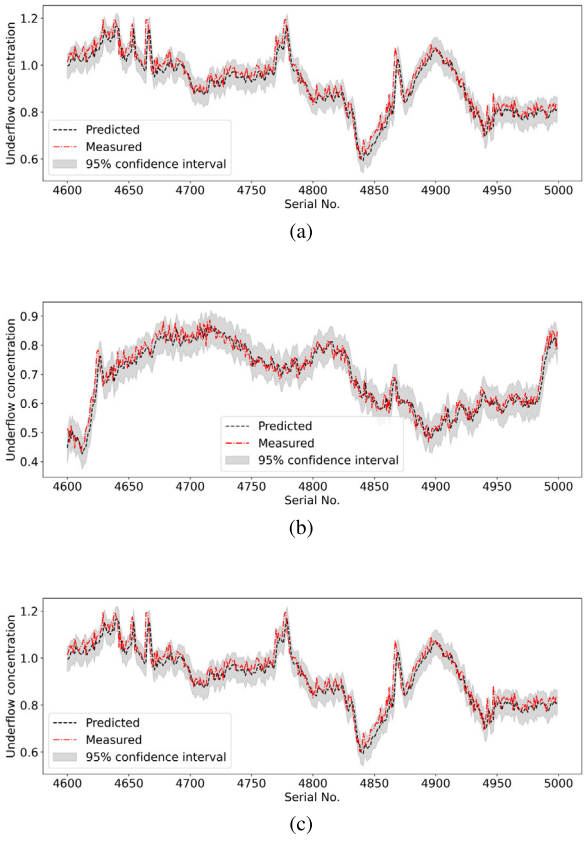


Fig. 7. ARMA model for industrial underflow prediction, a noticeable time lag can be seen in these results. (a) Validation dataset; (b) Testing dataset.

the root mean square error (about-96%). Fig. 8 evaluates the measured evaluation loss indexes comparison for different algorithms with GRU, LSTM, ANN, Ours, and RMSE loss is also illustrated. The traditional ANN has the largest loss during the training process. Figs. 9 and 10 further verify the prediction efficacy of the proposed framework. All the experiments are conducted based on the 95% confidence interval. In Fig. 9, compared to the LSTM and GRU, our proposed DualLSTM acquires smoother prediction performance in the training dataset. Fig. 9(e) outperforms Fig. 9(a)–(d) in the testing dataset performance, which means our proposed DualLSTM achieves the highest accurate prediction. Fig. 10 compares the DualLSTM's whole performance in the training, validation and testing datasets. It shows that the proposed DualLSTM gives a more stable and robust prediction (The confidence interval is smaller than BiLSTM), which means underflow concentration prediction is more confidential.

As shown in Table 2, compared to the other competitive methods, DualLSTM shows the best performance with the lowest RMSE and MAE, 0.2234 and 0.1735 respectively. The SVR with a multi-polynomial activation function achieves the second-best performance, compared with the basic LSTM and GRU network. However, if we use other kernels, like linear kernel or Gaussian kernel, the RMSE 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 additive, the performance improved by a satisfactory amount which means that the attention mechanism learns other complex representations with the different attention in the underflow concentration process variables. The multi-layer attention in those modules also helped to improve the overall prediction performance.

On the other side, the addition of the average moving filter in the industrial case study shows that the prediction accuracy has improved by a large margin (approximately 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 DualLSTM, the front of the 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 fashion. The other prediction model is also 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,

Table 2
Evaluation indexes for the different competitive prediction methods.

Methods	RMSE	MAE
RNN	0.6712 ± 0.0124	0.3488 ± 0.0082
GRU	0.4634 ± 0.0078	0.2499 ± 0.0345
LSTM	0.7533 ± 0.0030	0.5119 ± 0.0079
SVR Yuan, Qi, et al. (2020)	0.8133 ± 0.0002	0.6003 ± 0.0013
LGB Jing, Hu, Guo, Wang, and Chen (2020)	0.8091 ± 0.0092	0.6083 ± 0.0118
VAEWGAN Hou, Sun, Shen, and Qiu (2019)	0.7985 ± 0.0116	0.5847 ± 0.0164
GSTAE Prastyo, Nabila, Lee, Suhermi, and Fam (2019)	0.7735 ± 0.0068	0.5296 ± 0.0097
SS-PdeepFM Ren, Wang, Laili, and Zhang (2021)	0.7650 ± 0.0036	0.5206 ± 0.0073
MSWR-LRCN Chen, Zhang, and Zhang (2022)	0.7571 ± 0.0033	0.5235 ± 0.0092
LSTM-DeepFM Ren et al. (2021)	0.7497 ± 0.0026	0.5091 ± 0.0071
MPA-RNN Geng et al. (2022)	0.7515 ± 0.0034	0.2074 ± 0.0072
DA-RNN Qin et al. (2017)	0.851 ± 0.0029	0.2322 ± 0.0052
Attention	0.9086 ± 0.0021	0.2445 ± 0.0031
DSTP-RNN Liu, Gong, Yang, and Chen (2020)	0.8496 ± 0.0051	0.2254 ± 0.0042
Suggested DualLSTM	**0.2234 ± 0.0021**	**0.1731 ± 0.0013**

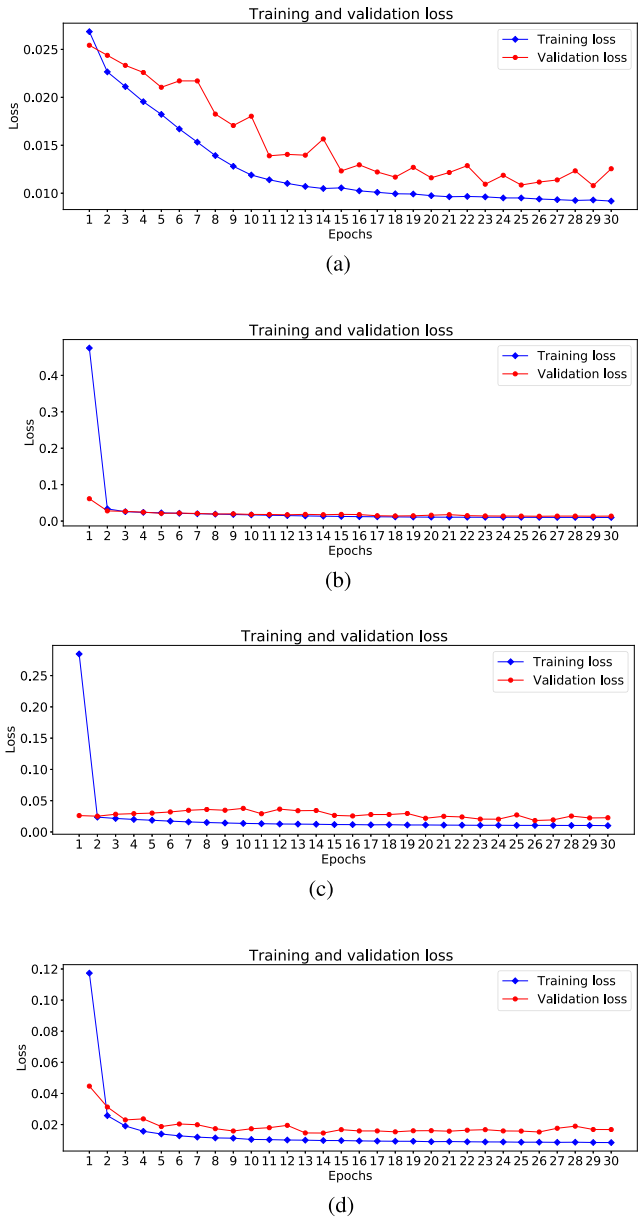


Fig. 8. The measured evaluation loss indexes comparison for different algorithms with (a) GRU (b) LSTM (c) BiLSTM (d) Ours and RMSE loss illustration.

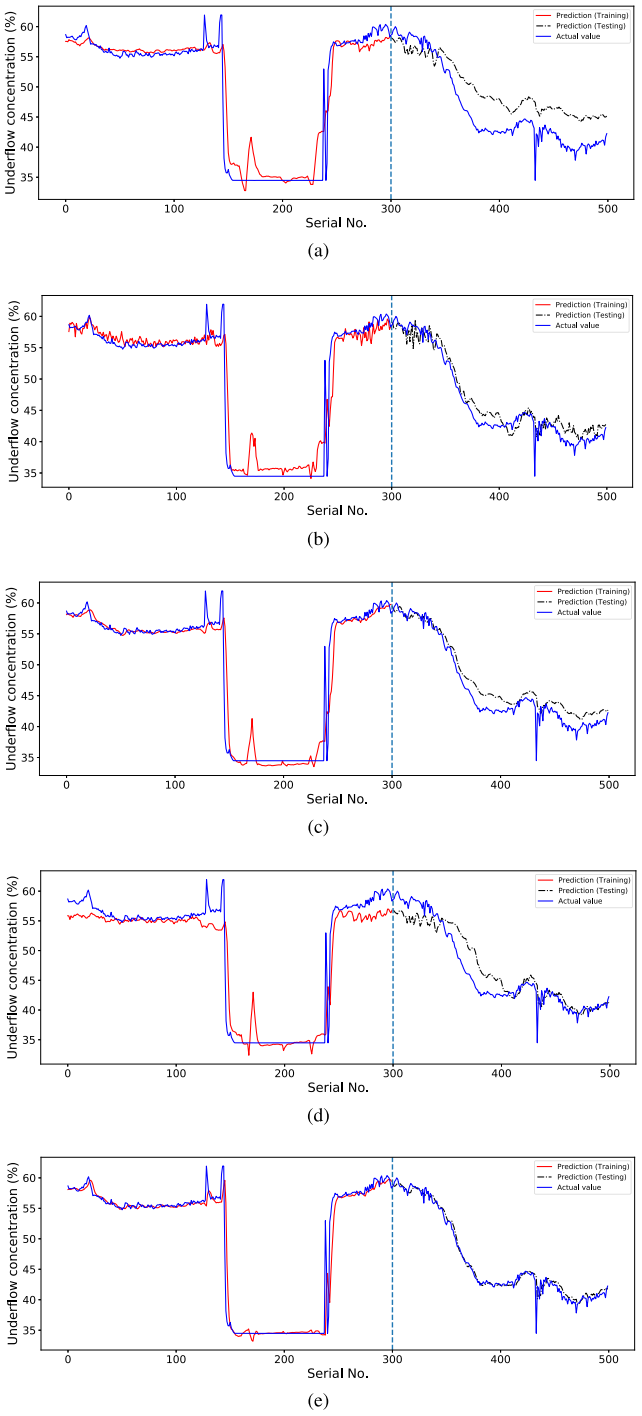


Fig. 9. The underflow prediction comparison with competitive algorithms. (a) ANN; (b) LSTM (c) BiLSTM; (d) GRU; (e) Ours.

respectively. Our DualLSTM 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 achieved in this proposed architecture.

5. Conclusion

This work provides a high-efficiency key-quality underflow prediction model for the deep cone thickener system to address the challenges of underflow prediction in industrial CTS plants. In the industrial case, compared with other competitive algorithms, the proposed DualLSTM proposed a new architecture with feedforward and inverse dual BiLSTM, and the final MAE and RMSE are 0.2234 and 0.1731, respectively. The proposed algorithms additionally take into account time-invariant elements and improved the prediction performance by a large margin. DCT's underflow concentration prediction assesses the presented algorithms' industrial feasibility and effectiveness. The DualLSTM can be extended to implement some other similar process industrial domains such as nonferrous, process manufacturing, and chemical production. For further research, we will continue to investigate this model and apply it to optimal control, as well as establish a comprehensive intelligent integrated platform to promote the development of the smart mining and paste-filling industry.

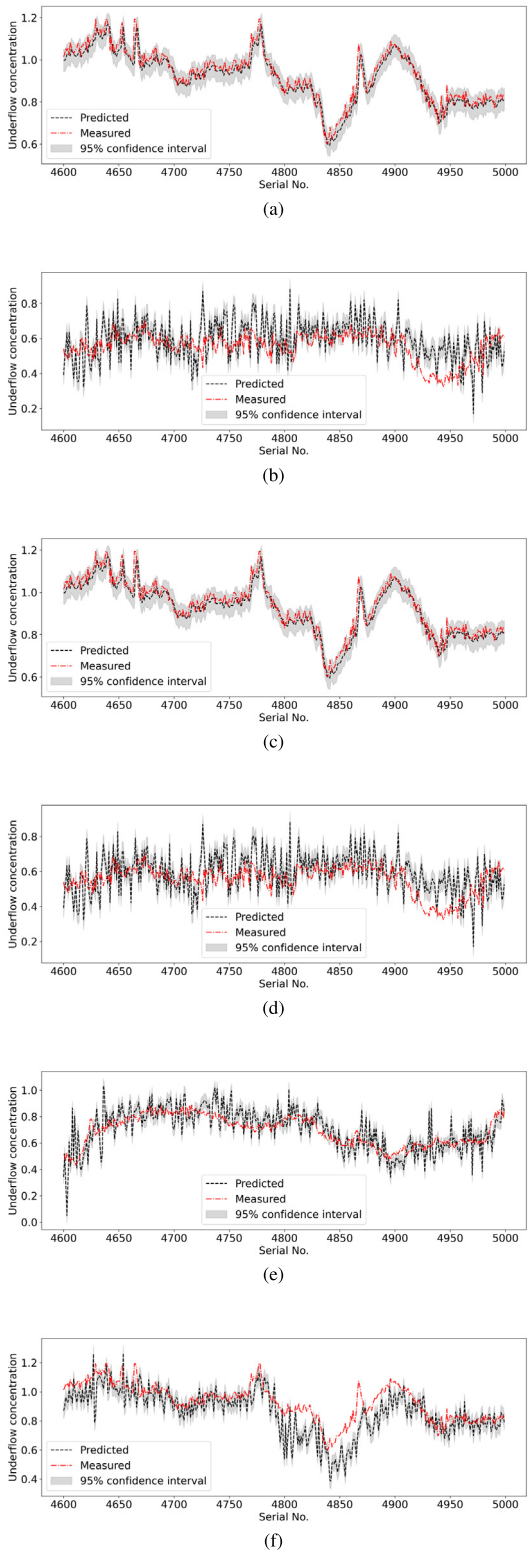


Fig. 10. The proposed DualLSTM underflow prediction with competitive algorithms BiLSTM comparison. (a) and (d) are Training datasets based on DualLSTM and BiLSTM; (b) and (e) are Validation datasets; (c) and (f) Testing datasets of DualLSTM and BiLSTM respectively.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the Italian Ministry of Education, University and Research through the Project "Department of Excellence LIS4.0-Lightweight and Smart Structures for Industry 4.0" and in part by the Horizon Marie Skłodowska-Curie Actions program (101073037) and China Scholarship Council (202006370101).

References

Barua, R., & Sharma, A. K. (2022). Dynamic Black Litterman portfolios with views derived via CNN-BiLSTM predictions. *Finance Research Letters*, *49*, Article 103111.

Chen, W., Chen, W., Liu, H., Wang, Y., Bi, C., & Gu, Y. (2022). A RUL prediction method of small sample equipment based on DCNN-BiLSTM and domain adaptation. *Mathematics*, *10*(7), 1022.

Chen, J., Li, X., Xiao, Y., Chen, H., & Zhao, Y. (2022). FRA-LSTM: A vessel trajectory prediction method based on fusion of the forward and reverse sub-network. arXiv preprint arXiv:2201.07606.

Chen, L., Zhang, W., & Ye, H. (2022). Accurate workload prediction for edge data centers: Savitzky–Golay filter, CNN and BiLSTM with attention mechanism. *Applied Intelligence*, *52*(11), 13027–13042.

Chen, Y., Zhang, D., & Zhang, W.-a. (2022). MSWR-LRCN: a new deep learning approach to remaining useful life estimation of bearings. *Control Engineering Practice*, *118*, Article 104969.

Fang, C., He, D., Li, K., Liu, Y., & Wang, F. (2022). Image-based thickener mud layer height prediction with attention mechanism-based CNN. *ISA Transactions*, *128*, 677–689.

Geng, J., Yang, C., Li, Y., Lan, L., & Luo, Q. (2022). MPA-RNN: a novel attention-based recurrent neural networks for total nitrogen prediction. *IEEE Transactions on Industrial Informatics*, *18*(10), 6516–6525.

Goh, H. H., He, R., Zhang, D., Liu, H., Dai, W., Lim, C. S., et al. (2022). A multimodal approach to chaotic renewable energy prediction using meteorological and historical information. *Applied Soft Computing*, *118*, Article 108487.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hou, X., Sun, K., Shen, L., & Qiu, G. (2019). Improving variational autoencoder with deep feature consistent and generative adversarial training. *Neurocomputing*, *341*, 183–194.

Hu, K., Wang, Y., Li, W., & Wang, L. (2022). CNN-BiLSTM enabled prediction on molten pool width for thin-walled part fabrication using Laser Directed Energy Deposition. *Journal of Manufacturing Processes*, *78*, 32–45.

Huan, W., Ting, L., Yuning, C., & Aixiang, W. (2019). Underflow concentration prediction model of deep-cone thickener based on data-driven. *The Journal of China Universities of Posts and Telecommunications*, *26*(6), 63.

Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. *Advances in Neural Information Processing Systems*, *28*.

Jiang, Y., & Yin, S. (2018). Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT. *IEEE Transactions on Industrial Informatics*, *15*(5), 2849–2858.

Jing, Y., Hu, H., Guo, S., Wang, X., & Chen, F. (2020). Short-term prediction of urban rail transit passenger flow in external passenger transport hub based on LSTM-LGB-DRS. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4611–4621.

Lei, Y., Karimi, H. R., Cen, L., Chen, X., & Xie, Y. (2021). Processes soft modeling based on stacked autoencoders and wavelet extreme learning machine for aluminum plant-wide application. *Control Engineering Practice*, *108*, Article 104706.

Lei, Y., Karimi, H. R., & Chen, X. (2022). A novel self-supervised deep LSTM network for industrial temperature prediction in aluminum processes application. *Neurocomputing*, *502*, 177–185.

Li, C., Tang, G., Xue, X., Saeed, A., & Hu, X. (2019). Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Transactions on Sustainable Energy*, *11*(3), 1370–1380.

Liang, T., Chai, C., Sun, H., & Tan, J. (2022). Wind speed prediction based on multi-variable Capsnet-BiLSTM-MOHHO for WPCCC. *Energy*, *250*, Article 123761.

Lin, W., Zhang, B., Li, H., & Lu, R. (2022). Multi-step prediction of photovoltaic power based on two-stage decomposition and BILSTM. *Neurocomputing*, *504*, 56–67.

Liu, Y., Gong, C., Yang, L., & Chen, Y. (2020). DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Systems with Applications*, *143*, Article 113082.

Liu, C., Wang, K., Wang, Y., Xie, S., & Yang, C. (2021). Deep nonlinear dynamic feature extraction for quality prediction based on spatiotemporal neighborhood preserving SAE. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–10.

Liu, C., Wang, K., Wang, Y., & Yuan, X. (2021). Learning deep multimanifold structure feature representation for quality prediction with an industrial application. *IEEE Transactions on Industrial Informatics*, *18*(9), 5849–5858.

Prastyo, D. D., Nabila, F. S., Lee, M. H., Suhermi, N., & Fam, S.-F. (2019). VAR and GSTAR-based feature selection in support vector regression for multivariate spatio-temporal forecasting. In *Soft computing in data science: 4th international conference, SCDS 2018, Bangkok, Thailand, August 15-16, 2018, proceedings 4* (pp. 46–57). Springer.

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. arXiv preprint arXiv:1704.02971.

Ren, L., Wang, T., Laili, Y., & Zhang, L. (2021). A data-driven self-supervised LSTM-DeepFM model for industrial soft sensor. *IEEE Transactions on Industrial Informatics*, *18*(9), 5859–5869.

Shuaiyi, L., Wang, K., Zhang, L., & Wang, B. (2022). Global-local integration for GNN-based anomalous device state detection in industrial control systems. *Expert Systems with Applications*, *209*, Article 118345.

Sun, S., Sun, J., Wang, Z., Zhou, Z., & Cai, W. (2022). Prediction of battery SOH by CNN-BiLSTM network fused with attention mechanism. *Energies*, *15*(12), 4428.

Takács, I., Patry, G. G., & Nolasco, D. (1991). A dynamic model of the clarification-thickening process. *Water Research*, *25*(10), 1263–1271.

Tan, C. K., Setiawan, R., Bao, J., & Bickert, G. (2015). Studies on parameter estimation and model predictive control of paste thickeners. *Journal of Process Control*, *28*, 1–8.

Wang, P., Yin, Y., Deng, X., Bo, Y., & Shao, W. (2022). Semi-supervised echo state network with temporal–spatial graph regularization for dynamic soft sensor modeling of industrial processes. *ISA Transactions*, *130*, 306–315.

Xiao, D., Xie, H., Jiang, L., Le, B. T., Wang, J., Liu, C., et al. (2020). Research on a method for predicting the underflow concentration of a thickener based on the hybrid model. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 13–26.

Xie, W., Wang, J., Xing, C., Guo, S., Guo, M., & Zhu, L. (2020). Variational autoencoder bidirectional long and short-term memory neural network soft-sensor model based on batch training strategy. *IEEE Transactions on Industrial Informatics*, *17*(8), 5325–5334.

Yuan, Z., Hu, J., Wu, D., & Ban, X. (2020). A dual-attention recurrent neural network method for deep cone thickener underflow concentration prediction. *Sensors*, *20*(5), 1260.

Yuan, X., Li, L., Shardt, Y. A., Wang, Y., & Yang, C. (2020). Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, *68*(5), 4404–4414.

Yuan, X., Qi, S., Wang, Y., & Xia, H. (2020). A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. *Control Engineering Practice*, *104*, Article 104614.

Zhang, C., Ma, H., Hua, L., Sun, W., Nazir, M. S., & Peng, T. (2022). An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. *Energy*, *254*, Article 124250.

Zhang, B., Xiong, D., & Su, J. (2018). Neural machine translation with deep attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(1), 154–163.

Zhu, X., Damarla, S. K., Hao, K., Huang, B., Chen, H., & Hua, Y. (2023). ConvLSTM and self-attention aided canonical correlation analysis for multi-output soft sensor modeling. *IEEE Transactions on Instrumentation and Measurement*.