Heliyon 10 (2024) e32901

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Prediction of steelmaking process variables using K-medoids and a time-aware LSTM network

Ruixuan Zheng [a], Yanping Bao [a,*], Lihua Zhao [b,**], Lidong Xing [a]

[a] State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing, Beijing, 100083, China
[b] School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, Beijing, 100083, China

ARTICLE INFO

*Keywords:*
Multi-model soft sensor modeling
K-medoids
Time-aware network
LSTM
Process industry
Variable prediction

ABSTRACT

A new method is required to address the challenge of predicting process parameters in high-temperature, high-pressure industrial processes. This study proposes a multi-model Long Short-Term Memory (LSTM) network prediction algorithm with irregular time interval sequences to predict the silicon yield in converter steelmaking. The experimental results demonstrate that this algorithm performs better than comparable neural network models in classifying high-dimensional, redundant industrial production data with noise and outliers. The algorithm is evaluated using data from a steel plant. The proposed algorithm has lower errors for predicting the alloy yield than other neural network models. An average mean absolute error (*MAE*) of less than 0.01 confirms the algorithm's feasibility and practicality.

## 1. Introduction

Numerous cutting-edge technologies have been developed in recent years to monitor, regulate, and optimize the process parameters of industrial production [1–4]. However, the majority of these techniques monitor vital quality indicators in real time, such as the levels of chemical composition, gas concentrations, and extremely high temperatures. More importantly, challenging measurement settings, expensive analyzers, and measurement delays make it challenging for hardware sensors to obtain crucial quality characteristics in real time in industrial operations [5–7].

Soft sensors are frequently used to forecast crucial quality characteristics that are challenging to monitor in real time using conventional sensors [8]. This method establishes a relationship between difficult-to-measure and quantifiable variables to predict quality variables that cannot be measured in real time. This strategy is used to monitor, control, and optimize industrial production processes [9,10].

Two categories of soft sensor models are mechanism and data-driven models [11]. The former uses the law of conservation of mass, kinetics, thermodynamics, material and heat balance, and chemical reactions to derive mathematical models [12]. The latter utilizes data samples to establish mathematical models [13]. Data-driven models have been preferentially used in industrial production processes due to their high accuracy, stability, and prediction speed.

Numerous data-driven models have been developed to predict process parameters in steel industry production, indicating that appropriate data-driven models achieve high prediction accuracy. Zhou et al. developed a unique knee-guided predictive evolutionary

* Corresponding author.
** Corresponding author.
*E-mail addresses:* baoyp@ustb.edu.cn, byp976431@163.com (Y. Bao), zhaolihua@metall.ustb.edu.cn (L. Zhao).

https://doi.org/10.1016/j.heliyon.2024.e32901
Received 28 May 2024; Received in revised form 11 June 2024; Accepted 11 June 2024
Available online 14 June 2024

algorithm that preserves non-dominated solutions around the knee and in border areas. This method reduces the cost of maintaining a large and diverse population during the evolution and has been used for dynamic coil order allocation in the steel industry [14]. Ji et al. proposed a hybrid machine learning and genetic algorithm-based method (MGH) to predict deviations in the width of a hot-rolled steel strip [15]. Jiang et al. used a data-driven model based on multi-level feature fusion to categorize the silicon content variations in a blast furnace online [16]. Feng et al. incorporated an adaptive fuzzy proportional-integral-derivative (PID) controller with a nonlinear compensation term. This method improved the system's interference resistance and achieved stable control of the crystallizer liquid level during casting [17]. Yuan et al. created a Case-Based Reasoning model based on the Heat Transfer Calculation (CBR-HTC) by integrating mechanism and data-driven models. The approach was used to predict the temperature of molten steel at the end of a ladle furnace (LF) and significantly improved the prediction accuracy [18]. Wang et al. used the explicit finite difference method to create a continuous prediction model for the carbon content of a 120 T converter based on the three-stage decarburization theory. The model's prediction accuracy for an end-point carbon content with an error of less than or equal to 0.02 % was 85 % [19]. Wu et al. improved the topology of a neural network and predicted the Charpy V-notch impact energy of low-carbon steel. They compared the performances of limit learning machines, three-layer neural networks, and deep neural networks (DNNs) [20]. Xin et al. developed a hybrid model based on expert control and a DNN to predict the temperature of molten steel in an LF. The hybrid model's accuracy was 99.4 % in the temperature range of −5 to 5 °C [21]. Zhou et al. used a data-driven approach to determine the relationship between an anomalous production process and product quality. A convolutional neural network and an autoencoder model were developed to replace manual visual screening for identifying inclusions in cord steel automatically. Production data were used to verify the effectiveness of the automatic anomaly detection method. The recall rate was 93.06 % [22]. Kim et al. predicted the strength and toughness of thick steel plates using a Bayesian neural network (BNN) model and evaluated uncertainty. They successfully employed the technique using a steelmaking production data set of Pohang Iron and Steel Company (POSCO) firms [23].

Researchers have provided many reliable solutions for predicting process variables in the steel industry. However, more research is required due to the highly nonlinear and stochastic nature of the production process [24–26]. Fig. 1 shows the Basic Oxygen Furnace (BOF) steelmaking steps and the addition of ferroalloys. The alloy yield is a key production indicator because it determines the amount of ferroalloy to be added and affects the steel quality. The accurate prediction of the alloy yield is crucial to ensure the high quality of steel products and low production costs.

The production of ferroalloys consumes significant amounts of energy, requiring from 4000 to 10000 kWh of electricity per ton of ferroalloy. The global annual production of ferroalloys exceeds 30 million tons, with 90 % consumed by steelmaking enterprises. Reducing ferroalloy waste and improving utilization efficiency can significantly reduce energy and resource consumption [27]. The alloy recovery rate is a key parameter in the steelmaking process and determines the amount of ferroalloy to be added. The alloy yield represents the absorption rate after adding the alloy to the molten steel. This variable cannot be directly measured and must be predicted. A 1 % prediction deviation results in a 1 % waste of ferroalloys and an increase in energy consumption. Equation (1) is used to calculate the alloy yield. The *Si* yield is the difference between the composition of the steel liquid after the addition of ferroalloys and the final composition. Improving the prediction accuracy of the yield rate ensures that the composition of the steel products is in an appropriate and narrow range. In the steelmaking process, there are typically only one to two opportunities to add ferroalloys; thus, the accuracy of predicting the alloy yield must be high. The final steel composition steel, the oxidizability of the molten steel, and the

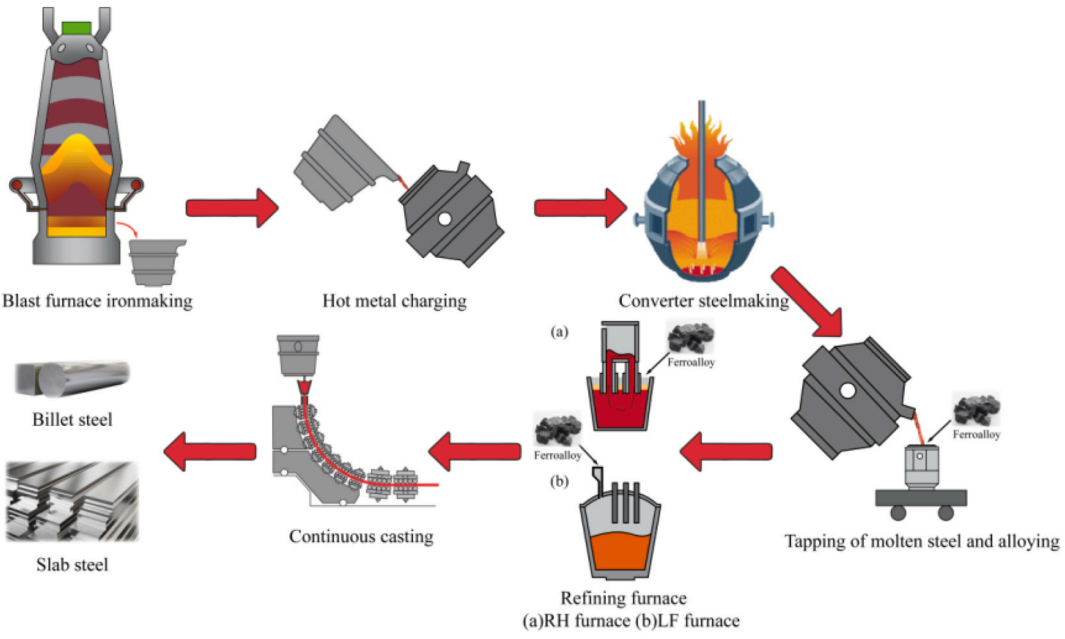


**Fig. 1.** Schematic diagram of the traditional steel manufacturing.

ambient temperature significantly affect the alloy yield. The sub-lance system has been widely used recently to obtain contact measurements during converter steelmaking. A probe is immersed in the high-temperature molten pool to measure the oxygen and carbon contents, the molten steel temperature, and other parameters. A data-driven model can be established using data obtained from the sub-lance system to predict the alloy yield. Therefore, we use data obtained during converter steelmaking to establish a prediction model for the *Si* yield.

$$Y_{e_x} = \frac{M_{steel}\left(C_{e_x} - C'_{e_x}\right)}{\sum_{i=1}^{N}\left(M_{F_i} C_{F_{i,e_x}} / 100\right)} \times 100\% \tag{1}$$

where $Y_{e_x}$ is the yield of element $X$ (%), $M_{steel}$ is the weight of molten steel (kg), $C_{e_x}$ is the $X$ element content in the molten steel after alloying (%), $C'_{e_x}$ is the $X$ element content in the molten steel before alloying (%), $M_{F_i}$ is the amount of the $i^{\text{-th}}$ ferroalloy (kg), and $C_{F_{i,e_x}}$ is the proportion of element $X$ in the $i^{\text{-th}}$ ferroalloy (%).

The primary problems addressed in this study are as follows.

1 ***Different working conditions:*** Converter steelmaking occurs under various working conditions, such as the different oxygen-blowing systems and slag formation routes. The raw material and the process conditions affect the outcome. A single model with one set of parameters is unsuitable to predict the outcome of industrial processes under various operating conditions. The high complexity and redundancy of historical data obtained from iron and steelmaking processes make it challenging to categorize and simulate working conditions;
2 ***Time series problem:*** Converter steelmaking occurs continuously for 24 h in the absence of accidents. The production cycle is similar for each heating stage (generally 20 min). Therefore, converter production data are time-series data with specific intervals. It is difficult to determine the relationship between variables in time-series data.
3 ***Complex process dynamics and black box model:*** Converter steelmaking is a complex process with physical and chemical changes, such as oxygen blowing and stirring; thus, the mechanism model is too complex for process control. A data-driven model is a black-box model, and some physical and chemical laws cannot be explained. Moreover, a data-driven model depends on data quality. Process control cannot be performed when the input variables are highly uncertain. Therefore, data-driven and mechanism models do not meet industrial requirements.

Due to these difficulties, we propose a hybrid model consisting of multi-mode mechanism and data-driven models for predicting the process parameters of converter steelmaking. Representative features are collected from high-dimensional redundant industrial data. The data are categorized into different working conditions using clustering to increase the flexibility of the prediction model. A Time-aware Long-Short-Term Memory (T-LSTM) network and different prediction models for the alloy yield in converter steelmaking are established. The T-LSTM neural network is a variant of the LSTM that considers the time interval in time-series data. It distinguishes between short-term and long-term memory and modifies the effect of the short-term memory depending on the gap between the input variables' values. Data with a long time interval have a negligible influence, whereas data with a short time interval have a large influence on the model.

The proposed T-LSTM neural network can deal with various working conditions during converter steelmaking and considers the time interval in time-series data. The lifetime of a converter is limited, and the refractory materials in the converter must be replaced periodically. In addition, the converter must be shut down for maintenance, scheduling, or safety reasons. These conditions change the dynamics and other technical conditions in the furnace. The T-LSTM neural network reduces the influence of older historical data and focuses more on recent historical data, improving the accuracy of the prediction model.

The novelty of this work is to predict difficult-to-measure process parameters of continuous industrial processes (such as steelmaking) using time-series data and machine learning methods.

The remainder of this paper is organized as follows. Section 2 presents the clustering algorithms and the T-LSTM network. Section 3 introduces the T-LSTM network with K-medoids. Section 4 provides the results and discussion, and Section 5 concludes this work and suggests future strategies to improve the proposed approach.

## 2. Clustering algorithm and T-LSTM network

### 2.1. K-means and K-medoids

Clustering has been widely used in several fields, including image segmentation, machine vision, speech recognition, information retrieval, and industrial processes. It has been used in geological research [28] to identify the lithology of ores, especially for crucial solid metal mineral resources. Cluster analysis is also widely used in the medical field [29]. Researchers can use it to distinguish different working conditions in industrial processes [30,31].

The most common clustering algorithms are K-means and K-medoids. The former is the simplest and most widely used clustering method. Its primary advantage is its speed and interpretability. However, this algorithm is highly sensitive to outliers, which affect the cluster's mean value, resulting in a large deviation between the mean and the majority of the data in the cluster. Therefore, this method is prone to a local optimal solution. Converter steelmaking data typically have outliers. The K-medoids algorithm prevents the

influence of outliers, is robust to noise, and has a high calculation speed. Therefore, we compare the accuracy of the K-means and K-medoids methods for classifying converter steelmaking data.

### 2.2. Time-aware long short-term memory network

The LSTM neural network is a variant of the Recurrent Neural Network (RNN). Its advantage over the RNN is that it is not prone to gradient disappearance and explosion for long time-series data describing adaptive processing [32]. Fig. 2 displays the computational flow of the LSTM network. It contains cells and gates.

The calculating procedure is as follows. The first step is to select data and construct a candidate vector $\widetilde{c}_t$ using the input gate $i_t$, where:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ \widetilde{c}_t &= tan\, h(W_c x_t + U_c h_{t-1} + b_c) \end{aligned} \tag{2}$$

The forget gate $f_t$ determines which outdated information from the old cell vector should be deleted. This process is expressed as follows:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{3}$$

The new cell vector $c_t$ is updated by including the new information $\widetilde{c}_t$ and deleting outdated information $c^*_{t-1}$:

$$c_t = f_t \cdot c^*_{t-1} + i_t \cdot \widetilde{c}_t \tag{4}$$

where $\cdot$ denotes the pointwise multiplication of two vectors.

The hidden vector $h_t$ and the output vector $\widehat{y}_t$ are computed by the output gate $o_t$ after filtering the new cell vector. Equation (5) describes this process:

$$\begin{aligned} o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \cdot tan\, h(c_t) \\ \widehat{y}_t &= \sigma\left(V h_t + b_y\right) \end{aligned} \tag{5}$$

where $x_t$ represents the current input equivalent weight matrices $W_i$, $W_c$, $W_f$, $W_o$, and the bias vectors of $V$ are denoted as $b_i$, $b_c$, $b_f$, $b_o$.

Equations (2)–(5) describe a standard LSTM unit. It is assumed that the time interval between the sequences is the same. However, furnace maintenance and replacement of refractory materials are required during converter steelmaking; thus, the converter is shut down periodically. Therefore, the time interval between the heating stages may range from hours to months. The T-LSTM considers this time interval. The following contents are added to the LSTM to create the T-LSTM [33]:

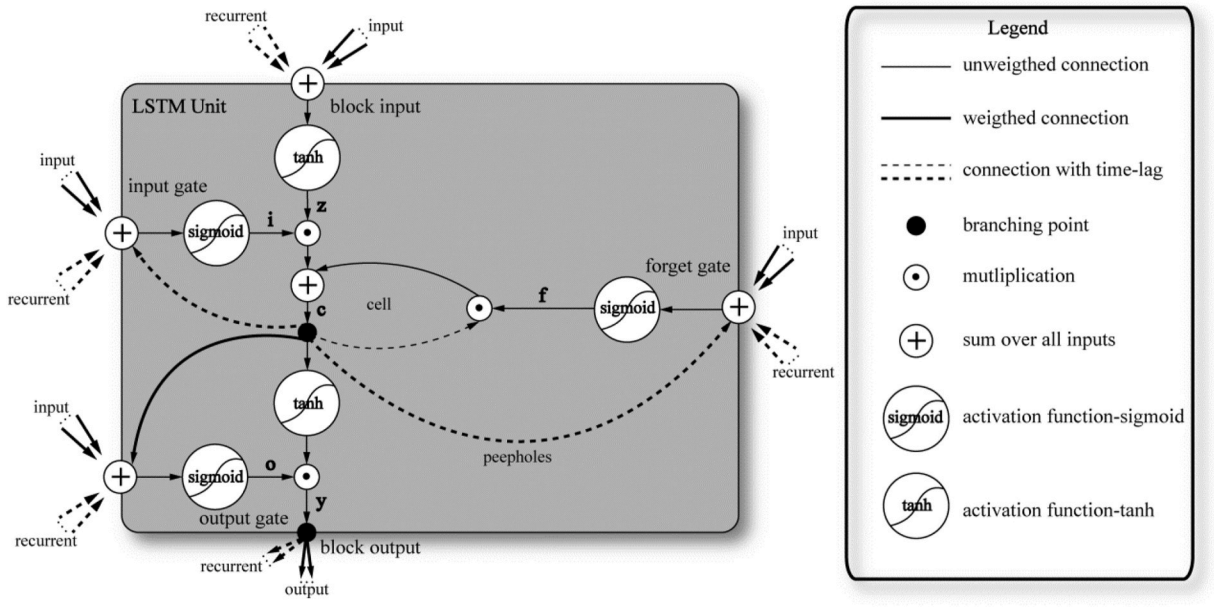


**Fig. 2.** The structure of the LSTM network.

$$\begin{aligned} c^s_{t-1} &= \tanh(W_d c_{t-1} + b_d) \\ \widehat{c}^s_{t-1} &= c^s_{t-1} \cdot g(\Delta t) \\ c^T_{t-1} &= c_{t-1} - c^s_{t-1} \\ c^*_{t-1} &= c^T_{t-1} - \widehat{c}^s_{t-1} \end{aligned} \tag{6}$$

where $c^s_{t-1}$ represents the short-term memory, $\widehat{c}^s_{t-1}$ represents the discounted short-term memory, $c^T_{t-1}$ represents the long-term memory, $c^*_{t-1}$ represents the adjusted memory, $b_d$ denotes the network bias vector of the subspace, $W_d$ represent the network weight matrices of the subspace, $\Delta t$ is the time difference between $x_{t-1}$ and $x_t$, and $g(\cdot)$ is a heuristic decay function whose value decreases with increasing $\Delta t$ value.

## 3. Time-aware long short-term memory with K-medoids

Time-series models are useful in the steelmaking industry. They can predict key metrics, such as production output, energy consumption, and market demand, enabling better planning and decision-making. Researchers have applied time-series models to predict the process parameters of BOF steelmaking and achieved satisfactory results [34–37]. Prediction models based on first principles or machine learning have become relatively mature [38,39] but they have numerous shortcomings, such as dealing with time-series problems. Unlike conventional machine learning approaches, the LSTM can handle time-series data. The T-LSTM has been extensively employed to predict critical parameters in different engineering domains [40,41]. This model handles irregular time intervals using subspace decomposition. It uses a time decay function to reduce the amount of historical data that reflect the passage of time. K-medoids are used in the T-LSTM network to create a novel soft sensor: the T-LSTM with K-medoids. This sensor identifies different working conditions using process variables and predicts the alloy yield in converter steelmaking. Fig. 3 shows the T-LSTM with K-medoids network structure.

### 3.1. Cleaning of BOF steelmaking data

The alloy yield is a crucial output metric in steelmaking. We used TensorFlow 2.0 to assess the proposed method for predicting the alloy yield. The samples were obtained from industry data derived from several sensors. The data require preprocessing due to a substantial amount of irrelevant or duplicate information.

The dataset was collected from a steelmaking plant in China. Data collection was conducted by the programmable logic controller (PLC) system and various instruments and devices. Data transmission was facilitated by the manufacturing execution system (MES) production system, and the data were stored in an Oracle database. The data were normalized. The changes in the parameter values

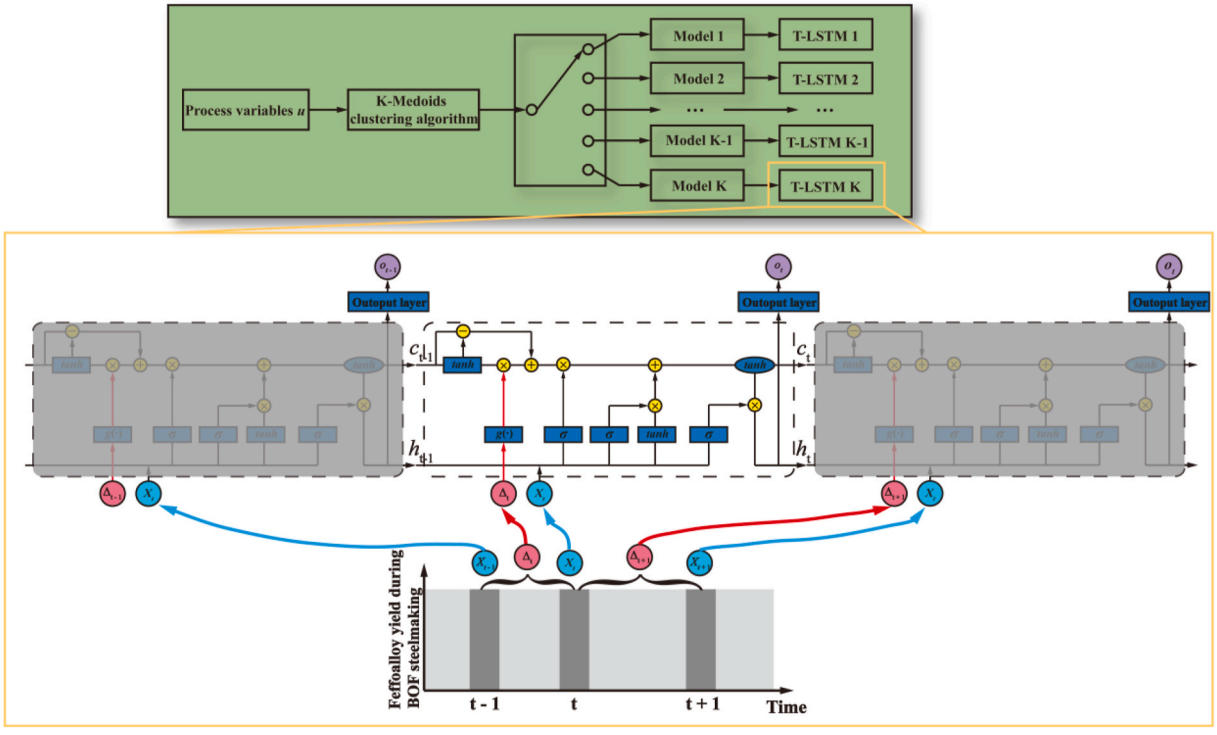


**Fig. 3.** The structure of the T-LSTM network with K-medoids.

during one week are shown in Fig. 4. The data have many outliers. A boxplot was used to detect outliers, as shown in Fig. 5. The data below Q1-1.5IQR and above Q3+1.5IQR were considered outliers.

Boxplots are widely used for detecting outliers in sample data and assessing the skewness and the tail weight of the data. The data distribution and the descriptive statistics after cleaning are shown in Fig. 6 and Table 1. The loss of molten steel during steelmaking was −20.9–8.4 t, the weight of the molten steel was 46.2–55.1 t, and the temperature was 1601–1666 °C. The final carbon and manganese contents were, respectively, 0.03–0.1 % and 0.04–0.17 %, and the silicon yield was 68–100 %. Deleting outliers can affect the integrity of the samples. It is necessary to replace the outliers with reasonable values that conform to the pattern. We used the average of neighboring points to replace outliers (Equation (7)). After data processing, 80 % of the data were used as the training set, and the remaining 20 % was used as the test set.

$$\overline{X}(n,p)=\frac{X(n-1,p)+X(n+1,p)}{2} \tag{7}$$

### 3.2. Correlation analysis

The process variables were screened by the reaction mechanism, and Pearson correlation analysis and significance tests were used to analyze the correlation between the process variables and the alloy yield [42]. Equation (8) describes the calculation of the Pearson correlation coefficient. The ranking of the input variables based on their effect on the alloy yield is $U_2 > U_3 > U_5 > U_1 > U_4$ (Fig. 7).

$$R_{xy}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}} \tag{8}$$

A $p$-value of less than 0.01 indicated high significance, and a value of less than 0.05 indicated significance. The results of the correlation analysis are listed in Table 2. The $p$-values between $Y$ and $U_1$, $U_2$, $U_3$, $U_4$, and $U_5$ are less than 0.01, indicating high significance between the silicon yield and the loss of liquid steel during steelmaking, the final carbon and manganese contents, and the temperature and weight of liquid steel. The normalized values of $U_1$, $U_2$, $U_3$, $U_4$, and $U_5$ were used as input variables of the T-LSTM with the K-medoids model.

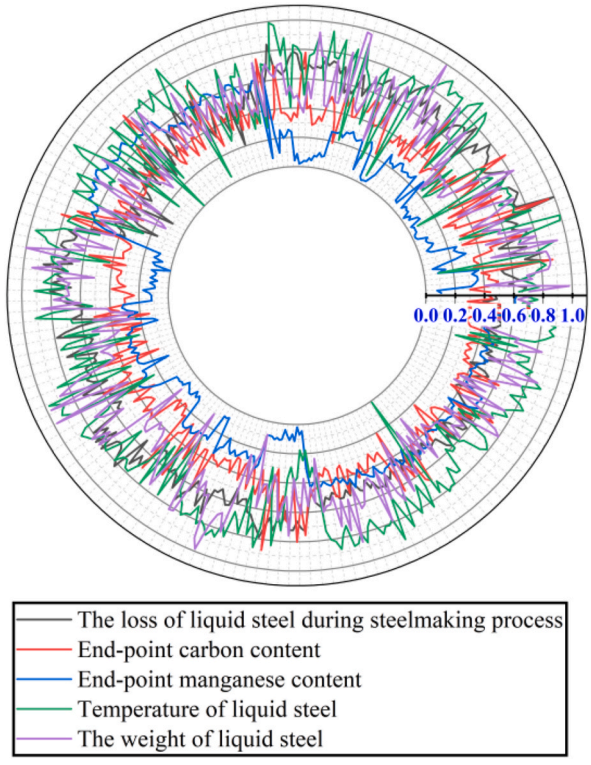


**Fig. 4.** Changes in the steelmaking process variables during one week.

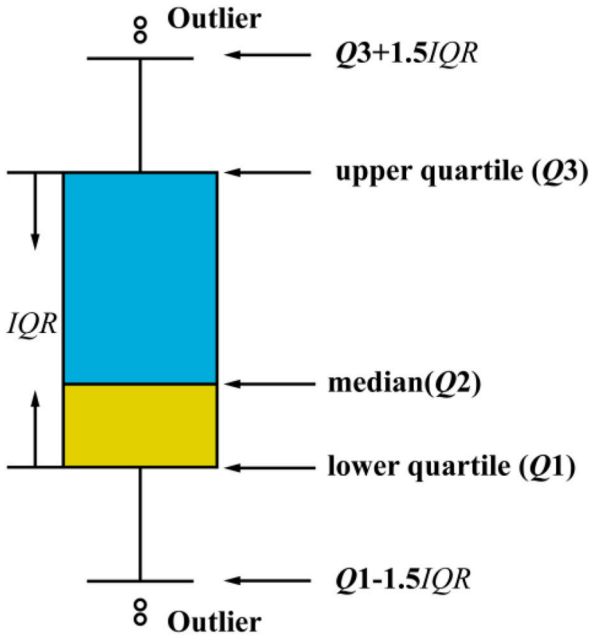


**Fig. 5.** Outlier detection using a boxplot.

### 3.3. ADF unit root and white noise tests

The time series must be stable over time, i.e., the mean and variance should not change significantly. Therefore, we assessed the stability of the data [43] using the augmented Dickey-Fuller (ADF) test, an extension of the Dickey-Fuller test, to determine the stationarity of the time series. It tests whether a unit root is present in the time series. If not, the time series is stable; otherwise, the time series is unstable.

The autoregressive equation is expressed as follows [44]:

$$y_t = \gamma y_{t-1} + \mu + \varepsilon_t \tag{9}$$

where $y_t$ is the sequence value of the data at time $t$, $y_{t-1}$ is the sequence value of the data at time $t$-1, $\gamma$ is the lag coefficient, $\mu$ is the lag constant, and $\varepsilon_t$ is the error term at time $t$.

If $\gamma = 1$, a unit root exists, which is the $H_0$ hypothesis of the ADF test. The significance levels are 10 %, 5 %, and 1 % at confidence intervals of 90 %, 95 %, and 99 %, respectively. White noise refers to data variability that cannot be explained by regression models. Therefore, a white noise test, the Ljung-Box test, was conducted. It tests for the existence of lag correlation in a time series. The expression is as follows [45]:

$$Q_{(m)} = n(n+2)\sum_{i=1}^{m}\frac{\widehat{p}_i}{n-i} \tag{10}$$

where $n$ is the number of data samples, $m$ is a randomly chosen natural number, and $\widehat{p}_i$ is the autocorrelation coefficient of the $i$th-order lag.

The following alternative and null hypotheses are used in the Ljung-Box test:

■ **$H_0$:** The data are independently distributed, i.e., the correlation coefficient is 0.■

If the null hypothesis is true, $Q_{(m)}$ follows a chi-square distribution with $m$ degrees of freedom. If the null hypothesis is accepted, the data sequence contains white noise; otherwise, it is assumed that the sequences are correlated.

The results of the ADF and Ljung-Box tests are listed in Table 3. The ADF test results show that the test statistic is considerably smaller than the three crucial values, and the $p$-value is close to zero. Thus, the null hypothesis is rejected, and the time-series data are stable. The null hypothesis is also rejected in the Ljung-Box test because the $p$-value is substantially lower than the threshold of 0.05, indicating that the time series data exhibit temporal autocorrelation. The results show that the time-series data of the converter alloy yield meets the requirements for using the T-LSTM with the K-medoids model.

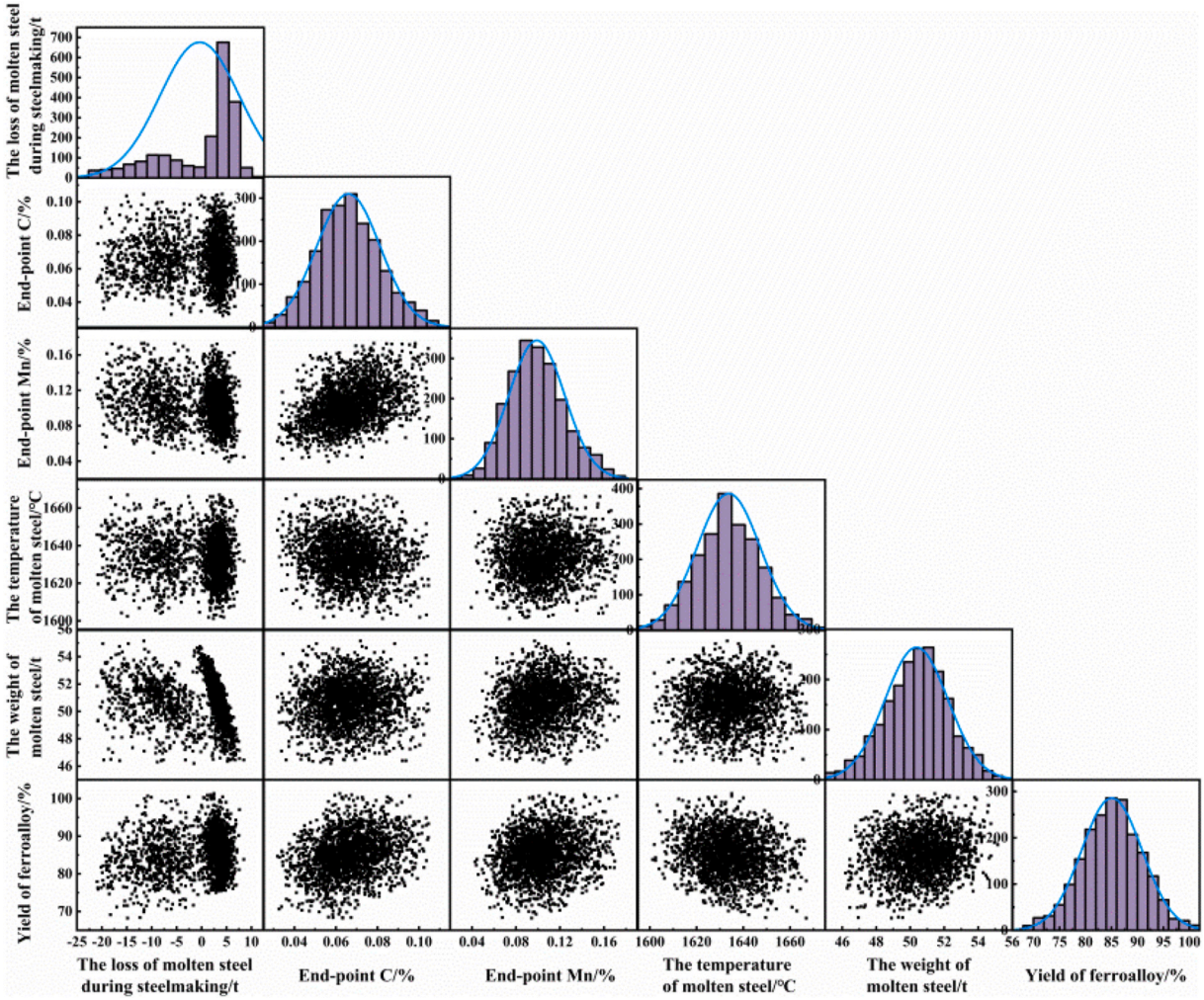


**Fig. 6.** Scatter plot of $U_1$, $U_2$, $U_3$, $U_4$, and $U_5$ versus $Y$.

**Table 1**
List of variables used in steelmaking.

| Input variable | Description | Range | Mean |
|---|---|---|---|
| $U_1$ | The loss of liquid steel during steelmaking process [t] | [-20.9,8.4] | −1.0 |
| $U_2$ | End-point carbon content [%] | [0.03,0.1] | 0.06 |
| $U_3$ | End-point manganese content [%] | [0.04,0.17] | 0.1 |
| $U_4$ | Temperature of liquid steel [°C] | [1601,1666] | 1634 |
| $U_5$ | The weight of liquid steel [t] | [46.2,55.1] | 50.6 |
| $Y$ | Yield of alloy [%] | [68,100] | 85 |

## 4. Experimental results and discussion

### 4.1. Evaluation of cluster models using the confusion matrix

The accuracy of the K-means and K-medoids methods for cluster analysis was compared. The number of clusters must be known before applying both methods [46]. We used the Calinski-Harabasz scores and the silhouette coefficient to determine the $K$ value. The Calinski-Harabasz score is defined as follows [29]:

$$S_{CH} = \frac{N-K}{K-1} \frac{\sum_{i=1}^{K} \|c_i - \overline{U}\|_2^2}{\sum_{i=1}^{K} \sum_{u_j \in cl_i} \|c_i - u_j\|_2^2} \tag{11}$$

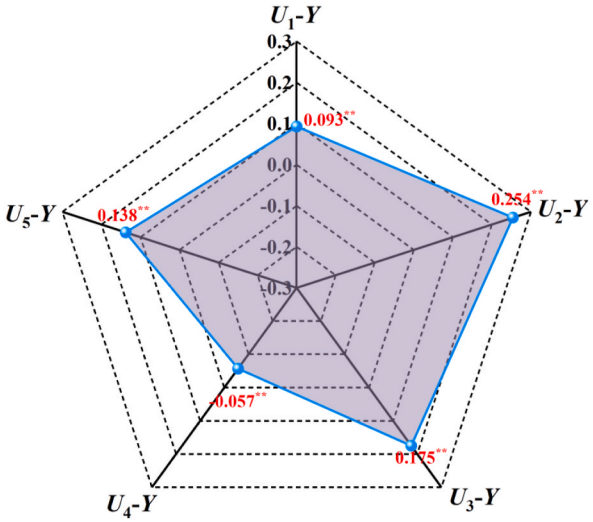


**Fig. 7.** Correlation coefficient matrix. Where, $U_1$ denotes the loss of steel liquid during the steelmaking process, t; $U_2$ denotes end-point carbon content, %; $U_3$ denotes end-point manganese content, %; $U_4$ denotes temperature of liquid steel, °C; $U_5$ denotes the weight of liquid steel, t; $Y$ denotes yield of alloy, %.

**Table 2**
Results of correlation analysis between input and output variables.

| | $U_1$–$Y$ | $U_2$–$Y$ | $U_3$–$Y$ | $U_4$–$Y$ | $U_5$–$Y$ |
|---|---|---|---|---|---|
| $R$ | 0.093 | 0.254 | 0.175 | −0.057 | 0.138 |
| $p$-value | $2.7 \times 10^{-5}$** | $3.2 \times 10^{-31}$** | $2.3 \times 10^{-15}$** | 0.01** | $4.4 \times 10^{-10}$** |

Where, $U_1$ denotes the loss of steel liquid during the steelmaking process, t; $U_2$ denotes end-point carbon content, %; $U_3$ denotes end-point manganese content, %; $U_4$ denotes temperature of liquid steel, °C; $U_5$ denotes the weight of liquid steel, t; $Y$ denotes yield of alloy, %.

**Table 3**
Results of ADF and white noise tests.

| parameter type | Value |
|---|---|
| ADF Statistic | −12.4057 |
| $p$-value of ADF | $4.4556 \times 10^{-23}$ |
| ADF critical values 1 % | −3.431 |
| ADF critical values 5 % | −2.862 |
| ADF critical values 10 % | −2.567 |
| $p$-value of white noise detection | $4.0051 \times 10^{-25}$ |

where $K$ is the number of clusters, $N$ denotes the number of data samples, $\|x\|_2$ denotes the Euclidean norm of the vector $x$, and $cl_i$ denotes the set of sample points that correspond to cluster center $c_i$.

The silhouette coefficient is defined as follows [29]:

$$S_{SC} = \frac{1}{N}\sum_{i=1}^{N}\frac{b_{i_k} - a_{i_k}}{max(b_{i_k}, a_{i_k})} \tag{12}$$

where $a_{i_k}$ represents the cohesion between the sample point and other points in the same cluster, and $b_{i_k}$ represents the separation between the sample point and other points in the adjacent cluster.

The Calinski-Harabasz scores and the silhouette coefficient are high for well-defined clustering solutions. If the cohesion factor is low, the separation factor is high, indicating low cohesiveness and high separation. The silhouette coefficient has a range of [−1, 1]. A higher value indicates a better clustering performance.

Thus, large Calinski-Harabasz scores and silhouette coefficients are desirable. The scores for 2 to 20 cluster were computed. The results are shown in Fig. 8. The Calinski-Harabasz score and silhouette coefficient have the maximum values for 4 clusters for the K-means and K-medoids methods. Therefore, we used 4 classes for data aggregation.

We evaluated the performance of the K-means and K-medoids models to choose the optimum one. Since four classes were used, the classification results for two classes are not applicable [47,48]. Therefore, the accuracy rate, recall rate, and $F_1$-*score* were used as

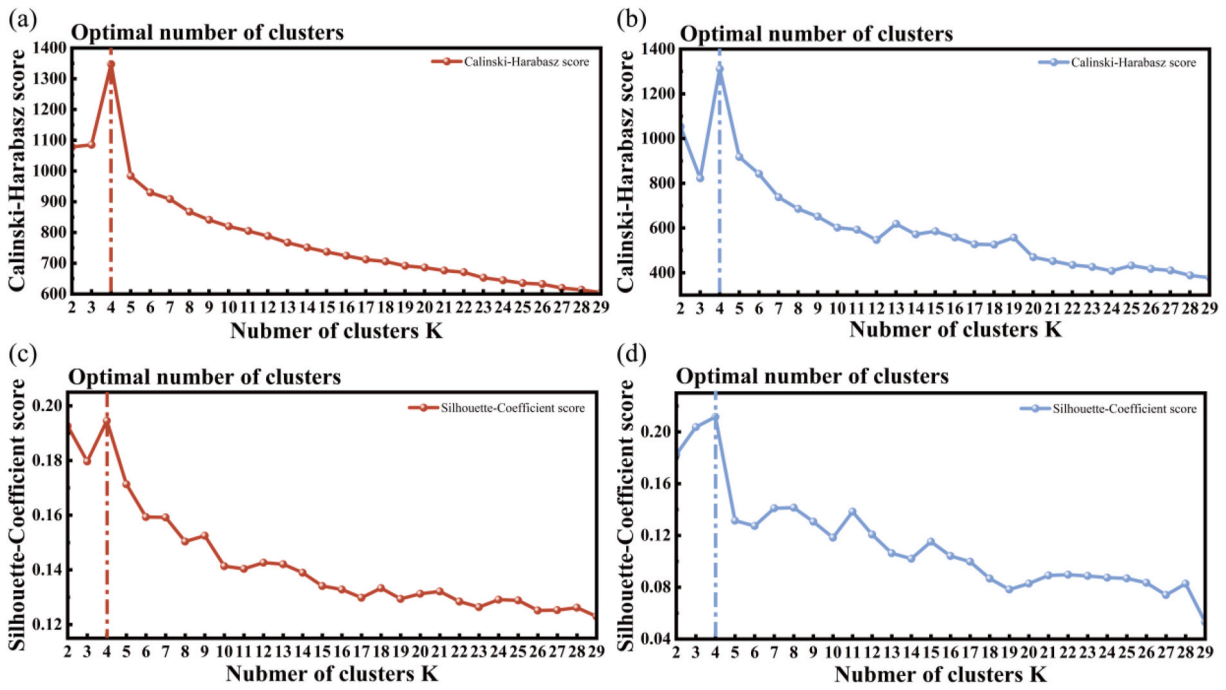


**Fig. 8.** The values of K based on different metrics: (a) Calinski-Harabasz score for K-Means; (b) Calinski-Harabasz score for K-Medoids; (c) silhouette coefficient for K-Means; (d) silhouette coefficient for K-Medoids.

metrics. The results are listed in Table 4, where TP represents the true positive, FP represents the false positive, TN represents the true negative, and FN represents the false negative. The higher the accuracy, recall, and $F_1$-*score* values, the better the model's classification performance.

We used the K-means and K-medoids methods to classify 2399 samples. Five characteristic variables of the converter steelmaking data set were used to evaluate the models' classification performance. The category label is divided into four grades (1–4) based on the alloy yield. The classification results are listed in Table 5, and the confusion matrix is shown in Fig. 9. The accuracy of the K-medoids (K-means) classification is 88.94 % (86.54 %), with a recall rate of 88.83 % (86.41 %) and an $F_1$-score of 0.8889 (0.8647). Thus, the K-medoids method has higher performance and was used to distinguish the working conditions. The classification results obtained from the K-medoids method are listed in Table 6. The box plots of the variables for the clusters are shown in Fig. 10.

### 4.2. *Comparison of the predicted alloy yield derived from different prediction methods*

The T-LSTM model was compared with five other neural network models: backpropagation neural network (BP), radial basis function neural network (RBF), RNN, gated recurrent unit neural network (GRU), and LSTM. The samples were split into a training data set (70 %) and a test set (30 %). Table 7 displays the training settings. The model performance was evaluated using the coefficient of determination ($R^2$), mean absolute error (*MAE*), mean square error (*MSE*), root mean square error (*RMSE*), and mean absolute percentage error (*MAPE*). They were calculated as follows:

**Table 4**
Performance metrics for classifications.

| Metric | Formula |
|---|---|
| Recall of Class $C_i$ | $R(C_i) = \frac{TP(C_i)}{TP(C_i) + FN(C_i)}$ |
| Precision of Class $C_i$ | $P(C_i) = \frac{TP(C_i)}{TP(C_i) + FP(C_i)}$ |
| Recall of Macro-average | $R_{Macro} = \frac{1}{N}\sum_{i=1}^{n} R(C_i)$ |
| Precision of Macro-average | $P_{Macro} = \frac{1}{N}\sum_{i=1}^{n} P(C_i)$ |
| $F_1$-*score* of Class $C_i$ | $F_1 - score = 2 \cdot \frac{P_{Macro} \cdot R_{Macro}}{P_{Macro} + R_{Macro}}$ |

Where, *TP* stands for True Positive, *TN* stands for True Negative, *FP* stands for False Positive, *FN* stands for False Negative, and *N* stands for the number of samples.

**Table 5**
Confusion matrix for the K-medoids and K-means classifications.

| K-medoids | | Prediction | | | | $P_{Macro}$ | $R_{Macro}$ | $F_1$-score |
|---|---|---|---|---|---|---|---|---|
| | | cluster 1 | cluster 2 | cluster 3 | | | | |
| reference | cluster 1 | 516 | 3 | 0.888 | 4 | 0.8894 | 0.8883 | 0.8889 |
| | cluster 2 | 23 | 553 | 5 | 18 | | | |
| | cluster 3 | 14 | 88 | 495 | 3 | | | |
| | cluster 4 | 14 | 7 | 12 | 567 | | | |
| K-means | | Prediction | | | | $P_{Macro}$ | $R_{Macro}$ | $F_1$-score |
| | | cluster 1 | cluster 2 | cluster 3 | cluster 4 | | | |
| reference | cluster 1 | 529 | 10 | 54 | 7 | 0.8654 | 0.8641 | 0.8647 |
| | cluster 2 | 14 | 489 | 49 | 47 | | | |
| | cluster 3 | 2 | 80 | 510 | 8 | | | |
| | cluster 4 | 24 | 16 | 15 | 545 | | | |

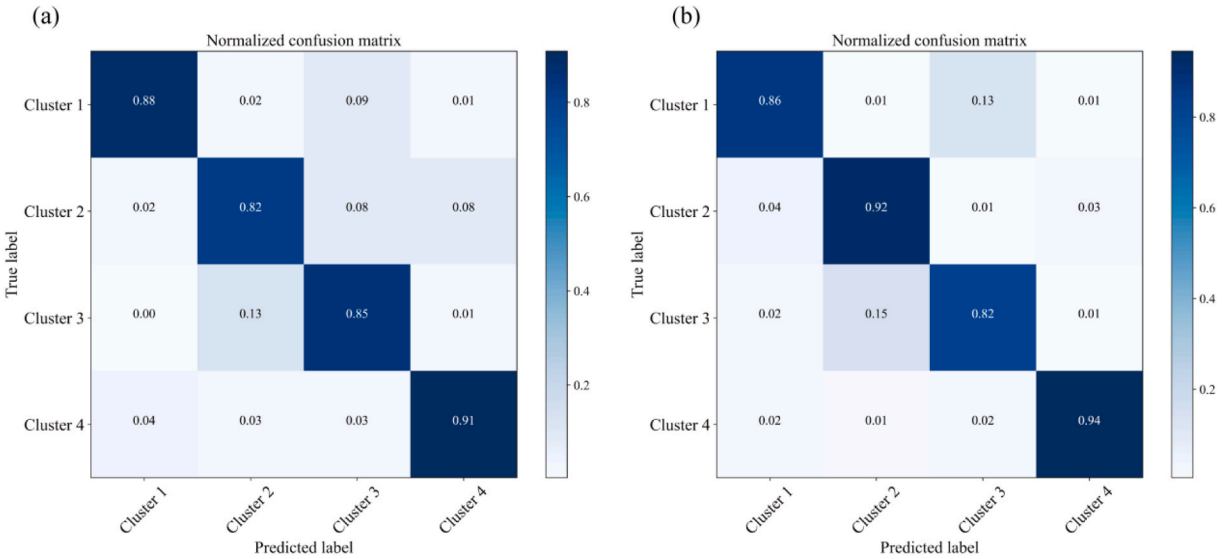


**Fig. 9.** Confusion matrix for the (a) K-means and (b) K-medoids classifications.

**Table 6**
The average values of parameters for models with different numbers of clusters.

| | Proportion of samples [%] | $U_1$ [t] | $U_2$ [%] | $U_3$ [%] | $U_4$ [°C] | $U_5$ [t] | $Y$ [%] |
|---|---|---|---|---|---|---|---|
| cluster 1 | 26.9 | 2.51 | 0.055 | 0.088 | 1638 | 50.08 | 81.54 |
| cluster 2 | 24.0 | −11.45 | 0.065 | 0.110 | 1635 | 51.15 | 83.78 |
| cluster 3 | 24.8 | 3.02 | 0.071 | 0.098 | 1624 | 49.85 | 87.22 |
| cluster 4 | 24.3 | 1.34 | 0.075 | 0.123 | 1637 | 51.54 | 88.19 |

Where, $U_1$ denotes the loss of steel liquid during the steelmaking process, t; $U_2$ denotes end-point carbon content, %; $U_3$ denotes end-point manganese content, %; $U_4$ denotes temperature of liquid steel, °C; $U_5$ denotes the weight of liquid steel, t; $Y$ denotes yield of alloy, %.

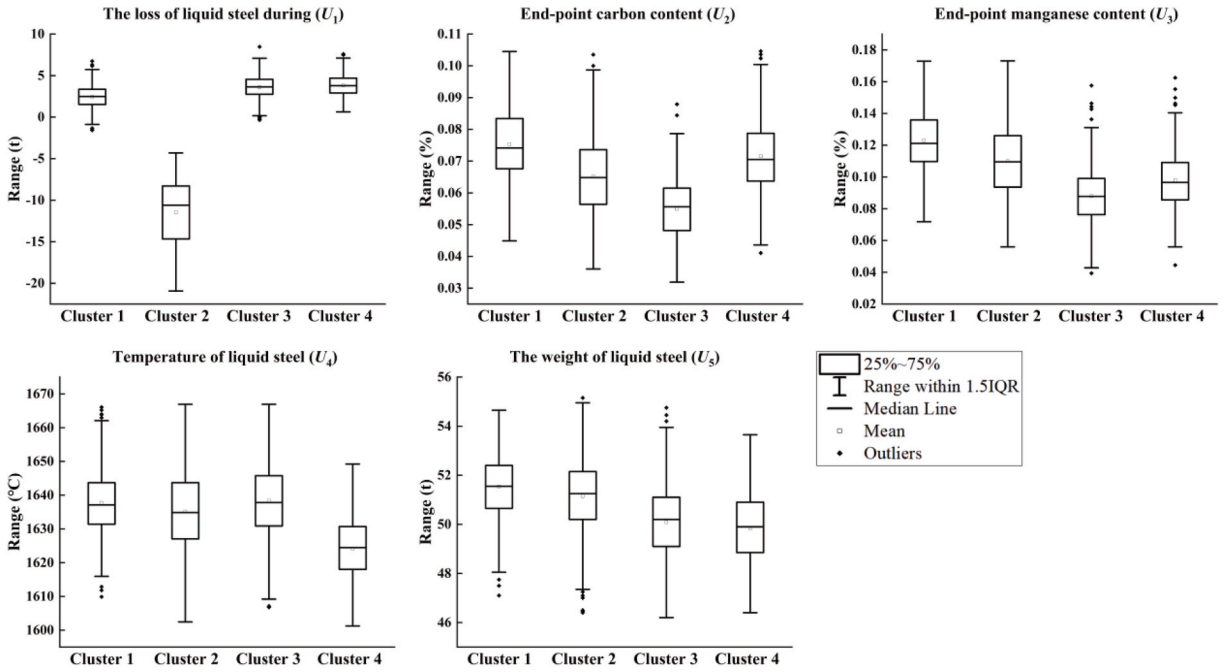


**Fig. 10.** Box plots of variables for the clusters derived from the K-medoids method.

**Table 7**
Training parameters of different models.

| Model | Experimental parameters | value |
|---|---|---|
| Bi-LSTM/GRU/RNN/LSTM | Hidden layer size | 64 |
| | No. of sample, time-steps, features | (14000, 40, 6) |
| | Learning rate | 0.05 |
| | Optimizer | Adam |
| | Batch size | 16 |
| | Epoch | 200 |
| BP | The learning velocity | 0.01 |
| | The maximum number of trainings | 1000 |
| | Training requirements precision | 0.000004 |
| RBF | Spread of radial basis functions | 0.5 |
| | Training requirements precision | 0.00004 |

$$
\begin{aligned}
R^2 &= 1 - \frac{\sum_{i=1}^{n}\left(y_i^{pre} - y_i^{act}\right)^2}{\sum_{i=1}^{n}\left(y_i^{pre} - \bar{y}\right)^2} \\
MAE &= \frac{1}{n}\sum_{i=1}^{n}\left|y_i^{pre} - y_i^{act}\right| \\
MSE &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i^{pre} - y_i^{act}\right)^2 \\
RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{pre} - y_i^{act}\right)^2} \\
MAPE &= \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i^{pre} - y_i^{act}}{y_i^{act}}\right|
\end{aligned}
\tag{13}
$$

where $n$ is the number of samples, $y_i^{pre}$ is the predicted value, $y_i^{act}$ is the true value, and $\bar{y}$ is the average of the true value.

The predicted alloy yield obtained from different models is shown in Fig. 11 and Table 8. The findings show that the seven prediction models have a reasonably high prediction accuracy, indicating that the process factors affect the alloy yield. The $R^2$ of the BP neural network is 0.633, and the *MAE*, *RMSE*, *MSE,* and *MAPE* are 0.0587, 0.1065, 0.0113, and 0.00768, respectively. The $R^2$ of the RBF neural network is 0.6613, and the *MAE*, *RMSE*, *MSE,* and *MAPE* are 0.0594, 0.1008, 0.0101, and 0.002234, respectively. The RBF

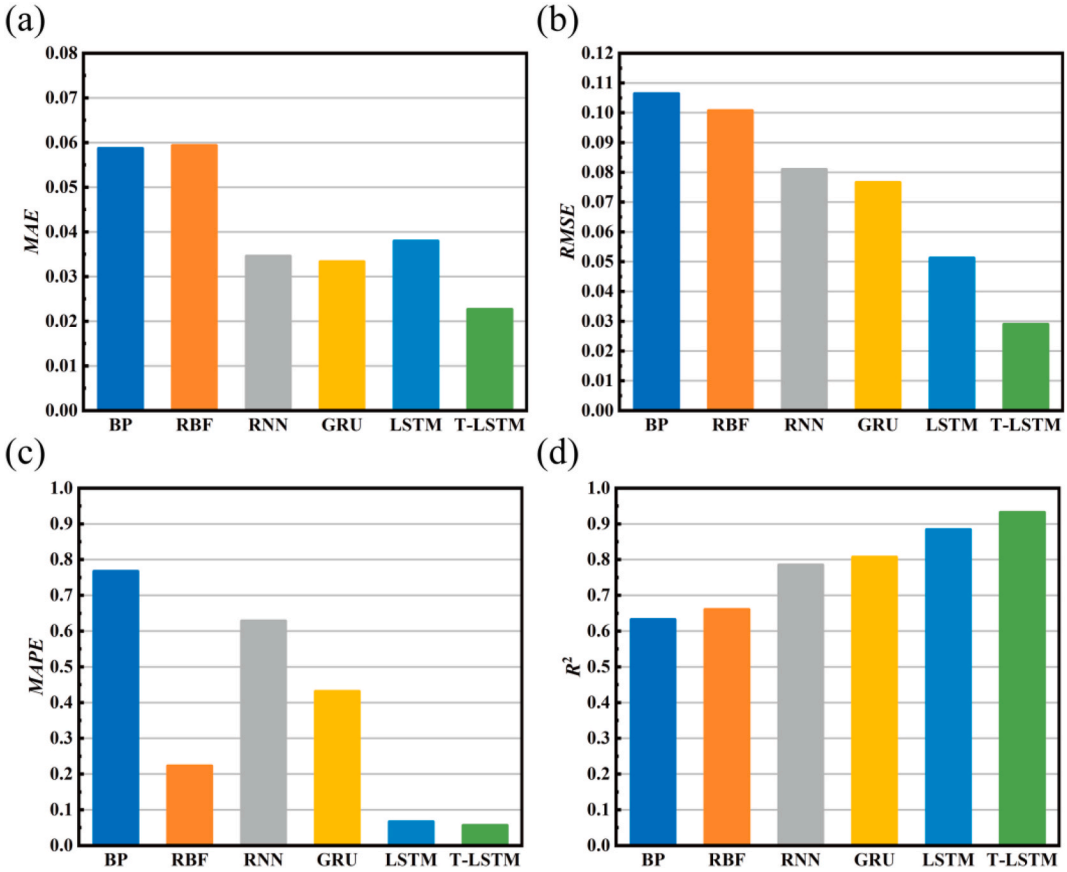


**Fig. 11.** Error metrics and $R^2$ of single-model prediction for different methods for the test set: (a) *MAE*; (b) *RMSE*; (c) *MAPE*; (d) $R^2$.

**Table 8**
Performance of different models.

| Method | *MAE* | *RMSE* | *MSE* | *MAPE*(%) | $R^2$ |
|---|---|---|---|---|---|
| BP | 0.0587 | 0.1065 | 0.0113 | 0.7680 | 0.6330 |
| RBF | 0.0594 | 0.1008 | 0.0101 | 0.2234 | 0.6613 |
| RNN | 0.0346 | 0.0810 | 0.0065 | 0.6289 | 0.7855 |
| GRU | 0.0334 | 0.0767 | 0.0058 | 0.4321 | 0.8076 |
| LSTM | 0.0380 | 0.0513 | 0.002632 | 0.0673 | 0.8847 |
| T-LSTM | 0.0227 | 0.0290 | 0.000841 | 0.0572 | 0.9327 |

model has higher performance than the BP model, but its accuracy is not as high as expected; therefore, it cannot be used to predict the alloy output.

The T-LSTM model has higher prediction accuracy than the conventional feedforward neural network models because it is more suitable for processing and analyzing time-series data.

The ranking of the models based on the prediction accuracy from large to small is T-LSTM, LSTM, GRU, and RNN. The RNN is the basis of the LSTM and GRU and is highly effective in predicting sequential data, but it has shortcomings. It multiplies the hidden layer output at time ($t$-1) by the $w$ matrix, adds the product to the value after multiplying the input at time t by u, and passes the result to the next time step using a nonlinear approach (tanh or *relu*). Gradient disappearance or explosion is likely after several multiplications. Thus, the RNN does not provide the expected outcome. It has the lowest prediction accuracy among the time-series models. The LSTM and GRU have higher accuracy. The LSTM includes forget, input, and output gates with different weights. Therefore, it is not prone to gradient disappearance like the RNN during training. The GRU has an update gate instead of forget and input gates, and the memory unit and hidden layer have reset gates, simplifying the network structure.

The GRU has exhibited better performance than the LSTM in some studies [49]. In this experiment, the prediction accuracy is higher for the LSTM and GRU than for the RNN. The $R^2$ is 0.8847 for the LSTM neural network and 0.8076 for the GRU neural network, indicating a better performance of the former than the latter. The T-LSTM considers the time interval of the time series. Its *MAE*, *RMSE*, *MSE,* and *MAPE* are 0.0227, 0.029, 0.000841, and 0.000572, respectively, which are the minimum values among all prediction

models. Its $R^2$ is the largest (0.9327), indicating that this model has the highest prediction accuracy. Therefore, the T-LSTM is selected as the core of the multi-model to predict the alloy yield.

In summary, we used the K-medoids method to classify the converter steelmaking data into four categories. Then, T-LSTM prediction models were established, trained, and tested. It is necessary to assess the residuals to determine whether the prediction model is appropriate. The residual trend is shown in Fig. 12. The cumulative probability distribution shows that the points are located approximately on a straight line. The histogram shows a normal distribution. The scatter plot of the residuals and the predicted and observed values indicates random fluctuations of the residuals. The residual plot demonstrates that the prediction model meets the requirements. Table 9 and Fig. 13 display the prediction results. They show a higher accuracy of the T-LSTM neural network after applying K-medoids clustering. The T-LSTM 4 has the optimum performance. The *MAE*, *RMSE*, *MSE*, and *MAPE* are, respectively, 0.0084, 0.0201, 0.000404, and 0.000489. They are higher than those of the T-LSTM 1 (0.0137, 0.0225, 0.000506, and 0.0523 %), T-LSTM 2 (0.0095, 0.0277, 0.000767, and 0.000701), and T-LSTM 3 (0.0084, 0.0271, 0.000734, and 0.000838). The results show that the four T-LSTM models can learn the parameter characteristics of converter steelmaking and accurately predict the alloy yield.

This model has been used in multiple steel plants in China and has achieved good results. The online performance metrics are listed in Table 10. The results demonstrate that the T-LSTM model can be used alone or in conjunction with other models and K-medoids clustering to predict the alloy yield accurately. The accuracy of network prediction can be further improved. In addition, this model can estimate hard-to-measure variables, such as alloy yield, that affect the monitoring, control, and optimization of other industrial processes.

## 5. Conclusions

The T-LSTM with K-medoids clustering was proposed as a multimodal data-driven method for predicting the process variables of industrial processes. This model uses deep learning to extract representative features from raw industrial data. The characteristics are used to classify the working environments and develop neural network models. A comparison of the clustering performance showed that the K-medoids method provided higher classification accuracy than the K-means method for high-dimensional, redundant, and industrial production data containing noise and outliers. The T-LSTM model had higher prediction accuracy than six other neural network models, including BP, RBF, RNN, LSTM, GRU, and LSTM. Most importantly, the T-LSTM network with K-medoids clustering had higher prediction accuracy than the T-LSTM network. A case study was used to validate the T-LSTM with the K-medoids method to

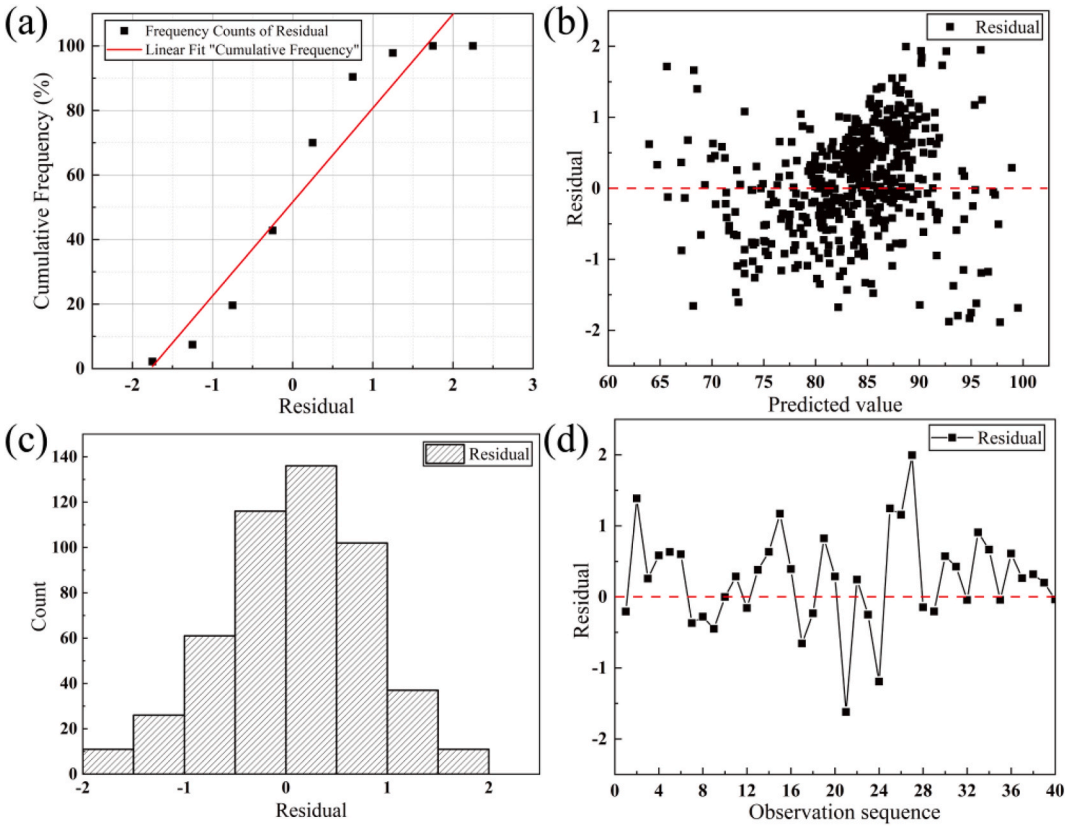


**Fig. 12.** Residual trend of prediction model: (a) cumulative probability distribution of residuals; (b) Residual and predicted values; (c) Histogram of residuals; (d) Residuals and observed values.

**Table 9**
The performance indicators of T-LSTM under four working conditions.

| Method | *MAE* | *RMSE* | *MSE* | *MAPE* |
|---|---|---|---|---|
| T-LSTM non clustering | 0.0113 | 0.0265 | 0.000676 | 0.000725 |
| T-LSTM 1 | 0.0137 | 0.0225 | 0.000506 | 0.000523 |
| T-LSTM 2 | 0.0095 | 0.0277 | 0.000767 | 0.000701 |
| T-LSTM 3 | 0.0084 | 0.0271 | 0.000734 | 0.000838 |
| T-LSTM 4 | 0.0084 | 0.0201 | 0.000404 | 0.000489 |
| Mean of T-LSTM 1-4 | 0.0100 | 0.0244 | 0.000603 | 0.000638 |

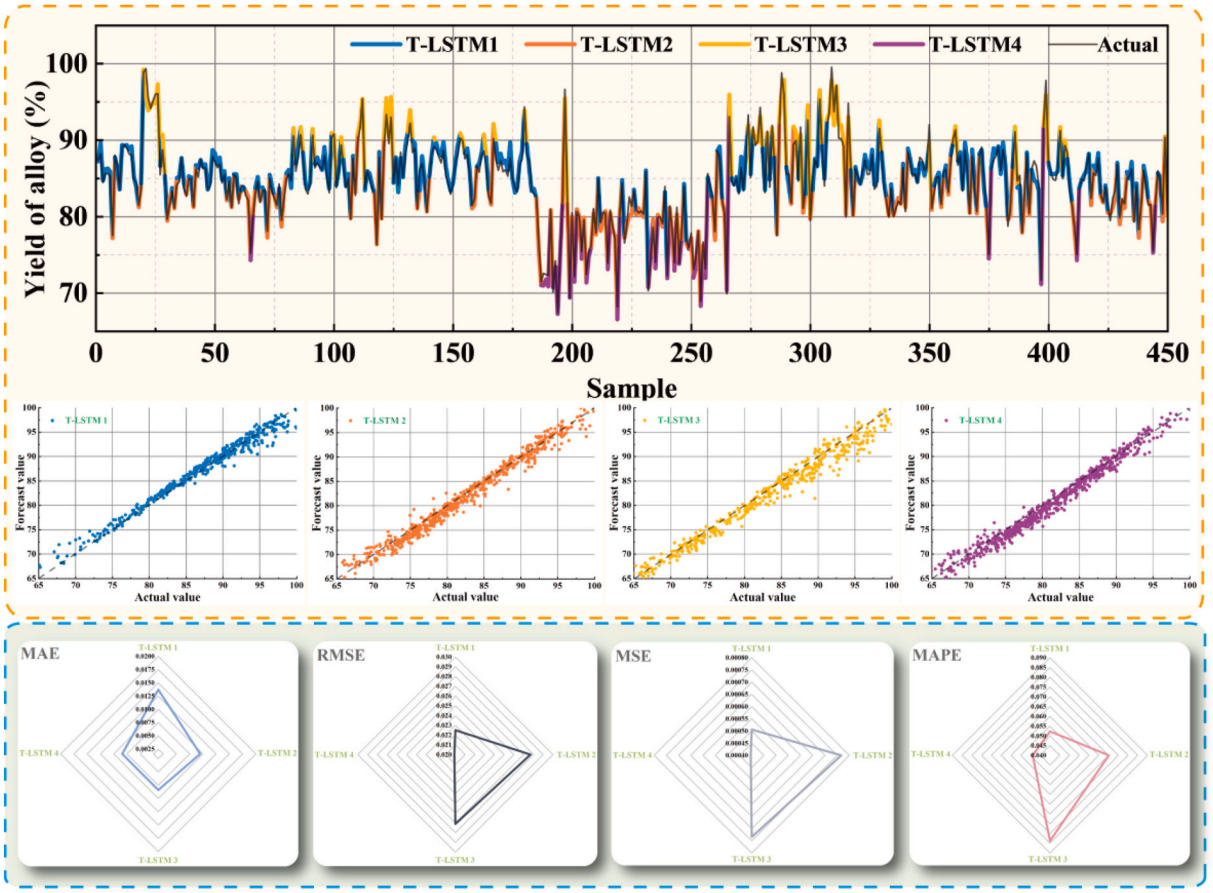


**Fig. 13.** Pre-training effects of the T-LSTM network with the K-medoids prediction model.

**Table 10**
The online performance of T-LSTM model in industrial applications.

| Steel Plant | Main steel products | Predictive accuracy of *Si* yield within ±2 % (%) | Control rate before application (%) | Control rate after application (%) |
|---|---|---|---|---|
| A | Special steel | 81.1 | 46 (Control rate of high-quality products) | 79 |
| B | Construction steel | 85.7 | 91-94 (Control rate of internal) | 95–98 |

predict the *Si* yield in converter steelmaking. This method can be used to model industrial processes using easy-to-measure variables to predict hard-to-measure variables. This study has some limitations. Different input variables of the model should be examined based on domain expertise, and the model should be assessed for use in different industrial processes. Furthermore, the neural network may require adjustment of the hyperparameters according to the training data.

A future study will focus on using larger data sets for K-medoids clustering, solving the RNN gradient problem in the LSTM, and reducing the computational complexity of computing the four fully connected layers in the LSTM cells.

## Data availability statement

The data in this study was not deposited into a publicly available repository because it is confidential.

## CRediT authorship contribution statement

**Ruixuan Zheng:** Writing – original draft, Visualization, Validation, Software. **Yanping Bao:** Resources, Project administration, Methodology, Conceptualization. **Lihua Zhao:** Supervision. **Lidong Xing:** Supervision.

## 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 supported by the National Natural Science Foundation of China (52174297). The authors wish to express their gratitude to the foundation for providing financial support.

## References

[1] C. Shang, F. Yang, D. Huang, W. Lyu, Data-driven soft sensor development based on deep learning technique, J. Process Control 24 (3) (2014) 223–233, https://doi.org/10.1016/j.jprocont.2014.01.012.
[2] Q. Sun, Z. Ge, A survey on deep learning for data-driven soft sensors, IEEE Trans. Ind. Inf. 17 (9) (2021) 5853–5866, https://doi.org/10.1109/TII.2021.3053128.
[3] K.B. Newhart, R.W. Holloway, A.S. Hering, T.Y. Cath, Data-driven performance analyses of wastewater treatment plants: a review, Water Res. 157 (2019) 498–513, https://doi.org/10.1016/j.watres.2019.03.030.
[4] L. Yao, Z. Ge, Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application, IEEE Trans. Ind. Electron. 65 (2) (2017) 1490–1498, https://doi.org/10.1109/TIE.2017.2733448.
[5] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, Comput. Chem. Eng. 33 (4) (2009) 795–814, https://doi.org/10.1016/j.compchemeng.2008.12.012.
[6] X. Yuan, Y. Wang, C. Yang, W. Gui, Stacked isomorphic autoencoder based soft analyzer and its application to sulfur recovery unit, Inf. Sci. 534 (2020) 72–84, https://doi.org/10.1016/j.ins.2020.03.018.
[7] S. Gupta, S. Ray, A.N. Samanta, Nonlinear control of debutanizer column using profile position observer, Comput. Chem. Eng. 33 (6) (2009) 1202–1211, https://doi.org/10.1016/j.compchemeng.2008.12.009.
[8] X. Yuan, L. Li, Y.A.W. Shardt, Y. Wang, C. Yang, Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development, IEEE Trans. Ind. Electron. 68 (5) (2020) 4404–4414, https://doi.org/10.1109/TIE.2020.2984443.
[9] X. Yuan, C. Ou, Y. Wang, C. Yang, W. Gui, A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process, IEEE Transact. Neural Networks Learn. Syst. 32 (8) (2019) 3296–3305, https://doi.org/10.1109/TNNLS.2019.2951708.
[10] Y. Jiang, S. Yin, J. Dong, O. Kaynak, A review on soft sensors for monitoring, control, and optimization of industrial processes, IEEE Sensor. J. 21 (11) (2020) 12868–12881, https://doi.org/10.1109/JSEN.2020.3033153.
[11] F.A.A. Souza, R. Araújo, J. Mendes, Review of soft sensor methods for regression applications, Chemometr. Intell. Lab. Syst. 152 (2016) 69–79, https://doi.org/10.1016/j.chemolab.2015.12.011.
[12] Y. Feng, T. Zhang, A.P. Sah, L. Han, Z. Zhang, Using appearance to predict pedestrian trajectories through disparity-guided attention and convolutional LSTM, IEEE Trans. Veh. Technol. 70 (8) (2021) 7480–7494, https://doi.org/10.1109/TVT.2021.3094678.
[13] Y. Sun, F. Haghighat, B.C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, Energy Build. 221 (2020) 110022, https://doi.org/10.1016/j.enbuild.2020.110022.
[14] F. Zou, G.G. Yen, L. Tang, A knee-guided prediction approach for dynamic multi-objective optimization, Inf. Sci. 509 (2020) 193–209, https://doi.org/10.1016/j.ins.2019.09.016.
[15] Y. Ji, S. Liu, M. Zhou, Z. Zhao, X. Guo, L. Qi, A machine learning and genetic algorithm-based method for predicting width deviation of hot-rolled strip in steel production systems, Inf. Sci. 589 (2022) 360–375, https://doi.org/10.1016/j.ins.2021.12.063.
[16] K. Jiang, Z. Jiang, Y. Xie, Z. Chen, D. Pan, W. Gui, Classification of silicon content variation trend based on fusion of multilevel features in blast furnace ironmaking, Inf. Sci. 521 (2020) 32–45, https://doi.org/10.1016/j.ins.2020.02.039.
[17] Y. Feng, M. Wu, X. Chen, L. Chen, S. Du, A fuzzy PID controller with nonlinear compensation term for mold level of continuous casting process, Inf. Sci. 539 (2020) 487–503, https://doi.org/10.1016/j.ins.2020.06.024.
[18] F. Yuan, A. Xu, M. Gu, Development of an improved CBR model for predicting steel temperature in ladle furnace refining, Int. J. Miner. Metall. Mater. 28 (2021) 1321–1331, https://doi.org/10.1007/S12613-020-2234-6.
[19] D. Wang, F. Gao, L. Xing, J. Chu, Y. Bao, Continuous prediction model of carbon content in 120 t converter blowing process, Metals 12 (1) (2022) 151, https://doi.org/10.3390/met12010151.
[20] S. Wu, J. Yang, G. Cao, Prediction of the Charpy V-notch impact energy of low carbon steel using a shallow neural network and deep learning, Int. J. Miner. Metall. Mater. 28 (8) (2021) 1309–1320, https://doi.org/10.1007/s12613-020-2168-z.
[21] Z. Xin, J. Zhang, J. Zheng, Y. Jin, Q. Liu, A hybrid modeling method based on expert control and deep neural network for temperature prediction of molten steel in LF, ISIJ Int. 62 (3) (2022) 532–541, https://doi.org/10.2355/isijinternational.ISIJINT-2021-251.
[22] Y. Zhou, K. Xu, F. He, Z. Zhang, Application of time series data anomaly detection based on deep learning in continuous casting process, ISIJ Int. 62 (4) (2022) 689–698, https://doi.org/10.2355/isijinternational.ISIJINT-2021-372.
[23] C. Kim, J.Y. Shin, J. Roh, H.J. Hwang, Prediction on the distributions of the strength and toughness of thick steel plates based on bayesian neural network, Steel Res. Int. 93 (6) (2022) 2100566, https://doi.org/10.1002/srin.202100566.
[24] X. Huang, X. Fan, X. Chen, G. Yang, M. Gan, Bed permeability state prediction model of sintering process based on data mining technology, ISIJ Int. 56 (12) (2016) 2113–2117, https://doi.org/10.2355/isijinternational.ISIJINT-2016-193.
[25] Z. Ding, J. Zhang, Y. Liu, Ensemble non-Gaussian local regression for industrial silicon content prediction, ISIJ Int. 57 (11) (2017) 2022–2027, https://doi.org/10.2355/isijinternational.ISIJINT-2017-251.
[26] X. Song, Y. Meng, C. Liu, Y. Yang, D. Song, Prediction of molten iron temperature in the transportation process of torpedo car, ISIJ Int. 61 (6) (2021) 1899–1907, https://doi.org/10.2355/isijinternational.ISIJINT-2020-335.

[27] R. Zheng, Y. Bao, L. Zhao, L. Xing, Method to predict alloy yield based on multiple raw material conditions and a PSO-LSTM network, J. Mater. Res. Technol. 27 (2023) 3310–3322, https://doi.org/10.1016/j.jmrt.2023.10.046.
[28] J. Jing, S. Ke, T. Li, T. Wang, Energy method of geophysical logging lithology based on K-means dynamic clustering analysis, Environ. Technol. Innov. 23 (2021) 101534, https://doi.org/10.1016/j.eti.2021.101534.
[29] R.S. Moorthy, P. Pabitha, Prediction of Parkinson's disease using improved radial basis function neural network, Comput. Mater. Continua (CMC) 68 (3) (2021) 3101–3119, https://doi.org/10.32604/cmc.2021.016489.
[30] X. Shi, Y. Li, Y. Yang, B. Sun, F. Qi, Multi-models and dual-sampling periods quality prediction with time-dimensional K-means and state transition-LSTM network, Inf. Sci. 580 (2021) 917–933, https://doi.org/10.1016/j.ins.2021.09.056.
[31] Z. Feng, Y. Li, B. Sun, C. Yang, T. Huang, A multimode mechanism-guided product quality estimation approach for multi-rate industrial processes, Inf. Sci. 596 (2022) 489–500, https://doi.org/10.1016/j.ins.2022.02.041.
[32] T. Liang, Q. Zhao, Q. Lv, H. Sun, A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers, Energy 230 (2021) 120904, https://doi.org/10.1016/j.energy.2021.120904.
[33] I.M. Baytas, C. Xiao, X. Zhang, F. Wang, A.K. Jain, J. Zhou, Patient subtyping via time-aware LSTM networks, in: Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, 2017, pp. 65–74, https://doi.org/10.1145/3097983.3097997.
[34] M. Gu, A. Xu, H. Wang, Z. Wang, Real-time dynamic carbon content prediction model for second blowing stage in BOF based on CBR and LSTM, Processes 9 (11) (2021) 1987, https://doi.org/10.3390/pr9111987.
[35] D. Hong, W. Han, C.-H. Yim, Convolutional recurrent neural network to determine whether dropping slag dart fills the exit hole during tapping in a basic oxygen furnace, Metall. Mater. Trans. B 52 (6) (2021) 3833–3845, https://doi.org/10.1007/s11663-021-02299-z.
[36] L. Klotz, G. Lenz, N. Antoniassi, R. Borges, T. Nunes, Multivariate analysis of the main operational variables involved in steel producing on BOF using time series tools, J. Inst. Eng. 30 (1) (2024) 70–80, https://doi.org/10.56801/MME1027.
[37] X. Liu, X. Qu, X. Xie, S. Li, Y. Bao, L. Zhao, Predicting alloying element yield in converter steelmaking using t-SNE-WOA-LSTM, Processes 12 (5) (2024) 974, https://doi.org/10.3390/pr12050974.
[38] T. Okura, I. Ahmad, M. Kano, S. Hasebe, H. Kitada, N. Murata, High-performance prediction of molten steel temperature in tundish through gray-box model, ISIJ Int. 53 (1) (2013) 76–80, https://doi.org/10.2355/isijinternational.53.76.
[39] I. Ahmad, M. Kano, S. Hasebe, H. Kitada, N. Murata, Gray-box modeling for prediction and control of molten steel temperature in tundish, J. Process Control 24 (4) (2014) 375–382, https://doi.org/10.1016/j.jprocont.2014.01.018.
[40] L. Chen, Z. Tian, S. Zhou, Q. Gong, H. Di, Attitude deviation prediction of shield tunneling machine using Time-Aware LSTM networks, Transp. Geotech. 45 (2024) 101195, https://doi.org/10.1016/j.trgeo.2024.101195.
[41] H.J. Park, M.S. Lee, D. Il Park, S.W. Han, Time-aware and feature similarity self-attention in vessel fuel consumption prediction, Appl. Sci. 11 (23) (2021) 11514, https://doi.org/10.3390/app112311514.
[42] I. Jebli, F.-Z. Belouadha, M.I. Kabbaj, A. Tilioua, Prediction of solar energy guided by pearson correlation using machine learning, Energy 224 (2021) 120109, https://doi.org/10.1016/j.energy.2021.120109.
[43] R. Zhang, N.H. Chan, Nonstationary linear processes with infinite variance GARCH errors, Econom. Theor. 37 (5) (2021) 892–925, https://doi.org/10.1017/S0266466620000377.
[44] K. Worden, I. Iakovidis, E.J. Cross, New results for the ADF statistic in nonstationary signal analysis with a view towards structural health monitoring, Mech. Syst. Signal Process. 146 (2021) 106979, https://doi.org/10.1016/j.ymssp.2020.106979.
[45] T. Lee, Wild bootstrap Ljung–Box test for residuals of ARMA models robust to variance change, J. Korean Surg. Soc. 51 (2022) 1005–1020, https://doi.org/10.1007/s42952-022-00172-6.
[46] A. Meiseles, L. Rokach, Source model selection for deep learning in the time series domain, IEEE Access 8 (2020) 6190–6200, https://doi.org/10.1109/ACCESS.2019.2963742.
[47] D. Zhang, J. Wang, X. Zhao, Estimating the uncertainty of average F1 scores, in: Proc. 2015 Int. Conf. Theory Inf. Retr., 2015, pp. 317–320, https://doi.org/10.1145/2808194.2809488.
[48] I. Markoulidakis, G. Kopsiaftis, I. Rallis, I. Georgoulas, Multi-class confusion matrix reduction method and its application on net promoter score classification problem, in: Proc. 14th PErvasive Technol. Relat. To Assist. Environ. Conf, 2021, pp. 412–419, https://doi.org/10.1145/3453892.3461323.
[49] M.J. Hamayel, A.Y. Owda, A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms 2 (4) (2021) 477–496, https://doi.org/10.3390/ai2040030. Ai.