Predicting Precipitation Rate in Alumina Production using Machine Learning

Long Duan[1], Shuai Shao[2] and Yanfang Zhang[3]
1. Assistant Engineer
2. Engineer
3. Professor
Zhengzhou Non-ferrous Metals Research Institute of CHALCO, Zhengzhou, China
Corresponding author: zyy_zyf@rilm.com.cn
https://doi.org/10.71659/icsoba2024-aa018

Abstract

The Bayer process for alumina production is complex and lengthy. Accurately and timely predicting process indicators and determining important parameters is the key to optimizing this process. This study focuses on the precipitation process during alumina production, achieving accurate and timely prediction of precipitation rate by using machine learning. Equipment and material data were collected over a 3-month period in the precipitation area of an alumina plant. The data was pre-processed and a time-series dataset was established with a total of 535 data samples (divided into training set, validation set, and testing set in a 3:1:1 ratio). Principal component analysis (PCA), partial least squares (PLS) and Pearson correlation analysis (Pearson) were used to select effective parameters. Four prediction models, including linear regression (LR), support vector machine (SVM), back propagation neural network (BP), and convolution neural network (CNN), were established based on effective parameters data. The coefficients of determination for the prediction model established in this study were all greater than 0.8 on the testing set. The results showed that by using mass production data to establish a machine learning model, we can accurately and timely predict the precipitation rate in the Bayer process. This provides a theoretical basis for predicting process indicators for the entire alumina production process and lays the foundation for the intelligent production of alumina.

Keywords: Alumina, Precipitation Rate, Machine Learning, Prediction Model.

1. Introduction

The Bayer process has advantages in low energy consumption and high product quality. In the Bayer process, the precipitation process is very important. Its main function is to precipitate aluminium hydroxide from the sodium aluminate solution via a sedimentation process. The goal of the precipitation process is to achieve the precipitation rate as close to the upper limit as possible while ensuring the qualified particle size of aluminium hydroxide. Precipitation efficiency directly affects the efficiency in production [1].

It is difficult to obtain the precipitation rate accurately and timely due to the complex and lengthy process, large number of devices and equipment, and the complex material transformation process of precipitation. Moreover, the process has the complexity of large inertia, nonlinearity, and many influencing factors and dependencies. Relying solely on human experience and mechanistic models cannot provide effective production control plans, resulting in low efficiency in the precipitation process.

With the progress of sensor technology and the continuous advancement of factory digitization, a large amount of production data can be obtained in the production process. Many scholars and process engineers have begun to study how to make full use of this production data, extract important values, predict key process indicators, and then guide production regulation [2–5].

In the aluminium smelting industry, there has been much research on prediction of process indicators for aluminium production by electrolysis [6-8]. However, in the Bayer industry, prediction of process indicators in alumina production mainly focus on the research of mechanistic and empirical models, with only few input parameters selected for modelling. The accuracy of these models is generally low, and the prediction results cannot guide real-time production regulation [9, 10].

Zhang et al. [11] constructed a time-series dataset using data from the first and last precipitation tank, and seed crystal size. Based on machine learning technology, this accurately and timely predicted the crystal seeds particle size attenuation. However, the device data and material data generated by each device in the whole production process are related to the process indicators. Using all available production data can make the predicted results to the actual production, but the increase of data items introduces more redundant information. Therefore, it is necessary to choose appropriate data processing methods and adopt efficient machine learning techniques to establish the prediction model.

In this study, the device data and material data of the precipitation process of an alumina plant were used to construct a time-series dataset, and a precipitation rate prediction model was established based on machine learning to guide production control, and ultimately improve efficiency in production with reduced costs.

2. Data and Methods

2.1 Data Acquisition

The data used in this study came from the Excellent Technology Centre of Chinalco, which supplied device and material data from the precipitation process in a production line for three months (September 2023 to November 2023). The precipitation process included 14 precipitation tanks, six heat exchangers, a classifying hydrocyclone, and a horizontal table filter. The precipitation period of the production line was 44 hours in a batch process.

The device data included data for 14 precipitation tanks and 6 heat exchangers. The material data included data for the precipitation raw liquid and the precipitation mother liquid. Table 1 shows the detailed data.

Table 1. Detailed data.

Data source	Data item	Data frequency
Precipitation tanks	Temperature (°C)	Every hour
	Stirring current (A)	
	Liquid level (m)	
Heat exchangers	Flow rate (m^3/h)	
	Feed pressure on heat source (MPa)	
	Feed pressure on cold source (MPa)	
	Outlet temperature of heat source (°C)	
	Outlet temperature of cold source (°C)	
Precipitation raw liquid & Precipitation mother liquid	α_k (Molar ratio of Na_2O_K to Al_2O_3)	Every four hours
	Suspended matter (g/L)	
	Al_2O_3 (g/L)	
	Na_2O_K (g/L)	
	Na_2O_T (total, g/L)	

Precipitation slurry flows from tank F1 to F14. Heat exchangers are arranged in order from H1 to H6, placed between tanks from F2 to F8. The classifying hydrocyclone is located in the F11 precipitation tank, and the graded underflow is sent to the horizontal table filter along with slurry from F13 and F14. The filter cake enters the seeds tank and is recycled to precipitation. Filtrate enters the evaporation process, and the overflow is returned to F13 and F14.

2.2 Dimensionality Reduction Methods

Principal Component Analysis (PCA) is a widely used dimensionality reduction method. PCA transforms data into a set of linearly independent variables through orthogonal transformation, reducing the dimensionality of the data while preserving as much specificity as possible of the original data.

Partial least squares (PLS) is suitable for solving the problem of multicollinearity between the predictive variable (independent variable) and the response variable (dependent variable). By searching for the basic relationship between predictive variable and response variable, PLS can extract the most useful components for prediction.

Pearson correlation analysis (Pearson) is a method of measuring the degree of linear correlation between two variables. In the process of dimensionality reduction, Pearson can be used to identify and exclude features with low correlation with the target variable.

2.3 Prediction Model Methods

Linear regression (LR) is a common prediction method. LR assumes that there is a linear correlation between the target value and the features, which satisfies a multivariate linear equation. Prediction is achieved by solving the linear equation.

Support Vector Machine (SVM) can be used for quantitative and qualitative analysis. SVM has strong generalization ability, it obtains stable prediction results by maximum decision boundaries.

Artificial Neural Network (ANN) is a mathematical model that mimics the behavioural characteristics of biological neural networks and performs distributed parallel information processing. ANN relies on the complexity of the system, adjusting the relationships between a large number of internal nodes. ANN can learn any nonlinear function. Back propagation neural network (BP) is a multi-layer feed forward ANN trained using an error back propagation algorithm, and is the most widely used form of ANN. The common BP structure consists of four layers: input layer, hidden layer, activation layer, and output layer.

Convolutional neural networks (CNN) perform better than shallow networks in feature selection, feature extraction, prediction, and classification. CNN has two advantages: (1) local sense, where CNN only senses local elements of the data, and then fuses this local information in higher-level networks to obtain all information of the data; and (2) weight sharing, where CNN reduces the complexity of the network and the number of weights. CNN has been widely applied in multiple fields, and the common CNN structure consists of six layers: input layer, convolutional structure, activation layer, pooling layer, fully connected layer, and output layer.

3. Experiment

3.1 Experiment Environment

Table 2 shows the configuration environment in which all of the data processing and modelling in this study were conducted.

Table 2. Configuration environment.

Type of configuration	Configuration info
Operating system	Ubuntu 22.04
Program language	Python
Development framework	PyTorch
CPU	Intel i9-13900K
GPU	NVIDIA GeForce RTX 4090 D (24 GB)
Memory	64 GB

3.2 Data Pre-Processing

The collected device and material data were found to have missing data. In this study, data cleaning was carried out to remove columns with overall missing data, and data interpolation (polynomial order three) was applied to columns with partial missing data. After data cleaning, the device data at each time node contained 29 data items, and the material data at each time node contained 10 data items.

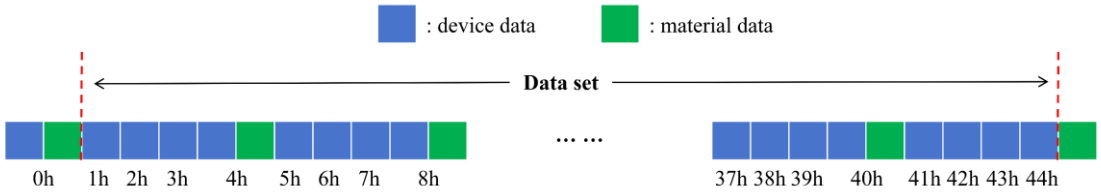

Figure 1. Data set.

In order to simulate the production more realistically, this study used a time-series dataset based on the complete precipitation process, with each piece of data containing a complete precipitation period (44 hours) of data. As shown in Figure 1, each data set contained 44 hours of device data and 10 points of material data sampled every 4 hours (with the last data point excluded), giving a total of 1276 device data items and 100 material data items. The time-series dataset consisted of 535 valid data samples, which were divided into training set, validation set, and testing set in a 3:1:1 ratio. The precipitation rate of each data set was calculated by Formula (1) from the mother liquor at the end of the precipitation period and the original liquor at the beginning of the precipitation period. In this manuscript, the precipitation rate is the molar ratio of precipitated Al_2O_3 to total Al_2O_3, represent the production efficiency of the precipitation process.

$$P = \frac{{\alpha_k}^a - {\alpha_k}^b}{{\alpha_k}^a} \tag{1}$$

Where

P Precipitation rate

α_k Molar ratio of Na_2O_K to Al_2O_3

${\alpha_k}^a$ α_k of precipitation mother liquid

${\alpha_k}^b$ α_k of precipitation raw liquid

3.3 Data Dimensionality Reduction

PCA, PLS and Pearson were used to reduce the dimensionality of time-series dataset, reduce redundant and noisy information, retain important indicators, and obtain the modelling dataset. In this research, PLS, PCA, and Pearson were all realized using the function library (scikit-learn 1.5.1) in Python. In order to reduce the optimization time of the modelling, the range of adjustable parameters for three dimensionality reduction methods were limited. The parameters of PCA

included the number of principal components and the singular value decomposition method. The number of principal components included 2, 3,..., and 200. The singular value decomposition method included 'auto', 'full', 'arpack', and 'randomized'. The parameters of PLS included the number of principal components, maximum number of iterations, and convergence criterion. The number of principal components included 2, 3,..., and 200. The maximum number of iterations included 1e3, 1e4, 1e5, and 1e6. The convergence criterion was set to 1e-6. The parameter of Pearson was the number of optimal parameters, the number of optimal included 2, 3,..., and 250.

3.4 Establishment of Prediction Model

LR, SVM, BP, and CNN were used to construct precipitation rate prediction models. In this study, model designs were conducted for four models.

- The LR model had a simple structure and did not require model design.
- The parameters of SVM included kernel function, kernel coefficient, regularization coefficient, convergence criterion, maximum number of iterations, and tolerance. The kernel functions included 'linear', 'poly', 'rbf', and 'sigmoid'. The kernel coefficients included 1e-8, 1e-7,..., and 1e5. The regularization coefficients include 1e-8, 1e-7,..., and 1e8. The convergence criterion included 1e-3, 1e-4, 1e-5, and 1e-6. The maximum number of iterations included 1e3, 1e4, 1e5, and 1e6. The tolerance was set to 0.
- The parameters of both the BP and CNN included learning rate, number of iterations, network structure, optimizer, and loss function. The learning rate included 1e-2, 1e-3, 1e-4, and 1e-5. The number of iterations was set to 500. Figures 2 and 3 show the network structure diagram for BP and CNN respectively. The optimizer used Adam. The loss function used mean square loss.

As in Figure 2, the BP model designed in this study consists of an input layer, four hidden layers, and an output layer. The input layer refers to the modelling data input. The hidden layer consisted of a fully connected layer and an activation layer, using 'ReLU' activation function. The number of neurons in the four fully connected layers was 1024, 512, 256, and 128, respectively. The output layer consisted of one fully connected layer, which outputs the predicted precipitation rate.

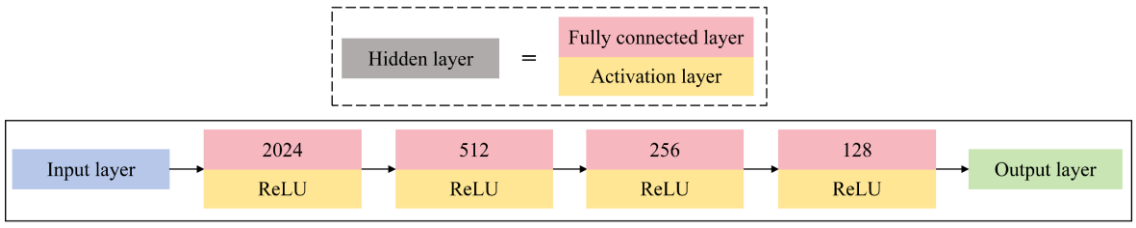

Figure 2. The structure of BP.

As in Figure 3, the CNN model designed in this study consisted of an input layer, four convolutional structures, a flatten layer, and an output layer. The input layer refers to the modelling data input. The convolutional structure consisted of a one-dimensional (1D) convolutional layer, an activation layer, a batch normalization layer, and an 1D max pooling layer. The activation function of the activation layer was 'ReLU'. The number of convolution kernels for four 1D convolutional layers was 512, 256, 128, and 64, respectively, with the kernel size was 1×3. The kernel size of the 1D maximum pooling layer was 1×3. The flatten layer flattened the multidimensional data features of the last convolutional structure into 1D data. The output layer consisted of two fully connected layers and an activation layer, using the 'ReLU' activation function. The first fully connected layer had 256 neurons. The second fully connected layer outputs the predicted precipitation rate.

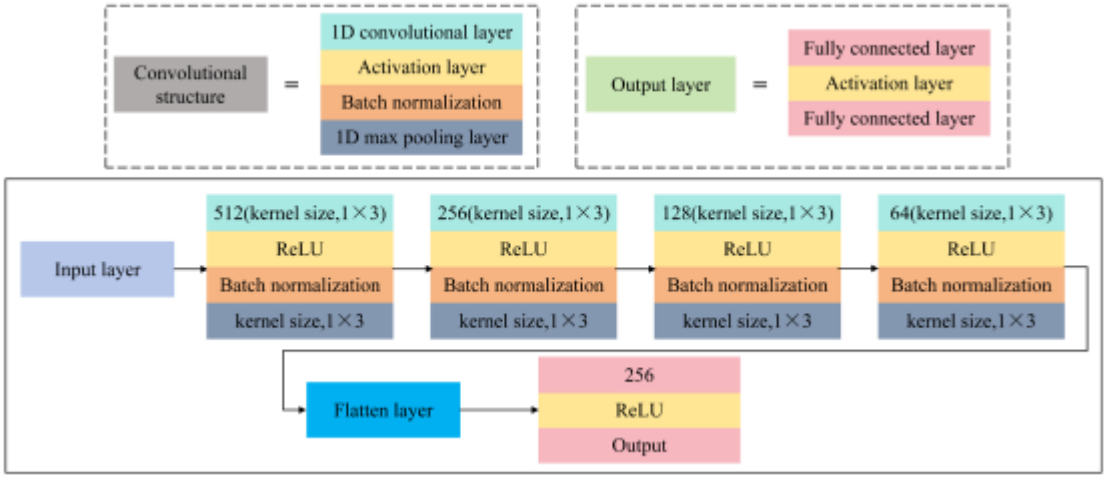

Figure 3. The structure of CNN.

In this study, the coefficient of determination R^2 was used to evaluate the prediction model. It represents the degree of fit of the prediction model to the dataset. The greater the R^2, the better the prediction model fits, R^2 is calculated by Formula (2).

$$R^2(P, P') = 1 - \frac{\sum_{i=0}^{n}(P_i - P'_i)^2}{\sum_{i=0}^{n}(P_i - \overline{P})^2} \quad (2)$$

Where

P *True* precipitation rate
P' *Predicted* precipitation rate
n Number of data points

4. Results and Discussion

4.1 Results of Data Pre-Processing

As shown in Figure 4, the precipitation rate throughout the three-month study period (September 2023 to November 2023) is constantly changing. The average precipitation rate over the three months was 0.5046. Increases and decreases in precipitation rate are both temporal and not abrupt.

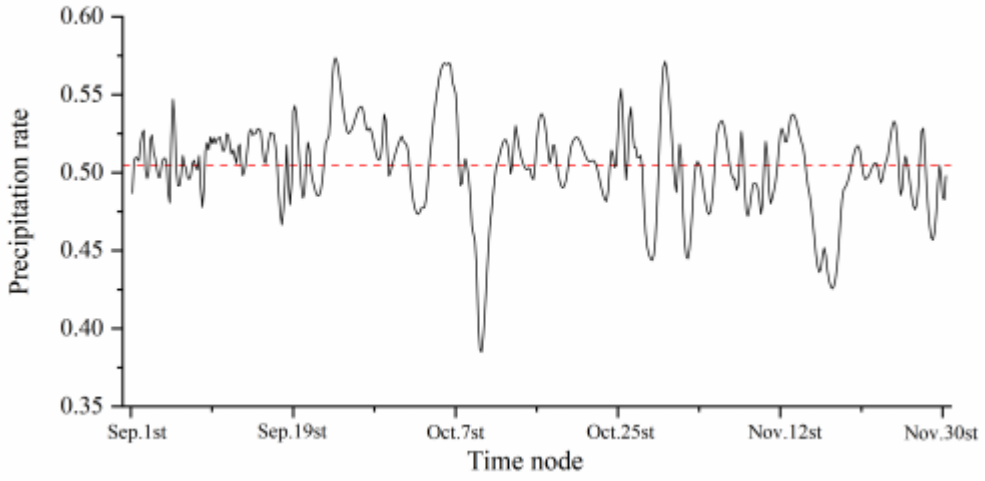

Figure 4. Precipitation rate.

4.2 Results of Data Dimensionality Reduction

PCA, PLS, and Pearson obtained the optimal parameters after parameter tuning using grid search. The parameters of the PCA included 87 principal components and the 'full' decomposition method. The parameters of the PLS included 18 principal components, 1e4 maximum iterations, and 1e-6 convergence criterion. The optimal number of parameters for Pearson was 34.

For PCA, Figure 5 shows the loadings (the contribution rates of each data) to the first principal component of PCA. The key features were stirring current of precipitation tank and outlet temperature of the cold source. As the precipitation process progressed, the contribution rate of the outlet temperature of the cold source gradually decreased.

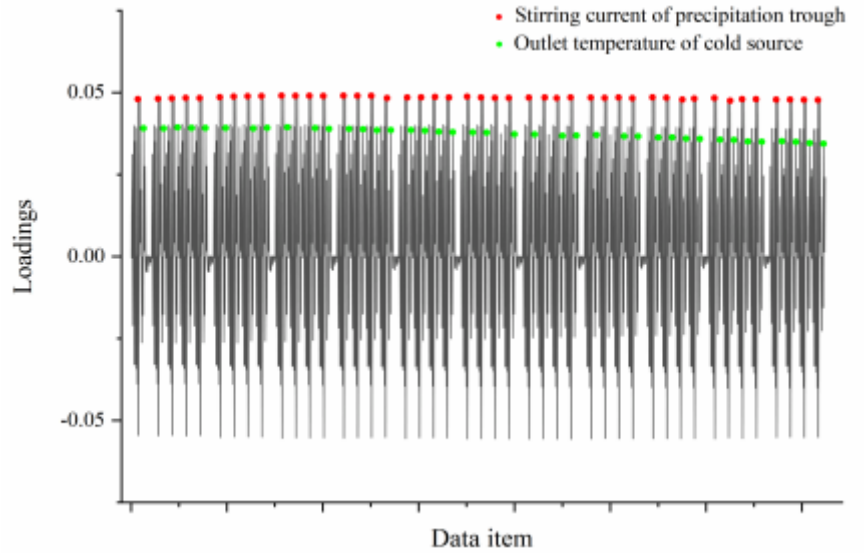

Figure 5. Loading of PCA.

For PLS, Figure 6 shows the contribution rates of each data item to the first principal component of PLS. The key features were outlet temperature of the cold source, Al_2O_3 of precipitation mother liquid, and Na_2O_K of precipitation mother liquid. As the precipitation process progressed, the contribution rates of each key feature were relative unchanged.

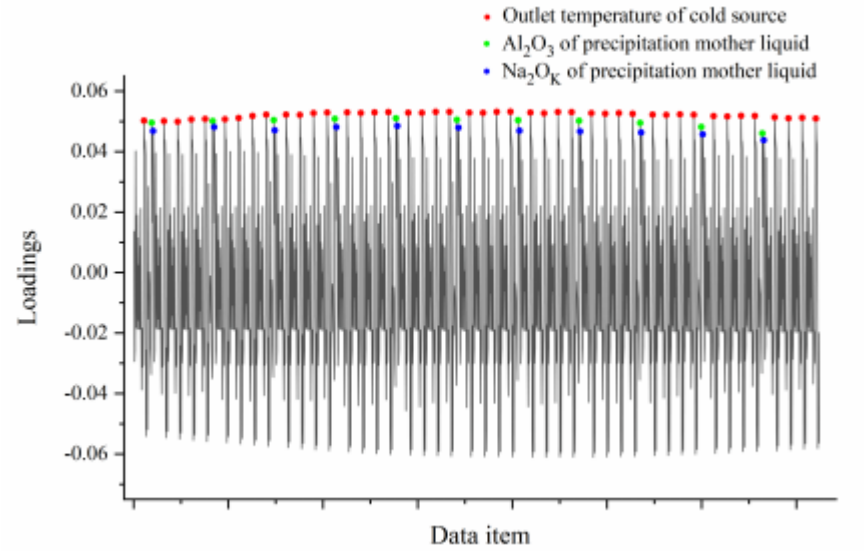

Figure 6. Loading of PLS.

For Pearson, Figure 7 shows the scores (the importance of feature) of each data item after calculating the correlation with the precipitation rate. The key features were α_k of precipitation mother liquid, α_k of precipitation raw liquid, and outlet temperature of the cold source. As the precipitation process progressed, the scores of the α_k of precipitation mother liquid increased, the scores of the α_k of precipitation raw liquid decreased, and the scores of the outlet temperature of the cold source gradually increased.

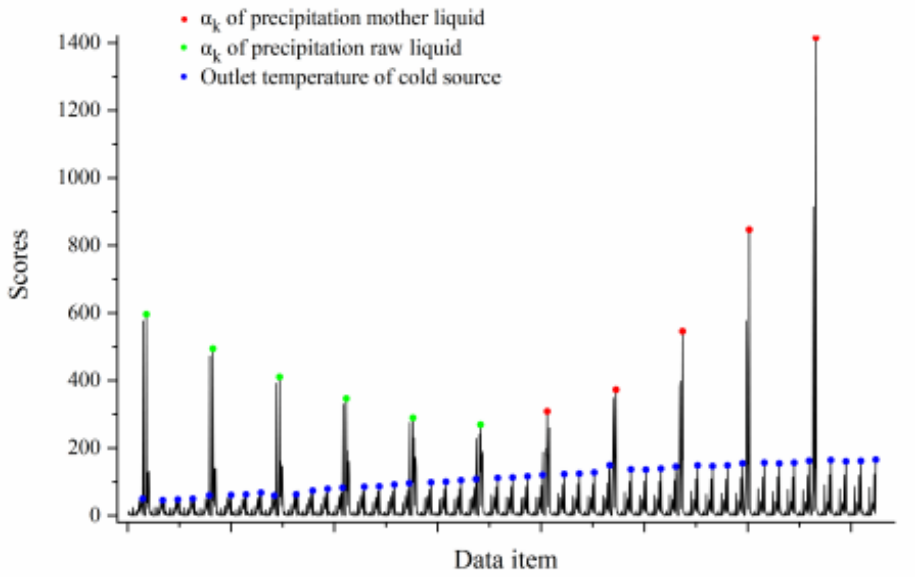

Figure 7. Scores of Pearson.

4.3 Results of Prediction Model

LR, SVM, BP and CNN obtained the optimal parameters after parameter tuning using grid search. Parameter tuning was not required for LR. The three other prediction methods were tuned to obtain the optimal parameters.

- The parameters of SVM included the 'poly' kernel function, 1e-7 regularization coefficient, 1e2 kernel coefficient, 1e5 maximum iterations, 1e-4 convergence criterion, and zero tolerance.
- The parameters of BP included 1e-3 learning rate, 500 iterations, the structure in Figure 2, Adam optimizer, and mean square loss function.
- The parameters of CNN included 1e-4 learning rate, 500 iterations, and the structure in Figure 3, Adam optimizer, and mean square loss function.

Twelve optimal models were obtained by combining the three data dimensionality reduction methods and four prediction methods. Table 3 shows the R^2 of the testing set data in the 12 models. The R^2 of more than half of the prediction models were greater than 0.8, and the R^2 of three prediction models were greater than 0.9. The prediction accuracy of the models was relatively very good. The dimensionality reduction effect of PLS and Pearson were greater than PCA. The PLS reduced dataset was more suitable for nonlinear models (SVM, BP, and CNN), while the Pearson reduced dataset was more suitable for linear models (LR). The model performance of SVM and CNN were greater than LR and BP. CNN performed best in all dimensionality reduction datasets. LR and BP had overfitting issues in modelling, resulting in lower R^2 in the testing set.

Table 3. R^2 of prediction models.

Modelling method	R^2 of prediction model		
	PCA	PLS	Pearson
LR	0.5703	0.8017	**0.8870**
SVM	0.7934	**0.9277**	0.8885
BP	0.8036	**0.8994**	0.8866
CNN	0.8499	**0.9288**	0.9122

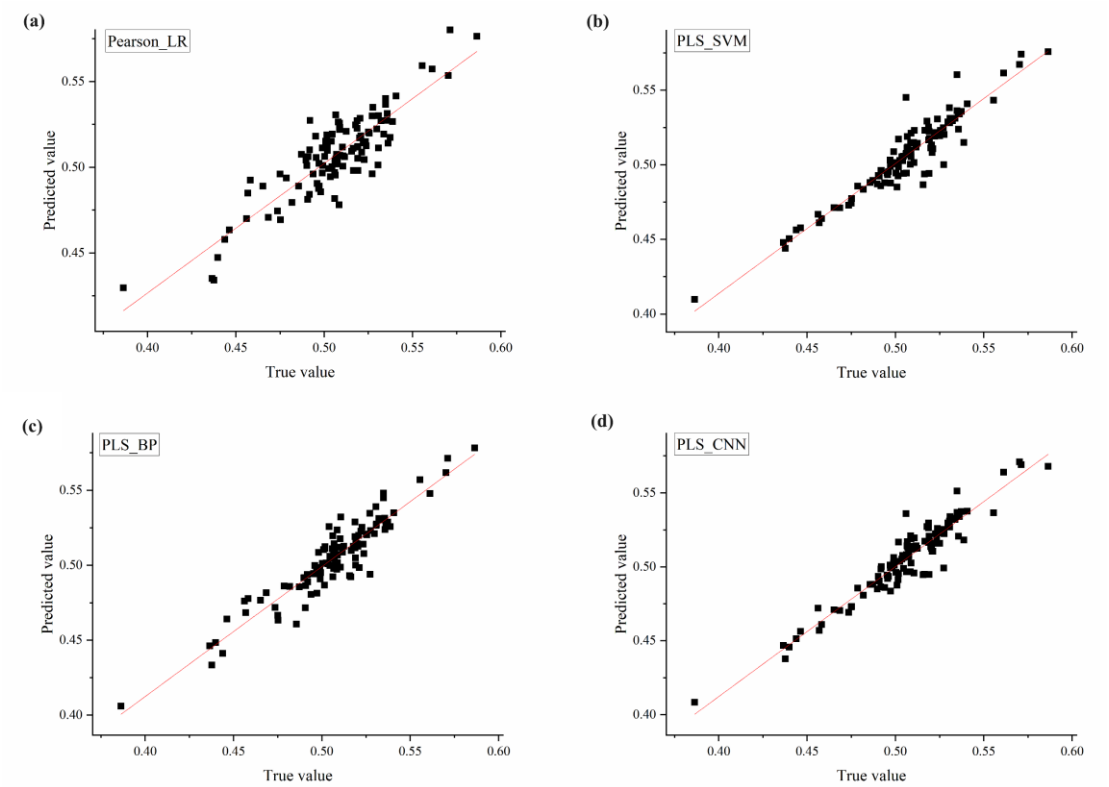

Figure 8. Results of the optimal prediction models. (a) Results of LR with Pearson, (b) results of SVM with PLS, (c) results of BP with PLS, (d) results of CNN with PLS.

Figure 8 shows the optimal model prediction results for each of LR, SVM, BP, and CNN, where the horizontal axis represents the true precipitation rate and the vertical axis represents the predicted precipitation rate. All four models obtained excellent fitting results. Figure 9 shows the performance of the best prediction model in this study – CNN with PLS.

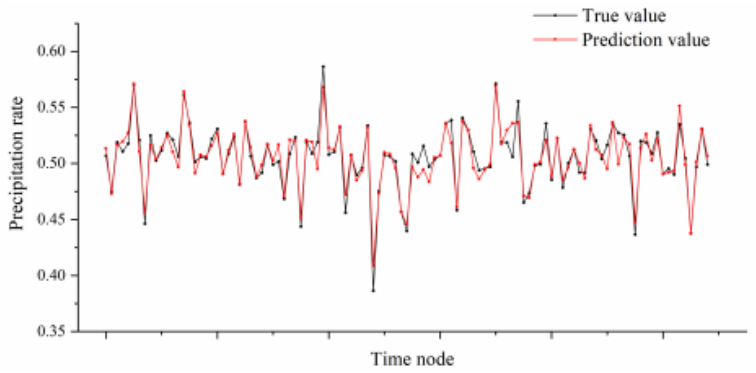

Figure 9. Prediction performance of CNN with PLS.

5. Conclusion

In this study, device and material data of the precipitation process in a production line for three months (September 2023 to November 2023) was collected, and a time-series dataset containing 535 valid data samples was obtained through data pre-processing. PCA, PLS, and Pearson were used to reduce the data dimensionality of the dataset to obtain the modelling dataset. The key features (outlet temperature of heat exchanger cold source, α_K, and Na_2O_K) were obtained through analysis of the dimensionality reduction results. Based on the modelling dataset, LR, SVM, BP, and CNN precipitation rate prediction models were established. The greatest R^2 was 0.9288, obtained by CNN with PLS. The results showed that using mass production data to establish machine learning model can accurately and timely predict the precipitation rate in the Bayer process. It provides a theoretical basis for predicting the process indicators of the entire alumina production process and lays the foundation for the intelligent production of alumina.

6. References

1. Shao Shuai, Study on Influencing Factors of Precipitation Ratio, *Proceedings of 41st International Conference of ICSOBA*, 5–9 November 2023, Dubai, UAE, *TRAVAUX* 52, 623–630.
2. Jinliang Ding et al., Data-based multiple-model prediction of the production rate for hematite ore beneficiation process, *Control Engineering Practice* 2015, 45, 219-229.
3. F. Nakhaei et al., Recovery and grade accurate prediction of pilot plant flotation column concentrate: Neural network and statistical techniques, *International Journal of Mineral Processing* 2012, 110-111, 140-154.
4. Chen Long et al., Data-driven Prediction on Performance Indicators in Process Industry: A Survey, *Acta Automatica Sinica* 2017, 43(6), 944-954.
5. Kai Gong et al., Data-Driven Prediction of Quartz Dissolution Rates at Near-Neutral and Alkaline Environments, *Frontiers in Materials* 2022, 9, 924834.
6. Ye Zhu et al., Temperature prediction of aluminum reduction cell based on integration of dual attention LSTM for non-stationary sub-sequence and ARMA for stationary sub-sequences, *Control Engineering Practice* 2023, 138, 105567.
7. Xiaoxue Wan et al., A novel multiple temporal-spatial convolution network for anode current signals classifcation, *International Journal of Machine Learning and Cybernetics* 2022, 13, 3299-3310.
8. Yongxiang Lei et al., A self-supervised temporal temperature prediction method based on dilated contrastive learning, *Journal of Process Control* 2022, 120, 150-158.
9. Marija Savic et al., ANFIS-Based Prediction of the Decomposition of Sodium Aluminate Solutions in the Bayer Process, *Chemical Engineering Communications* 2016, 203(8), 1053-1061.
10. Mohammadreza Baghoolizadeh et al., Using of artificial neural networks and different evolutionary algorithms to predict the viscosity and thermal conductivity of silica-alumina-MWCN/water nanofluid, *Heliyon* 2024, 10(4), e26279
11. Zhang Yufei et al., Precast method of digestion RSA based on BP neural network, *Light Metals* 2017, 9, 53-58.