ISBN: 978-981-18-7950-0
2023 the 13th International Workshop on Computer Science and Engineering (WCSE 2023)
doi: 10.18178/wcse.2023.06.049

Predicting Process Quality Indicators Based on a Transfer Learning Approach That Combines TCN and GRU Hybrid Networks

Bin Yi [1], Wenqiang Lin [1], Wenqi Li [1], Xiaohua Gao [1], Bing Zhou [1], Junjun Fang [2], Xiaoli Xu [2] and Jun Tang [1 +]

[1] Technology Center, China Tobacco Yunnan Industrial Co., Ltd.
[2] Affiliation Yuxi Cigarette Factory, Hongta Tobacco (Group) Co., Ltd.

Abstract. The evaluation of process quality indicators directly determines the product quality, and the prediction of process quality indicators plays a critical role in process industry product quality and production scheduling. In order to deeply explore effective information contained in massive time-series process data and improve the practicality of algorithm models under different working conditions, while ensuring the accuracy of process quality indicator prediction, this paper proposes a process quality indicator prediction algorithm based on a transfer learning approach that combines TCN (Temporal Convolutional Network) and GRU (Gated Recurrent Unit) hybrid networks. The algorithm constructs a continuous feature matrix as input by sliding windows over massive process data, operating data, and time data in a sequential manner. In terms of model construction, TCN is used to extract temporal features from sequence data and GRU is used to capture temporal dependencies in sequence data. To address the problem of training efficiency in actual production, a transfer learning strategy is introduced which greatly improves training efficiency while maintaining training accuracy. Furthermore, a case study using process data from a micro-cigarette production line in a tobacco factory verifies the correctness and effectiveness of the proposed algorithm.

Keywords: Process industry; TCN-GRU; Hybrid model; Transfer learning

1. Introduction

In modern industrial production, process manufacturing is an important production method with characteristics of good production continuity, numerous equipment, and complex coupling between variables. In order to ensure stable operation of the process manufacturing industry, real-time prediction and control of production quality have become a key issue [1-2]. With the continuous development of industrial Internet technology, a large amount of historical process data has been collected and accumulated, so data-based algorithms for predicting product quality in process production have been widely studied [3]. However, at present, the predictability, linkage, and autonomy levels of all process factors in the process manufacturing industry are still relatively low, which seriously restricts further improvement of the level of process manufacturing production. Therefore, accurate prediction algorithms for high-dimensional and multiscale process data are critical to supporting the smooth operation of the workshop [4-5].

Currently, scholars both domestically and abroad have conducted beneficial research on the accurate prediction of process quality in manufacturing industry production processes. Among them, reference [6] established a thermal rolling TRIP steel mechanical property prediction model based on an adaptive neural network and fuzzy inference system combination and verified through experiments that the model can accurately predict quality indicators such as tensile strength, yield strength, elongation, and retained austenite under given operating conditions. Reference [7] used a multi-output support vector regression method to simultaneously predict multiple quality indicators, conduct collaborative design of multiple process parameters, and realize concomitant optimization of process parameters to meet multi-objective quality requirements. Reference [8] trained a convolutional neural network and support vector machine combined with automotive instrument assembly process data collection of production data and effectively characterized the pointer deflection angle of the instrument quality, making accurate quality predictions. Reference [9]

[+] Corresponding author. Tel.: + 18208738630
E-mail address: juntang2013@163.com

proposed a meta-learning-based multi-confidence deep neural network that uses the immense potential of deep neural networks and meta-learning theory in high-dimensional information extraction and approximation, resulting in significant improvements in prediction accuracy and training efficiency. In addition, references [10-12] used a deep neural network with the introduction of time series in processing production problem predictions. These studies have found the importance of time series in data mining through analysis of different working conditions and further explored the superiority of recurrent networks in dealing with sequential tasks. The research results show that deep learning models have broad prospects for application in process manufacturing industrial production scenarios.

The research conducted by the aforementioned authors has achieved numerous successes in the corresponding research field. However, there are still some deficiencies when it is applied to the industrial scenario of process manufacturing, including: (1) insufficient considerations on the characteristics of the time series, non-linearity and temporal dependencies of production data in the process manufacturing industry; (2) the prediction accuracy and training efficiency of algorithms are problems that cannot be ignored in practical application.

Based on the above analysis, we discovered that production data in the process manufacturing industry has the characteristics of time series, non-linearity and temporal dependencies of process parameters. In order to improve accuracy in predicting quality indicators, a novel process quality indicator prediction algorithm is proposed. This algorithm integrates the advantages of TCN (Temporal Convolutional Networks) and GRU(Gated Recurrent Unit) networks. Specifically, the algorithm first eliminates non-steady-state data based on the characteristics of process production, then uses the TCN network to effectively extract the potential relationship between continuous and non-continuous data to form feature vectors, which are then used as input for the GRU network to achieve the prediction of quality indicators. To address the problem of low efficiency of the algorithm in actual production scenarios, the transfer learning strategy was introduced to significantly reduce training time while ensuring training accuracy. In a prediction experiment on the silk-making process in a certain workshop, our algorithm performed better than other time-series analysis algorithms, showing significant improvements in MAE and RMSE compared to state-of-the-art algorithms such as CNN-LSTM and CNN-BiGRU, with an improvement in fitting degree of about 3%.

2. Model Selection Analysis

In process manufacturing, the complexity of production line structure and equipment conditions increases the difficulty of prediction. Data features are becoming more diverse, covering multiple processing operations, complex processes, and various quality influencing factors. These data reveal the correlation and influence relationships between operational parameters, process parameters, and quality indicators. Meanwhile, changes in material flow, information flow, and energy flow have an impact on production data during circulation. In addition, the sampling period causes a periodic relationship between parameters and acquisition time, showing significant time dependency. Therefore, these characteristics must be considered when constructing process manufacturing prediction models. The time series of multiple processes makes each process essentially have complex correlational properties. In actual production situations, the long training time of hybrid network algorithms is a common problem. Furthermore, as production time elapses and process data changes, the accuracy of prediction models may gradually decline. Therefore, when establishing prediction models, these factors must be taken into account and corresponding optimization strategies should be adopted to improve model efficiency and accuracy.

Based on the above analysis, this article proposes a novel hybrid network algorithm based on transfer learning that integrates TCN and GRU for predicting quality indicators in process manufacturing. The primary improvements of this algorithm include:

(1) The hybrid neural network model introduces multiple different sizes of convolution kernels of time convolutional networks to extract multiscale features, better capture the changes in process sequence data, and increase residual connections and dilated convolutions to handle long-term dependencies, avoiding the gradient disappearance problems in traditional RNN and GRU.

(2) The hybrid neural network model introduces gated recurrent units to extract temporal dependencies of process parameters and mine inner hidden information of process data.

(3) The hybrid neural network introduces transfer learning strategy, enabling the model to quickly obtain useful information from the trained process manufacturing quality prediction algorithm, reducing training time and data size for new tasks, improving learning efficiency and accuracy, and enhancing the generalization ability and prediction performance of the model.

3. Transfer Learning-based TCN-GRU Hybrid Network Model

3.1. Quality indicator prediction model

To improve the fitting accuracy of the prediction model, historical process data in the collected data are processed using a smoothing method to eliminate data noise. Based on this, a quality indicator prediction algorithm for process manufacturing based on transfer learning that integrates TCN and GRU is proposed to address the time series, nonlinearity, and temporal dependency characteristics of production data in the process manufacturing industry. The algorithmic model structure is shown in Figure 1, which includes three parts: constructing a TCN network to extract multiscale process features better, capturing changes in process sequence data; constructing a GRU network module to extract temporal dependencies of process parameters; building a transfer learning strategy to promote model training speed.

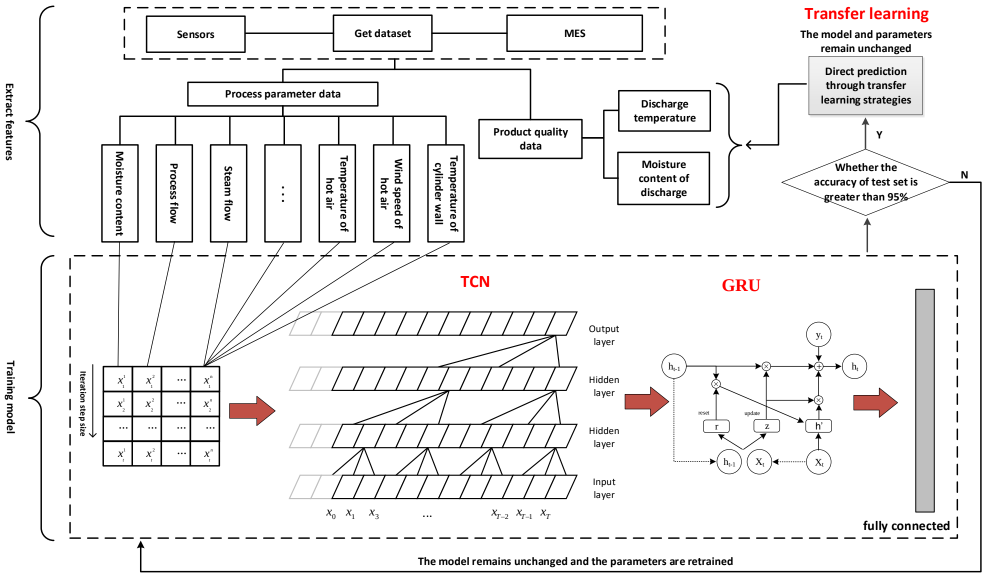


Fig. 1: Framework Diagram of Hybrid Network Model

3.2. Building Input and Output Data

Let the quality indicator sequence at any time in the process workshop be represented by $Y = (y_1, y_2, \dots, y_m)$, and the corresponding temporal data of the key process parameters be represented by $X = (x_1, x_2, \dots, x_m) = (x^1, x^2, \dots, x^n)^t$, which can be expanded into Equation (1). That is, taking n process parameters and t data points as input, and taking one quality indicator parameter as output, the original data is transformed into time series data.

$$X = \begin{bmatrix} x_1^1 \dots x_1^n \\ \dots \dots \dots \\ x_t^1 \dots x_t^n \end{bmatrix} \quad (1)$$

Here, Y represents the quality indicators, m represents the number of quality indicators, X represents the process parameters, and n and t represent the number of process parameters and the iteration step length between times, respectively.

3.3. TCN Neural Network Model

When considering the problem of long-distance impact in the process manufacturing process, the TCN network is introduced to perform convolution operations on the entire history of time series data, thereby capturing long-term dependency relationships in process time series data. The core idea of the TCN network is to capture local features at different time scales through convolutional operations while maintaining causality. In TCN, causal convolution is used and its convolution operation can be represented as:

$$x(t) = \sum\{j = 0\}^{\{k - 1\}} w(j) * x(t - j)) \qquad (2)$$

Here, $x'(t)$ represents the value of the input sequence at time step t, $w(j)$ represents the convolution kernel parameter, and k represents the size of the convolution kernel. Through causal convolution, we can ensure that the output sequence $x(t)$ depends only on the current and previous time steps t and t-j and not on future time steps.

In addition, to increase the network's receptive field and capture larger ranges of time series features, TCN typically uses dilated convolution. Specifically, when computing the output sequence $x(t)$, the input data points considered by dilated convolution are distributed at fixed intervals d. The dilated convolution formula is as follows:

$$x(t) = \sum\_\{j = 0\}^{\{k - 1\}} w(j) * x(t - j * d) \qquad (3)$$

Here, d is the dilation factor. When d=1, dilated convolution degenerates into ordinary convolution. By adjusting the value of d, we can control the network's ability to capture features at different time scales.

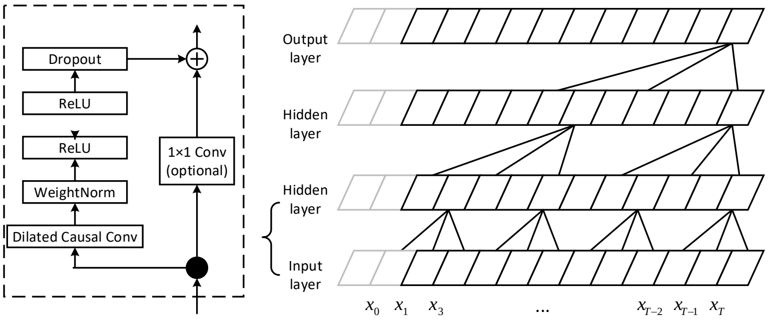


Fig. 2: TCN Network Architecture

3.4. GRU Neural Network Model

When considering the problem of long-distance impact in the process manufacturing process, the TCN network is introduced to perform convolution operations on the entire history of time series data, thereby capturing long-term dependency relationships in process time series data. The core idea of the TCN network is to capture local features at different time scales through convolutional operations while maintaining causality. In TCN, causal convolution is used and its convolution operation can be represented as:

Considering the strong correlation among time series data in process manufacturing lines, using recurrent neural networks to process deep time-sequenced data is highly efficient and can deeply mine the temporal correlation information in process data. The gated recurrent unit (GRU) as a type of recurrent neural network adopts gate mechanisms to control the flow and retention of process data information, thereby avoiding the problem of gradient disappearance caused by long-term dependencies. Meanwhile, through gate control, the model can selectively forget or preserve historical process data information, thereby learning long-term dependency relationships.

The basic unit of the GRU network contains update gates and reset gates in the basic unit of the more GRU network. The input x_t in the update gate together with the state memory unit h_t and the intermediate output h' determine the state memory unit to forget and retain information. The input x_t in the input gate is varied by the sigmoid and tanh functions respectively to jointly determine the retention vector in the state memory unit. The intermediate output h' is jointly determined by the reset gate r_t and is calculated as shown in equation (2-7).

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \qquad (4)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \qquad (5)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h' \qquad (6)$$

$$h' = tanh(W_{h'}[r_t \cdot h_{t-1}, x_t] + b_{h'}) \tag{7}$$

Here, z_t is the update gate, r_t is the reset gate, h_t is the memory, and h' is the candidate hidden layer for controlling forgetting and saving information. W_z, W_r, $W_{h'}$ is the weight matrix, b_z, b_r, $b_{h'}$ represents the bias value, and $\cdot$ is the dot product.

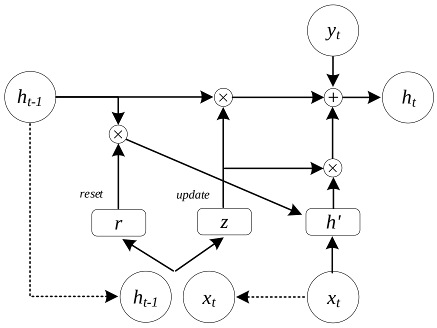


Fig. 3: GRU Network Architecture

3.5. Transfer Learning Strategies

Since in process manufacturing, the same processes in different production lines exhibit high similarity while there are significant differences in the production process data of the same processes at different times. In addition, deep neural network models require a long training time to mine the deep hidden features of process parameters. Therefore, this paper proposes to apply transfer learning to the training process of deep neural network models to enable the algorithm to adapt to the same process in different production lines and different time periods. The transfer learning method defined here includes retraining and transfer learning, which are based on the accuracy of the model's validation set.

Retraining (Model1) is used when the accuracy is below 96%. The network structure of the pre-trained model remains unchanged, all layers' network parameters are randomly initialized, and the model is retrained on the target domain after re-dividing the dataset.

Transfer learning (Mode2) is used when the accuracy is equal to or greater than 96%. This method shares the entire network structure and parameters of the pre-trained model, using the pre-trained model parameters as the initial values for the new model parameters, thereby improving the learning effect in the target domain.

4. Experimental Verification Analysis

4.1. Experimental Design

To verify the superiority of the CNN-GRU transfer learning algorithm model in predicting process parameters, a DELL workstation was used as the hardware platform in this study. The workstation is equipped with an Intel(R) Core (TM) i7-12600KF processor, 32.0GB RAM, and an RTX3060ti graphics card, and it runs on the Windows 10 operating system. In terms of software, Python programming language was used together with the open-source machine learning library Tensorflow and a GPU-supported computing engine to implement the application of the algorithm model.

4.2. Data Source and Preprocessing Rules

This study is based on the process data of the thin-plate drying process in Yuxi Cigarette Factory. A total of more than 200,000 data points from 54 batches were used as the data source to analyze and identify the impact of key process parameters on quality indicators through comparative analysis of various deep learning algorithms. An accurate correlation model between the cigarette-making process parameters and quality indicators was constructed. Based on the weight analysis results, 20 process parameters and 2 quality indicators were selected as the algorithm's dataset, including process parameters such as temperature measurement of cylinder wall (middle), negative pressure of dehumidification, steam flow rate of inlet HT, thin-plate condensate water temperature, actual value of material moisture content at inlet, HT steam pressure, and quality indicators such as discharge temperature and discharge moisture content.

The data was sourced from a cigarette manufacturing production line of a process manufacturing company, covering 54 batches from July to November 2022, with a total of over 200,000 records. As process manufacturing involves continuous production processes, special attention was paid to cleaning, raw materials, and downtime during the data preprocessing stage. When calculating non-steady-state indices, the head and tail of the data were trimmed according to Table 1 and followed certain rules for data deletion.

Table 1: Non-stationary index calculation data interception rules

Determining factors	Start counting conditions	End point counting conditions
Material accumulation	The first point where there is no change in the accumulated volume	The last point at which there is no change in the accumulated volume
Data at different moments	The value corresponding to 5 minutes with no change is the first point after that period	The value corresponding to a change in quantity is the first point where the change occurs
Head and tail data	Synchronization with the starting point of moisture content at the process exit	Synchronization with process inlet material flow end point

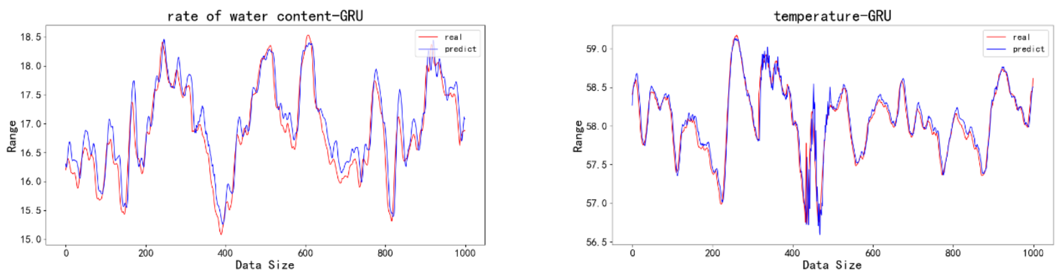
4.3. Melting Experiment

As for the parameters of the TCN-GRU neural network model, 20 process parameters and a time series length of 6 were chosen as inputs, and the data was split into training and test sets at a ratio of 7:3. The TCN network consists of 5 layers, with a filter size of 64 in the convolution layer and a convolution kernel size of 3. The TCN blocks are stacked once, and the distance between each element in the convolution kernel and other elements is set to (1, 2, 4, 8, 16). The Bi-LSTM network has 3 layers, with 128, 16, and 2 neurons respectively, and outputs the quality indicator prediction values through a fully connected layer. The batch size was set to 256, and the number of iterations was set to 30.

Table 2: Comparison of prediction results error of the model

Algorithm	Quality indicators	MSE	MAE	R2
GRU	Rate of water content	0.066648	0.200774	0.902
	Temperature	0.019266	0.143764	0.899
TCN	Rate of water content	0.010848	0.086440	0.978
	Temperature	0.006680	0.051526	0.969
TCN-GRU	Rate of water content	0.009108	0.055419	0.985
	Temperature	0.005522	0.041675	0.978

To verify the effectiveness of the model network structure in improving prediction accuracy, a melting experiment was designed with two comparison models, the TCN and GRU networks. Under the same experimental conditions, the above models were trained, and the current predicted values and actual values of the two quality indicators were compared. The prediction performance of the models was evaluated using three indicators: mean absolute error, mean squared error, and coefficient of determination (R2). The experimental results are shown in Table 2 and Figure 4. From the comparative results in Table 2, it can be seen that the proposed TCN-GRU neural network algorithm has the best prediction performance, with a fitting degree of 0.985 and 0.978 for the two quality indicators, respectively, and the fitting values of the algorithm are higher than those of other models. The mean squared errors are 0.009108 and 0.005522, respectively, and the error values are smaller than those of other models, proving the superiority of the neural network model's prediction performance and the effectiveness of the network structure.



(a) (b)

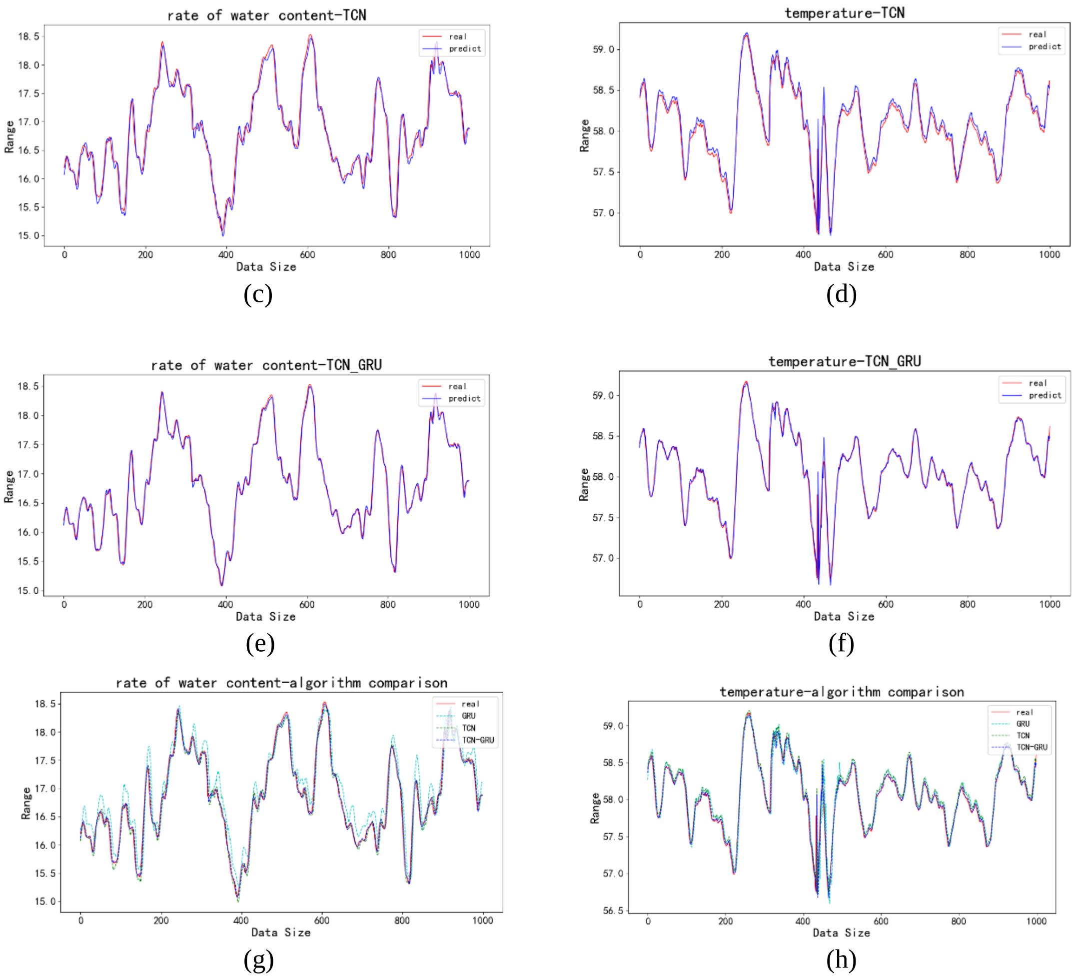


Fig. 4: Comparison of predicted values by different algorithm

In Figure 4, to more clearly see the fitting accuracy of the three methods, this paper selected 1000 data points for demonstration. Figures a and b represent the comparison between the predicted and actual values of the quality indicators using GRU; Figures c and d represent the comparison using TCN; Figures e and f represent the comparison using the proposed model; and Figures g and h represent the comparison among all four networks. From the graph, it can be seen that the TCN-GRU has the best fitting effect.

4.4. Comparative Experiment

This section designed a comparative experiment to compare the influence of network structures on algorithm accuracy between the CNN-LSTM network [13] and the CNN-BiGRU network [14]. Under the same experimental conditions, the current predicted values and actual values of the two quality indicators were compared, and the model's prediction performance was evaluated using three indicators: mean absolute error, mean squared error, and coefficient of determination (R2). The experimental results are shown in Table 3, which demonstrate that the TCN-GRU neural network algorithm had the best prediction performance, with fitting degrees of 0.985 and 0.978 for the two quality indicators, respectively. The mean squared errors were 0.009108 and 0.005522, respectively, and the error values were smaller than other models, proving the superiority of the neural network model's prediction performance and the effectiveness of the network structure.

Table 3: Comparison of prediction results error of the model

Algorithm	Quality indicators	MSE	MAE	R2
CNN-LSTM	Rate of water content	0.038534	0.153924	0.938
	Temperature	0.014871	0.090343	0.923
CNN-BiGRU	Rate of water content	0.024067	0.113712	0.961
	Temperature	0.007871	0.060343	0.953

Algorithm	Quality indicators	MSE	MAE	R2
TCN-GRU	Rate of water content	0.009108	0.055419	0.985
	Temperature	0.005522	0.041675	0.978

In Figure 5, to more clearly observe the fitting accuracy of the three methods, this paper selected 1000 data points for demonstration. Figures a and b represent the comparison between predicted and actual values of the quality indicators using CNN-LSTM; Figures c and d represent the comparison using CNN-BiGRU; Figures e and f represent the comparison using the proposed model; and Figures g and h represent the comparison among all four networks. From the graph, it can be seen that the TCN-GRU has the best fitting effect.

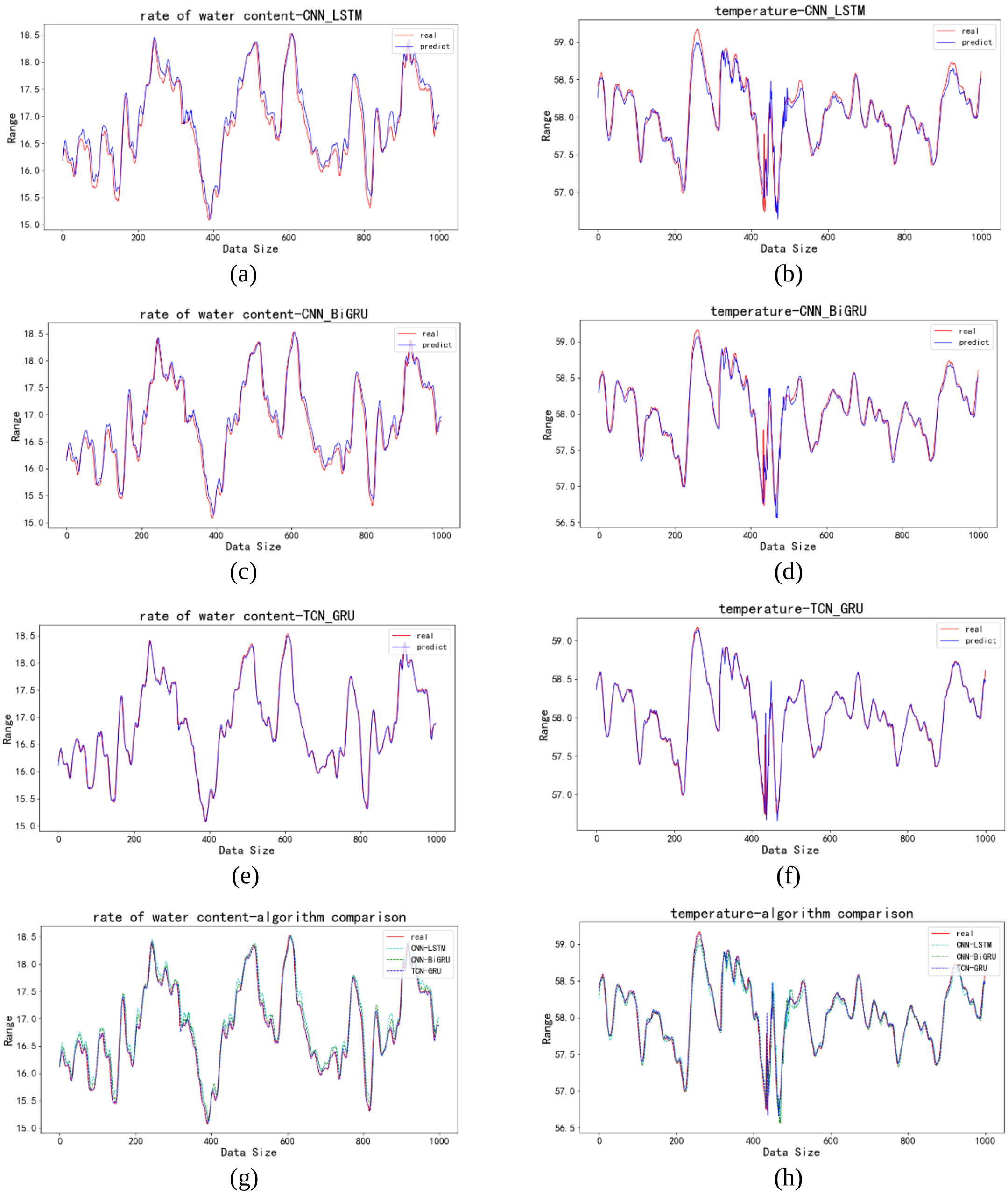


Fig. 5: Comparison of predicted values by different algorithm

5. Conclusion

To meet the high accuracy and timeliness requirements for process parameter regulation in complex production processes, this paper proposes a process manufacturing quality indicator prediction method based on transfer learning fusion of a Time Convolutional Network (TCN) and Gated Recurrent Unit (GRU). This method aims to achieve accurate prediction of processing quality at different time points during the same process. The research results show that the proposed transfer learning fusion method of TCN and GRU can accurately predict process quality by extracting temporal coupling features under equipment parameter changes and material variations, with an average accuracy of 0.985 and 0.978, respectively. In addition, due to the constantly changing processing conditions of the production line, the proposed transfer learning method significantly reduces model training time while ensuring accuracy, meeting the real-time requirements for process manufacturing quality prediction and parameter regulation. The results of this study are of great significance for improving the production efficiency and process quality of process manufacturing lines, and can be applied and promoted in relevant process manufacturing industries. In future research, external factors such as material source fluctuations will be taken into consideration to further improve the robustness of process quality prediction.

6. Acknowledgments

This research was supported by Key Scientific and Technological Projects of Yunnan China Tobacco Industry Co., Ltd (Grant no. 2022GY02)

7. References

[1] CRH Márquez, Ribeiro C C . Shop scheduling in manufacturing environments: a review[J]. International Transactions in Operational Research, 2022, 29(6):3237-3293.

[2] Jwab C , Cx D , Jie Z , et al. Big data analytics for intelligent manufacturing systems: A review[J]. Journal of Manufacturing Systems, 2021.

[3] Chaudhry I A , Khan A A . A research survey: review of flexible job shop scheduling techniques[J]. International Transactions in Operational Research, 2015, 23(3):551-591.

[4] Kai, Zhong, Min, et al. Data-Driven Based Fault Prognosis for Industrial Systems:A Concise Overview[J]. IEEE/CAA Journal of Automatica Sinica, 2020, v.7(02):19-34.

[5] Jiao K , Wang S . Research on Computer Information Processing Technology in the Era of Big Data[J]. Modern Industrial Economy and Informationization, 2016(iceiti).

[6] S. Hore, S.K. Das, S. Banerjee, and S. Mukherjee, An adaptiveneuro-fuzzy inference system-based modelling to predict mechanicalproperties of hot-rolled TRIP steel[J]. Ironmaking Steelmaking,2017,44(9):656-665.

[7] Yi-fan Yan,Zhi-min Lü.Multi-objective quality control method for cold-rolled products oriented to customized requirements[J].International Journal of Minerals Metallurgy and Materials,2021,28(08):1332-1342.

[8] Burbaum B , Jokisch T . METHOD FOR PRODUCING A GAS TURBINE COMPONENT:, EP3525962A1[P]. 2019.

[9] Elbaz K , Shen S L , Zhou A , et al. Prediction of Disc Cutter Life During Shield Tunneling with AI via the Incorporation of a Genetic Algorithm into a GMDH-Type Neural Network[J]. Engineering, 2021, 007(002):P.238-251.

[10] Shen Z , Zhang Y , Lu J , et al. A novel time series forecasting model with deep learning[J]. Neurocomputing, 2020, 396:302-313.

[11] Bai Y , Xie J , Wang D , et al. A manufacturing quality prediction model based on AdaBoost-LSTM with rough knowledge[J]. Computers & Industrial Engineering, 2021, 155(5):107227.

[12] Lara-Benitez P , Carranza-Garcia M , Riquelme J C . An Experimental Review on Deep Learning Architectures for Time Series Forecasting[J]. International Journal of Neural Systems, 2020.

[13] Qi X , Zheng X , Chen Q . A short term load forecasting of integrated energy system based on CNN-LSTM[C]// 2020:01032.

[14] Niu D , Yu M , Sun L , et al. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism[J]. Applied Energy, 2022, 313:118801-.