Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 13, No. 3, 2024

Research on CNN-LSTM-Attention based surface temperature prediction model for fermented grains

Ao Zhang [1, 2], Xianguo Tuo [1, 2] *, Rui Yao [1, 2]

[1] School of Automation and Information Engineering Sichuan University of Science & Engineering, Yibin 644000, China
[2] Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China

* **Corresponding author**: Xianguo Tuo (Email: 2533606401@qq.com)

Abstract: The prediction of the temperature on the surface of the fermented grains during the steaming process holds significant importance in improving the alcohol yield and quality. To accurately forecast the temperature variations on the surface of the fermented grains, a predictive model based on CNN-LSTM-Attention is proposed. Leveraging convolutional neural networks, latent features of the temperature data on the fermented grains surface are extracted. These extracted feature vectors, representing a time series, are then inputted into a long short-term memory network to further capture the temporal characteristics of the sequence. Finally, employing an attention mechanism, the influential features are highlighted to achieve precise prediction of the fermented grains surface temperature. Historical temperature data from the steaming process of the fermented grains is utilized, and experimental comparisons are conducted with other neural network prediction models. The results demonstrate that the CNN-LSTM-Attention model achieves optimal root mean square error of 1.282 and average absolute error of 0.647, surpassing other models. The experimental findings substantiate the superior accuracy of the CNN-LSTM-Attention model in forecasting the changing trend of the fermented grains surface temperature.

Keywords: Baijiu brewing process; Prediction of fermentation grains temperature; Attention mechanism; CNN; LSTM architecture.

1. Introduction

Chinese Baijiu industry boasts a profound history of brewing, dating back thousands of years, making it one of the most traditional alcoholic beverages in China. Throughout the extensive course of time and continuous innovation, Chinese Baijiu has gradually developed a distinctive solid-state fermentation process [1]. Solid state brewing is the process of gradually placing the spirits into retort barrels, where the ethanol and other aromatic substances are extracted by distillation [2]. The key to producing premium-quality liquor lies in the quality of the distillation process, which, in turn, depends directly on the excellence of the steaming process [3]. The steaming process is an indispensable and crucial step in the entire brewing process. It determines the state and quality of the raw materials during distillation and significantly influences the taste and quality of the liquor. Therefore, to produce high-quality liquor, one must master the steaming process, carrying out the brewing process meticulously and scientifically to ensure the quality of the distilled liquor. The process of distilling liquor is the process of spreading the fermented grains little by little and evenly into the retort barrels for distillation. [4]. In the past, this task was entirely performed manually, as workers would swiftly distribute the fermentation grains in the area where steam emerged to avoid any losses caused by the wine steam [5]. In recent years, many baijiu companies have started to use retorting robots for retorting operations, for example, Wuhan Fenjin Intelligent Machine Co, Ltd. [6]. Robot retorting is done through infrared thermal imager to obtain the surface temperature of the fermentation grains to determine the point of steam, i.e., the retorting robot can only "see the steam" on the fermentation grains, and cannot make the decision of spreading the material before the steam comes out to further reduce the wine loss [7, 8]. By accurately predicting when and where wine vapor will emerge, robots can plan their paths in advance, thereby reducing wine losses. Therefore, accurate prediction of the time and area of wine vapor emergence is important for improving both wine yield and wine quality.

Predicting the surface temperature of fermented grains is essentially a time series prediction problem, and it is necessary to predict the temperature value in the future according to the time series in the historical data. Currently, time series forecasting methods encompass statistical methods, machine learning methods, and their combinations, including linear prediction methods, nonlinear model-based methods, and deep learning-based methods. Representative models based on linear prediction methods include autoregressive models (AR models), moving average models (MA models), and their combination, autoregressive moving average models (ARMA models). These models all require that temperature data be stationary and cannot guarantee prediction accuracy in the face of complex temperature changes. Nonlinear prediction methods include support vector machine regression models (SVM), artificial neural network models (ANN) [9], decision tree models (DT) [10], among others. These models exhibit stronger fitting capabilities compared to linear models and can capture more complex data relationships. However, they suffer from high computational costs, poor interpretability, and a need for large amounts of data. Deep learning-based prediction methods, in contrast to previous approaches, possess enhanced feature extraction capabilities, making them well-suited for the task of predicting fermented grains temperature. Given the dynamic nature of the surface temperature, recurrent neural networks (RNNs) are particularly suitable for modeling the temporal changes in the fermented grains surface, as RNNs can handle long input sequences, thus ensuring the ability to learn time series [11]. However, traditional RNNs have drawbacks such as long-term dependencies, which can lead to issues like "exploding gradients" and "vanishing gradients." The introduction of long short-term memory (LSTM) to some extent alleviates these problems encountered during RNN training [12]. Haoshi Lin et al. [13] utilized an LSTM model based on an attention mechanism to predict the changes in the fermented grains surface temperature and optimized the hyperparameters within the LSTM network using Bayesian algorithms, achieving favorable prediction results. However, this model has long computation time and poor real-time performance. Zhang Jie et al. [14] employed an improved GRU model to predict the fermentation steam, which slightly reduced computation time but did not significantly enhance accuracy. Current research has shown promising results for deep learning in predicting fermented mash temperature data, but existing prediction models are too singular, mainly focusing on incorporating attention mechanisms, while rarely considering composite models and optimizing the model structure to improve prediction accuracy.

2. Principles of the CNN-LSTM-Attention Model

2.1. CNN Model

CNN, known as Convolutional Neural Network, is a deep neural network architecture widely applied in domains such as image processing and natural language processing. With its characteristics of weight sharing and local connectivity, CNN efficiently extracts feature information, reducing model complexity and modeling parameters. This implies that CNN can effectively handle large-scale data, improve computational efficiency, and possess excellent generalization capabilities [15].

CNN consists of convolutional layers, pooling layers, and fully connected layers. The convolutional layer performs convolutional computations on input data, effectively extracting data features. The pooling layer reduces the dimensionality of input features, preserving important information while reducing computational complexity and helping prevent overfitting. The fully connected layer further integrates information from the convolutional and pooling layers, mapping it to output signals through the output layer [16]. Figure 1 illustrates the network architecture of a one-dimensional CNN.

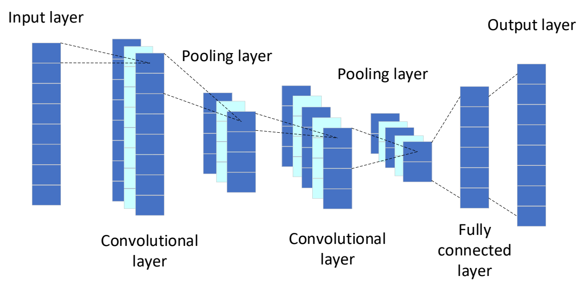

Figure 1. Structure of CNN

2.2. LSTM Architecture

LSTM (Long Short-Term Memory) is an improved variant of recurrent neural networks (RNN) originally proposed by Hochreiter and Schmidhuber [17]. It introduces gating mechanisms and memory units, effectively addressing the issues of gradient vanishing, gradient exploding, and long-term dependencies in RNN. In capturing the long-term dependencies of time series, LSTM demonstrates remarkable performance. By incorporating three types of gating units (input gate, forget gate, and output gate) and a memory unit, LSTM controls the flow of information. The input gate determines whether new input information should be integrated into the memory unit, the forget gate determines which old memories should be forgotten, and the output gate determines the contents of the outputted memory unit. This gating mechanism enables LSTM to selectively update and propagate information, thus capturing long-term dependencies in time series more effectively [18]. The structural diagram of an LSTM unit is depicted in the accompanying Figure 2.

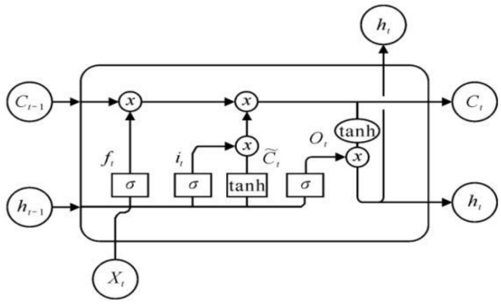

Figure 2. Structure of LSTM network memory cell

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

$$i_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_t) \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{5}$$

$$h_t = O_t \times \tanh(C_t) \tag{6}$$

where:$h_{t-1}$is the hidden state at time t - 1 ;h_t is the hidden state at time t ;C_t is the cell state at moment t ;$\tilde{C}_t$ denotes the immediate cell state at moment t ;x_t is the input vector at time t ;i_t, f_t, O_t are the input gate output, forget gate output at time t respectively and output gate output ;σ is the Sigmoid activation function ;W and b denote the weight matrix and bias, respectively.

2.3. Attention Mechanism

The attention mechanism [19] is a technique that allows for the identification of relevant segments within an input sequence while generating corresponding parts in the output sequence. The merits of self-attention lie in its ability to capture intricate relationships between different segments in the input sequence, without the need for a fixed local attention window like traditional convolutional neural networks. This attribute makes it highly effective in handling sequence data tasks. By simulating the human brain's allocation of attentional resources, the attention mechanism focuses on crucial information, highlighting the impact of important details, thereby enhancing the accuracy of the model. This mechanism effectively improves the LSTM network's performance by mitigating information loss caused by excessively long sequences.

3.Predictive model

The prediction model network structure is shown in Figure

3, The vector of the highest temperature, lowest temperature, average temperature, steam pressure, humidity of fermented grains, laying thickness, height and formula of distiller grains in the process is input, The vector extracts features through convolution operation and LSTM layer, then realizes multi-dimensional feature weight extraction through the attention mechanism layer, and finally enters the fully connected layer and outputs the temperature prediction results.

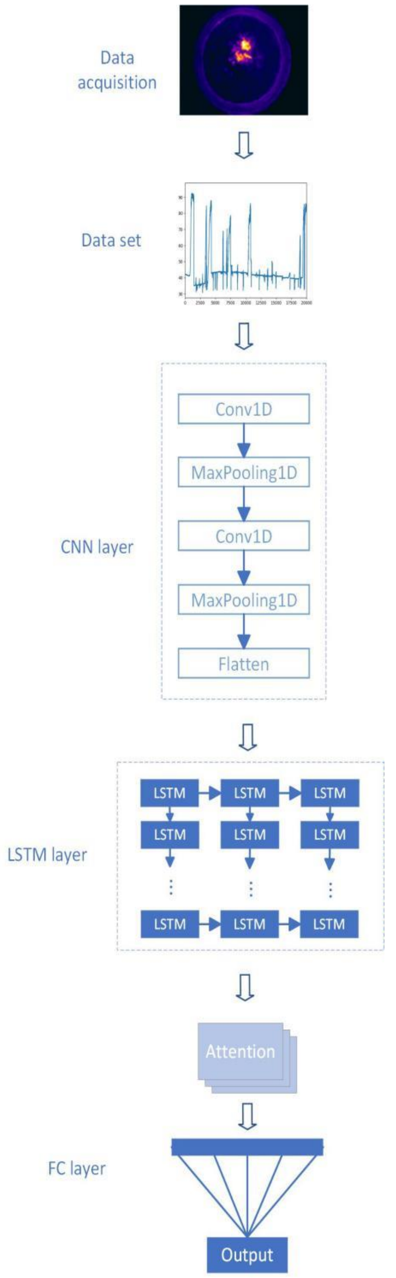

Figure 3. CNN-LSTM-Attention fermented grains temperature prediction network model

4. Experimental Validation and Result Analysis

4.1. Sample Processing

The experiment utilized data derived from 108 fermentation videos and additional parameters collected during the brewing process of a certain winery. These parameters include steam pressure, humidity of fermented grains, thickness of paving material, retort height and distiller's grains formula. To illustrate, images were extracted from the videos at a frequency of one image every 10 frames. For instance, in a 40-minute fermentation video, approximately 2400 image data points were generated. Then, in the picture with the steamer pot as the diameter external rectangle, the rectangle is divided into regions according to 6x6, and the highest temperature value, average temperature value and minimum temperature value of each area are derived. However, due to the lack of temperature data in the four corners of the divided regions, these data points were excluded. Consequently, a new time series temperature data set, denoted as X={X1,X2,……,X32}, was obtained as depicted in Figure 4.

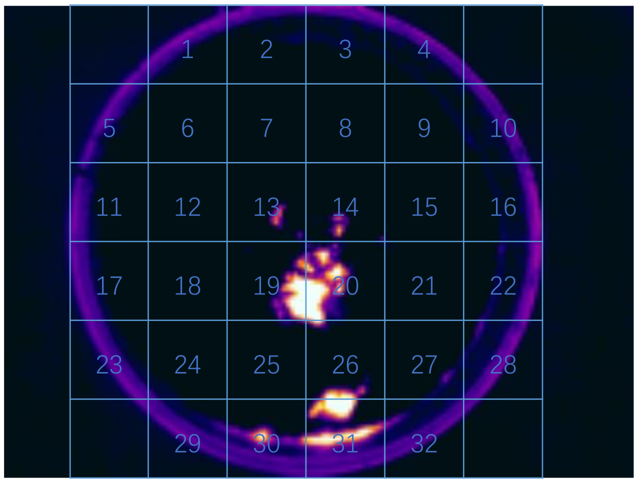

Figure 4. Surface temperature image of wine spirit

Training the model with unprocessed raw data leads to reduced training speed and compromised prediction results. To mitigate the impact of raw data and enhance prediction progress and model convergence, this study employs the method of maximum-minimum normalization to preprocess the temperature data, constraining it within the range of [0, 1]. The formula for this normalization process is as follows:

$$x' = \frac{x-\min(x)}{\max(x)-\min(x)} \tag{7}$$

In the given equation, x represents the original data, x' denotes the normalized data, min(x) represents the minimum value within the data samples, and max(x) represents the maximum value within the data samples. The original time series data of the fermented grains temperature and the corresponding normalized data are depicted in Figure 5 and Figure 6, respectively.

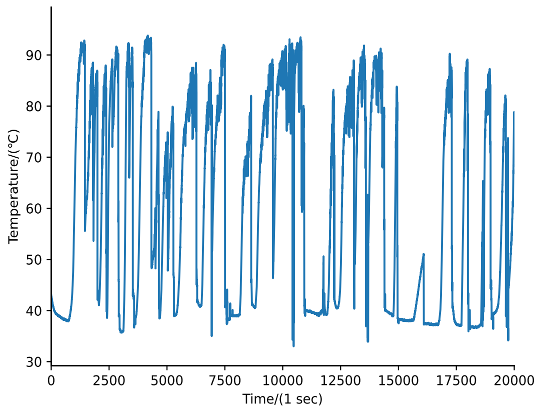

Figure 5. Original temperature

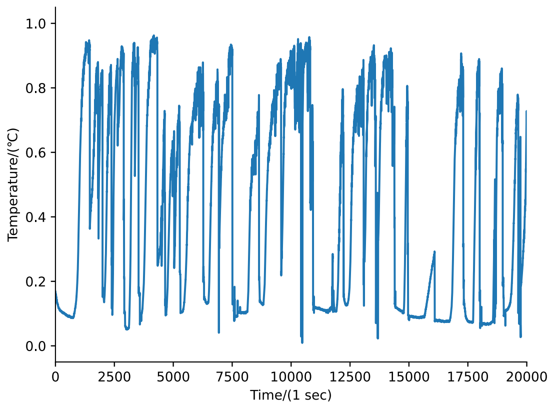

Figure 6. Normalized temperature

4.2. Evaluation metrics

In order to evaluate the prediction accuracy of the model, the root mean square error, mean absolute error and certainty coefficient are used as the model evaluation indexes, and the calculation formula of each index is as follows:

Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N}(y_i - \hat{y}_i)}{N}}, \in [0, +\infty) \tag{8}$$

Absolute Mean Error (MAE):

$$\mathrm{MAE} = \frac{\sum_{n=1}^{N}|y_i - \hat{y}_i|}{N}, \in [0, +\infty) \tag{9}$$

Correlation coefficient (R2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}, \in [0, +\infty) \tag{10}$$

In the equation, n represents the number of samples; yi denotes the true value of the temperature of the wine residue; $\hat{y}_i$ signifies the predicted value of the temperature of the wine residue; and $\bar{y}_i$ represents the mean value of the wine residue temperature.

4.3. Model Parameter Configuration

The proposed predictive model, CNN-LSTM-Attention, in this paper has a time step of 12. This means that it takes the historical temperature data of the previous 12 seconds as input and predicts the temperature data for the next 1 second. The model's configuration is based on empirical knowledge and is continuously adjusted based on the obtained results. The specific parameter settings are as follows: two convolutional layers with ReLU activation function, both with 64 filters and a filter size of 2; two max pooling layers with a pool size of 6; one fully connected layer; three LSTM network layers with 128, 64, and 32 units respectively; an attention layer incorporated, followed by a final connection to the fully connected layer. The batch size is set to 256, each epoch is set to 200 iterations, mean squared error is used as the objective function, and the optimization algorithm employed is the adaptive learning rate Adam algorithm.

4.4. Analysis of Experimental Results

To validate the accuracy of the CNN-LSTM-attention model in predicting the temperature of the fermentation grains, this study compares it with the LSTM model, GRU model, and CNN-LSTM model. The experimental results are presented in Figures 7(a), 7(b), 7(c), and 7(d).

From Figure 7, it is evident that the CNN-LSTM-attention model outperforms the other three models in terms of prediction performance. The LSTM model shows significant discrepancies between predicted and actual values, rendering its performance less than ideal. The GRU model improves the accuracy of predictions, but it still exhibits larger errors in regions with significant fluctuations in fermentation grains temperature. In these temperature-volatile regions, the CNN-LSTM model displays smaller prediction errors compared to the GRU model and achieves better overall performance. Furthermore, in comparison to the CNN-LSTM model, the CNN-LSTM-attention model exhibits a more ideal fit to the temperature curve and achieves greater precision in predictions.

Table 1 presents the evaluation results of the four prediction models. Through comparative analysis of the prediction errors in the table, it is evident that the proposed CNN-LSTM-attention model yields the smallest values for RMSE and MAE, while achieving the highest value for R2. Thus, it demonstrates the best predictive performance. Compared to the LSTM model, the CNN-LSTM-attention model reduces RMSE by 1.096, MAE by 1.011, and increases R2 by 0.015.

In comparison to the GRU model, the CNN-LSTM-attention model reduces RMSE by 0.573, MAE by 0.619, and increases R2 by 0.009. When compared to the CNN-LSTM model, the CNN-LSTM-attention model reduces RMSE by 0.28, MAE by 0.351, and increases R2 by 0.003. Hence, the proposed model in this study outperforms the other models in predicting the temperature of the fermentation grains.

Table 1. Three Scheme comparing

Model	**RMSE**	**MAE**	**R2**
LSTM	2.178	1.658	0.981
GRU	1.855	1.266	0.987
CNN-LSTM	1.562	0.998	0.993
CNN-LSTM-Attention	1.282	0.647	0.996

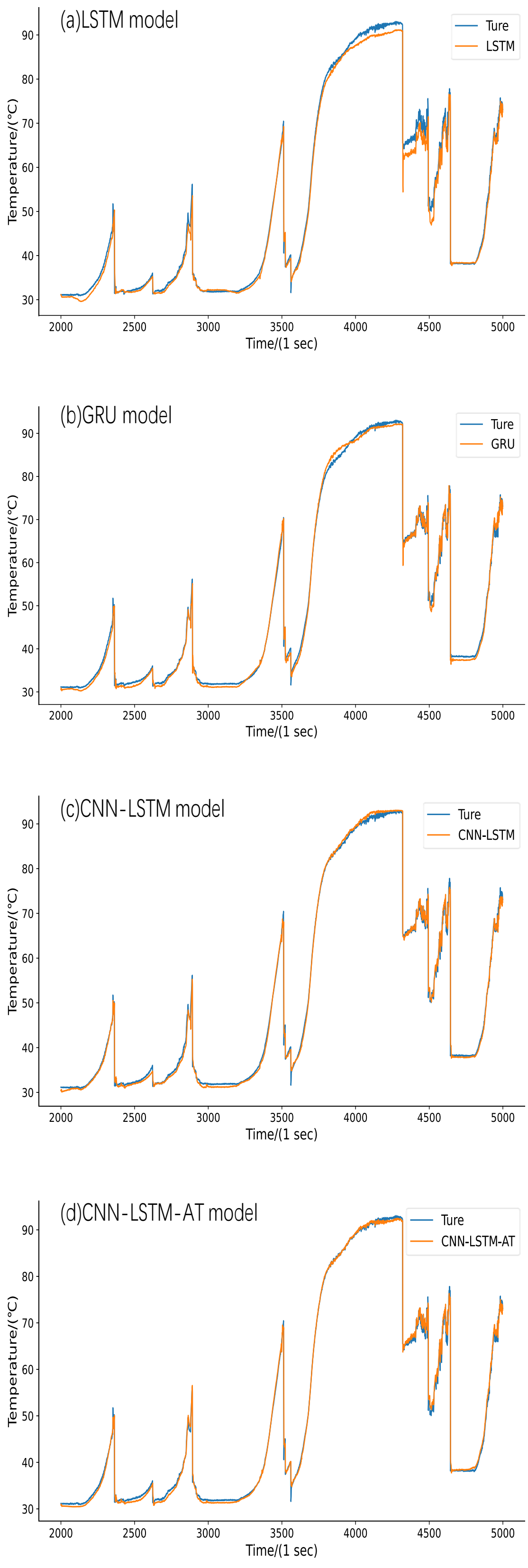

Figure 7. Comparison of prediction results by models

5.Conclusion

In view of the phenomenon of air leakage and running steam in the process of liquor brewing, accurate prediction of the temperature on the surface of the fermentation grains is instrumental in reducing liquor losses, enhancing liquor yield and quality, and improving the economic benefits of Baijiu enterprises. In this paper, we propose a CNN-LSTM model based on the attention mechanism to predict the future fermentation grains temperature by combining the historical fermentation grains temperature during the retorting process and other retorting parameters. Through experimental comparisons, it is evident that the CNN-LSTM-attention prediction model proposed in this paper exhibits higher predictive accuracy compared to the conventional LSTM, GRU, and CNN-LSTM models. It also demonstrates good fitting to the actual temperature of the fermentation grains, thereby validating the effectiveness of the proposed approach in predicting the temperature of the fermentation grains.

References

[1] Yang Bo, The innovation of traditional solid-state fermentation technology in China: the example of traditional solid-state brewing of Fenjiu. Brewing Science and Technology, 2021(07): pp. 106-109.

[2] Yang, et al. Study on the effect of retort barrel height and structural design on distillation efficiency and quality of spirits. Brewery Technology, 2012(10): pp. 94-98.

[3] Liu, W.B. et al, Progress in the application of upper retort robot in white wine brewing. Journal of Sichuan Light Chemical University (Natural Science Edition), 2023. 36(01): pp. 24-32.

[4] Li D. & Li G. H., Relationship between top retort distillation technology and quality of liquor production. Brewery Technology, 2012(01): pp. 65-66+69.

[5] Wei, Jingjun et al. Study on the effect of "eight moves of Yijiu on retort" on distillation effect and quality of white wine. Brewing, 2014. 41(06): pp. 50-52.

[6] Qing Yihui et al. Kinematic analysis and trajectory study of the upper retort robot. Food and Machinery, 2020. 36(12): pp. 70-73.

[7] Li L. et al. Study on the application of wine spirits on retort robot in the brewing production of soy sauce type white wine. Modern Food, 2023. 29(02): pp. 127-129.

[8] Li Yao et al. Structural analysis of wine spirits on retort robot and its application in the production of clear spiced white wine brewing. Brewing Science and Technology, 2021(12): pp. 95-99+107.

[9] Najwa, M.R.N., et al., Comparison between Regression Models, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in River Water Quality Prediction. Processes, 2022. 10(8).

[10] Abreu, L.R., et al., A decision tree model for the prediction of the stay time of ships in Brazilian ports. Engineering Applications of Artificial Intelligence, 2023. 117(PB).

[11] Yu, X. T. & Liu, P., Long-term and short-term memory network-convolutional neural network (LSTM-CNN) based PM_(2.5) concentration prediction in Beijing. Environmental Engineering, 2020. 38(06): pp. 176-180+66.

[12] X., W. and L. W., Time Series Prediction Based on LSTM-Attention-LSTM Model. IEEE Access, 2023. 11: p. 48322-48331.

[13] Hao, Shilin, Research on LSTM-based temperature prediction model for upper retort wine spirits, 2020, Hubei University of Technology. Page 63.

[14] Zhang, J. & Zhang, Z. P., Research on the improved GRU-based prediction model for liquor steam operation. Journal of Hubei University of Technology, 2022. 37(04): pp. 43-48+61.

[15] Chen, J., Y. Li and S. Zhang, Fast Prediction of Urban Flooding Water Depth Based on CNN−LSTM. Water, 2023. 15(7): p. 1397.

[16] Yan, R., et al., Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Systems with Applications, 2021. 169(prepublish).

[17] Sepp Hochreiter, jürgen Schmidhuber, Long Short-Term Memory.Neural Comput 1997; 9(8):1735-1780.

[18] Jia, X., et al., Prediction of Sea Surface Temperature in the East China Sea Based on LSTM Neural Network. Remote Sensing, 2022. 14(14).

[19] Tianjun, X., et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.