Journal of Computer Science and Digital Technology, 2025, 1(1), 11-20
https://doi.org/10.71204/by8g5g40

Dairy Product Production Prediction Based on BiLSTM-Attention model

Xiangyuan Qi [1], Guanglei Qiang [1,*], Fujiang Yuan [1]

[1] School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China; qiangguanglei_qgl@163.com

*** Correspondence:**
Guanglei Qiang

qiangguanglei_qgl@163.com

Received: 6 April 2025 /Accepted: 16 April 2025 /Published online: 21 April 2025

Abstract

With the rapid development of the dairy industry, accurate prediction of production is crucial to optimizing production plans. To this end, this paper proposes a prediction model that combines a bidirectional long short-term memory network (BiLSTM) with an attention mechanism (Attention). BiLSTM can effectively capture long-term and short-term dependencies in time series, while the attention mechanism can dynamically focus on the features of key time points, thereby improving prediction accuracy. Experiments show that the BiLSTM-Attention model improves the accuracy of dairy production prediction compared to traditional regression analysis and a single LSTM model, especially when processing long time series data. This study provides an effective solution for the accurate prediction of dairy production.

Keywords: Dairy Product Yield; Yield Prediction; BiLSTM; Attention; LSTM

1. Introduction

The dairy industry is an important part of global food production and consumption, covering a variety of products such as milk, yogurt, and cheese, which are widely used in daily life. With the increase in consumer demand for dairy products, the scale of dairy production continues to expand. How to efficiently and accurately predict the output of dairy products has become an important issue in production management and supply chain optimization (Zanchi, 2025). Accurate output forecasting not only helps manufacturers to reasonably arrange production plans and reduce inventory backlogs, but also effectively avoids supply shortages and ensures that market demand is met in a timely manner. However, the forecasting of dairy output faces many challenges. First, dairy production is affected by many factors, including weather changes, animal health, feed supply, policy adjustments, and market demand fluctuations. Traditional forecasting methods, such as linear regression models and time series analysis, often fail to fully capture the

interrelationships between these complex factors and the nonlinear characteristics in the data, resulting in limited forecasting effects (Sanjulián et al, 2025).

In recent years, the performance of deep learning technology in time series forecasting has received widespread attention. In particular, long short-term memory networks (LSTM) have been widely used in various forecasting tasks due to their excellent time-dependent modeling capabilities. Murphy (2014) used the MLR model and the nonlinear autoregressive model with exogenous input (NARX) to predict the total daily milk production of a herd in different forecast periods, thereby significantly reducing the error rate. Nguyen (2020) studied milk production prediction by using different machine learning methods (support vector machine, random forest, neural network), and finally the experiment showed that the support vector machine is the best compromise between accuracy and computational cost. Vithitsoontorn (2022) proposed a method for forecasting demand for dairy production plans that combines LSTM and ARIMA. The results show that both statistical methods and deep learning methods are reliable and suitable for demand forecasting, but there is no single best optimization algorithm. The ARIMA model performs best on a few sequences with small fluctuations by using an average straight line to predict future trends. The LSTM model can better capture the seasonal characteristics of the sequence, especially in sequences with strong trends, where LSTM performs better than ARIMA. By training the model on monthly data, LSTM is able to provide lower prediction errors. Deshmukh (2016) proposed using ARIMA and VAR time series models to predict India's milk production. The results showed that the prediction rate was more accurate and predicted that milk production would reach 160 million tons by 2017. Khan (2025) a novel application combining the long short-term memory (LSTM) algorithm with seasonal mean interpolation (SMI) is proposed; experiments show that the model has good performance and potential in improving the accuracy of wind power forecasting. Başarslan (2025) proposed a MC&M-BL model, which effectively improves the accuracy and other correlation coefficients by combining the convolutional neural network (CNN) for image feature extraction with the bidirectional long short-term memory network (BiLSTM) for sequential data processing. Gao (2025) using two different models is proposed: a self-attention-assisted bidirectional long short-term memory model and a bidirectional long short-term memory model based on a multi-head attention mechanism; these models consider spatial continuity and adaptively adjust the weights of each step to improve the classification results using the attention mechanism; the study shows that the proposed bidirectional long short-term memory model based on a multi-head attention mechanism can improve the classification performance. Chen (2019) uses BiLSTM-CRF and CNN to improve the prediction accuracy and is applicable to more scenarios.

In summary, traditional machine learning algorithms and LSTM still have certain limitations in capturing long-term dependencies. To overcome these problems, this paper proposes a prediction model based on a combination of a bidirectional long short-term memory network (BiLSTM) and an attention mechanism (Attention). BiLSTM can capture both forward and backward information of a sequence, while the attention mechanism can highlight important information at critical moments through dynamic weight allocation, thereby further improving prediction accuracy.

2. Based on BiLSTM-Attention Prediction Model

2.1. LSTM

Long Short-Term Memory (LSTM) is a special type of recurrent neural network (RNN) that aims to solve the gradient vanishing and gradient exploding problems that traditional RNNs often encounter when processing long time series data (Aslan et al, 2021). In standard RNNs, as information is back-propagated over multiple iterations, the gradient may become very small (gradient vanishing) or very large (gradient exploding) (Qiang, 2025), which leads to performance degradation when the network learns long-term dependencies. To solve this problem, LSTM introduces a unique gating mechanism to effectively control the transmission of information flow and is able to capture important long-term dependency information in longer time series (Yuan, 2025).

The key innovation of LSTM lies in its special gating structure, including the forget gate, input gate, and output gate. The forget gate determines which information should be discarded from the network's memory unit. Its output value is between 0 and 1, with 0 indicating complete forgetfulness and 1 indicating complete retention (Lu et al, 2021). The input gate controls how the current input information is updated. It generates new candidate information by combining the input data at the current moment and the hidden state at the previous moment, which will be stored in the cell state. The output gate is responsible for controlling the flow of information extracted from the memory (Xu et al, 2019). It determines the output of the network at the current moment through the hidden state at the previous moment and the current cell state. In addition, the LSTM model transmits long-term memory through the design of the cell state, which can be continuously updated during the operation of the entire network. Unlike standard RNNs, LSTM ensures that only key information can be retained for a long time through the cell state through a gating mechanism and passed to the next moment at each time step. In this way, LSTM can effectively avoid the loss of information in long sequences, thereby capturing long-term dependencies in time series data(Sherstinsky, 2020).

These structural advantages of LSTM enable it to show remarkable results in many tasks that require processing time-dependent information. For example, in the fields of natural language processing, speech recognition, and financial forecasting, LSTM has shown excellent performance. Compared with traditional statistical models or shallow learning algorithms, LSTM has a stronger ability to capture complex nonlinear patterns and time dependencies in data. Therefore, LSTM has become an important tool in time series analysis and forecasting tasks and is widely used in various fields. The LSTM model diagram is shown in Figure 1.

The specific calculation process is as follows:

$$f_t = \sigma\left[w_f^x x_t + w_i^h h_{t-1} + b_f\right] \tag{1}$$

$$i_t = \sigma\left[w_i^x x_t + w_i^h h_{t-1} + b_i\right] \tag{2}$$

$$o_t = \sigma\left[w_o^x x_t + w_o^h h_{t-1} + b_o\right] \tag{3}$$

$$c_t = f_t c_{t-1} + i_t \tanh(\left[w_c^x x_t + w_c^h h_{t-1} + b_c\right] \tag{4}$$

$$h_t = o_t \tanh(c_t) \tag{5}$$

In the formula, w_i^x、w_f^x、w_o^x、w_c^x is the input weight matrix, w_i^h、w_f^h、w_o^h、w_c^h is the recursive weight matrix, b_n is the bias, x_t is the current input value, c_t is the long-term memory, h_t is the short-term memory, σ is the sigmoid activation function, and tanh () is the tanh activation function.

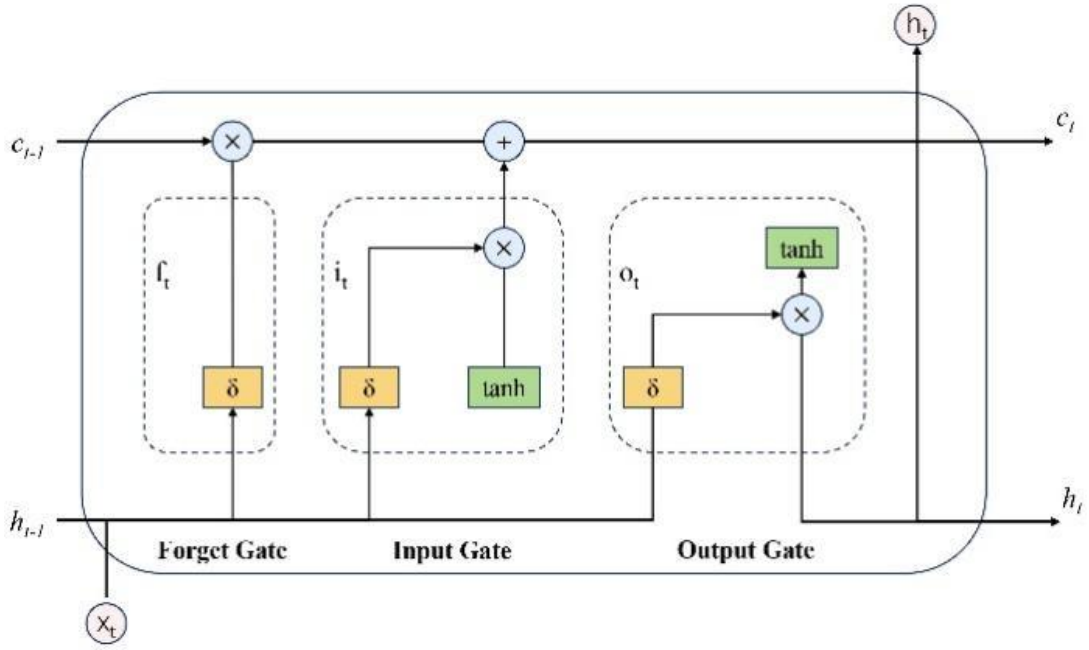

Figure 1. LSTM Model

2.2. BiLSTM

Bidirectional long short-term memory (BiLSTM) (Liu & Guo, 2019) is an extended form of LSTM. By combining information flows in both the forward and backward directions, it can more comprehensively capture contextual information in sequence data. As shown in Figure 2, BiLSTM processes forward and backward sequence data in parallel and integrates the hidden states of both, thereby enhancing the model's ability to understand the dependencies between the front and back of the sequence. This bidirectional structure makes it particularly effective in tasks such as natural language processing and speech recognition, and can more accurately capture complex contextual features.

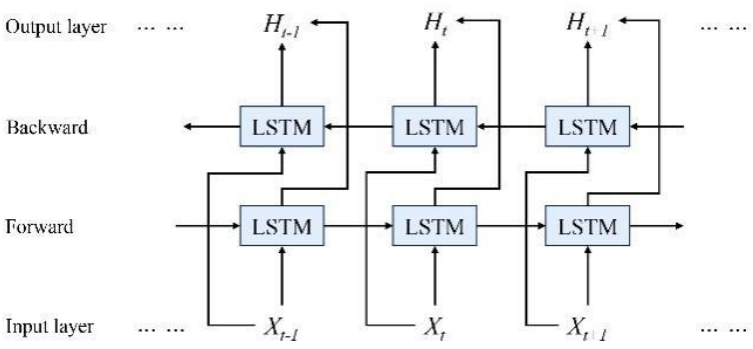

Figure 2. BiLSTM Model

2.3. Attention Model

This paper introduces the Attention mechanism into the BiLSTM model (Zhang et al, 2023), aiming to enhance the model's ability to capture key information in the input sequence by dynamically assigning weights. BiLSTM first performs bidirectional encoding on the input sequence to generate a hidden state sequence containing contextual information; then, the Attention mechanism calculates the attention weight of each hidden state, selects important features and generates a context vector (Shan et al, 2021). This improves the model's ability to model long-distance dependencies and enhances task performance, which is significant in the prediction of dairy production in this paper (Kavianpour P, 2023). The model diagram is shown in Figure 3. The specific calculation process of the Attention mechanism is as follows:

$$s(x, q) = x^T q \tag{6}$$

$$\alpha_n = \frac{\exp(s(x_n, q))}{\sum_{j=1}^{N} (s(x_j, q))} \tag{7}$$

$$C = \sum_{n=1}^{N} \alpha_n x_n \tag{8}$$

In the formula, the input feature information x refers to the output of the BiLSTM neural network, that is, the hidden state sequence obtained after bidirectional encoding. The query vector q is used to interact with the input features, and the importance score of each hidden state is calculated through the attention score function $s(\mathrm{x}, q)$. The attention distribution α_n is obtained by normalizing these scores through Softmax, and finally the context vector C is generated by weighted summation as the output of the model(Li, 2019). This mechanism enables the model to dynamically focus on the key information in the input sequence, thereby improving task performance.

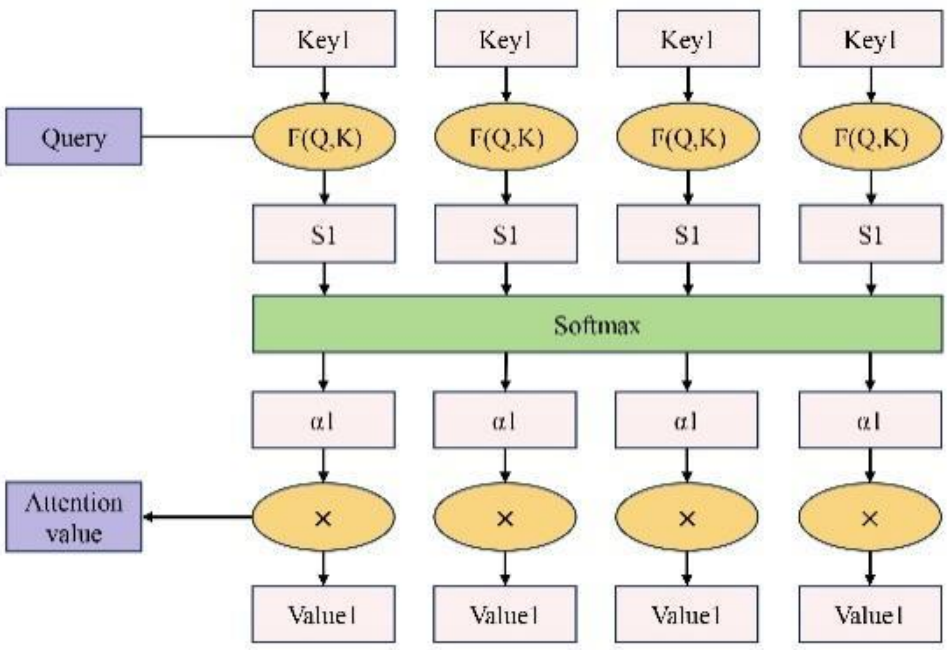

Figure 3. Attention Model

3. Results and Analysis

3.1. Data preparation

The time range is from January 1989 to November 2024, and the relevant data of the research variables are all from the National Bureau of Statistics. The experiment uses the relevant data of dairy production in the past 35 years as the research object, with a total of 390 samples.

3.2. Experimental environment

This model is implemented using the PyTorch framework. During the training process, the dropout method is used to effectively prevent overfitting. The experimental environment and configuration are shown in Table 1. During the training process, the batch size is set to 32, the training round is 100, the optimizer is Adam, and the learning rate is set to 0.001

Table 1. Experimental environment configuration table

Name	Version
CPU	Intel(R) Xeon(R) E5-2680 v4 @ 2.40GHz
GPU	RTX 3090 24GB
Programming language	Python3.10
Operating system	Ubuntu 22.04

3.3. Experimental procedures

The experimental process of this paper includes four main steps: data preprocessing, model construction, model training and optimization, and result evaluation.

(1) In the data preprocessing stage, the original data is filled with missing values, outliers are detected and normalized, and the sliding window method is used to construct the time series data set. Then the training set and the test set are divided in a ratio of 8:2 and converted into a tensor format suitable for deep learning.

(2) In the model construction stage, BiLSTM-Attention is selected as the core prediction model. The model extracts the long-term dependency features of the time series by a bidirectional LSTM layer, and assigns weights to different time steps through the attention mechanism to enhance the information contribution of key moments. The network structure consists of a BiLSTM layer, an Attention mechanism, a fully connected layer and an output layer, and the mean square error (MSE) is used as the loss function. The specific formula is shown below.

$$loss = \frac{1}{n}\sum_{i=1}^{n}(\bar{y}_i - y_i) \tag{9}$$

(3) In the model training and optimization stage, the Adam optimizer was used to iteratively train the model. The initial learning rate was set to 0.001, and the early stopping mechanism was used to prevent overfitting. During the model training process, batch training and gradient

clipping techniques were used to optimize parameter updates and improve the stability of the model.

(4) In the result evaluation stage, the root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the prediction performance of the model, and a comparative analysis was performed with the traditional LSTM and BiLSTM models to verify the effectiveness of the BiLSTM-Attention model in the dairy yield prediction task.

3.4. Data preprocessing

In order to ensure the accuracy and stability of model training, this study systematically preprocessed the data. First, the original data was tested for missing values, and a small number of missing values were filled by linear interpolation, while outliers were screened and corrected according to the 3σ principle. Secondly, all variables were normalized by min-max to eliminate the influence between different dimensions and improve the convergence speed of the model, as shown in the following formula. Then, a sliding window data set was constructed, based on time series features, with data from the past n periods as input, to predict the target variable for the next period. Finally, the data was divided into a training set and a test set in a ratio of 8:2, and converted into a tensor format for subsequent model training and evaluation, which is suitable for deep learning models.

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

In the formula, x is the original data, and X is the normalized new data.

3.5. Experimental Results

This paper evaluates the dairy yield prediction model based on BiLSTM-Attention and compares it with the traditional LSTM model. The experimental results are shown in Table 2.

Table 2. Comparison of model prediction performance

Model predictions	MAE	RMSE
LSTM	5.1008	9.4679
BiLSTM	5.0610	9.2315
BiLSTM-Attention	4.8954	9.0233

This experiment uses MAE and RMSE as core evaluation indicators. The results show that the BiLSTM-Attention model is superior to traditional LSTM and BiLSTM in prediction accuracy, with lower error and stronger generalization ability. This is because BiLSTM captures the dependencies between time series through bidirectional propagation and improves feature representation capabilities, but there may still be problems of information redundancy or dilution of key features. To optimize feature extraction, the BiLSTM-Attention mechanism introduces an attention mechanism to automatically assign weights to different time steps, enhance attention to key information, and effectively reduce redundant interference, thereby improving the model's prediction ability.

By comparing the curves of the predicted values and the true values, the BiLSTM-Attention model can better follow the changing trend of dairy production and accurately capture seasonal fluctuations and sudden fluctuations. The visual analysis is shown in Figure 4.

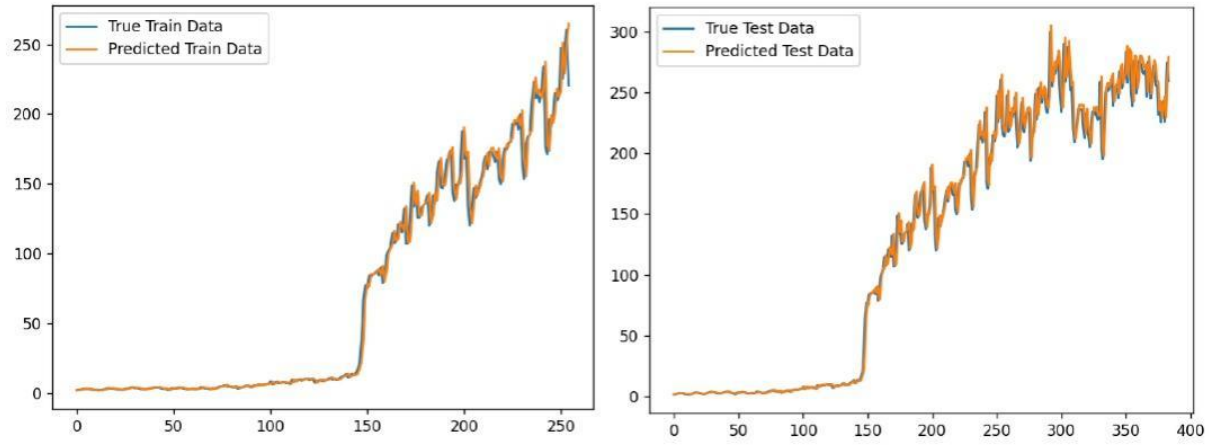

Figure 4. Visualization result graph

The visualization results further prove the significant advantages of the BiLSTM Attention model in prediction accuracy: its prediction curve has a higher degree of fit with the true value, the overall error is more concentrated, and the deviation and fluctuation amplitude are also smaller. The main reason for this is that BiLSTM bidirectional parallel encoding takes into account both historical trends and future trends at each moment, enabling the model to respond to sudden fluctuations and potential long-term laws at the same time; the Attention mechanism highlights the key moments and features that are highly relevant to the prediction target through adaptive weight allocation, effectively suppressing the interference of noise and outliers; the cross-time connection of the attention layer provides a direct gradient channel for long-distance dependencies, which not only retains global information but also alleviates the gradient attenuation of deep networks; in addition, the sparse distribution of attention weights has a natural regularization effect, reducing the risk of overfitting irrelevant information, thereby further improving the stability and generalization ability of the model on different samples. Through this model, the focus on key features is effectively improved, the interference of irrelevant information is reduced, and the model's prediction on complex time series data is more accurate and has better generalization ability.

4. Conclusions

This paper proposes a dairy production prediction model based on BiLSTM-Attention, and verifies its advantages in time series prediction tasks by comparing it with the traditional LSTM model. Experimental results show that the BiLSTM-Attention model can effectively capture the long-term dependencies and seasonal fluctuations in dairy production data, significantly improving the prediction accuracy. Compared with other models, this model shows lower errors in indicators such as mean square error (MSE) and root mean square error (RMSE), indicating its reliability and accuracy in practical applications. Future research can further optimize the model

structure and combine more feature engineering methods to improve the performance of the model in more complex scenarios.

Author Contributions:

Conceptualization,X.Q., G.Q., F.Y.; methodology,X.Q., G.Q., F.Y.; software,X.Q., G.Q., F.Y.; validation,X.Q., G.Q., F.Y.; formal analysis,X.Q., G.Q., F.Y.; investigation,X.Q., G.Q., F.Y.; resources,X.Q., G.Q., F.Y.; data curation,X.Q., G.Q., F.Y.; writing—original draft preparation,X.Q., G.Q., F.Y.; writing—review and editing,X.Q., G.Q., F.Y.; visualization,X.Q., G.Q., F.Y.; supervision,X.Q., G.Q., F.Y.; project administration,X.Q., G.Q., F.Y.; funding acquisition,X.Q., G.Q., F.Y. All authors have read and agreed to the published version of the manuscript.

Funding:

This research was funded by 2024 Taiyuan Normal University Graduate Education Innovation Project (SYYJSYC-2475).

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of Interest:

The authors declare no conflict of interest.

References

Aslan, M. F., Unlersen, M. F., Sabanci, K., & Durdu, A. (2021). CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. Applied Soft Computing, 98, 106912.

Başarslan, M. S. (2025). MC &M-BL: a novel classification model for brain tumor classification: multi-CNN and multi-BiLSTM. The Journal of Supercomputing, 81(3), 1-25.

Chen, T., Xu, R., He, Y., et al. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 72, 221-230.

Deshmukh, S. S., & Paramasivam, R. (2016). Forecasting of milk production in India with ARIMA and VAR time series models. Asian Journal of Dairy and Food Research, 35(1), 17-22.

Gao, Y., Tian, M., Grana, D., Xu, Z., & Xu, H. (2025). Attention mechanism-assisted recurrent neural network for well log lithology classification. Geophysical Prospecting, 73(2), 628-649.

Kavianpour, P., Kavianpour, M., Jahani, E., & Ramezani, A. (2023). A CNN-BiLSTM model with attention mechanism for earthquake prediction. The Journal of Supercomputing, 79(17), 19194-19226.

Khan, S., Muhammad, Y., Jadoon, I., Awan, S. E., & Raja, M. A. Z. (2025). Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting. Applied Soft Computing, 170, 112765.

Li, Y., Zhu, Z., Kong, D., Han, H., & Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. Knowledge-Based Systems, 181, 104785.

Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing, 337, 325-338.

Lu, W., Li, J., Wang, J., & Qin, L. (2021). A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications, 33(10), 4741-4753.

Murphy, M. D., O'Mahony, M. J., Shalloo, L., French, P., & Upton, J. (2014). Comparison of modelling techniques for milk-production forecasting. Journal of dairy science, 97(6), 3352-3363.

Nguyen, Q. T., Fouchereau, R., Frenod, E, et al. (2020). Comparison of forecast models of production of dairy cows combining animal and diet parameters. Computers and Electronics in Agriculture, 170, 105258.

Qiang, G. (2025). Scalability improvement and empirical analysis of PBFT consensus mechanism in blockchain. Electronic Components and Information Technology, 9(01),192-195.

Sanjulián, L., Fernández-Rico, S., González-Rodríguez, N., et al. (2025). The Role of Dairy in Human Nutrition: Myths and Realities. Nutrients, 17(4), 646.

Shan, L., Liu, Y., Tang, M., et al. (2021). CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. Journal of Petroleum Science and Engineering, 205, 108838.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

Vithitsoontorn, C., & Chongstitvatana, P. (2022). Demand forecasting in production planning for dairy products using machine learning and statistical method. In 2022 international electrical engineering congress (iEECON), 1-4.

Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. Ieee Access, 7, 51522-51532.

Yuan, F., Huang, X., Zheng, L., et al. (2025). The Evolution and Optimization Strategies of a PBFT Consensus Algorithm for Consortium Blockchains. Information, 16(4), 268.

Zanchi, M., La Porta, C. A., et al. (2025). Influence of microclimatic conditions on dairy production in an Automatic Milking System: Trends and Time-Series Mixer predictions. Computers and Electronics in Agriculture, 229, 109730.

Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using CNN-BiLSTM-Attention model. Mathematics, 11(9), 1985.