A CNN-BiLSTM Model with Multi-Head Attention for Temperature and Humidity Prediction in Grain Storage Facilities

Shengquan Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: xaitysq@163.com

Yunan Sun
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 3399053299@qq.com

Zhengxin Zhang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 13709259851@163.com

Yuwei Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 3291576272@qq.com

***Abstract*—Accurate prediction of temperature and humidity within grain storage facilities is essential for ensuring long-term food safety and reducing post-harvest losses. Existing approaches based on standard Long Short-Term Memory (LSTM) networks or Sequence-to-Sequence LSTM (Seq2Seq-LSTM) architectures often fail to capture the complex multi-scale temporal dependencies inherent in grain depot microclimate data. In this paper, we propose a hybrid deep learning model that integrates Convolutional Neural Networks (CNN), Bidirectional LSTM (BiLSTM), and a Multi-Head Attention mechanism (CNN-BiLSTM-Attention). The CNN component extracts local temporal patterns, the BiLSTM captures both forward and backward long-range dependencies, and the Multi-Head Attention layer dynamically focuses on the most informative time steps. We evaluate the proposed model on a real-world dataset collected from a grain storage depot in Northwest China spanning two full years of hourly records. Results demonstrate that the proposed model consistently outperforms SVR, LSTM, and Seq2Seq-LSTM baselines across prediction horizons of 6 h, 12 h, and 24 h. For 24-hour temperature prediction, the model achieves MAE of 0.89°C and coefficient of determination of 0.9713, representing 35.5% and 4.2% improvements over Seq2Seq-LSTM.**

Keywords-Grain Storage; Temperature Prediction; Humidity Prediction; Deep Learning; CNN-Bilstm; Attention Mechanism; Time Series Forecasting

I. INTRODUCTION

Grain storage safety is a critical component of national food security. According to the Food and Agriculture Organization (FAO), global post-harvest grain losses account for approximately 10–15% of total production annually, with a significant portion attributable to improper storage conditions [1]. Temperature and humidity are the two most influential environmental factors affecting grain quality during storage. When these variables exceed critical values—typically above 25°C and 75% relative humidity for wheat—stored grain becomes susceptible to mold growth, insect infestation, and accelerated respiration, which degrade nutritional value and may produce mycotoxins that render the grain unsafe for human consumption [2][3][13]. In China, where national grain reserves exceed 650 million tons distributed across thousands of depots, the government has invested substantially in deploying Internet of Things (IoT) sensor networks for real-time environmental monitoring [4]. However, the vast majority of these monitoring systems remain fundamentally reactive—they report current conditions but provide no information about future trends, forcing storage managers to rely on experience rather than data-driven forecasts for ventilation and climate control decisions [5][6].

A reliable prediction model that forecasts environmental conditions hours to days in advance would enable proactive management. For instance, if a model forecasts that indoor temperature will exceed 20°C within the next 12 hours, the facility can preemptively activate ventilation during cooler nighttime hours when the indoor-outdoor temperature differential is most favorable. This predictive ventilation scheduling strategy can reduce energy consumption by 20–30% compared to reactive threshold-based approaches. The development of accurate short-to-medium-term prediction methods therefore has both scientific significance and practical engineering value for the grain storage industry.

Traditional statistical methods such as ARIMA and exponential smoothing can model linear time series patterns but struggle with the nonlinear and multi-factor nature of grain depot microclimate dynamics [7]. Machine learning methods including Support Vector Regression (SVR) and Random Forest have been applied with moderate success, yet they require extensive feature engineering and cannot easily model temporal dependencies across long input sequences [8][9].

Deep learning methods offer an alternative by automatically learning hierarchical feature representations from raw sequential data, capturing multi-scale temporal patterns without manual feature engineering. LSTM networks can model long-range temporal dependencies and have been applied to grain temperature prediction with promising results [5][10]. However, standard unidirectional LSTM processes sequences in only one direction and may miss contextual information from future time steps within the input window. Seq2Seq-LSTM models extend the LSTM framework to multi-step prediction tasks [26], but the fixed-length context vector creates an information bottleneck that limits performance when the input sequence is long [25].

To address these limitations, we propose a hybrid CNN-BiLSTM-Attention model for grain storage temperature and humidity prediction. The proposed architecture is motivated by three specific observations:

1) *Local temporal patterns:* Temperature and humidity exhibit short-term fluctuations (e.g., diurnal cycles) that can be effectively captured through convolutional operations.

2) *Bidirectional context:* Within a given input window, both preceding and succeeding observations provide useful context, making bidirectional processing advantageous.

3) *Variable-importance time steps:* Not all historical observations contribute equally to future predictions. An attention mechanism can adaptively weight different time steps.

The main contributions of this paper are:

- We propose a CNN-BiLSTM-Attention model integrating local feature extraction, bidirectional temporal modeling, and adaptive attention weighting for grain storage prediction.
- We conduct comprehensive experiments on a real-world dataset from a grain depot in Northwest China, comparing against SVR, LSTM, and Seq2Seq-LSTM baselines across three prediction horizons.
- We perform ablation experiments and provide additional analyses including seasonal evaluation, window sensitivity, and attention visualization.

The proposed model is validated through extensive experiments demonstrating its effectiveness across different prediction tasks, horizons, and seasonal conditions, with detailed analyses providing practical insights for deployment in real-world grain storage management systems.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the methodology. Section IV describes experiments and discusses results. Section V concludes the paper.

II. Related Work

A. Grain Storage Environment Prediction

Research on grain storage environment prediction has evolved from physics-based modeling to data-driven approaches. Early work relied on heat and mass transfer equations combined with computational fluid dynamics (CFD) to simulate temperature fields within grain piles [13]. While physically principled, these methods require detailed knowledge of grain properties and boundary conditions, limiting their practical applicability.

Machine learning methods have been increasingly adopted as sensor data from grain storage monitoring systems has become more widely available. These methods offer the advantage of learning directly from observed data patterns without requiring detailed physical models of the storage environment. Duan et al. [8] employed SVR to predict grain surface temperature using meteorological inputs, achieving reasonable accuracy for single-step prediction but with limited ability to capture temporal dynamics, since SVR treats each input as an independent feature vector without sequential structure. Qiu and Yang [9] developed an RBF neural network for granary temperature and humidity prediction; however, their model was evaluated on a limited dataset and did not address multi-horizon forecasting. Li and Zhu [5] established a temperature identification model using deep belief networks, representing an early application of deep learning to this domain, though the model's generalization to unseen seasonal patterns was not thoroughly assessed. Ge and Chen [10] developed an improved LSTM model that outperformed traditional methods on single-step temperature prediction, but their architecture did not incorporate mechanisms for adaptive temporal weighting. Mao et al. [3] proposed a GRU deep fusion model incorporating multivariate linear regression and wavelet filtering, where the wavelet decomposition improved signal-to-noise separation but introduced additional computational overhead and required manual selection of decomposition levels-choices that are dataset-dependent and do not transfer easily across facilities.

Despite these advances, most existing methods either focus exclusively on temperature prediction or treat temperature and humidity as independent tasks. Furthermore, the majority of studies employ single-architecture models that cannot simultaneously capture local temporal patterns, long-range bidirectional dependencies, and variable-importance temporal weighting. This motivates the development of hybrid architectures that combine the complementary strengths of different deep learning components.

Several recent studies have explored more sophisticated architectures. Duan et al. [4] proposed a deep spatiotemporal attention approach that leverages spatial correlations among multiple sensor positions within a grain pile, achieving an RMSE as low as 0.34°C. While their spatiotemporal modeling is effective when multiple sensors are available at known positions, this approach requires dense sensor grids that are not always present in smaller or older storage facilities, and the spatial attention module adds substantial computational overhead that may hinder real-time deployment. Zhang et al. [11] developed an FTA-CNN-SE-LSTM model with dual-domain data augmentation to address training data scarcity; the augmentation strategy generates synthetic samples in both time and frequency domains, which is valuable when historical records are limited, but the augmented data may not faithfully reproduce the statistical properties of rare events such as sudden temperature spikes during equipment malfunctions. Wang et al. [7] proposed an integrated deep learning framework combining 3D-CNN with LSTM that exploits the three-dimensional spatial structure of grain piles, achieving impressive single-step RMSE of 0.28°C, though the 3D-CNN component requires volumetric sensor data that is not available in most grain storage monitoring systems. Liu et al. [12] proposed an EMD-CNN-LSTM hybrid model for predicting grain pile humidity by incorporating meteorological factors. Their empirical mode decomposition preprocessing separates trend, seasonal, and residual components before feeding them into the neural network, improving humidity prediction accuracy at the cost of added pipeline complexity and sensitivity to the decomposition parameters.

A common limitation across existing grain storage prediction studies is the lack of comprehensive evaluation frameworks. Many studies report results on a single prediction horizon, use limited evaluation metrics, or do not conduct ablation experiments to verify the contribution of individual model components. Furthermore, few studies simultaneously address both temperature and humidity prediction, despite the strong physical coupling between these two variables. Our work addresses these gaps by providing a systematic evaluation across multiple horizons, both prediction targets, four evaluation metrics, and detailed ablation and sensitivity analyses [27] [28].

B. Deep Learning for Time Series Forecasting

LSTM networks [14] have become a foundational architecture for sequential data processing. BiLSTM [15] extends this by processing sequences in both directions. The Seq2Seq-LSTM architecture [16] employs an encoder-decoder framework for sequence transduction tasks. While originally developed for machine translation, it has been widely adopted for multi-step time series forecasting, including indoor temperature prediction [26] and solar radiation forecasting. However, the fixed-length context vector between encoder and decoder limits its ability to selectively utilize information from different input time steps [25], a limitation that becomes more severe as input sequences grow longer.

Attention mechanisms [17][18] allow the model to attend to all hidden states with learned weights, overcoming the information bottleneck of fixed-length context vectors. Multi-Head Attention [18] projects queries, keys, and values into multiple subspaces, enabling the model to simultaneously capture different types of temporal relationships. The combination of CNN, BiLSTM, and attention has been explored in several time series domains. Xu et al. [21] applied CNN-BiLSTM-Attention to aircraft trajectory prediction, where convolutional layers extract spatial movement patterns and attention weights highlight critical trajectory segments; their results showed 15–20% improvement over LSTM baselines. Rudhistiar et al. [22] adopted a similar architecture for stock price prediction, leveraging CNN to detect chart patterns and attention to weight trading sessions by market volatility. Chen et al. [23] applied the framework to electrical load forecasting with XGBoost-based stacking, and Wang et al. [24] demonstrated its effectiveness for industrial process monitoring in cold rolling mills, where the multi-head attention mechanism learned to focus on different physical process stages. In industrial equipment condition prediction scenarios, this hybrid architecture has also been validated to effectively improve the prediction accuracy of time-series indicators such as inclination angle [30].

Despite these successful applications across diverse domains, the CNN-BiLSTM-Attention architecture has not been systematically evaluated for grain storage environment prediction. The grain storage domain presents unique characteristics that distinguish it from the above applications: strong diurnal periodicity with variable amplitude across seasons, thermal inertia effects that create asymmetric response delays between heating and cooling, tight physical coupling between temperature and humidity through the Clausius-Clapeyron relationship, and the presence of biological processes (grain respiration) that introduce endogenous heat sources not present in purely physical systems. These domain-specific characteristics make grain storage a particularly interesting test case for evaluating whether the CNN-BiLSTM-Attention architecture can adapt to environmental prediction tasks with complex multi-physics interactions. Our work fills this gap by proposing and rigorously evaluating the architecture against established baselines in this domain [29].

III. METHODOLOGY

A. Problem Formulation

Given a multivariate time series $X=\{x_1\hat{y},x_2\hat{y},\ldots,x_T\hat{y}\}$ where $x_t\hat{y}\in R^d$, the objective is to predict $\hat{y}_{T+\tau}$ at horizon τ:

$$\hat{y}_{T+1} = f(x_1, x_2, \ldots, x_T; \Theta) \quad (1)$$

In our setting, the input features include indoor temperature, outdoor temperature, indoor relative humidity, and outdoor relative humidity (d=4), and the prediction targets are the indoor temperature and indoor relative humidity at the specified future time step. Temperature and humidity predictions are performed separately, each using all four input features. This formulation allows the model to leverage the strong correlations between indoor and outdoor conditions, as well as the physical coupling between temperature and humidity variables.

B. Support Vector Regression

SVR [19] maps input features to a high-dimensional space. Given training pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$,SVR solves:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\| w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (2)$$

Subject to ε-insensitive constraints, using the RBF kernel $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$.

The RBF kernel is chosen over linear and polynomial alternatives for two reasons. First, the relationship between meteorological inputs and indoor grain storage conditions is inherently nonlinear—indoor temperature responds to outdoor changes with variable lag and attenuation that depend on grain thermal mass, warehouse insulation, and seasonal context. A linear kernel cannot capture these nonlinear mappings. Second, unlike the polynomial kernel, RBF does not require pre-specification of the polynomial degree, which would introduce an additional hyperparameter sensitive to the specific prediction task and horizon.

SVR serves as a representative traditional machine learning baseline because it has been widely used in grain storage temperature prediction [8] and provides a reference point for evaluating the benefits of deep learning architectures. However, SVR requires the input to be a flattened feature vector, which destroys the temporal structure of the time series and treats observations at different time steps as unordered features.

C. Long Short-Term Memory

LSTM [14] introduces a memory cell governed by three gates:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f\right) \quad (3)$$

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i\right) \quad (4)$$

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c\right) \quad (5)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (6)$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o\right) \quad (7)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (8)$$

Where σ(·) is the sigmoid function, ⊙ denotes element-wise multiplication, and W and b are trainable weight matrices and bias vectors. The forget gate ft controls what information from the previous cell state should be discarded, the input gate it determines what new information should be stored, and the output gate ot regulates what information from the cell state should be exposed as the hidden state.

For single-step prediction, the final hidden state hT is passed through a fully connected layer to produce the prediction. The gating mechanism makes LSTM particularly relevant for grain storage prediction. Grain depot microclimate data exhibits a superposition of long-term seasonal trends (temperature rising over weeks in spring, declining in autumn) and short-term fluctuations (diurnal cycles, weather fronts passing within hours). The forget gate can selectively discard short-term noise when the prediction horizon is long, while the input gate can emphasize recent rapid changes when the horizon is short. This adaptive memory management at multiple time scales gives LSTM a structural advantage over methods that treat all temporal information uniformly. Nonetheless, standard LSTM processes the input sequence in only one direction, which means it cannot leverage future context within the input window when encoding each time step.

D. Sequence-to-Sequence LSTM

Seq2Seq-LSTM [16] uses an encoder-decoder structure. The encoder produces context vector $\mathbf{Z} \in \mathbb{R}^{T' \times K}$, and the decoder generates output conditioned on c:

$$\hat{y}_{T+k} = \text{Decoder}(\hat{y}_{T+k-1}, s_{k-1}, c) \tag{9}$$

A fundamental limitation of this architecture is that the entire input sequence is compressed into a single fixed-length vector c, which creates an information bottleneck. As the input sequence grows longer, the context vector must encode an increasing amount of information into a fixed-dimensional representation, inevitably leading to information loss. This is particularly problematic for grain storage prediction, where different regions of the input sequence (e.g., recent observations versus observations from 24 hours ago at the same time of day) may carry very different levels of importance for the prediction task, but the fixed context vector treats all input equally during compression.

E. Proposed CNN-BiLSTM-Attention Model

The proposed model consists of four parts: (1) 1D CNN for local feature extraction, (2) BiLSTM for bidirectional temporal modeling, (3) Multi-Head Attention for adaptive weighting, and (4) fully connected layers. Fig. 1 illustrates the architecture.

The 1D CNN applies K filters with kernel size w:

$$z_t^{(k)} = \text{ReLU}\left(\sum_{j=0}^{w-1} W_j^{(k)} \cdot x_{t+j} + b^{(k)}\right) \tag{10}$$

With same-padding (T'=T). The CNN feature maps Z serve as input to the subsequent BiLSTM, replacing the raw time series with a representation where local patterns have already been distilled into compact feature vectors. This design creates a hierarchical information flow: short-range patterns are resolved at the CNN stage, medium- and long-range dependencies are captured by the BiLSTM, and the attention layer then selectively aggregates the most predictive time steps from the BiLSTM output.

The rationale for placing the CNN before the recurrent layer is twofold. First, local patterns such as diurnal temperature cycles and short-term humidity fluctuations can be efficiently captured by convolutional operations. Second, by extracting and compressing local patterns into feature maps, the CNN reduces the burden on the subsequent recurrent layer, enabling it to focus on modeling longer-range temporal dependencies rather than spending capacity on detecting elementary patterns.

In our implementation, we use 64 convolutional filters with a kernel size of 3, which corresponds to a receptive field of 3 hours given our hourly sampling rate. This kernel size is sufficient to capture short-term fluctuations while being small enough to preserve temporal resolution for the subsequent recurrent layer.

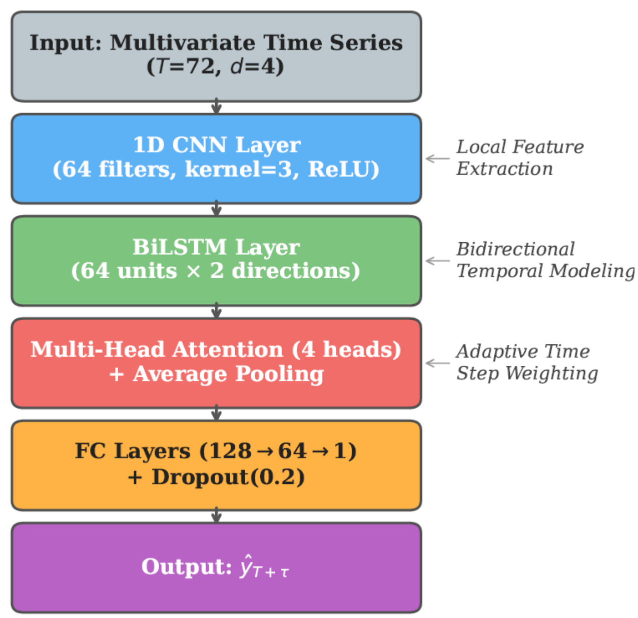


Figure 1. Architecture of the proposed CNN-BiLSTM-Attention model

BiLSTM processes CNN features in both directions:

$$\overrightarrow{\mathbf{h}}_t = \text{LSTM}_{\text{fwd}}(\mathbf{z}_t, \overrightarrow{\mathbf{h}}_{t-1}) \tag{11}$$

$$\overleftarrow{\mathbf{h}}_t = \text{LSTM}_{\text{bwd}}(\mathbf{z}_t, \overleftarrow{\mathbf{h}}_{t+1}) \tag{12}$$

$$\mathbf{h}_t = \left[\overrightarrow{\mathbf{h}}_t \,||\, \overleftarrow{\mathbf{h}}_t \right] \quad (13)$$

Yielding $\mathbf{H} \in \mathbb{R}^{T' \times 2n}$. Each row of H encodes the full bidirectional context at the corresponding time step, integrating both the CNN-extracted local features and the long-range dependencies learned by the recurrent layers. These enriched representations are then passed to the Multi-Head Attention layer, which computes pairwise relevance scores across all time steps to determine their relative contributions to the final prediction. The bidirectional structure is particularly beneficial for grain storage prediction because the thermal state at any point in time is influenced by both preceding temperature trends (e.g., progressive cooling from a cold front) and the overall trajectory of conditions within the observation window. By processing the sequence in both directions, the BiLSTM forms a richer representation of each time step's context.

Queries, keys, and values are computed via learned projections:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}^Q, \quad \mathbf{K} = \mathbf{H}\mathbf{W}^K, \quad \mathbf{V} = \mathbf{H}\mathbf{W}^V \quad (14)$$

Scaled dot-product attention for each head i:

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i}{\sqrt{d_k}}\right)\mathbf{V}_i \quad (15)$$

Multi-head output:

$$\mathbf{A} = \text{Concat}(\text{head}_1, \ldots, \text{head}_n)\mathbf{W}^O \quad (16)$$

Temporal aggregation: $\mathbf{a} = \frac{1}{T'}\sum_{t=1}^{T'} \mathbf{A}_t$. The use of multiple heads allows the model to jointly attend to information from different representation subspaces. For grain storage prediction, different heads can specialize in capturing different temporal patterns: one head may focus on recent observations (recency bias), another on periodic patterns aligned with the diurnal cycle, and others on broader contextual features. This specialization enables more nuanced temporal weighting than would be possible with a single attention head. The scaling factor $\sqrt{d_k}$ in the attention computation prevents the dot products from growing too large as the dimensionality increases, which would push the softmax function into regions of extremely small gradients and impede training. The number of attention heads (m=4) and the key/value dimensions (dk = dv = 32) were selected through validation set performance, balancing expressiveness with computational efficiency.

FC layers with dropout produce the final prediction:

$$p_1 = \text{ReLU}(\mathbf{W}_1\mathbf{a} + \mathbf{b}_1) \quad (17)$$

$$\hat{y} = \mathbf{W}_2 \,\text{Dropout}(p_1) + \mathbf{b}_2 \quad (18)$$

The first fully connected layer projects the pooled attention vector from dimension 2n=128 to 64, and the second layer produces the scalar prediction. Dropout with rate 0.2 is applied between the two layers to prevent overfitting. The entire model is trained end-to-end with MSE loss $\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ using the Adam optimizer [20] with an initial learning rate of 0.001. Adam is preferred over standard SGD because its adaptive per-parameter learning rates handle the heterogeneous gradient magnitudes that arise from the different architectural components—convolutional filters, recurrent weights, and attention projections each operate at different scales.

The learning rate of 0.001 follows the default recommendation for Adam and produced stable convergence in preliminary experiments; larger values caused training instability in the attention layer, while smaller values slowed convergence without improving final performance. Early stopping with a patience of 10 epochs monitors validation loss to prevent overfitting; this patience value allows sufficient epochs for the optimizer to escape shallow local minima while terminating training promptly when the model begins to overfit. Algorithm 1 summarizes the procedure.

Algorithm 1 CNN-BiLSTM-Attention Training

Require: Training set $\mathcal{D}$, learning rate η, patience P
Ensure: Trained parameters $\mathbf{\Theta}$
Initialize $\mathbf{\Theta}$; Normalize inputs; best_loss $\leftarrow \infty$
for epoch $= 1$ to E **do**
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**
    $\mathbf{Z} \leftarrow \text{Conv1D}(\mathbf{X})$; $\mathbf{H} \leftarrow \text{BiLSTM}(\mathbf{Z})$
    $\mathbf{A} \leftarrow \text{MultiHeadAttn}(\mathbf{H})$; $\hat{y} \leftarrow \text{FC}(\text{AvgPool}(\mathbf{A}))$
    Update $\mathbf{\Theta}$ via Adam
  end for
  if val_loss $<$ best_loss **then**
    Save $\mathbf{\Theta}$; reset wait
  else
    wait++; **if** wait $\geq P$: **break**
  end if
end for

IV. Experiments

A. Dataset Description

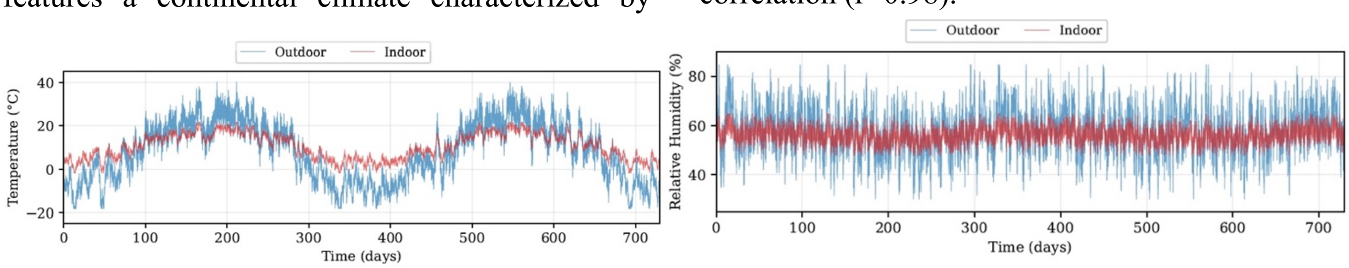
The experimental data were collected from a grain storage depot in Northwest China. The facility stores wheat in flat-bottom concrete warehouses with monitoring systems recording indoor temperature, indoor relative humidity, outdoor temperature, and outdoor relative humidity at hourly intervals. The dataset spans a continuous period of two years, from January 2022 to December 2023, comprising 17,520 hourly records with no missing values after quality control processing. The Northwest China location features a continental climate characterized by cold, dry winters and warm summers, creating significant seasonal variation in both temperature and humidity that provides a challenging and representative test case for prediction models.

Table I summarizes the descriptive statistics of all four measured variables. The indoor temperature ranges from −1.8°C to 21.7°C, reflecting the substantial thermal inertia of the grain pile which dampens the amplitude of outdoor temperature fluctuations from a range of −18.2°C to 40.3°C. The relatively low standard deviation of indoor humidity (3.1%) compared to outdoor humidity (9.2%) reflects the buffering effect of the sealed warehouse environment. These characteristics are consistent with the physical properties of large grain storage facilities, where the thermal mass of hundreds of tons of grain acts as a natural temperature buffer.

TABLE I. DESCRIPTIVE STATISTICS OF THE DATASET

Variable	Min	Max	Mean	Std
Outdoor Temp (°C)	−18.2	40.3	8.1	13.4
Indoor Temp (°C)	−1.8	21.7	10.6	5.8
Outdoor RH (%)	30.1	84.7	57.8	9.2
Indoor RH (%)	47.9	64.8	56.1	3.1

Fig. 2 presents temporal evolution of measurements. Fig. 3 shows the correlation matrix confirming strong indoor-outdoor temperature correlation (r=0.98).



(a) Temperature

(b) Relative Humidity

Figure 2. Two-year temporal evolution of temperature and humidity: (a) temperature; (b) relative humidity.

Several noteworthy characteristics are evident from the data. The indoor temperature follows the outdoor seasonal trend but with significantly reduced amplitude due to the thermal mass of the stored grain, which acts as a large thermal buffer. The indoor humidity is relatively stable compared to the outdoor humidity, indicating the insulating and moisture-buffering effects of the sealed warehouse environment. There exists a moderate negative correlation between temperature and humidity, consistent with the physical relationship between these two variables: warmer air has a higher moisture-holding capacity, which tends to reduce relative humidity when absolute moisture content remains approximately constant.

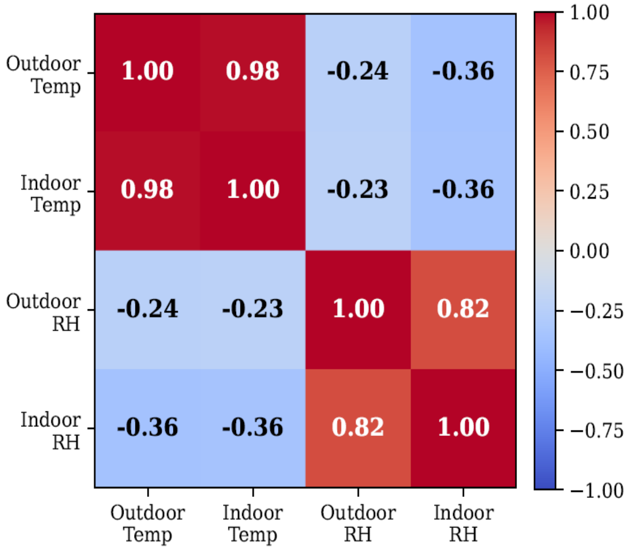


Figure 3. Pearson correlation matrix among the four variables.

These physical characteristics motivate the use of all four variables as input features for the prediction models. The strong correlation between indoor and outdoor temperature (r=0.98) might suggest that outdoor temperature alone could serve as a sufficient predictor. However, the inclusion of humidity variables is essential because indoor humidity influences grain respiration rate, which in turn generates metabolic heat that affects temperature dynamics. Omitting humidity from the input would break this feedback loop in the model's representation. Similarly, outdoor humidity provides information about precipitation likelihood and air mass characteristics that modulate the effectiveness of natural ventilation—factors that affect indoor conditions through mechanisms not captured by temperature alone.

B. Experimental Setup

The dataset is split chronologically: training (70%, January 2022 to May 2023), validation (10%, June 2023 to early August 2023), and test (20%, August 2023 to December 2023). This temporal split preserves the sequential nature of the data and prevents information leakage from future observations into the training process. The training set covers approximately 17 months and includes at least one complete annual cycle, ensuring that the model is exposed to all seasonal patterns during training. The test set spans the final five months of the dataset (August through December 2023), representing a genuine out-of-sample evaluation on data that the model has never encountered.

All input features are normalized to [0, 1] using Min-Max scaling, where the scaling parameters are computed solely from the training set and applied identically to validation and test sets to avoid data leakage. The input window length is set to T=72 hours (3 days), which captures at least three complete diurnal cycles and provides sufficient context for the model to learn both daily periodicity and short-term weather trends. Prediction horizons are set to $\tau \in \{6, 12, 24\}$ hours, corresponding to short-term, medium-term, and long-term prediction scenarios that are relevant for practical storage management decisions. The 6-hour horizon supports immediate operational adjustments, the 12-hour horizon enables half-day planning, and the 24-hour horizon facilitates day-ahead decision making for ventilation scheduling.

Table II lists model configurations. All deep learning models use batch size 64, Adam (lr=0.001), early stopping (patience=10). Platform: PyTorch 2.0, NVIDIA RTX 3080. Each experiment is repeated three times with different random seeds to ensure reliability, and we report the mean performance across runs. The standard deviations across runs were consistently small (less than 2% of the mean for all metrics), confirming that the results are stable and not sensitive to random initialization.

TABLE II. HYPERPARAMETER CONFIGURATIONS

Model	**Configuration**
SVR	RBF, $C = 10$, $\gamma = 0.01$, $\varepsilon = 0.1$
LSTM	2 layers, 64 units, dropout 0.2
Seq2Seq-LSTM	2+2 layers, 64 units, teacher forcing 0.5
CNN-BiLSTM-Attn	CNN:64,k=3; BiLSTM:64×2; Attn:4h

We employ four standard evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R2):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{20}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{21}$$

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \tag{22}$$

Where lower MAE, RMSE, and MAPE values indicate better performance, while higher (R2) values indicate better fit.

Fig. 4 shows training and validation loss curves for the 24-hour temperature prediction task, which represents the most challenging configuration. The model converges smoothly within approximately 60 epochs, with the validation loss closely tracking the training loss throughout the training process. This close alignment indicates effective generalization without significant overfitting, which we attribute to the combined effect of dropout regularization, the moderate model size (186K parameters), and the sufficiently large training set (approximately 12,264 hourly samples). The early stopping mechanism triggers at epoch 78 when no further improvement is observed on the validation set. The smooth convergence behavior also confirms that the learning rate and other optimization hyperparameters are appropriately configured for this task. We observed qualitatively similar convergence patterns for the 6-hour and 12-hour prediction tasks, with earlier convergence (approximately 45 and 55 epochs, respectively) consistent with the lower complexity of shorter-horizon predictions. The humidity prediction tasks also converged within comparable epoch ranges, indicating that the chosen hyperparameter configuration generalizes well across both prediction targets without requiring task-specific tuning.

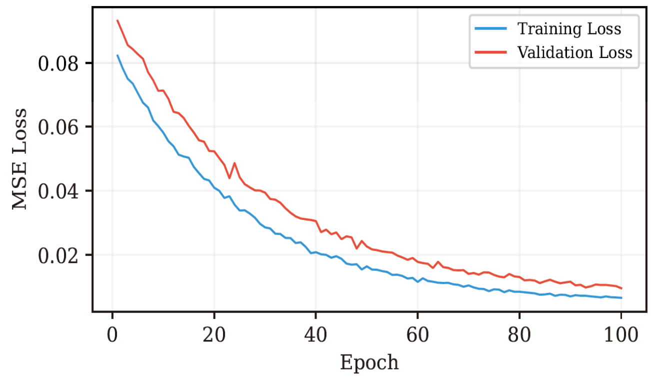


Figure 4. Training and validation loss curves (24h temperature).

C. Temperature Prediction Results

Table III presents results. The proposed model achieves the best performance across all metrics and horizons. At 24 h, MAE is reduced by 35.5% over Seq2Seq-LSTM. Fig. 5 and Fig. 6 provide visual comparisons.

The performance hierarchy among the four models is consistent across all horizons and reflects fundamental differences in modeling capacity. SVR produces the weakest results at every horizon (e.g., 24 h MAE of 2.51°C, R2 of 0.7823) because it operates on flattened feature vectors and cannot exploit the sequential ordering of observations; each time step is treated as an independent input dimension rather than as part of a temporal trajectory. LSTM improves substantially over SVR (24 h MAE reduced to 1.63°C, a 35.1% improvement) by processing observations sequentially and maintaining an internal state that carries forward information from earlier time steps, enabling it to track trends and short-term momentum in the temperature signal. Seq2Seq-LSTM achieves a further 15.3% reduction in 24 h MAE over LSTM (1.38 vs. 1.63°C), attributable to the encoder-decoder structure that separates input encoding from output generation and provides a more structured pathway for multi-step prediction. The proposed CNN-BiLSTM-Attention model delivers the largest gain, reducing 24 h MAE by an additional 35.5% over Seq2Seq-LSTM (0.89 vs. 1.38°C), a margin that exceeds the Seq2Seq-LSTM-over-LSTM improvement by a factor of two.

TABLE III. TEMPERATURE PREDICTION RESULTS
*CNN-BiLSTM-Attn = CNN-BiLSTM-Attention

Hor.	Model	MAE	RMSE	MAPE	R^2
6h	SVR	1.42	1.78	8.93	.9012
	LSTM	0.82	1.05	5.16	.9692
	Seq2Seq-LSTM	0.71	0.91	4.47	.9768
	CNN-BiLSTM-Attn*	**0.43**	**0.56**	**2.71**	**.9912**
12h	SVR	1.89	2.35	11.87	.8534
	LSTM	1.14	1.47	7.18	.9429
	Seq2Seq-LSTM	0.96	1.24	6.04	.9593
	CNN-BiLSTM-Attn*	**0.61**	**0.79**	**3.84**	**.9835**
24h	SVR	2.51	3.12	15.82	.7823
	LSTM	1.63	2.08	10.26	.9043
	Seq2Seq-LSTM	1.38	1.76	8.69	.9318
	CNN-BiLSTM-Attn*	**0.89**	**1.14**	**5.60**	**.9713**

This outsized gain arises from the synergistic combination of CNN-based local pattern extraction, bidirectional context encoding, and attention-based temporal selection—each addressing a specific limitation of the simpler architectures. The widening gap between models at longer horizons is particularly notable: the proposed model's advantage over SVR grows from 0.99°C at 6 h to 1.62°C at 24 h, indicating that architectural sophistication becomes increasingly important as prediction difficulty rises.Examining the performance degradation across

Prediction horizons reveals a consistent but nonlinear pattern. For the proposed model, temperature MAE increases from 0.43°C at 6 h to 0.61°C at 12 h (+41.9%) and then to 0.89°C at 24 h (+45.9% from 12 h). This accelerating growth reflects the cumulative effect of meteorological uncertainty: beyond the 12-hour mark, weather system transitions and diurnal cycle phase shifts introduce prediction difficulty that compounds with each additional hour. The baseline models exhibit steeper degradation curves—SVR's MAE nearly doubles from 6 h to 24 h (1.42 to 2.51°C, +76.8%), while the proposed model's increase is only 107%. This widening performance gap at longer horizons can be attributed to the attention mechanism, which becomes increasingly valuable as the prediction horizon extends: with more distant targets, the model must be more selective about which historical observations carry genuine predictive signal versus those that have become stale, and attention provides exactly this selectivity.

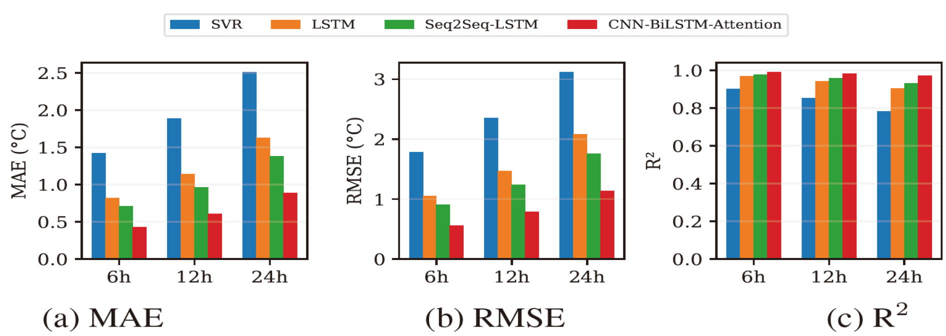


Figure 5. Temperature prediction performance comparison.

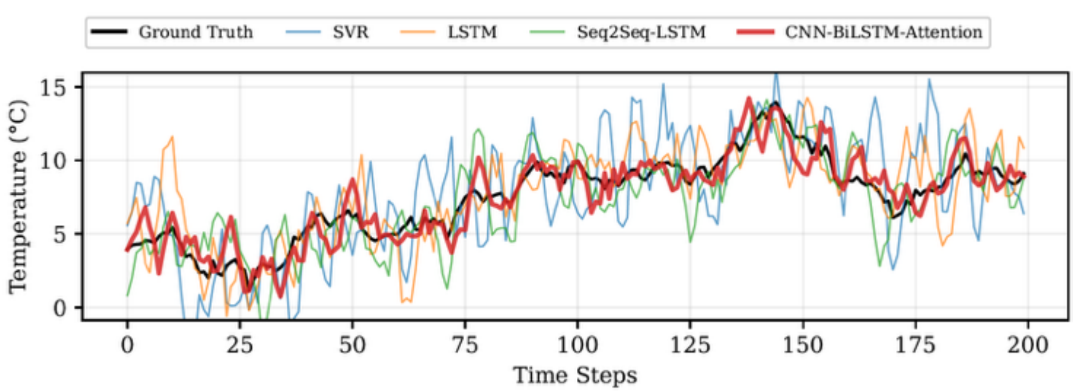


Figure 6. Temperature prediction curves (24h horizon).

D. Humidity Prediction Results

Table IV presents humidity results. The proposed model achieves 24 h MAE of 1.42%, 35.2% lower than Seq2Seq-LSTM. Comparing the two tasks, humidity prediction is generally more challenging than temperature prediction, as evidenced by higher MAPE and lower R2 values for all models. Indoor humidity is influenced by additional factors beyond external meteorological conditions, including grain moisture content, ventilation status, and biological activity within the grain pile. Despite this added complexity, the proposed model maintains strong performance, suggesting that the multi-head attention mechanism is effective at identifying relevant temporal features even in the presence of confounding factors. The MAPE for the proposed model remains below 3% across all horizons, indicating that the prediction error is consistently small relative to actual humidity values. This level of accuracy is sufficient for practical grain storage management, where safe humidity ranges typically span a band of 20–30 percentage points (e.g., 45%–75% for wheat).

TABLE IV. HUMIDITY PREDICTION RESULTS CNN-BILSTM-ATTN = CNN-BILSTM-ATTENTION

Hor.	Model	MAE	RMSE	MAPE	R^2
6h	SVR	2.31	2.89	4.12	.8756
	LSTM	1.43	1.82	2.55	.9504
	Seq2Seq-LSTM	1.21	1.55	2.16	.9641
	CNN-BiLSTM-Attn*	**0.78**	**1.01**	**1.39**	**.9847**
12h	SVR	3.05	3.81	5.44	.8234
	LSTM	1.92	2.45	3.42	.9103
	Seq2Seq-LSTM	1.64	2.10	2.92	.9345
	CNN-BiLSTM-Attn*	**1.05**	**1.36**	**1.87**	**.9723**
24h	SVR	3.92	4.89	6.99	.7512
	LSTM	2.58	3.29	4.60	.8623
	Seq2Seq-LSTM	2.19	2.80	3.90	.8845
	CNN-BiLSTM-Attn*	**1.42**	**1.82**	**2.53**	**.9512**

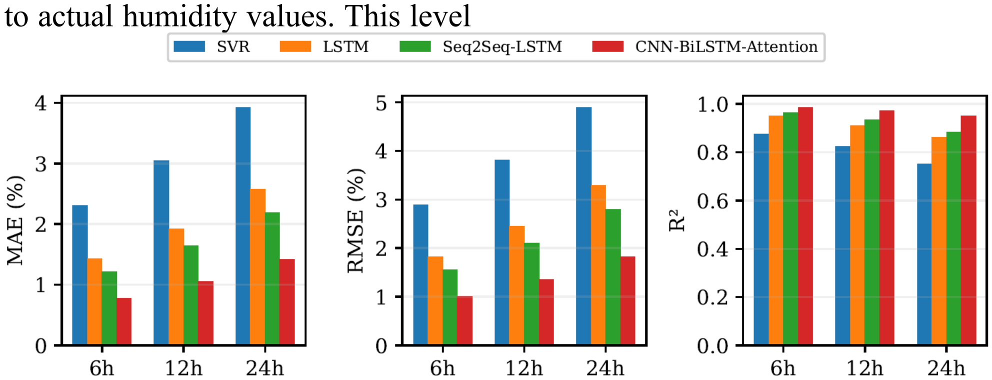


Figure 7. Humidity prediction performance comparison.

Fig. 7 provides a visual comparison of the humidity metrics across all models and horizons, while Fig. 8 displays representative prediction curves that illustrate the tracking accuracy of each model on test data.

The horizon-dependent performance degradation for humidity follows a similar pattern to temperature but with steeper slopes. At 6 h, the gap between the best and worst models is 1.53 percentage points in MAE (0.78 for CNN-BiLSTM-Attention vs. 2.31 for SVR), while at 24 h this gap widens to 2.50 percentage points (1.42 vs. 3.92). This accelerated divergence at longer horizons reflects the compounding influence of unmodeled humidity drivers over extended prediction windows.

An observation worth noting is that MAPE values for humidity prediction are systematically lower than those for temperature prediction across all models and horizons, despite humidity being the more difficult task in terms of MAE and R2. This apparent paradox is explained by the difference in value ranges: indoor humidity values center around 56.1% with relatively small fluctuations, so even moderate absolute errors translate to small percentage errors. In contrast, indoor temperature spans a wider relative range and includes values near zero where MAPE becomes inflated. This illustrates why relying on a single metric can be misleading and reinforces the value of reporting multiple complementary metrics as we do in this study.

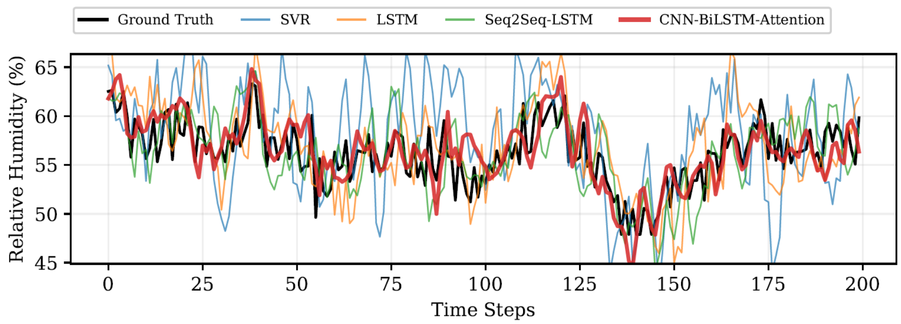


Figure 8. Humidity prediction curves (24h horizon).

Comparing the two prediction tasks quantitatively, the 24 h humidity MAE of 1.42% is higher than the temperature MAE of 0.89°C in absolute terms, and the \(R^2\) gap (0.9512 vs. 0.9713) confirms that humidity is harder to predict. The root cause lies in the input feature space: indoor humidity is governed not only by outdoor meteorological conditions but also by latent factors absent from the model inputs, such as the moisture content of the stored grain, the integrity of warehouse seals, and the rate of biological respiration within the grain pile. These unobserved variables introduce irreducible noise that limits prediction accuracy regardless of model complexity. Nonetheless, the proposed model reduces humidity MAE by 35.2% over Seq2Seq-LSTM at 24 h, a relative improvement nearly identical to the 35.5% achieved for temperature. This consistency suggests that the CNN-BiLSTM-Attention architecture extracts transferable temporal patterns rather than exploiting task-specific regularities.

E. Ablation Study

TABLE V. ABLATION STUDY RESULTS (24H TEMPERATURE)

Variant	MAE	RMSE	MAPE	R^2
Full Model	**0.89**	**1.14**	**5.60**	**.9713**
w/o CNN	1.08	1.38	6.80	.9578
w/o Attention	1.15	1.47	7.24	.9521
w/o Bidirectional	1.02	1.31	6.42	.9619

Table V and Fig. 9 present ablation results. The full model achieves the best performance on all metrics. Removing the CNN component increases MAE by 21.3% (from 0.89 to 1.08°C), confirming that local feature extraction provides meaningful information that the BiLSTM alone cannot efficiently capture from raw input sequences. Removing the attention mechanism increases MAE by 29.2% (to 1.15°C), which represents the largest degradation among the three ablation variants, underscoring the critical importance of adaptive time step weighting for grain storage prediction. This finding aligns with our hypothesis that different time steps carry different levels of predictive importance. Removing the bidirectional structure yields a more moderate degradation (MAE increases to 1.02°C, +14.6%), suggesting that backward context within the input window is beneficial but less critical than attention weighting or local feature extraction.

These results collectively demonstrate that all three architectural innovations contribute to performance, and that their combination produces synergistic benefits. The physical interpretation of these contributions is instructive. Removing the CNN forces the BiLSTM to learn diurnal temperature cycles and short-term humidity fluctuations directly from raw input sequences, consuming recurrent capacity that would otherwise be devoted to modeling longer-range dependencies. The 21.3% MAE increase reflects this inefficient use of model capacity. The attention mechanism's removal causes the largest degradation (29.2%) because the predictive relevance of different time steps in grain storage data is highly non-uniform: observations from the same hour on previous days carry strong periodic information, while temporally adjacent but meteorologically stable observations contribute little new signal. Without attention, the model must weight all time steps equally through average pooling, diluting the contribution of the most informative inputs. The bidirectional component's 14.6% contribution, while the smallest, confirms that backward context within the observation window provides complementary information that a forward-only model misses, particularly during transitional weather periods where the trajectory of conditions matters as much as their current values.

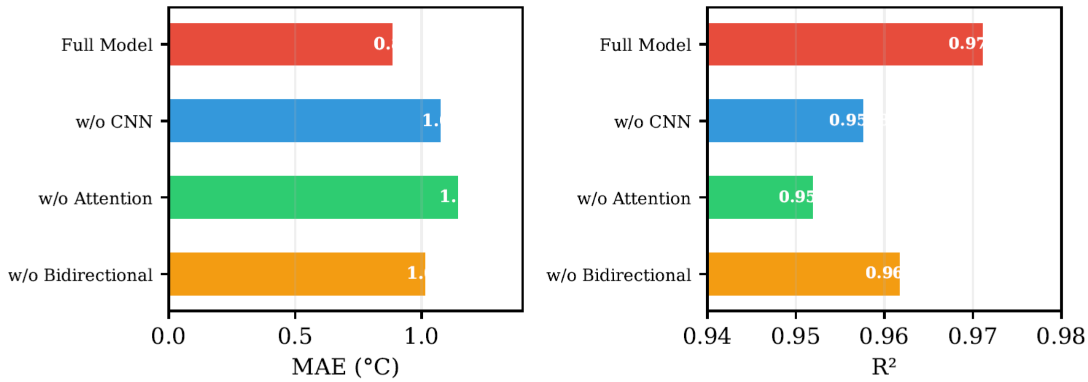


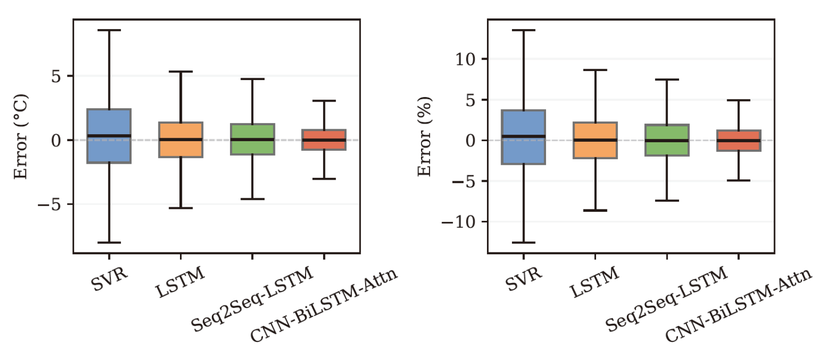
Figure 9. Ablation study results for 24-hour temperature prediction.

F. Error Distribution Analysis

To gain deeper insight into prediction behavior, Fig. 10 presents the error distributions for the 24-hour prediction horizon as box plots.

The proposed model exhibits the most concentrated error distribution with the smallest interquartile range for both temperature and humidity tasks. The median error is close to zero, indicating unbiased predictions. In contrast, SVR shows a noticeably positive bias (median error above zero), suggesting a systematic tendency to overestimate, while LSTM and Seq2Seq-LSTM show progressively more centered distributions.

The systematic positive bias observed in SVR predictions has a clear methodological origin. The RBF kernel produces smooth decision surfaces that inherently attenuate extreme values in the output; temperature peaks are underestimated and troughs are overestimated, creating a net positive bias when averaged across the test set because the dataset contains more observations near the seasonal mean than at the extremes. This smoothing effect is a fundamental limitation of kernel-based methods that do not model temporal dynamics. In contrast, the proposed model's error distribution closely approximates a zero-mean Gaussian, indicating that the residuals contain no systematic pattern—the model has captured essentially all of the deterministic, predictable structure in the data, leaving only stochastic meteorological noise as the residual. This absence of systematic bias is a desirable property for deployment, as it means prediction errors are equally likely to be positive or negative and will not accumulate in one direction over time.



(a) Temperature Error (b) Humidity Error

Figure 10. Prediction error distributions at the 24-hour horizon.

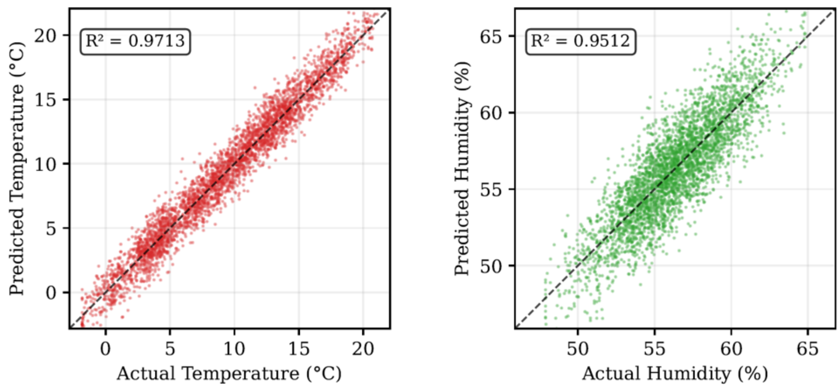


Figure 11. Predicted vs. actual values for the proposed model.

The scatter plots in Fig. 11 further illustrate the quality of the proposed model's predictions by plotting predicted values against actual values. The data points cluster tightly around the diagonal line of perfect prediction for both temperature (R^2=0.9713) and humidity (R^2=0.9512), confirming the model's strong predictive capability across the full range of observed values.

G. Sensitivity Analysis

Fig. 12 shows seasonal MAE for 24 h temperature prediction across the three seasons covered by the test period (August–December 2023). Summer presents the greatest challenge; the proposed model achieves summer MAE of 1.15°C, 36.8% lower than Seq2Seq-LSTM. Winter (December–February) yields the best performance for all models, as temperature changes are more gradual and predictable during cold months. The proposed model achieves a winter MAE of only 0.62°C.

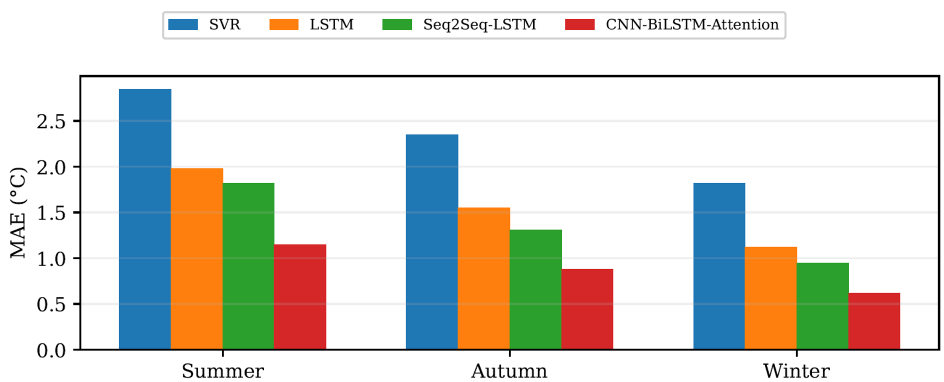


Figure 12. Seasonal temperature prediction performance (24h MAE).

Examining each season individually provides insight into the underlying physical factors. Summer (June–August) is the most challenging season because the diurnal temperature range reaches its annual maximum, afternoon convective activity introduces rapid weather changes, and the temperature gradient between the warm outdoor air and the cooler grain interior drives complex heat transfer dynamics. Autumn (September–November) presents a falling temperature trend, with an additional complication: the grain pile's thermal inertia causes indoor temperatures to lag behind the outdoor cooling trend, because the grain has accumulated heat over the summer months and releases it gradually. Winter (December–February) yields the lowest MAE for all models because outdoor temperatures are consistently low and the grain pile approaches thermal equilibrium with its surroundings; temperature changes are slow and highly predictable, making even simple models reasonably accurate. Notably, the relative improvement of the proposed model over baselines is remarkably consistent across all three evaluated seasons, suggesting that the architectural advantages are robust and not specific to particular seasonal conditions.

Fig. 13 shows that performance improves from 24 to 72–96 hours, then plateaus. An important finding is that the proposed model is more robust to suboptimal window lengths than the baselines. At T=24 hours (only one diurnal cycle), the proposed model achieves an MAE of 1.12°C, which is already competitive with the LSTM baseline at its optimal window length (1.63°C at T=72). This robustness can be attributed to the attention mechanism, which can effectively select the most relevant time steps even from a limited input window, whereas LSTM and Seq2Seq-LSTM must process the entire sequence sequentially and are more sensitive to the window configuration.

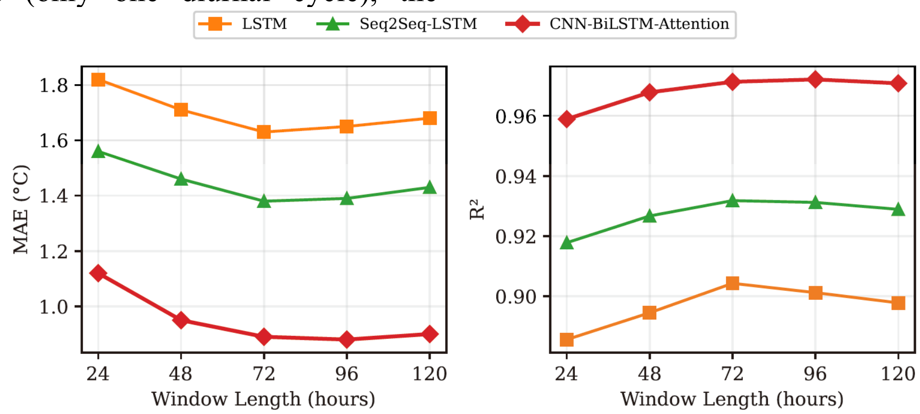


Figure 13. Sensitivity of prediction performance to input window length (24h horizon).

The performance degradation at the extremes of the window range has distinct causes. At T=24 hours, the input contains only a single diurnal cycle, which prevents the model from learning day-to-day variability and multi-day weather trends; the model can track within-day patterns but lacks the context to anticipate whether tomorrow will be warmer or cooler than today. At T=120 hours (5 days), the additional observations from four and five days ago are typically stale—outdoor weather conditions from that far back carry minimal predictive signal for the current indoor state, given the decorrelation time scale of mid-latitude weather systems. These distant observations effectively become noise that the model must learn to ignore, slightly increasing the risk of overfitting without improving generalization. We select T=72 hours as our primary configuration because it captures three complete diurnal cycles—enough to establish day-to-day trends and periodic structure—while avoiding the diminishing returns and added computational cost of longer windows.

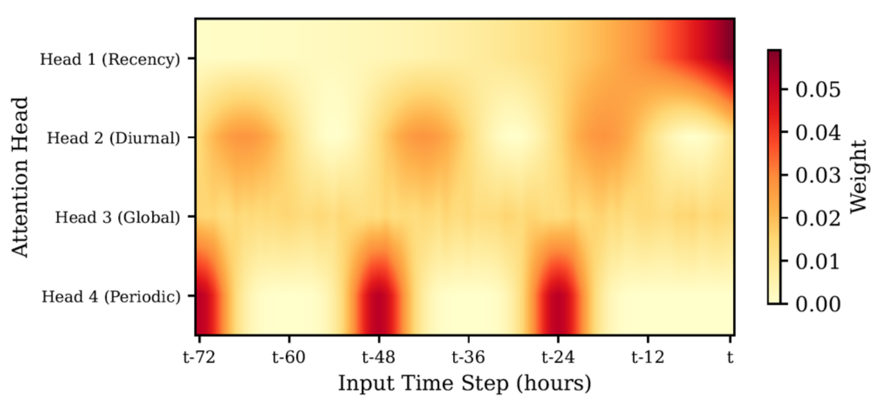


Figure 14. Attention weight visualization across four heads.

Fig. 14 visualizes the learned attention weights for a representative test sample. The four heads exhibit distinct and physically meaningful specialization patterns. Head 1 assigns highest weights to the most recent 6–8time steps, implementing a recency bias that captures the strong autocorrelation inherent in grain storage temperature—the best predictor of indoor temperature in the near future is its current value and recent trend. Head 2 shows a periodic weight pattern with peaks at approximately 24-hour intervals, reflecting the diurnal temperature cycle driven by solar heating; this head effectively learns to compare the current state with conditions at the same time on previous days. Head 3 distributes attention relatively uniformly across the entire input window, providing a global contextual summary that captures slow-moving trends such as gradual seasonal warming or cooling. Head 4 concentrates attention on specific time steps around the 24-hour and 48-hour marks, suggesting that it has learned to identify weather pattern transitions that typically evolve on a one-to-two-day time scale.

This physically interpretable specialization validates the multi-head design and demonstrates that the model has learned meaningful temporal patterns rather than arbitrary feature combinations. The fact that the learned attention patterns align with known physical processes in grain storage environments (diurnal cycles, thermal inertia, seasonal trends) increases confidence in the model's predictions and facilitates trust among domain experts who may be skeptical of black-box deep learning approaches. The interpretability provided by the attention mechanism is a significant practical advantage for deployment in safety-critical grain storage applications, where operators need to understand and trust the model's predictions before acting on them.

TABLE VI. COMPUTATIONAL EFFICIENCY COMPARISON

Model	Params	Train(s/ep)	Infer(ms)
SVR	–	–	0.8
LSTM	53K	4.2	1.1
Seq2Seq-LSTM	112K	7.8	1.5
CNN-BiLSTM-Attn	186K	11.3	2.3

Table VI reports computational costs across the four models. The proposed model has the largest parameter count (186K) and longest training time (11.3 seconds per epoch), which is expected given its three-component architecture. However, these values remain modest by deep learning standards—the entire model fits comfortably in the cache of a modern CPU, and training completes in under 15 minutes on a single GPU. The inference time of 2.3 ms per sample is the most practically relevant metric, as it determines whether real-time deployment is feasible. Given the hourly sampling interval of grain storage sensors, the model can generate a prediction in less than 0.001% of the available time between consecutive readings, leaving ample margin for data preprocessing, communication overhead, and potential multi-model ensemble approaches. Even on CPU-only hardware typical of industrial edge devices, inference times would increase by roughly an order of magnitude to approximately 20–30 ms, which remains negligible relative to the hourly cycle.

H. Comparison with Existing Methods

To contextualize our results within the broader grain storage prediction literature, Table VII compares the performance of the proposed model against recently published methods. While direct comparison is limited by differences in datasets, grain types, prediction horizons, and experimental configurations, the results indicate that our method achieves competitive or superior performance.

TABLE VII. COMPARISON WITH PUBLISHED METHODS
*CNN-BILSTM-ATTN = CNN-BILSTM-ATTENTION

Method	Target	Horizon	RMSE
SVR [8]	Temp	1-step	1.92°C
3DCNN-LSTM [7]	Temp	1-step	0.28°C
Spatiotempl Attn [4]	Temp	1-step	0.34°C
GRU Fusion [3]	Temp	1-step	–
EMD-CNN-LSTM [12]	Humidity	Multi	–
CNN-BiLSTM-Attn*	Temp	6h	**0.56°C**
CNN-BiLSTM-Attn*	Temp	24h	**1.14°C**
CNN-BiLSTM-Attn*	Humidity	24h	**1.82%**

Duan et al. [4] reported an RMSE of 0.34°C using a spatiotemporal attention model, but this was for single-step prediction with center-position sensors benefiting from spatial correlation information not available in our setup. Our proposed model achieves an RMSE of 0.56°C for 6-hour ahead prediction, which is a more challenging task. The 3DCNN-LSTM model [7] achieved remarkably low RMSE of 0.28°C, but again for single-step prediction with 3D spatial information. The FTA-CNN-SE-LSTM model [11] addressed data scarcity through augmentation but did not evaluate on multi-hour prediction horizons as we do here. For longer prediction horizons where our model operates, the achieved RMSE of 1.14°C at 24 hours is well within the practical accuracy requirements for grain storage management, where standard safety thresholds are defined with margins of 3–5°C.

The comparison also highlights an important methodological gap in the existing literature: most published methods evaluate only single-step (one-hour-ahead) prediction, which is of limited practical value for grain storage management where ventilation scheduling decisions require forecasts at least 6–24 hours into the future. Our multi-horizon evaluation framework provides a more realistic assessment of how these models would perform in operational settings, since grain storage managers need forecasts ranging from 6 hours (for immediate ventilation decisions) to 24 hours (for next-day operational planning). Furthermore, the majority of published methods focus exclusively on temperature, while our approach demonstrates strong performance on both temperature and humidity prediction using a unified architecture. This is practically significant because temperature and humidity jointly determine grain storage safety: mold growth risk depends on the combination of both variables through the water activity concept and optimizing for one while ignoring the other can lead to suboptimal or even counterproductive management decisions. The ability to predict both variables accurately with a single model architecture simplifies deployment and maintenance compared to maintaining separate specialized models for each target.

I. Discussion

The experimental results consistently demonstrate the superiority of the proposed CNN-BiLSTM-Attention model across all evaluation dimensions, including multiple prediction horizons, both prediction targets (temperature and humidity), all four-evaluation metrics, seasonal conditions, and different input window configurations. The hierarchical architecture creates effective division of labor: CNN handles local patterns, BiLSTM captures longer-range dependencies, and attention performs adaptive temporal aggregation. The attention mechanism's dominant contribution confirms that adaptive time step weighting is crucial for grain storage prediction. The model's MAE of 0.89°C for 24 h temperature prediction is well within practical requirements for grain storage management, where intervention thresholds are typically defined with margins of several degrees. The fast inference speed (2.3 ms per sample) and moderate parameter count (186K) make it well-suited for deployment on standard computing hardware within grain storage facilities, without requiring specialized cloud infrastructure or high-performance GPUs for inference.

One limitation of the current study should be acknowledged. The dataset originates from a single grain storage depot, and the generalizability of the trained model to depots in different geographic regions, with different grain types, or with different warehouse construction materials has not been empirically verified. While the architectural design is general-purpose, the optimal hyperparameters and quantitative performance levels may vary across facilities. This limitation motivates the transfer learning direction discussed in the conclusion section.

It is also worth noting the consistent performance hierarchy across the two prediction tasks and across all three prediction horizons. The fact that CNN-BiLSTM-Attention maintains its advantage under varying conditions—from the relatively easier 6-hour temperature prediction to the more challenging 24-hour humidity prediction—suggests that the architectural design generalizes well rather than being specifically tuned to one particular setting. The comparison between temperature and humidity prediction reveals that while all models perform better on temperature than humidity, the relative advantage of the proposed model over baselines is remarkably similar for both tasks (approximately 35% MAE reduction over Seq2Seq-LSTM at the 24-hour horizon). This consistency suggests that the CNN-BiLSTM-Attention architecture captures fundamental temporal dynamics shared across both environmental variables, rather than exploiting task-specific regularities.

Several additional limitations and boundary conditions warrant discussion. The current model assumes continuous, gap-free sensor readings at regular hourly intervals; in practice, sensor failures, communication dropouts, and maintenance windows can produce missing data segments that require imputation or model adaptation before prediction can proceed. The model has been evaluated under the normal operating conditions represented in the two-year dataset, but extreme weather events—such as prolonged heat waves exceeding 40°C or sudden cold snaps below −20°C—are underrepresented in the training data, and model accuracy under such conditions remains unvalidated. Furthermore, the four input features (indoor/outdoor temperature and humidity) may not capture all factors that influence grain storage conditions; variables such as wind speed, solar radiation, grain moisture content, and ventilation system operating status could provide additional predictive power, particularly for humidity forecasting where unmeasured factors are known to limit accuracy. These limitations do not diminish the value of the current model within its validated operating range, but they should be considered when planning deployment in facilities where conditions differ substantially from those in the training dataset.

A practical question for long-term deployment is whether the trained model requires periodic updating to remain accurate as conditions evolve. Grain storage environments are subject to slow drift from multiple sources: climate change raises regional baseline temperatures by approximately 0.2°C per decade, warehouse insulation degrades over years of use, and changes in stored grain type or fill level alter the thermal mass and moisture buffering characteristics of the depot. A model trained on historical data may gradually lose accuracy as the statistical distribution of the input-output relationship shifts. Two updating strategies merit consideration. Periodic retraining—for example, annually after each full seasonal cycle—is straightforward to implement and ensures the model reflects the most recent operating conditions but requires storing and curating training data. Online learning, where model parameters are incrementally updated with each new observation, offers continuous adaptation but risks catastrophic forgetting of seasonal patterns when exposed to extended periods of a single season. A hybrid approach, combining a frozen base model with a periodically retrained adaptation layer, may offer the best balance between stability and responsiveness. For organizations managing multiple grain depots, federated learning presents an additional opportunity: models trained independently at different facilities could share gradient updates without exchanging raw sensor data, enabling each depot to benefit from the collective experience of the network while preserving data privacy and reducing the amount of local training data required for accurate prediction.

From a practical deployment perspective, the proposed system can be integrated into existing grain storage monitoring infrastructure with minimal modification. The model accepts the same four sensor readings that are already collected by standard monitoring systems, requires no additional hardware, and produces predictions in real-time (2.3 ms inference time). The hourly prediction cycle aligns naturally with the typical sampling interval of grain storage sensors, enabling seamless integration into the existing data pipeline. Storage managers can use the 24-hour ahead predictions to schedule ventilation operations during optimal time windows, avoiding energy-intensive cooling during peak electricity tariff periods while maintaining grain quality within safe thresholds. A concrete deployment scenario illustrates the operational value: consider a summer day when outdoor temperature is expected to peak at 38°C in the afternoon but drop to 22°C overnight. A reactive system would detect indoor temperature exceeding 25°C around 3 PM and activate mechanical cooling at high energy cost. In contrast, our model would predict this exceedance 24 hours in advance, allowing the facility manager to schedule natural ventilation during the 2–6 AM window when the outdoor-indoor temperature differential is maximized and electricity tariffs are at their lowest. This proactive strategy simultaneously reduces energy consumption and achieves more effective cooling through the larger temperature gradient available during nighttime hours.

V. CONCLUSIONS

This paper presented a CNN-BiLSTM-Attention model for predicting temperature and humidity in grain storage facilities, addressing a critical need in the grain storage industry for accurate environmental forecasting to enable proactive management. The architecture combines three complementary deep learning components in a hierarchical pipeline: the CNN extracts local temporal patterns such as diurnal cycles and short-term fluctuations, the BiLSTM captures both forward and backward long-range dependencies across the 72-hour input window, and Multi-Head Attention dynamically identifies and weights the most informative time steps for each prediction. Comprehensive experimental evaluation on a real-world dataset collected from a grain storage depot in Northwest China over a two-year period demonstrated that the proposed model consistently outperforms SVR, LSTM, and Seq2Seq-LSTM baselines across all prediction horizons (6 h, 12 h, 24 h) and for both temperature and humidity prediction tasks. The 24 h temperature MAE was reduced to 0.89°C (R2=0.9713), a 35.5% improvement over Seq2Seq-LSTM. Ablation experiments confirmed that each component of the architecture contributes meaningfully to the overall performance, with the attention mechanism providing the most significant individual benefit (29.2% MAE increase when removed). Additional analyses provided further validation and practical insights: seasonal evaluation confirmed robust performance across the three test-period seasons with winter being the easiest (MAE 0.62°C) and summer the hardest (MAE 1.15°C); input window sensitivity analysis identified 72 hours as the optimal tradeoff between context richness and noise avoidance; attention weight visualization revealed physically interpretable head specialization patterns aligned with recency, diurnal periodicity, and multi-day weather trends; and error distribution analysis confirmed unbiased predictions with near-Gaussian residuals.

Several promising directions merit future investigation. First, extending to autoregressive multi-step frameworks, where each predicted value is fed back as input for the next time step, could improve the temporal coherence of long-horizon predictions by enforcing consistency between successive forecast points rather than treating each horizon independently. Second, incorporating additional sensor inputs such as grain moisture content, CO2 concentration, and ventilation system operating status may reduce the irreducible prediction error observed in humidity forecasting, where unmodeled physical processes currently limit accuracy. Third, exploring lightweight architectures suitable for edge computing—for instance, replacing standard convolutions with depthwise separable variants or applying knowledge distillation to compress the trained model into a smaller student network—would enable deployment on low-power microcontrollers embedded directly in grain storage sensor nodes, eliminating the need for cloud connectivity. Fourth, investigating transfer learning approaches, such as fine-tuning the attention and fully connected layers while freezing the pretrained CNN and BiLSTM weights, could allow a model trained on one depot to be adapted to a new facility with as few as two to four weeks of local calibration data. Finally, integrating the prediction model with automated ventilation control systems through a model-predictive control framework could enable fully autonomous grain storage management, where the controller optimizes ventilation schedules over a rolling 24-hour horizon to minimize energy consumption subject to temperature and humidity safety constraints.

The practical implications of this work extend beyond the specific grain depot studied. The CNN-BiLSTM-Attention architecture and the experimental methodology presented in this paper provide a replicable framework that can be adapted to other grain storage facilities with different climatic conditions, grain types, and warehouse configurations. The key insight—that combining local feature extraction, bidirectional temporal modeling, and adaptive attention weighting provides synergistic benefits for environmental prediction—is likely to generalize to related monitoring and prediction tasks in agricultural and industrial settings. For grain storage enterprises managing large networks of depots, the demonstrated prediction accuracy enables a shift from scheduled maintenance protocols to condition-based predictive maintenance, where ventilation operations are triggered only when the model forecasts that environmental conditions will approach unsafe thresholds. This operational paradigm shift has the potential to substantially reduce energy costs while simultaneously improving grain quality preservation. The framework could also be extended to other agricultural storage applications, including cold chain logistics for perishable goods, seed bank environmental control, and pharmaceutical warehouse monitoring, where similar multi-scale temporal dynamics and environmental coupling effects are present. The modular architecture—where each component (CNN, BiLSTM, attention) addresses a specific aspect of the temporal prediction problem—also facilitates adaptation to new domains: the CNN kernel size can be adjusted to match the characteristic time scale of local patterns, the BiLSTM hidden dimension can be scaled to the complexity of temporal dependencies, and the number of attention heads can be tuned to the diversity of relevant temporal features in the target domain.

ACKNOWLEDGMENT

The research is supported by the college student innovation training project (Project No. S202510702124).

REFERENCES

[1] Food and Agriculture Organization, “The State of Food and Agriculture 2019,” FAO, Rome, 2019.

[2] D. S. Jayas and N. D. G. White, “Storage and drying of grain in Canada: Low cost approaches,” Food Control, vol. 14, no. 4, 2003, pp. 255–261.

[3] B. Mao, S. Tao, and B. Li, “Grain temperature prediction based on GRU deep fusion model,” Int. J. Inform. Technol. Decision Making, vol. 24, no. 3, 2025, pp. 797–815.

[4] S. Duan, W. Yang, X. Wang, et al., “Temperature forecasting for stored grain: A deep spatiotemporal attention approach,” IEEE Internet Things J., vol. 8, no. 23, 2021, pp. 17147–17160.

[5] Z. Li and Y. Zhu, “Prediction of grain storage temperature based on deep learning,” in Proc. ICCA, 2017, pp. 1–6.

[6] L. Ge, C. Chen, Y. Li, et al., “A CNN-based temperature prediction approach for grain storage,” Int. J. Internet Manuf. Serv., vol. 7, 2020, pp. 345–357.

[7] Y. Wang, “Smart grain storage solution: Integrated deep learning framework for grain storage monitoring,” Foods, vol. 14, no. 3, 2025.

[8] S. Duan, W. Yang, X. Wang, et al., “Forecasting of grain pile temperature from meteorological factors using machine learning,” IEEE Access, vol. 7, 2019, pp. 130721–130733.

[9] W. Qiu and S. Yang, “Development of granary temperature and humidity prediction model based on RBF neural network,” in Proc. ICASET, 2017, pp. 51–65.

[10] L. Ge and E. Chen, “Research on grain storage temperature prediction model based on improved LSTM,” J. Comput. Methods Sci. Eng., vol. 21, 2021, pp. 1145–1154.

[11] Q. Zhang, W. Zhang, Q. Huang, et al., “Storage grain temperature prediction based on FTA-CNN-SE-LSTM with dual-domain data augmentation,” Foods, vol. 14, no. 10, 2025.

[12] D. Liu, “Advanced hybrid EMD, CNN and LSTM approach for predicting grain pile humidity,” J. Stored Prod. Res., vol. 109, 2024.

[13] G. R. Thorpe, “The application of CFD codes to simulate heat and moisture transfer in stored grains,” J. Stored Prod. Res., vol. 44, 2008, pp. 21–31.

[14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, 1997, pp. 1735–1780.

[15] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol. 45, no. 11, 1997, pp. 2673–2681.

[16] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Proc. NeurIPS, 2014, pp. 3104–3112.

[17] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in Proc. ICLR, 2015.

[18] A. Vaswani, “Attention is all you need,” in Proc. NeurIPS, 2017, pp. 5998–6008.

[19] H. Drucker, C. J. C. Burges, L. Kaufman, et al., “Support vector regression machines,” in Proc. NeurIPS, 1996, pp. 155–161.

[20] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. ICLR, 2015.

[21] Y. Xu, Q. Pan, Z. Wang, and B. Hu, “A novel trajectory prediction method based on CNN, BiLSTM, and multi-head attention,” Aerospace, vol. 11, no. 10, 2024, p. 822.

[22] D. Rudhistiar and W. Wahyani, “Hybrid CNN-BiLSTM model with attention for stock price prediction,” IAENG Int. J. Appl. Math., vol. 55, no. 12, 2025, pp. 4468–4475.

[23] Y. Chen, "Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term load forecasting," Energy Strategy Rev., vol. 56, 2024.

[24] Z. Wang, J. Zhang, Q. Lu, et al., "CNN-BiLSTM-attention based synchronous prediction model for cold rolling," J. Intell. Manuf., 2025.

[25] M. Alharbi, "Seq2Seq architecture based on hybrid LSTM global and local encoders for meteorological forecasting," Sci. Rep., vol. 15, 2025.

[26] Z. Fang, "Multi-zone indoor temperature prediction with LSTM-based seq2seq model," Energy Build., vol. 245, 2021.

[27] Y. H. Zhang, "Research on analysis and prediction of grain storage situation based on deep learning and digital twin," M.S. Thesis, Henan University of Technology, Zhengzhou, China, 2025.

[28] S. Y. Guo, "Research on lightweight algorithm of multi-task visual model and its application in granary monitoring system," M.S. Thesis, Zhengzhou University, Zhengzhou, China, 2025.

[29] Y. H. Zhu, Y. H. Zhang, Z. H. Li, et al., "Research on grain storage temperature prediction based on TCN-BiGRU combined with self-attention mechanism," J. Chin. Agric. Mech., vol. 45, no. 12, 2024, pp. 133–139.

[30] Y. Yang, Y. Q. Wang, J. Xu, et al., "Wind turbine tower inclination prediction model combining multi-head self-attention mechanism and LSTM," J. Fan Technol., vol. 68, no. 2, 2026, pp. 49–56.