RESEARCH ARTICLE

Spatiotemporal prediction model of pest infestation based on deep learning in low temperature grain storage

Fei Han[1, 2], Zhijie Leng[1, *]

[1]College of Engineering, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang, China. [2]School of Economics and Management, Mudanjiang Normal University, Mudanjiang, Heilongjiang, China.

Received: March 24, 2026; **accepted:** July 3, 2026.

Low temperature mechanical refrigeration and sealed insulation technologies provide guarantees for green grain storage. However, temperature, humidity, and gas conditions inside grain piles exhibit spatial heterogeneity, making it difficult to identify timely and assess proactively local high-risk pest areas. To address this issue, this study designed a framework for constructing spatiotemporal pest samples in low temperature grain storage and established a multi-step spatiotemporal pest risk prediction model using a convolutional long short-term memory network as the backbone and incorporating channel-space-time joint attention and a multi-task loss function. The results showed that the proposed model achieved an accuracy of 85.12% for predicting pest risk levels at key locations and 97.56% at the warehouse scale. For 1-, 3-, and 7-day-ahead prediction tasks, the hit rates of high-risk grids reached 90.97%, 92.30%, and 83.20%, respectively, outperforming the benchmark models. Under different seasonal, vertical, and typical regional spatial scenarios, the average error between predicted risk indices and measured values was generally controlled below 0.10. The study indicated that a spatiotemporal pest prediction model based on multi-source monitoring data fusion and deep learning could finely depict the spatiotemporal evolution pattern of pests in low temperature grain storage, providing valuable references for improving grain storage safety management.

Keywords: low temperature; grain storage; pest; spatiotemporal prediction; deep learning; warehouse monitoring; risk warning.

[*]**Corresponding author:** Zhijie Leng, College of Engineering, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China. Email: jenniferleng@163.com.

Introduction

With the continuous improvement of national requirements for grain quality and safety, low temperature grain storage has become an important technical route for green grain storage. Mechanical refrigeration and sealed insulation reduce the grain-pile temperature, suppress pest metabolism, and support long term quality maintenance [1, 2]. However, grain piles are heterogeneous porous media, and temperature, humidity, gas concentration, and pest activity often vary across vertical layers and horizontal locations [3]. Therefore, pest prevention in low temperature flat warehouses requires not only monitoring of average warehouse conditions but also fine scale assessment of local spatiotemporal risk [4, 5].

Existing research on stored grain pest monitoring has gradually shifted from manual inspection to sensor-based monitoring, intelligent detection, and deep learning-based identification. Fürstenau *et al*. deployed semiochemical traps inside and outside four farms for two years and proposed an integrated field-warehouse early warning approach, showing that several stored-grain pests were active both inside and outside storage environments [6]. Mendoza *et al*. developed an online pest detection system based on cameras, Jetson Nano, and deep learning models to improve the timeliness of identifying multiple stored-grain pests in grain facilities [7]. Müller-Blenkle *et al*. proposed the "Beetle Sound Tube" acoustic detection system, which embedded sensors in grain bulks for continuous collection and automatic analysis of insect sounds, providing a feasible method for detecting hidden low-density infestations [8]. These studies demonstrate that multi-source sensing and automated detection enhance the early warning capabilities of stored-grain pest monitoring. Driven by advancements in deep learning, image-based pest recognition has advanced significantly. Jamali *et al*. constructed a pest monitoring model based on YOLOv7 and YOLOv8 to automatically detect red flour beetles and lesser grain borers with the optimal configuration achieving a 97.5% recall rate on a rice dataset [9]. Ong *et al*. proposed an insect identification framework based on MobileNetV2 for sticky-trap images and found that the recognition performance was closely related to trap color and image background conditions [10]. Xia *et al*. combined U-Net with attention mechanisms to segment pest-damaged areas and count damage patches, obtaining a coefficient of determination of 0.879 between model outputs and ground measurements [11]. These studies demonstrate the effectiveness of deep learning in pest image recognition, target localization, and damaged-area extraction. Although previous studies improved the accuracy and automation of pest monitoring, they focused primarily on macro-level presence detection, species recognition, acoustic or image-based identification, and static spatial assessment. In low temperature grain storage, the internal conditions of grain piles are invisible from the outside, and pest activity is governed by coupled variations in temperature, relative humidity, CO_2 concentration, vertical layering, and local positioning. Although mechanical refrigeration and sealed insulation technologies suppress overall pest activity, significant spatial differences still exist inside grain piles [12, 13]. Existing single-sensor, single-point, or image-oriented models cannot sufficiently characterize the continuous spatiotemporal evolution of hidden pest risks inside grain piles. Consequently, current methods fail to provide forward looking spatial decision support for localized cooling, targeted ventilation, and precision pest treatment.

This study constructed a multi-step spatiotemporal pest risk prediction model for low temperature grain storage by developing a spatiotemporal sample construction framework to integrate environmental and pest monitoring data, mapping discrete monitoring points onto gridded warehouse representations. A convolutional long short-term memory (LSTM) network was used as the prediction backbone, and channel-space-time joint attention and multi-task learning were introduced to improve the identification of high-risk grids and critical historical moments. This study transformed sparse multi-source monitoring data into continuous spatiotemporal risk grids and established a deep learning-based modeling paradigm for proactive pest prevention in low temperature grain storage, ultimately advancing the safety and intelligence of grain storage management. The proposed model could finely depict the evolution of pest risks inside grain piles and provide interpretable prediction results for warehouse scale risk assessment and key location warning.

Materials and methods

Multi-source data fusion and spatiotemporal pest sample construction in low temperature grain storage

This study proposed a framework for spatiotemporal pest sample construction and multi-source data fusion (ST-SC) in a low temperature flat grain storage warehouse using mechanical refrigeration and sealed insulation. For unified notation, the warehouse contained *P* monitoring points. The environmental feature vector at the *P* monitoring point at time *t* was denoted as $x_{t,p} = [T_{t,p}, H_{t,p}, C_{t,p}, L_{t,p}, \dots]^\top$, where $T_{t,p}$ was the grain pile temperature (℃), $H_{t,p}$ was the relative humidity (%), $C_{t,p}$ was CO_2 concentration (ppm), and $L_{t,p}$ was the vertical layer of the grain pile where the monitoring point was located. Pest monitoring data included pest activity intensity and pest count. The study concatenated environmental and pest information into an extended feature vector as follows.

$$z_{t,p} = [x_{t,p}^\top, \rho_{t,p}]^\top \tag{1}$$

where $\rho_{t,p}$ was the pest density at time *t* and monitoring point *P*. $z_{t,p}$ was the basic feature for subsequent spatial interpolation and gridding, integrating both environmental factors and pest intensity. Multi-point collaborative monitoring of the environment and pest situation in low temperature flat warehouses and silos relied on a unified sensor layout and data acquisition system. In the monitored warehouse, the pest surveillance system consisted of mechanical refrigeration and sealed-insulation facilities, environmental sensors, pest monitoring devices, and a data acquisition and storage unit. Temperature, relative humidity, and CO_2 sensors were arranged along the horizontal and vertical directions of the grain pile, while pest monitoring devices were placed near grain pile boundaries and representative internal points. These devices were connected to the data acquisition and control system through an industrial communication network, forming synchronized multi-source observations for subsequent gridded modeling. Multi-source data from monitoring could be irregular in the sampling rate with missing values and abrupt outliers. To enhance the robustness of the model predictions, this study normalized the original data and detected outliers for each variable with the processed observations being expressed as follows.

$$\tilde{x}_{t,p}^{(k)} = \frac{x_{t,p}^{(k)} - \mu_k}{\sigma_k} \tag{2}$$

where μ_k and σ_k were the mean and standard deviation of feature *k* in the training set. Considering minor timestamp offsets across different sensors, the study aligned multi-source data through time resampling and smoothed pest sequences using a moving average [14]. Before entering the modeling stage, raw environmental and pest sequences were processed through outlier detection and correction, missing-value interpolation, time alignment, pest sequence smoothing, and standardization. The preprocessing pipe identified abnormal values for each environmental and pest variable and corrected them to reduce abrupt measurement noise. Missing records caused by sensor interruption were then interpolated, and timestamp offsets among different sensors were unified through resampling. A moving average smoothed the pest density sequence was finally applied. This study standardized all variables to obtain unified time-series data for subsequent spatial gridding. To capture pest spatial distribution within a single warehouse, the three-dimensional warehouse space was discretized into $H \times W$ regular grids, each (i, j) representing a local grain pile volume. The spatial coordinates of monitoring point *P* were denoted as $s_p = (x_p, y_p, z_p)$ and assigned to the nearest grid set $\Omega_{i,j}$. The study used inverse distance weighting interpolation to map monitoring point data to warehouse grids. The feature of each grid unit was calculated as below.

$$\begin{cases} g_{t,i,j} = \dfrac{\sum_{p \in \Omega_{i,j}} w_{p,i,j} z_{t,p}}{\sum_{p \in \Omega_{i,j}} w_{p,i,j}} \\ w_{p,i,j} = \dfrac{1}{d_{p,i,j}^{\alpha} + \varepsilon} \end{cases} \tag{3}$$

where $g_{t,i,j}$ was the integrated feature vector of grid (i, j) at time *t*. $w_{p,i,j}$ was the interpolation weight from monitoring point *P* to grid (i, j). $d_{p,i,j}$ was the Euclidean distance from monitoring point *P* to the grid center. The explicit form of the interpolation weight was shown below.

$$w_{p\to g} = \frac{d_{pg}^{-\alpha}}{\sum_{p'\in\mathcal{P}_g} d_{p'g}^{-\alpha}} \quad (4)$$

where $w_{p\to g}$ was the normalized interpolation weight of monitoring point P for grid unit g. d_{pg} was the Euclidean distance from the monitoring point to the grid center. α was the distance decay exponent. $\mathcal{P}_g$ was the set of monitoring points involved in interpolating grid g. The study constructed positional encoding vectors using normalized grid coordinates. The two-dimensional position vector was expressed as follows.

$$e_{i,j} = [\frac{i}{H}, \frac{j}{W}, (\frac{i}{H})^2, (\frac{j}{W})^2]^\top \quad (5)$$

where $e_{i,j}$ was the position encoding of grid (i,j). The study concatenated environmental-pest features with position encoding to obtain the complete feature vector of each grid unit as below.

$$f_{t,i,j} = [g_{t,i,j}^\top, e_{i,j}^\top]^\top \quad (6)$$

where $f_{t,i,j}$ was the basic unit for subsequent spatiotemporal sample construction and model input. According to pest density grading standards related to grain storage pest control, pest density $\rho_{t,i,j}$ could be mapped to discrete risk levels $r_{t,i,j}$ as below.

$$r_{t,i,j} = \sum_{k=1}^{K} k \cdot 1(\theta_{k-1} < \rho_{t,i,j} \le \theta_k) \quad (7)$$

where $\{\theta_k\}_{k=0}^{K}$ was the pest density thresholds being set based on standards and expert experience. The study set risk levels ranging from 0 to 3, where 0 indicated almost no pests and 3 indicated high density. The construction of input-output tensors using a sliding window was expressed below.

$$X_t = \{F_{t-L+1}, \dots, F_t\}, Y_t = \{R_{t+1}, \dots, R_{t+H}\} \quad (8)$$

where F_t was the grid feature tensor at time t. L was the length of the historical window. H was the number of future prediction steps. R_t was the grid of risk level labels mapped from equation (6). This proposed ST-SC framework collected multi-source sequences including temperature, relative humidity, CO_2, and pest density, and preprocessed them through missing-value interpolation, outlier correction, and unified timestamp alignment. Monitoring-point features were then mapped to warehouse grids using inverse distance weighting, and normalized grid coordinates were encoded as position vectors. These position vectors were concatenated with environmental-pest features to generate grid representations. According to the historical and future prediction windows, the framework further generated spatiotemporal samples and risk labels based on pest density grading, achieving the transformation from raw monitoring data to model-ready tensors. The construction of the ST-SC framework not only achieved data alignment but also enabled the digital reconstruction of the grain pile microenvironment field. Given the complex anisotropy and invisibility within grain piles, the ST-SC framework employed spatial topological mapping to transform sparse monitoring data into spatiotemporal tensors that reflected physical distribution characteristics, thereby endowing the model with the capability to perceive gradient changes within the grain pile's internal environment.

Spatiotemporal prediction model combining ConvLSTM and attention mechanism

The ST-SC framework constructed unified pest samples within a single warehouse, providing input for the spatiotemporal prediction model. However, LSTM or convolutional neural network (CNN)-LSTM models built from single points or few layers could only capture temporal dependencies or coarse spatial features. Therefore, the study introduced a convolutional long short-term memory (ConvLSTM) backbone, which naturally incorporated local spatial correlations within its gated recurrent structure, making it suitable for gridded spatiotemporal pest sequences [15, 16]. The input of the l-th ConvLSTM layer at time t was denoted as $X_t^{(l)}$. The encoder-decoder ConvLSTM backbone could capture pest evolution in both time and space through a hierarchical sequence-to-sequence structure. In the encoder, several ConvLSTM layers and convolutional downsampling modules progressively extracted multi-scale spatiotemporal features and aggregated sequence information at the bottleneck layer. In the decoder, symmetric ConvLSTM layers and upsampling modules generated intermediate feature sequences for multiple future steps. The output head consisted of two branches including risk probability estimation and risk level classification, jointly predicting future pest risk grids. The decoder used symmetric ConvLSTM layers and upsampling structures to generate intermediate feature sequences for multiple future steps. Relying only on local receptive fields of convolution kernels, the model could not automatically focus on high-risk pest regions and critical time steps [17, 18]. To enhance key feature extraction, the study introduced an attention mechanism into ConvLSTM. The channel-space weighted feature map was updated as below.

$$U^{'}(i,j,c) = M_s(i,j) \cdot U_c(i,j,c) \tag{9}$$

where U_c was the channel-weighted feature map. M_s was the spatial attention map used to measure the importance of each grid position at the current time step. To improve the model's ability to identify hidden and localized pest risks, a channel-space-time joint attention module was embedded into the ConvLSTM backbone. Channel attention assigned adaptive weights to different monitoring variables and strengthened leading environmental indicators related to pest outbreaks such as CO_2 anomalies and localized temperature increases. Spatial attention further enhanced grids with higher pest occurrence probability, especially wall interfaces, grain corners, and other areas with active heat exchange, thereby reducing the dilution of local high-risk regions by the overall low temperature background. Temporal attention focused on key historical moments in the accumulation process from concealed pest activity to visible risk formation, enabling the model to capture time steps that strongly influenced future pest evolution. Let the encoder output hidden state sequence be $\{H_{t-L+1}, \ldots, H_t\}$, to emphasize key historical moments, the temporal attention mechanism calculated attention weights and weighted hidden states as follows.

$$\begin{cases} \alpha_\tau = \dfrac{exp(q^\top vec(H_\tau))}{\sum_{s=t-L+1}^{t} exp(q^\top vec(H_s))} \\ H_{att} = \sum_{\tau=t-L+1}^{t} \alpha_\tau H_\tau \end{cases} \tag{10}$$

where q was the learnable temporal query vector. $vec(\cdot)$ was the flattened tensor. α_τ was the temporal attention weights. H_{att} was the weighted sum of historical hidden states. The training objective considered both risk level classification and pest intensity regression with an L2 regularization term to suppress overfitting. The overall loss function was expressed as below.

$$\mathrm{L} = \lambda_1 \mathrm{L}_{\mathrm{WCE}} + \lambda_2 \mathrm{L}_{\mathrm{MSE}} + \lambda_3 \left\|\Theta\right\|_2^2 \tag{11}$$

where $\mathrm{L}_{\mathrm{WCE}}$ was the cross-entropy loss for four-class risk level classification. $\mathrm{L}_{\mathrm{MSE}}$ was the mean squared error between the predicted pest risk probability (sigmoid output) and the observed risk probability. Θ was the set of all trainable model parameters. λ_1, λ_2, λ_3 were weighting coefficients. To emphasize high-risk samples, larger class weights w_h were assigned to high-risk categories (level ≥ 2) in $\mathrm{L}_{\mathrm{WCE}}$ with default w_h= 0.1 and w_h= 0.2 used in sensitivity analysis for comparison. Based on the ST-SC spatiotemporal sample framework, the study used ConvLSTM as the backbone and introduced attention mechanisms to enhance the representation of high-risk pest regions and critical time steps, constructing the low temperature grain storage spatiotemporal pest prediction model (ST-CLSTM-Att) (Figure 1). The ST-CLSTM-Att model used training, validation, and test samples divided by warehouse and year. The model structure included a ConvLSTM encoder, spatiotemporal attention module, ConvLSTM decoder, and multi-task output layer, generating future multi-step pest risk grids and corresponding risk level predictions. During inference, the model received the most recent L

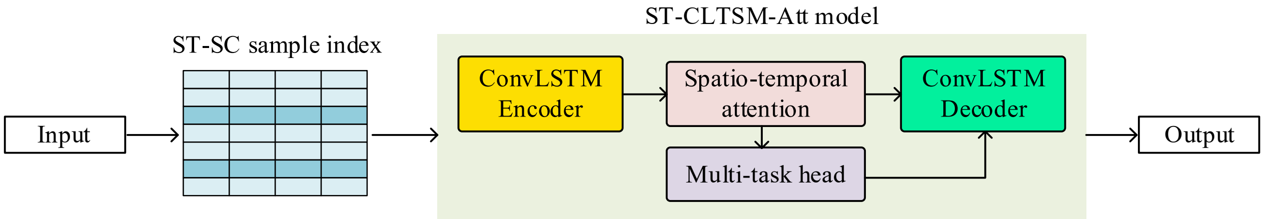

Figure 1. ST-CLSTM-Att model architecture.

steps of grid sequences X as input. After ConvLSTM encoding-decoding and spatiotemporal attention weighting, it output pest risk probability grids and corresponding risk level grids for H_f future steps. To facilitate intuitive judgment for managers, a comprehensive risk index for each grid during the prediction period could be calculated. The comprehensive risk index was expressed as follows.

$$\bar{R} = \frac{1}{H_f}\sum_{h=1}^{H_f} \mathrm{E}[R_{t+h}],\ \ \mathrm{E}[R] = \sum_{C=0}^{3} c \cdot p_c \qquad (12)$$

where R_{t+h} was the pest risk probability of grid (i, j) at time $t + h$. $\bar{R}$ was the average risk level of the grid within the future window. p_c was the four-class probability output of the risk level branch using Softmax (https://docs.pytorch.org/docs/2.12/generated/torch.nn.Softmax.html). $\mathrm{E}[R]$ was the expected value of the continuous risk level.

Study site, dataset, software sources, and experimental design

The monitoring dataset used in this study was collected from the No. 5 Low Temperature Flat Warehouse of the Zhengzhou Grain Reserve Depot in Zhengzhou, Henan, China, which employed mechanical refrigeration and sealed insulation technology. The internal temperature, relative humidity, CO_2 concentration, and multi-point pest monitoring in the warehouse were continuously recorded from October 2023 to September 2025. Sensors were deployed along the height and horizontal dimensions of the grain pile, consistent with the monitoring-point setup used in the ST-SC framework. Following the chronological division principle, the first 70% of data from the monitoring period was used as the training set, while the middle 15% of data was used as the validation set and the last 15% of data was used as the test set. Linear regression (LR) (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html), bidirectional long short-term memory (BiLSTM) (https://docs.pytorch.org/docs/stable/generated/torch.nn.LSTM.html), and residual YOLOv4 (Res-YOLOv4) implemented in PyTorch (https://pytorch.org/) were employed as comparison models. The hardware environment consisted of a Linux server equipped with an NVIDIA RTX 3090 graphics processor, Intel Xeon Silver 4210 processor, and 128 GB memory. The operating system was Ubuntu 20.04 LTS, and the acceleration environment used CUDA Toolkit 11.8 (https://developer.nvidia.com/cuda-11-8-0-download-archive) and cuDNN 8.7 (https://docs.nvidia.com/deeplearning/cudnn/archives/cudnn-870/index.html). To compare the ability of different models to reconstruct the spatial distribution pattern of pest infestation in low temperature grain storage at different prediction horizons, rasterized distributions of pest risk levels at future 1-day, 3-day, and 7-day time points were used as evaluation points. For Res-YOLOv4, connected high-risk areas with risk level ≥ 2 were treated as detection targets, and object detection results were rasterized and projected to calculate high-risk grid hit rates under the same spatial dimension. In addition, the warehouse-wide average comprehensive risk index and the high-risk raster hit rate were used to evaluate the temporal sensitivity and spatial warning ability of the models. To analyze the model's ability to characterize pest evolution

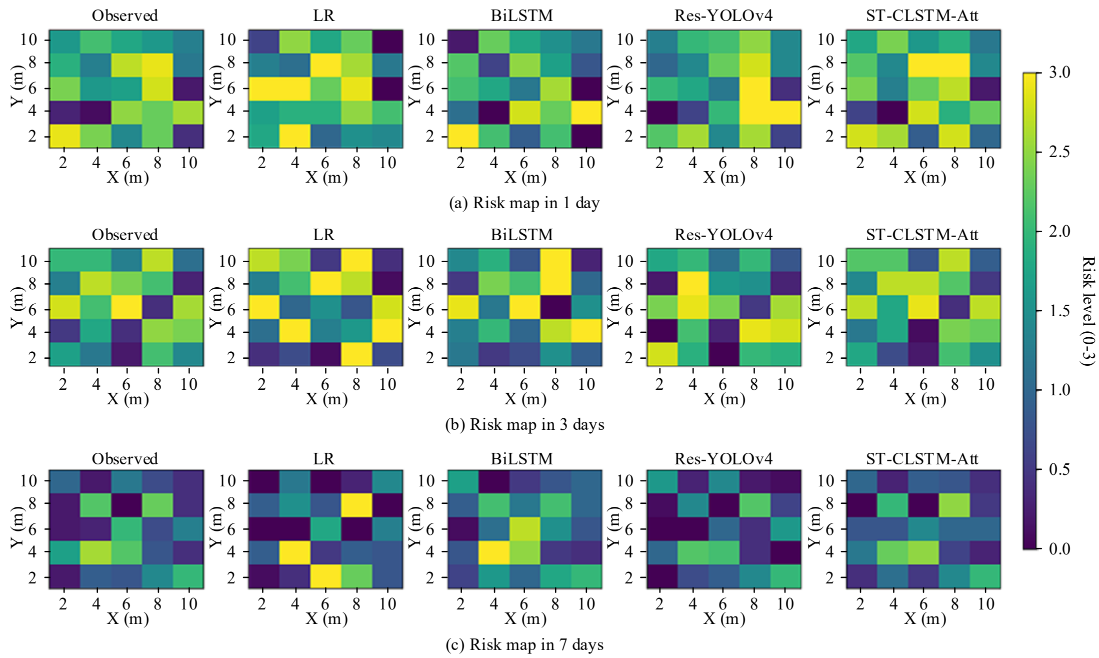

Figure 2. Comparison of pest spatial distribution at 1-day (**a**), 3-day (**b**), and 7-day (**c**) prediction time points.

under different external temperature and seasonal contexts, daily-scale risk indices in winter, transition season, and summer were used to compare the consistency between observed processes and predicted processes. To further evaluate spatial risk discrimination, pest risk statistics were calculated under three height levels and four typical warehouse area types across different seasons. Risk validation was also conducted at different height levels and representative horizontal regions within the warehouse. To quantitatively evaluate the contribution of the ST-SC framework and the spatiotemporal attention mechanism, ablation experiments were conducted using four variants including baseline ConvLSTM, ST-ConvLSTM with the ST-SC framework, ConvLSTM-Att with the attention module, and the complete ST-CLSTM-Att model. High-risk detection precision was compared under warehouse scale and key location evaluation conditions. Sensitivity analysis was further performed by changing the high-risk class weight from 0.1 to 0.2 to examine the robustness of different model variants under weight perturbation.

Results and discussion

Overall prediction performance and spatial alignment

The spatial alignment results of different models at 1-, 3-, 7-day prediction horizons demonstrated that, within 1 day, ST-CLSTM-Att model's fitted results for the location and intensity of high-risk areas aligned most closely with the observed values. Specifically, the expected risk level in the primary risk area in the bottom-left corner remained stable between 2.6 and 2.8, whereas the observed value was 2.8 (Figure 2a). For the 3-day prediction, ST-CLSTM-Att consistently produced high-risk zones above 2.0 with a central peak of 2.7, closely matching the observed distribution (Figure 2b). When the prediction horizon extended to 7 days, most comparison models showed misalignment of high-risk blocks and false scattered points, whereas ST-CLSTM-Att maintained relatively complete high-risk areas with levels of 2.3 - 2.6 in the upper and corner regions, closing to the observed raster levels of 2.3 - 2.5 (Figure 2c). The results indicated that ST-CLSTM-Att possessed a stronger ability to reconstruct the spatial distribution patterns of

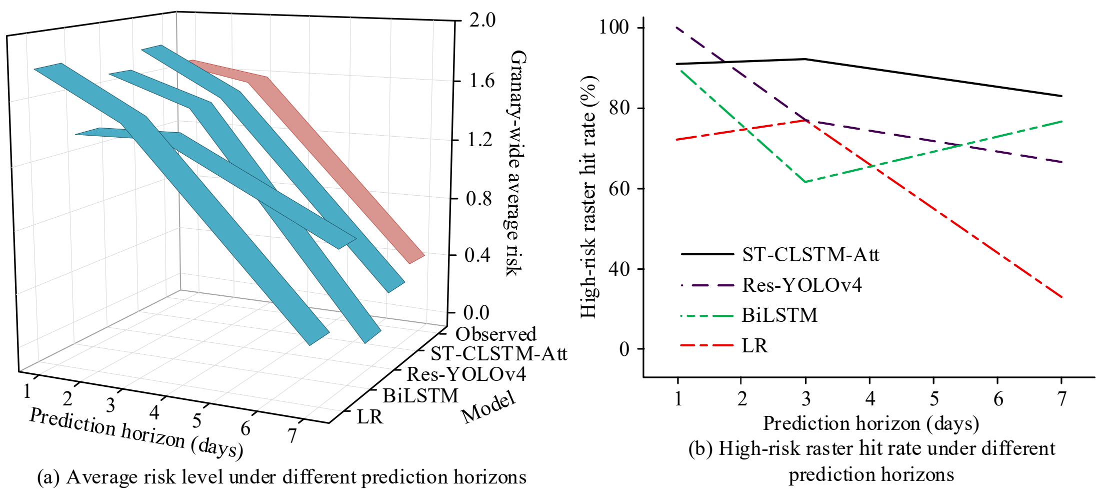

Figure 3. Comparison of risk level and high-risk identification capability.

pest infestation in low temperature grain storage than the benchmark models. This advantage was mainly attributed to the combination of local spatiotemporal dependency modeling and spatiotemporal attention weighting. The ConvLSTM module suppressed risk block drift by preserving local temporal continuity, while the attention module strengthened high-risk channels, key spatial regions, and critical historical moments. Therefore, local high-risk grids were less likely to be diluted by the overall low temperature background or by smoothed warehouse level trends. The temporal risk assessment and high-risk grid warning results showed that, starting from day 3, the warehouse wide average risk value of ST-CLSTM-Att consistently closely followed the observed trajectory. At day 7, the observed warehouse wide average risk value was 0.40, while the ST-CLSTM-Att predicted value was 0.31. Among the benchmark models, LR was closer to the observed values than BiLSTM and Res-YOLOv4, which was because that this indicator exhibited roughly monotonic and smooth changes over time, making it easier for linear regression to fit the overall trend (Figure 3a). ST-CLSTM-Att achieved a high-risk raster hit rate of 90.97% at day 1, maintained 92.30% at day 3, and still reached 83.2% at day 7. In contrast, the Res-YOLOv4 model had a high-risk raster hit rate of 66.64% at day 7, while LR dropped to 33.01% (Figure 3b), which resulted from the ST-SC framework unifying raster samples and strengthening key channels, making high-risk units less likely to be "diluted" by overall smoothing trends.

Practical prediction results under seasonal and spatial contexts

The seasonal prediction results showed that, during the winter phase, the overall risk was relatively low with observed risk indices mostly in the 0.3 - 1.0 range, and the ST-CLSTM-Att prediction trajectory nearly overlapped with the observed curve (Figure 4a). During the transition season, the risk level gradually increased over time with the observed curve reaching a plateau of 1.5 - 2.0 in the middle-to-late segment. ST-CLSTM-Att reached a peak risk index of 1.97, and its upward slope and plateau duration were consistent with the observations (Figure 4b). During summer, when pest infestation was high, the observed risk indices were mainly in the 1.5 - 2.8 range, and the ST-CLSTM-Att risk indicators from days 30 - 90 nearly overlapped with the actual observed trends (Figure 4c). To compare the risk level discrimination ability of different models under different operating condition combinations, this study compiled pest risk statistics under spatial contexts of three height levels and four area types across winter, transition season, and summer. The results

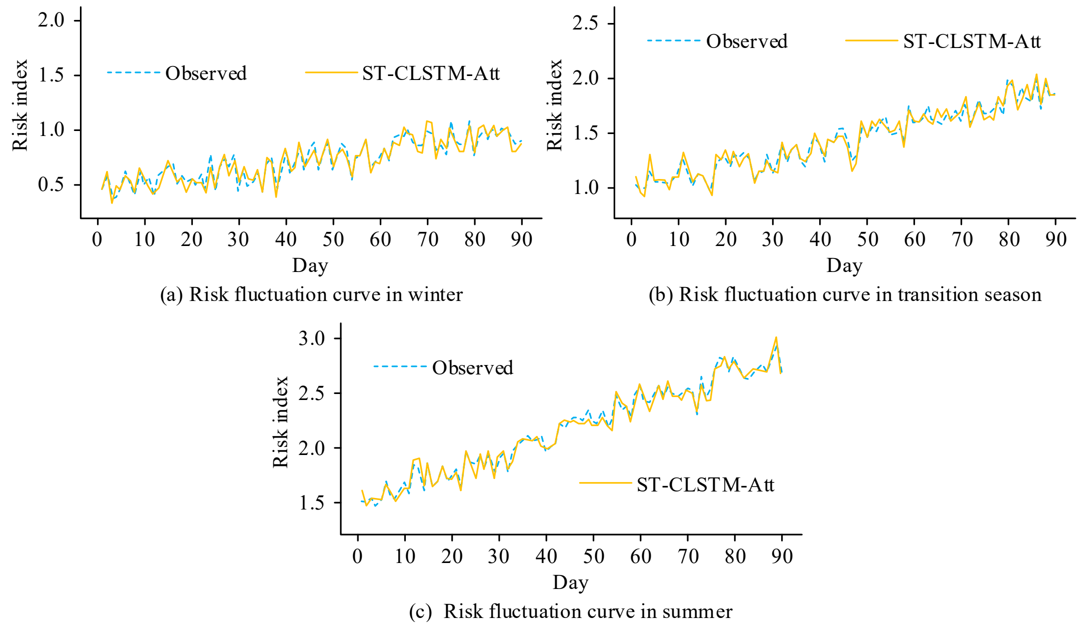

Figure 4. Pest risk evolution trends under seasonal conditions.

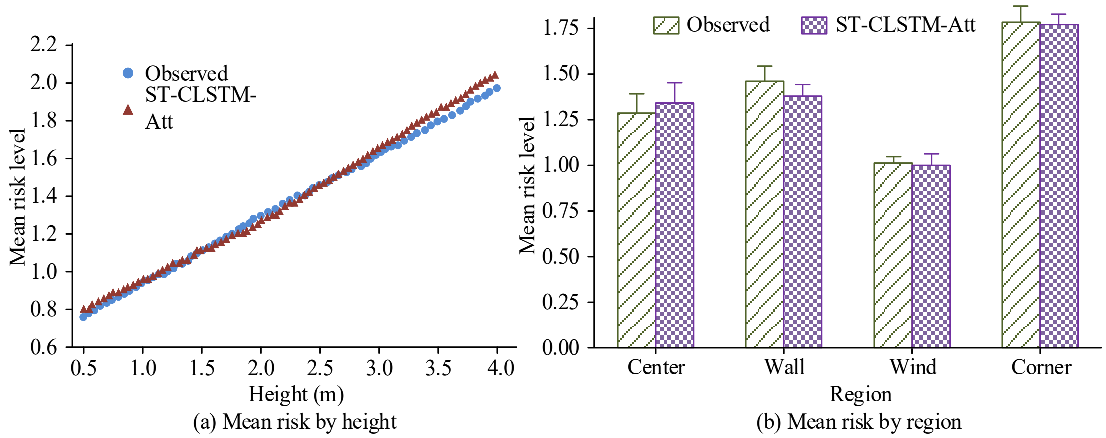

Figure 5. Comparison of model prediction accuracy under spatial heterogeneity.

showed that overall risk level demonstrated an upward trend with season and height, manifesting as summer > transition season > winter. For different heights in the grain warehouse, risk level followed 4.0 m > 2.0 m > 0.5 m. Further, the results indicated clear spatial heterogeneity inside the warehouse, where wall-adjacent areas and corner regions generally showed higher risks than the central region (Table 1). Under most seasonal and spatial contexts, the predicted risk values and high-risk proportions of ST-CLSTM-Att were close to the measured values, which suggested that the proposed model distinguished risk differences caused by seasonal temperature changes, vertical grain-pile stratification, and horizontal spatial heterogeneity. The stability mainly came from the multi-task loss, which jointly constrained risk level classification and risk intensity regression. The height-level and regional risk validation results indicated that the average risk values measured at 0.5 m, 2.0 m, and

Table 1. Pest risk under combinations of different seasons, heights, and warehouse regions.

Season	Height (m)	Region	Observed mean risk	Predicted mean risk	Observed high risk (%)	Predicted high risk (%)
Winter	0.5	Center	0.75	0.77	8.60	4.02
		Wall	0.85	0.91	10.25	4.18
		Wind	0.83	0.77	13.81	20.56
		Corner	0.82	0.77	0.50	0.64
	2.0	Center	0.90	0.86	3.41	6.32
		Wall	1.11	1.16	32.35	38.25
		Wind	0.94	0.99	10.72	11.37
		Corner	0.96	0.93	0.58	4.57
	4.0	Center	1.03	1.01	44.58	44.03
		Wall	1.17	1.21	30.21	34.72
		Wind	1.10	1.09	37.97	37.36
		Corner	1.19	1.22	29.69	31.86
Transition season	0.5	Center	1.10	1.13	39.94	44.70
		Wall	1.25	1.25	30.99	26.24
		Wind	1.16	1.14	22.64	23.45
		Corner	1.25	1.23	19.61	21.55
	2.0	Center	1.24	1.21	55.49	59.63
		Wall	1.44	1.47	48.73	45.09
		Wind	1.37	1.37	33.68	40.45
		Corner	1.38	1.39	64.16	62.85
	4.0	Center	1.45	1.46	47.89	49.62
		Wall	1.59	1.56	30.46	31.61
		Wind	1.43	1.38	36.89	39.42
		Corner	1.54	1.51	66.31	61.22
Summer	0.5	Center	1.41	1.44	36.02	42.28
		Wall	1.53	1.56	61.68	55.77
		Wind	1.49	1.46	46.75	50.95
		Corner	1.57	1.59	53.54	49.73
	2.0	Center	1.56	1.55	58.57	58.80
		Wall	1.67	1.67	65.39	59.33
		Wind	1.67	1.65	75.29	76.93
		Corner	1.67	1.66	68.24	72.15
	4.0	Center	1.73	1.79	91.23	86.41
		Wall	1.90	1.88	84.55	84.49
		Wind	1.73	1.73	57.91	53.48
		Corner	1.79	1.82	57.40	61.96

4.0 m were 0.79, 1.25, and 2.03, respectively, showing an increasing trend with grain pile height, while the corresponding ST-CLSTM-Att prediction values were 0.76, 1.28, and 1.96 with errors below 0.1 at all height levels (Figure 5a). The observed horizontal distribution indicated higher risks in wall adjacent areas and corners, while the central region had a lower risk level. The ST-CLSTM-Att predictions were 1.34, 1.37, 1.00, and 1.77, maintaining the same spatial relationship of higher risk near walls and corners, lower risk in the central region (Figure 5b). These results further confirmed that the proposed model encoded both vertical stratification and horizontal regional differences in low temperature grain storage.

Ablation analysis and robustness evaluation

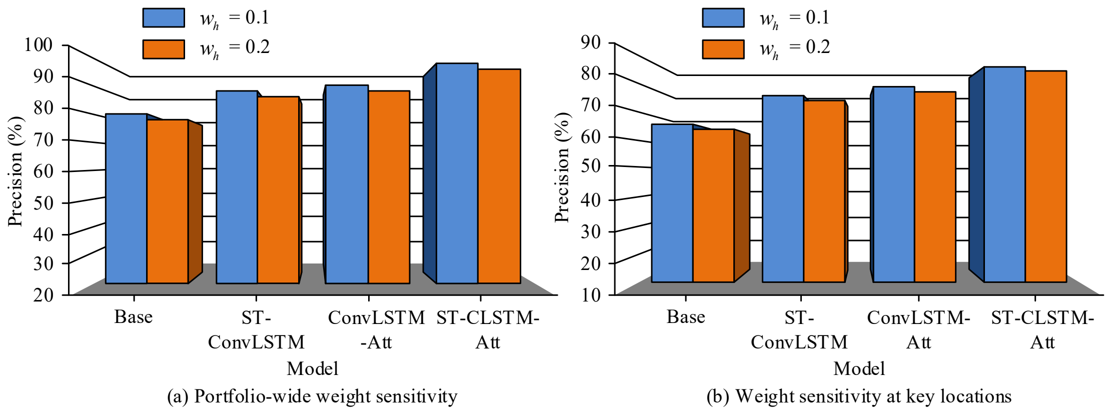

Figure 6. Performance comparison of ablation experiments.

The ablation results of different model variants showed that, when the high-risk class weight was set to 0.1, the high-risk detection precision of ST-CLSTM-Att increased from 79.72% for the baseline ConvLSTM to 97.56%, and the introduction of the attention mechanism produced the most significant gain with this variant improving by 10.22%. When the class weight increased to 0.2, the precision of each model showed only a slight decline with ST-CLSTM-Att still maintaining the highest precision of 95.74% (Figure 6a). For key locations, ST-CLSTM-Att achieved high-risk detection precision of 85.12% when the high-risk class weight was set to 0.1, which was higher than ST-ConvLSTM's 76.36%. When the class weight increased to 0.2, high-risk detection precision of each variant similarly showed a small declining trend with ST-CLSTM-Att still maintaining the highest level at 83.34% (Figure 6b). This decline was consistent with full-scale results, indicating that the proposed architecture had good robustness and ability to prioritize identification of high-risk areas under different spatial scales and weight perturbation conditions.

Conclusion

Although mechanical refrigeration and sealed insulation technology effectively suppressed overall pest populations compared to ambient-temperature ventilation warehouses, this microenvironment also caused pest distributions to exhibit stronger concealment and locality. To identify high-risk pest areas in advance during long-term operations, this study proposed the ST-SC framework and built the ST-CLSTM-Att spatiotemporal prediction model for pest infestation by combining it with ConvLSTM. The results of this study demonstrated that the proposed ST-CLSTM-Att model effectively provided multi-step risk assessments and identifies high-risk areas for pest infestation in low temperature grain storage. Compared with conventional point based or image oriented pest monitoring methods, the proposed framework transformed multi-source monitoring data into gridded spatiotemporal risk representations, which made the prediction results more directly applicable to localized ventilation, cooling, and pest treatment decisions. This was particularly important for low temperature flat warehouses because mechanical refrigeration and sealed insulation could suppress overall pest activity while also making local pest distribution more concealed and spatially heterogeneous. However, the current study had limitations because pest monitoring relied primarily on CO_2 concentrations and trap images. These unidimensional sensing capabilities made it difficult to precisely reconstruct the grain warehouse environment. Future work should expand in directions of multi-warehouse joint modeling, cross-regional transfer learning, and adaptive parameter updating to further improve the foresight and resource allocation efficiency of pest control.

References

1. Wang Q, Hou M, Qin Y, Lian F. 2024. temperature forecasting of grain in storage: An improved approach based on broad learning network. IEEE Access. 12:115112-115123.
2. Sun F, Gong C, Lyu Z. 2024. Grain storage temperature prediction based on chaos and enhanced RBF neural network. Sci Rep - UK. 14:24015.
3. Li X, Wu W, Guo H, Qiao X, Lu Y, Wu Y, *et al*. 2025. An interpretable temperature prediction method for grain in storage based on improved temporal fusion transformers. Comput Electron Agr. 236:110414.
4. Hou Y, Xiong L, Luo X, Han S, Tang X. 2025. Detection of pest infestation in stored grain using an electronic nose system optimized for sensor arrays. J Food Meas Charact. 19(1):439-452.
5. Ta-Phaisach N, Konvipasruang P, Attasopa K, Sringarm K, Chiu CI, Chanbang Y. 2024. Quantitative identification and sensor sensitivity analysis in detection of odor volatile compounds emitted by scaly grain mites, Pest Manag Sci. 80(9):4286-4296.
6. Fürstenau B. 2023. Early detection of stored-product pest insects in grain storage and in the field in Germany. J Kulturpflanz. 75(9-10):259-268.
7. Mendoza QA, Pordesimo L, Neilsen M, Armstrong P, Campbell J, Mendoza PT. 2023. Application of machine learning for insect monitoring in grain facilities. AI. 4(1):348-360.
8. Müller-Blenkle C, Simon U, Meyer R, Szallies I, Lorenz D, Prozell S, *et al*. 2023. A method for acoustic storage pest detection and its challenges. J Kulturpflanz. 75(9-10):235-247.
9. Jamali PV, Nambi E, Saravanan LMS, Chandrasekar V. 2025. Rice-YOLO: An automated insect monitoring in rice storage warehouses with the deep learning model. ACS Agric Sci Technol. 5(2):206-221.
10. Ong SQ, Høye TT. 2025. Trap color strongly affects the ability of deep learning models to recognize insect species in images of sticky traps. Pest Manag Sci. 81(2):654-666.
11. Xia L, Zhang R, Chen L, Li L, Yi T, Chen M. 2024. Monitoring the leaf damage by the rice leafroller with deep learning and ultra-light UAV. Pest Manag Sci. 80(12):6620-6633.
12. Denny DMT, Cerri CEP, Cherubin MR, Burnquist HL. 2025. Nature-based solutions: Sustainable development of Latin America. Glce. 3(4):325-337.
13. Harman RR, Morrison III WR, Ludwick D, Gerken AR. 2024. Predicted range expansion of *Prostephanus truncatus* (*Coleoptera: Bostrichidae*) under projected climate change scenarios. J Econ Entomol. 117(4):1686-1700.
14. Zhao S, Yao H, Lin C, Gao Y, Ding G. 2024. Multi-source-free domain adaptive object detection. Int J Comput Vision. 132(12):5950-5982.
15. Xu L, Cai R, Yu H, Du W, Chen Z, Chen N. 2024. Monthly NDVI prediction using spatial autocorrelation and nonlocal attention networks. IEEE J Stars. 17:3425-3437.
16. Majidizadeh A, Hasani H, Jafari M. 2024. Semantic segmentation of oblique UAV video based on ConvLSTM in complex urban area. Earth Sci Inform. 17(4):3413-3435.
17. Zhao C, Bai C, Yan L, Xiong H, Suthisut D, Pobsuk P, *et al*. 2024. AC-YOLO: Multi-category and high-precision detection model for stored grain pests based on integrated multiple attention mechanisms. Expert Syst Appl. 255:124659.
18. Wang N, Fu S, Rao Q, Zhang G, Ding M. 2025. Insect-YOLO: A new method of crop insect detection. Comput Electron Agr. 232:110085.