



## OPEN ACCESS

## EDITED BY

Andra-Sabina Neculai-Valeanu,  
Rural Development Research Platform  
Association, Romania

## REVIEWED BY

Haopu Li,  
Shanxi Agricultural University, China  
Hassan Seif Mluba,  
The University of Dodoma, Tanzania

## \*CORRESPONDENCE

Wangli Hao  
✉ haowangli@sxau.edu.cn

RECEIVED 28 July 2025

ACCEPTED 04 September 2025

PUBLISHED 02 October 2025

## CITATION

Hao W, Hu X, Xue Y, Shu H and Han M (2025)  
S\_TransNeXtM: a pig behavior recognition  
model based on the TransNeXtM and the  
sLSTM. *Front. Vet. Sci.* 12:1674842.  
doi: 10.3389/fvets.2025.1674842

## COPYRIGHT

© 2025 Hao, Hu, Xue, Shu and Han. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# S\_TransNeXtM: a pig behavior recognition model based on the TransNeXtM and the sLSTM

Wangli Hao<sup>1\*</sup>, Xinyuan Hu<sup>1</sup>, Yakui Xue<sup>1</sup>, Hao Shu<sup>1</sup> and  
Meng Han<sup>2,3</sup>

<sup>1</sup>College of Software, Shanxi Agricultural University, Jinzhong, Shanxi, China, <sup>2</sup>School of Information Science and Engineering, Shanxi Agricultural University, Jinzhong, Shanxi, China, <sup>3</sup>Hangzhou Dianzi University, Hangzhou, Zhejiang, China

Pig behavior recognition serves as a crucial indicator for monitoring health and environmental conditions. However, conventional pig behavior recognition methods are limited in their ability to effectively extract image features and analyze long sequence dependencies, ultimately reducing pig behavior recognition performance. To address these challenges, we propose a pig behavior recognition model S\_TransNeXtM which leverages both spatial and temporal information underlying the video. Specifically, an innovative backbone, named TransNeXtM, has been developed for the spatial domain. It incorporates a bio-inspired Aggregated Attention Mechanism, a Convolutional GLU, and a Mamba unit, which allows the model to capture more discriminative global and local features. For the temporal domain, the sLSTM is proposed to process sequence data by utilizing an exponential gating mechanism and a stabilizer state. This design allows the model to establish longer temporal sequence dependencies, outperforming conventional GRU and LSTM. Based on the above insights, the S\_TransNeXtM enhances the performance of pig behavior recognition. Experimental results demonstrate that the proposed S\_TransNeXtM model achieves the state-of-the-art performance in pig behavior recognition task. Consequently, the S\_TransNeXtM attains an accuracy of 94.53%, marking an improvement of up to 11.32% over previous benchmarks.

## KEYWORDS

S\_TransNeXtM, TransNeXtM, temporal sequence, pig behavior recognition, sLSTM

## Highlights

- Conventional pig behavior recognition methods are limited in their ability to effectively extract image features and analyze long sequence dependencies, ultimately reducing pig behavior recognition performance.
- A S\_TransNeXtM model for precise pig behavior recognition, by jointly optimizing both spatial and temporal domains.
- We first develop a new TransNeXtM module to deal with the spatial features underlying pig behavior videos.
- We ingeniously employ the sLSTM to investigate the temporal features of the corresponding pig behavior videos.

## 1 Introduction

Pig behavior recognition plays an important role in feeding management (1). Through accurate identification of pig behavior, we can understand its physiological and psychological needs, and then targeted adjustment of the feeding environment and feeding formula, improve production efficiency (2).

Therefore, pig behavior recognition is not only a key link in improving the level of feeding management, but also an important means of promoting the sustainable development of animal husbandry.

Traditional pig behavior recognition methods mainly rely on manual observation and sensor technology (3). While manual observation is intuitive, it is subjective, time-consuming, and limits the depth of exploration into pigs' physiological and psychological needs (4, 5). Sensor technology, using speed, infrared, and sound sensors, accurately captures pig movement for precise behavior identification and monitoring. However, sensor layout is complex and prone to detachment, potentially affecting data accuracy and continuity, and posing risks to pig growth and health.

Therefore, with the progress of science and technology, more and more researchers began to explore the use of computer vision, deep learning and other advanced technologies for pig behavior recognition (6, 7). These methods improve the accuracy and efficiency of identification, and provide strong technical support for the optimization of feeding management.

Hengyi et al. (8) pioneeringly integrated the Temporal Shift Module (TSM) into various mainstream 2D convolutional neural network architectures, including ResNet50, ResNeXt50, DenseNet201, and ConvNeXt-t. This innovation significantly enhanced the model's capability to recognize pig aggression behaviors, achieving an impressive accuracy rate of 95.69% in experiments.

Yue et al. (9) focused on behavior analysis within video sequences by introducing a hybrid model that integrates Convolutional Neural Networks (CNNs) with Gated Recurrent Units (GRUs). This model effectively leverages the spatial feature extraction capabilities of CNNs and the temporal sequence processing strengths of GRUs, achieving an accuracy rate of 94.8% in experiments.

Lili et al. (10) delved into pig expression recognition, integrating ASPP and CReToNeXt modules into the ASP-YOLOv5 model. This innovation optimized feature extraction and fusion processes, resulting in a mean Average Precision (mAP) of 93.2%.

Junjie et al. (11) focused on classifying interactive behaviors among pigs by designing a deep learning framework that fused Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM). Under fixed training set sizes, the framework was validated through multiple methods, including random validation, temporal blocking validation, and feeder blocking validation, achieving an average accuracy of 96.8%.

Ma et al. (12) achieved remarkable results in model optimization, proposing the optimized M-YOLOv4-C network model. This model adopted the lightweight MobileNet-v3 as its core architecture and incorporated depthwise separable convolution into YOLOv4's feature extraction network. This innovation not only boosted the model's accuracy to 98.15% but also achieved a detection speed of 106.3 frames per second while keeping the model size at 44.74 MB, enabling real-time applications for pig behavior recognition.

Zhang et al. (13) proposed a Transformer-based Neural Network (TNN) model that leveraged attention heatmap visualization techniques to precisely pinpoint and analyze critical image regions, demonstrating exceptional performance in piglet behavior recognition. Notably, even with reduced parameter

counts and computational complexity, the TNN model maintained outstanding recognition efficacy.

Despite some progress in deep learning for pig behavior recognition, challenges remain in accurately extracting image features and simultaneously analyzing long temporal sequence data.

The key contributions of this paper are 3 fold:

- This paper introduces a S\_TransNeXtM model for precise pig behavior recognition, by jointly optimizing both spatial and temporal domains. Specifically, it leverages TransNeXtM to capture more abundant global and local discriminative features in the spatial domain, facilitating the discrimination of subtle pig behavioral differences. Additionally, sLSTM (spatial Long Short-Term Memory) is creatively employed in the temporal domain to handle long sequence dependencies.
- We first develop a new TransNeXtM module to deal with the spatial features underlying pig behavior videos. Concretely, the module incorporates the Aggregated Attention Mechanism and Convolutional GLU (Convolutional Gate-controlled Linear Unit) for capturing global and local features, respectively. To further enhance its discriminative capability, the Mamba unit is innovatively introduced into the module. This design promotes the filtering of noise and redundant information that may interfere with discrimination. Consequently, it enable the model to capture subtle differences in pig behavior, thereby significantly improves the performance of the model.
- In this paper, we ingeniously employ the sLSTM to investigate the temporal features of the corresponding pig behavior videos. This module can effectively capture the long temporal dependencies through exponential gating and stabilized states. Thus, a notable enhancement of the model's performance can be obtained.

## 2 Materials and methods

### 2.1 Dataset

The pig behavior recognition data is collected from the pig breeding base of Nonglv Yuan Agriculture Co., LTD., Xiangfen County, Linfen City, Shanxi Province. Specifically, the data collection range from August 12, 2022 to September 25, 2022. The farm comprises six pig houses, each housing 6-month-old ternary breed pigs, with an average of 10 pigs per house. Concretely, cameras have been installed in all six pig houses, employing the Hikvision DS-2DE3Q120MY-T/GLSE. These cameras are mounted at a 45-degree angle to the side of each pig house, 3 meters above the ground. They are capable of capturing RGB color space video with a resolution of 1,920\*1,080 pixels and a sampling rate of 25 frames per second. A visual depiction of the camera's view inside one of the pig houses is presented in the [Figure 1](#).

Finally, after data collection, we obtain 1.5 TB of pig video data, which covers more than 5,000 video files. Subsequently, video clips with durations ranging from 5 to 10 s are selected. After this video preprocessing, we identify six pig behavior categories: Drinking, Eating, Fighting, Exploring, Lying, and Walking, as



FIGURE 1  
View of the camera inside some pig houses.

shown in Figure 2. Specifically, Drinking was characterized by oral suction at water sources. Eating by head-in-trough mastication. Fighting by aggressive pushing or biting for resource competition. Exploring constitutes sustained active contact exceeding 3 s with environmental elements (14). Lying by resting postures with extended or curled limbs. Walking by limb-alternation displacement for spatial movement (15).

Each category comprised approximately 450 videos, with the total dataset containing 2,755 videos. Each video consists of 24 image frames, which are arranged in temporal order to form a complete video segment. The inter-frame difference method was used to detect blurry frames, and a total of 3.7% of low-quality segments in the original data were eliminated. Within each category, the training and test sets were partitioned in an 8:2 ratio. Further, the detailed statistics of the number of each pig behavior are presented in Table 1.

## 2.2 S\_TransNeXtM

In order to improve the performance of pig behavior recognition, this paper proposes a novel model, S\_TransNeXtM, which leverages both spatial and temporal information. Specifically, the S\_TransNeXtM model which comprised of two modules: the TransNeXtM and the sLSTM(spatial Long Short-Term Memory) respectively, as shown in Figure 3. Following, we will elaborate on these two modules in detail.

### 2.2.1 TransNeXtM module

This subsection introduces a novel pig behavior recognition backbone module, named TransNeXtM, aiming to capture spatial information in the corresponding task.

As shown in Figure 4, the TransNeXtM mainly contains a four-stage hierarchical architecture (16). Specifically, the first three stages utilize the MAACG Block, which incorporates a Mamba (17) unit, an Aggregated Attention Mechanism, and a Convolutional GLU (Convolutional Gate-controlled Linear Unit) (18), respectively. Additionally, the MMHSACG Block is employed in the final stage, which includes a Mamba unit, a Multi-Head Self-Attention Mechanism (19), and a Convolutional GLU. Notably, in the fourth stage, the small size of the feature maps makes the application of Multi-Head Self-Attention mechanism particularly suitable. This mechanism can effectively capture more abundant features without significantly increasing the computation time.

#### 2.2.1.1 Mamba

In order to enhance the discrimination of the model in pig behavior recognition task, we introduce the Mamba (17) unit into the MAACG and MMHSACG Blocks of the TransNeXtM module, as shown in Figure 4, aiming to filter out irrelevant and redundant information. The Mamba unit we utilized can be formulated as follows:

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t \quad (1)$$

$$\bar{A} = \exp(\Delta A) \quad (2)$$

$$\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B \quad (3)$$

where  $h_t$  and  $h_{t-1}$  represent the state at the current and previous time steps, respectively. The parameters ( $\bar{A}$ ,  $\bar{B}$ ) are discrete parameters transformed by ( $\Delta$ ,  $A$ ,  $B$ ), as shown in Equations 2, 3. Specifically,  $\Delta$  denotes sampling time-scale,  $A$  indicates the state transition matrix,

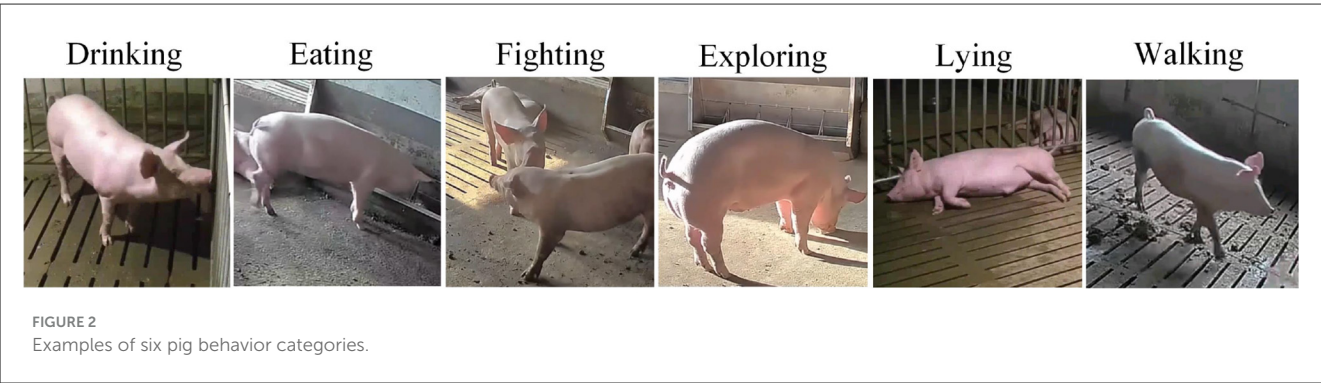
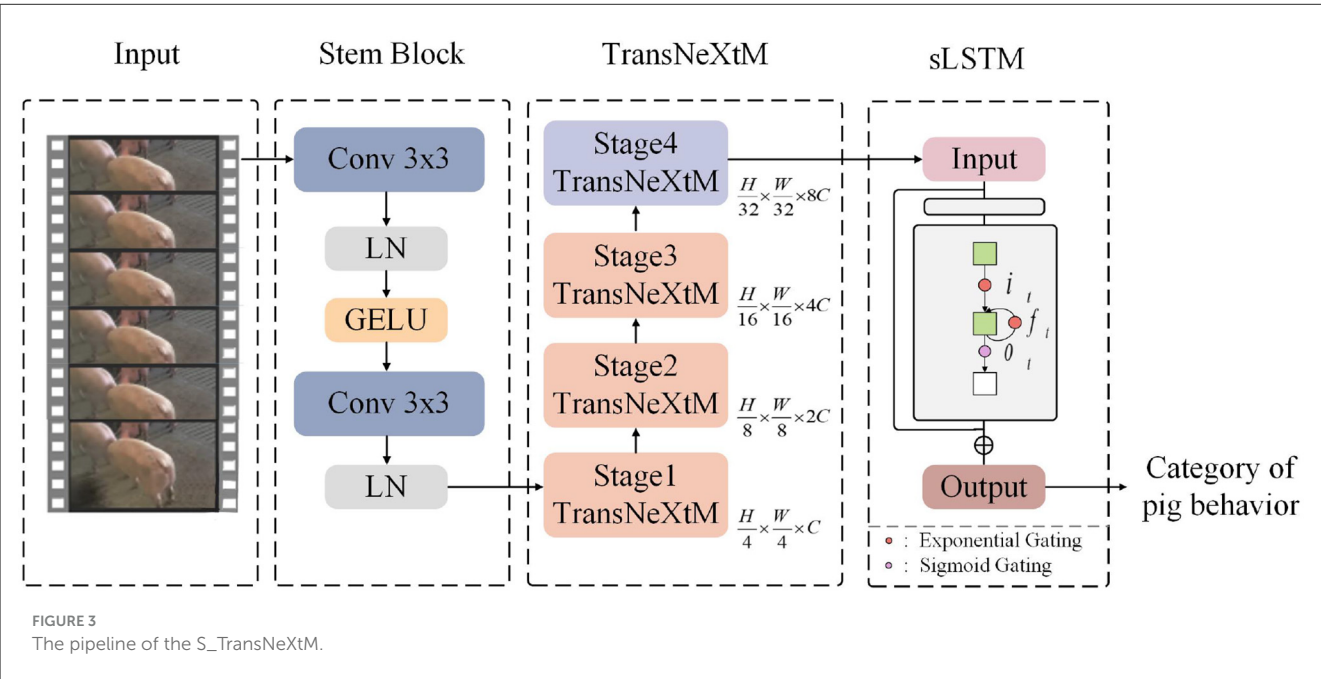


TABLE 1 Statistical of the number of videos for different categories in the dataset.

| Behavior category | Drinking | Eating | Fighting | Exploring | Lying | Walking |
|-------------------|----------|--------|----------|-----------|-------|---------|
| Number            | 426      | 489    | 438      | 485       | 474   | 443     |
| Training          | 341      | 391    | 350      | 388       | 379   | 354     |
| Test              | 85       | 98     | 88       | 97        | 95    | 89      |



$B$  is the transition matrix for input features, and  $I$  stands for the identity matrix.

$$y_t = Ch_t \tag{4}$$

where  $y_t$  represents the output feature, and the matrix  $C$  determines how the  $h_t$  is transformed into the  $y_t$ .

2.2.1.2 Aggregated Attention

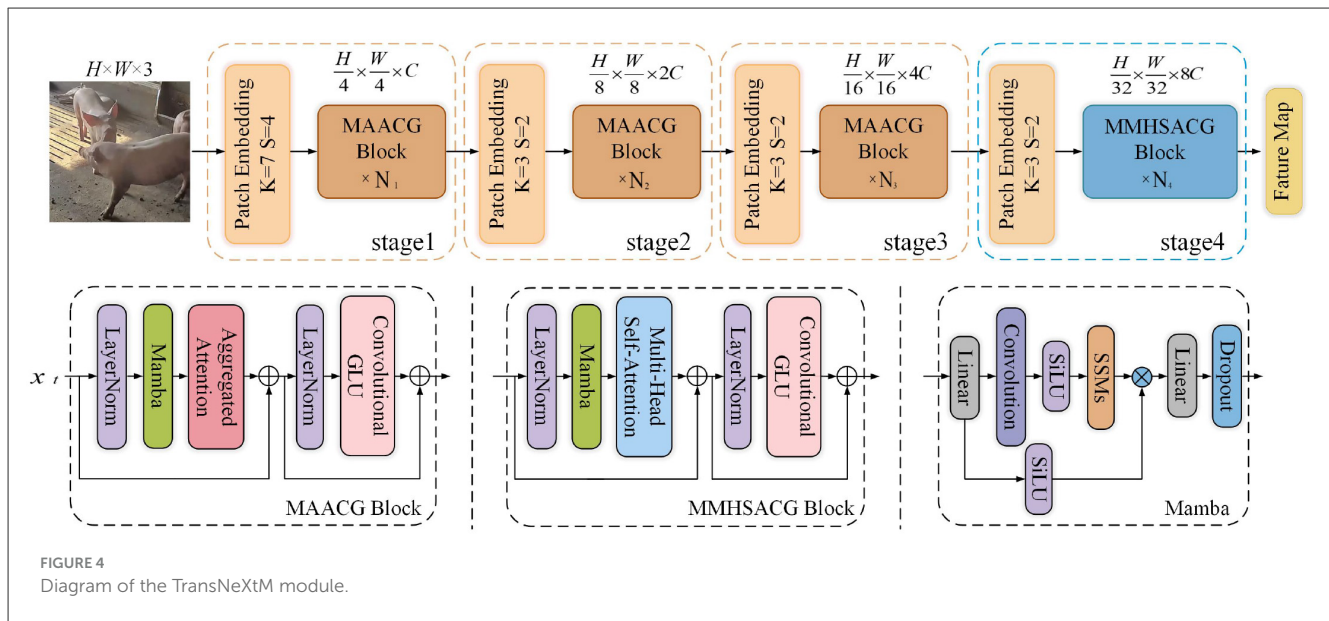
This mechanism aims to combine local sliding window attention and global pooling attention through a dual-path design (20). It simulates the flexibility of eye movements and the capacity to process multi-scale information. Consequently, it

enables the model to obtain more comprehensive features. The formulation of this mechanism is presented as follows:

$$F_{AA}(X_{(i,j)}) = (Z_{(i,j) \sim \rho(i,j)} + \widehat{Q}_{(i,j)}T)V_{\rho(i,j)} + Z_{(i,j) \sim \sigma(X)}V_{\sigma(X)} \tag{5}$$

where  $(i,j)$  indicates the coordinates of a center pixel,  $\rho(i,j)$  defines a set of pixels within a sliding window centered at  $(i,j)$ , this representation specifically focuses on local region. On the other hand,  $\sigma(X)$  represents a feature set obtained from a whole pooled feature map, which covers the information of the global region.  $Z_{(i,j) \sim \rho(i,j)}$  and  $Z_{(i,j) \sim \sigma(X)}$  respectively represent attention weight matrices computed based on the sliding window and the pooled features.  $\widehat{Q}$  is the query matrix,  $T$  denotes a learnable token, and  $V$  stand for the value matrix.





$$Z_{(i,j) \sim \rho(i,j)}, Z_{(i,j) \sim \sigma(X)} = \text{Split}(Z_{(i,j)}) \text{ with size}[k^2, H_p W_p] \quad (6)$$

where  $k \times k$  represents the fixed window size, and  $H_p \times W_p$  indicates the pooling size.

$$Z_{(i,j)} = \text{softmax}(\tau \log N * \text{Concat}(S_{(i,j) \sim \rho(i,j)}, S_{(i,j) \sim \sigma(X)} + B_{(i,j)})) \quad (7)$$

where  $\tau$  is a learnable variable initialized to  $\frac{1}{0.24}$ .  $N$  denotes the count of effective keys each query interacts with, and  $\text{Concat}$  stands for concatenation (21).  $S_{(i,j) \sim \rho(i,j)}$  and  $S_{(i,j) \sim \sigma(X)}$  reflect the relevance of local and global features, respectively. Additionally,  $B$  represents the bias.

$$S_{(i,j) \sim \rho(i,j)} = (\hat{Q}_{(i,j)} + Q_E) \hat{K}_{\rho(i,j)}^T \quad (8)$$

$$S_{(i,j) \sim \sigma(X)} = (\hat{Q}_{(i,j)} + Q_E) \hat{K}_{\sigma(X)}^T \quad (9)$$

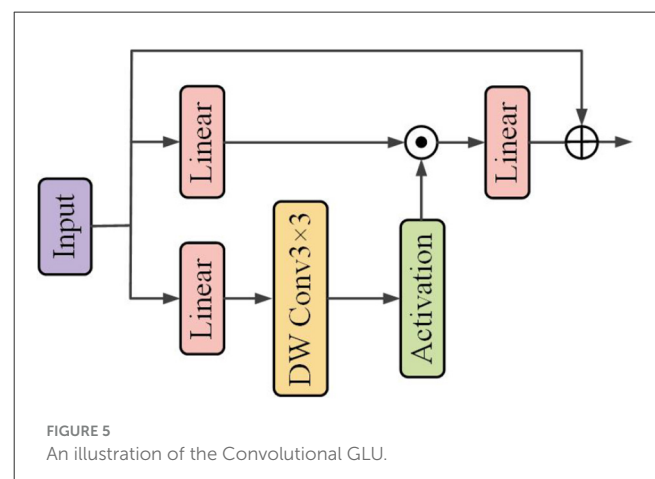
where  $\hat{Q}$  and  $\hat{K}$  stand for the query matrix and the key matrix, respectively.  $Q_E$  denotes the learnable parameter matrix.

### 2.2.1.3 Multi-Head Self-Attention

The Multi-Head Self-Attention mechanism maps the input sequence into multiple distinct representation heads (19). Each head then independently performs self-attention calculations on the input sequence. Consequently, this approach allows the mechanism to capture diverse levels of information within the input, thereby enhances the model's feature extraction capabilities.

The Multi-Head Self-Attention is calculated by the following equations.

$$\text{head}_u = \text{Attention}(QW_u^Q, KW_u^K, VW_u^V) \quad (10)$$



$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (11)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_u) W^O \quad (12)$$

where  $W_u^Q$ ,  $W_u^K$ ,  $W_u^V$ , and  $W^O$  denote the learnable weight matrix,  $d_k$  indicates the dimension of the key vector. Additionally,  $u$  is the number of heads.

### 2.2.1.4 Convolutional GLU

The Convolutional GLU integrates the Gated Linear Unit (GLU) with Depthwise Convolution (DW Conv), as shown in Figure 5. It enables each token to perform channel attention computation based on the image features of its nearest neighbors, thereby enhancing the model's capability in capturing local features.

## 2.2.2 sLSTM module

The temporal sequence module selected in this paper is sLSTM (22), as extended variant of LSTM (23). Its core improvement lies in the introduction of an exponential gated activation function (24) and a normalized state, enabling the sLSTM model to excel in capturing long temporal sequence dependencies. The specific formula is as follows:

$$c_t = f_t c_{t-1} + i_t z_t \quad (13)$$

where  $c_t$  represents the state held by the cell at time step  $t$ ,  $f_t$  and  $i_t$  correspond to the forget gate and the input gate, respectively.

$$z_t = \varphi(\tilde{z}_t), \tilde{z}_t = w_z^T x_t + r_z h_{t-1} + b_z \quad (14)$$

where  $z_t$  denotes the unit input, which is a candidate unit state calculated based on the current input and the previous state. Here,  $\varphi$  represents the activation function of the unit input gate,  $\tilde{z}_t$  is an intermediate variable,  $w$  denotes the weight matrix,  $r$  indicates a learnable parameter, and  $b$  represents the bias.

$$n_t = f_t n_{t-1} + i_t \quad (15)$$

where  $n_t$  denotes the normalized state, which represents equivalent to adding a large denominator to prevent overflow, as the exponential activation function may produce excessively large values.

$$h_t = o_t \tilde{h}_t, \tilde{h}_t = c_t / n_t \quad (16)$$

where  $h_t$  represents the hidden state and  $o_t$  denotes the output gate, while  $\tilde{h}_t$  indicates an intermediate variable. The  $n_t$  is employed to prevent overflow issues.

Subsequently, we proceed to introduce the output gates, input gates, and forget gates.

$$o_t = \sigma(\tilde{o}_t), \tilde{o}_t = w_o^T x_t + r_o h_{t-1} + b_o \quad (17)$$

where  $\tilde{o}_t$  indicates an intermediate variable.

$$\begin{aligned} i_t &= \exp(\log(i'_t) - m_t), \\ i'_t &= \exp(\tilde{i}'_t), \tilde{i}'_t = w_i^T x_t + r_i h_{t-1} + b_i \end{aligned} \quad (18)$$

where  $i_t$  denotes the stabilized input gate adjusted by  $m_t$ .  $\exp$  indicates the exponential function. Since the exponential function grows faster than the *sigmoid* function and is more sensitive to input changes, so that the model can capture the changes in input information faster.  $\tilde{i}'_t$  represents an intermediate variable.

$$\begin{aligned} f_t &= \exp(\log(f'_t) + m_{t-1} - m_t), \\ f'_t &= \sigma(\tilde{f}'_t), \tilde{f}'_t = w_f^T x_t + r_f h_{t-1} + b_f \end{aligned} \quad (19)$$

where  $f_t$  indicates the stabilized forget gate, it is also adjusted according to the value of  $m_t$ .  $\sigma$  represents the gated activation function *sigmoid*, and  $\tilde{f}'_t$  denotes an intermediate variable.

$$m_t = \max(\log(f'_t) + m_{t-1}, \log(i'_t)) \quad (20)$$

where  $m_t$  denotes the stabilizer state.  $\log$  indicates the exponential inverse operation, which is equivalent to using  $\log$  to degrade  $i'_t$  and  $f'_t$  to avoid overflow.

## 2.3 The loss function

To effectively train the proposed model S\_TransNeXtM, the common loss function for recognition tasks are employed. The detailed definition of the loss can be presented as follows:

$$Loss = - \sum_{n=1}^N \sum_{m=1}^M p_n^m \log(\hat{p}_n^m) \quad (21)$$

where  $N$  and  $M$  represent the number of samples and categories, respectively.  $p_n^m$  indicates the true label for the respective sample, and  $\hat{p}_n^m$  demonstrates the prediction label for the corresponding sample.

## 3 Experiments and analysis

A series of comprehensive experiments were conducted to evaluate the proposed model from multiple perspectives, with detailed designs and documentation systematically presented. All experimental procedures were executed on a hardware system configured with 125 GB RAM, an Intel i7-7800X CPU operating at 3.50 GHz, and an NVIDIA TITAN Xp GPU equipped with 12GB GDDR5X memory. The Ubuntu 20.04.6 LTS operating system was employed as the software foundation, while Python 3.12.4 served as the primary programming environment. Key libraries included PyTorch 2.1.2, Transformers 4.35.0, NumPy 2.0.1, and Pandas 2.2.2. Hyperparameters were systematically set as follows: batch size was configured to 4, initial learning rate was established at 0.001, and training was conducted over 200 epochs using the Adam optimizer.

In this experiment, all images were uniformly resized to 224 × 224 pixels. The input is [4, 24, 3, 224, 224], which represents batch\_size, num\_frames, channels, height and width respectively. The output is [4, 6], which represents batch\_size and num\_class, corresponding to the classification results of the pig behaviors.

### 3.1 Evaluate the effectiveness of the Mamba unit

To verify the performance of the Mamba (17) unit in S\_TransNeXtM, comparisons are conducted between models that with and without the Mamba unit. Specifically, several basic models such as Swin Transformer, ConvNeXt, and TransNeXt are utilized. Comparison results are presented in Table 2, where models with the

TABLE 2 Comparison of different models with and without Mamba.

| Model                    | Accuracy (%) | Loss          |
|--------------------------|--------------|---------------|
| Swin Transformer         | 90.51        | 0.4902        |
| <b>SwinM Transformer</b> | <b>92.00</b> | <b>0.4690</b> |
| ConvNeXt                 | 91.20        | 0.4880        |
| <b>ConvNeXtM</b>         | <b>92.35</b> | <b>0.4564</b> |
| TransNeXt                | 92.88        | 0.4777        |
| <b>TransNeXtM</b>        | <b>93.25</b> | <b>0.4521</b> |

The bold values indicate that the performance of this model is the best.

TABLE 3 Comparison of different models with and without the Aggregate Attention Mechanism.

| Model            | Accuracy (%) | Loss          |
|------------------|--------------|---------------|
| wo_AggAttn       | 93.04        | 0.4665        |
| <b>w_AggAttn</b> | <b>93.20</b> | <b>0.4521</b> |

The bold values indicate that the performance of this model is the best.

Mamba unit are written as SwinM Transformer, ConvNeXtM, and TransNeXtM, respectively.

Table 2 demonstrates that models with Mamba unit outperform their counterparts without it. Specifically, the SwinM Transformer achieves 92.00% accuracy, making a 1.49% improvement compared to the Swin Transformer. The accuracy of the ConvNeXtM is 92.35%, surpassing the ConvNeXt by 1.15%. The TransNeXtM demonstrates the highest accuracy of 93.25%, a 0.37% increase compared to the TransNeXt. Furthermore, the loss values of the SwinM Transformer, the ConvNeXtM, and the TransNeXtM are 0.4690, 0.4564, and 0.4521, respectively, representing reductions of 0.0212, 0.0316, and 0.0259 compared to their non-Mamba counterparts.

Models with Mamba unit achieve superior accuracy in pig behavior recognition. This is attributed to Mamba’s ability to filter noise and redundant information. Consequently, the model with Mamba unit is able to capture more discriminative features, leading to superior performance in the pig behavior recognition task.

### 3.2 Evaluate the effectiveness of the Aggregate Attention Mechanism

In order to assess the performance of the Aggregate Attention Mechanism in S\_TransNeXtM, a comparative analysis is carried out between models with and without this mechanism. Comparison results are demonstrated in Table 3, where the model with the Aggregate Attention Mechanism is denoted as w\_AggAttn(TransNeXtM) and the model without it is written as wo\_AggAttn, respectively.

The results in Table 3 indicates that the w\_AggAttn model demonstrates superior performance compared to the wo\_AggAttn model. Specifically, the accuracy of the TransNeXtM(w\_AggAttn), which employed the Aggregate Attention Mechanism, reached 93.20%, marking a 0.16% improvement over the wo\_AggAttn model. Furthermore, the loss value of the TransNeXtM is 0.4521, a reduction of 0.0144 compared to the wo\_AggAttn model.

TABLE 4 Comparison of different models with and without the Convolutional GLU.

| Model                      | Accuracy (%) | Loss          |
|----------------------------|--------------|---------------|
| wo_Convolutional GLU       | 92.29        | 0.4760        |
| <b>w_Convolutional GLU</b> | <b>93.20</b> | <b>0.4521</b> |

The bold values indicate that the performance of this model is the best.

The model utilizing the Aggregate Attention Mechanism demonstrates superior performance. This mechanism mimics the biological visual system, effectively processes information across various scales, and enhances the model’s global perception capabilities. Consequently, it enhances the model’s performance in pig behavior recognition task.

### 3.3 Evaluate the effectiveness of the Convolutional GLU

This subsection focuses on validating the impact of Convolutional GLU in the S\_TransNeXtM for pig behavior recognition. Specifically, we employ models both with and without the Convolutional GLU for this validation. The comparison results are presented in Table 4, where the model without the Convolutional GLU is written as wo\_Convolutional GLU, and the model with the Convolutional GLU is denoted as w\_Convolutional GLU(TransNeXtM). Notably, the wo\_Convolutional GLU model utilizes the original MLP module.

Table 4 shows that the w\_Convolutional GLU model performs better than the wo\_Convolutional GLU model. Specifically, the accuracy of the TransNeXtM is 93.20%, which is 0.91% higher than that of the wo\_Convolutional GLU model. The loss of the TransNeXtM is 0.4521, decreasing by 2.39% compared to the wo\_Convolutional GLU model.

In Figure 6a, the TransNeXtM exhibits superior accuracy compared to the model without the Convolutional GLU. Figure 6b illustrates that the loss of the TransNeXtM is lower. In conclusion, the TransNeXtM demonstrates improved performance in the task of pig behavior recognition.

The model with the Convolutional GLU outperforms the one that without, because the Convolutional GLU integrates the GLU with DW Conv. This integration enables each token to perform channel attention computation based on the image features of its nearest neighbors. Consequently, this mechanism enhances the model’s ability to capture local features, thereby improving overall performance in pig behavior recognition task.

### 3.4 Evaluate the effectiveness of temporal sequence module

In this section, we design two experiments to verify the effectiveness of the temporal sequence module. The first experiment is aimed at evaluating the effectiveness of different temporal sequence modules. The second experiment focuses on comparing models with and without the temporal sequence module.

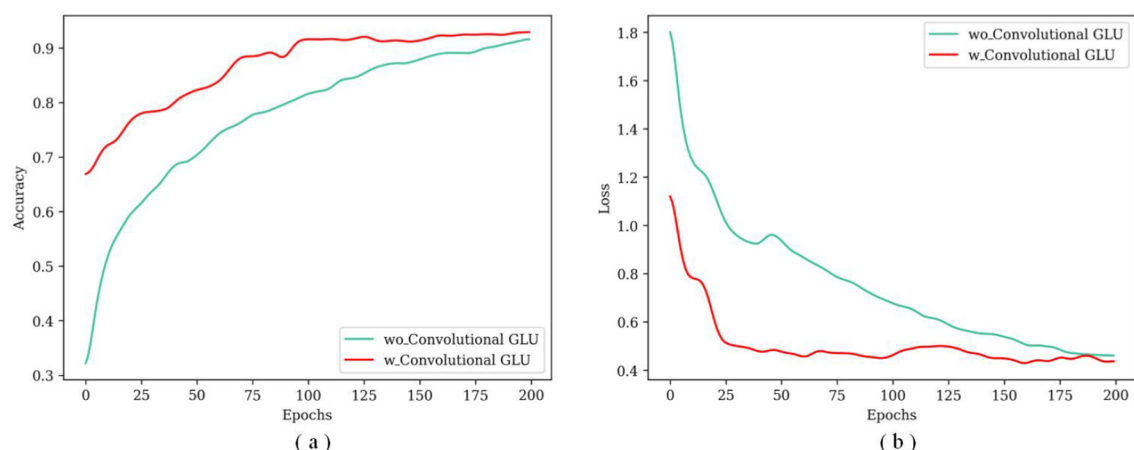


FIGURE 6

The accuracy and loss curves under different epochs for models with or without the Convolutional GLU. (a) Is for accuracy and (b) is loss.

TABLE 5 Comparison results of different temporal sequence modules.

| Model               | Accuracy (%) | Loss          |
|---------------------|--------------|---------------|
| TransNeXtM          | 93.25        | 0.4521        |
| TransNeXtM_GRU      | 93.42        | 0.5981        |
| TransNeXtM_LSTM     | 93.38        | 0.6579        |
| <b>S_TransNeXtM</b> | <b>94.53</b> | <b>0.3350</b> |

The bold values indicate that the performance of this model is the best.

### 3.4.1 Evaluate the effectiveness of different temporal sequence modules

To evaluate the effectiveness of different temporal sequence module, several models with diverse temporal module (GRU (18), LSTM (23) and sLSTM (22)) are utilized for comparative analysis. Specifically, these models are denoted as TransNeXtM\_GRU, TransNeXtM\_LSTM and S\_TransNeXtM(TransNeXtM\_sLSTM). The results are shown in Table 5.

Table 5 illustrates that the S\_TransNeXtM outperforms all other models in pig behavior recognition task. Specifically, the S\_TransNeXtM model reaches 94.53% accuracy, which is 1.11% and 1.55% higher than those of the TransNeXtM\_GRU and the TransNeXtM\_LSTM. Furthermore, its loss value is 0.3350, which is lower than those of the TransNeXtM\_GRU and the TransNeXtM\_LSTM.

To further validate the effectiveness of different temporal sequence modules, Figure 7 reports the accuracy and loss values of these models under different training epochs.

Figure 7a demonstrates that the accuracy of the S\_TransNeXtM exceeds those of both the TransNeXtM\_GRU and the TransNeXtM\_LSTM. Additionally, Figure 7b shows that the loss of the S\_TransNeXtM model during training epochs is the lowest among the three models. This further confirms the effectiveness of the sLSTM temporal sequence module.

From now on, we will default to considering the temporal sequence module mentioned as the most effective sLSTM.

### 3.4.2 Evaluate the effectiveness when with or without the temporal sequence module

This subsection focuses on validating the impact of temporal sequence module for pig behavior recognition. Concretely, several models including the SwinM Transformer, the ConvNeXtM, and the TransNeXtM, with or without temporal sequence module, are employed for this validation. The comparison results are detailed in Table 6.

Table 6 indicates that those models with temporal sequence module perform better than those without. Specifically, the accuracy of the S\_SwinM Transformer is 92.70%, exceeding the SwinM Transformer by 0.7%. The S\_ConvNeXtM achieves 93.70% accuracy, outperforming 1.35% compared to the ConvNeXtM. The S\_TransNeXtM achieves 94.53% recognition accuracy, representing an increase of 1.28% compared to the TransNeXtM. Furthermore, the S\_SwinM Transformer demonstrates a loss value of 0.4382, which is 0.0308 lower than that of the SwinM Transformer. The loss value of the S\_ConvNeXtM is 0.3900, a reduction of 0.0664 compared to the ConvNeXtM. The S\_TransNeXtM's loss value is 0.4521, representing a decrease of 0.0256 compared to the TransNeXtM.

The reason why S\_TransNeXtM achieves superior accuracy in analyzing pig behavior recognition is because the sLSTM with an exponential gated activation function and normalized state, which enables the model to effectively capture longer temporal sequence dependencies underlying the video. In contrast, although the GRU and the LSTM are also tools for processing temporal sequence data, they fail to reach the level of the sLSTM in terms of the depth of utilizing temporal dependency relationships, thereby affecting overall performance. Meanwhile, this unique design of sLSTM enhances the accuracy of the model's data processing capabilities, making models with the sLSTM achieve superior accuracy in pig behavior recognition.

In summary, for the task of pig behavior recognition, models with the sLSTM outperforms other models with different temporal sequence modules in terms of performance. Additionally, models with temporal sequence modules tend to exhibit better performance than without.



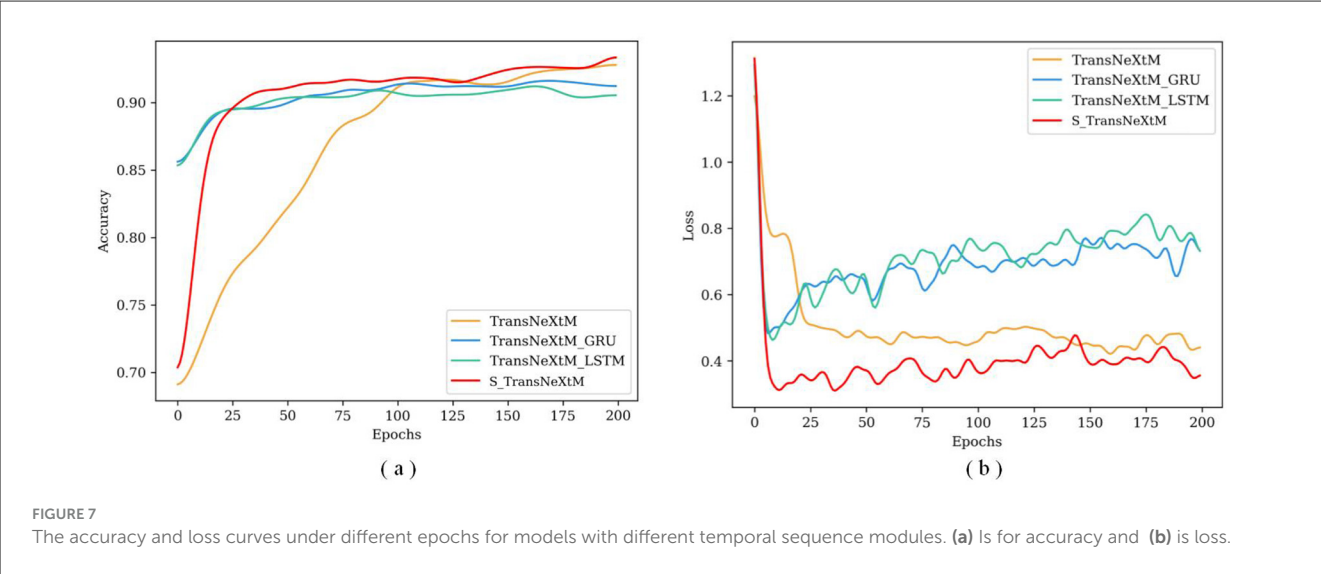


TABLE 6 Comparison results of models with or without temporal sequence module.

| Model                      | Accuracy (%) | Loss          |
|----------------------------|--------------|---------------|
| SwinM Transformer          | 92.00        | 0.4690        |
| <b>S_SwinM Transformer</b> | <b>92.70</b> | <b>0.4382</b> |
| ConvNeXtM                  | 92.35        | 0.4564        |
| <b>S_ConvNeXtM</b>         | <b>93.70</b> | <b>0.3900</b> |
| TransNeXtM                 | 93.25        | 0.4521        |
| <b>S TransNeXtM</b>        | <b>94.53</b> | <b>0.3350</b> |

The bold values indicate that the performance of this model is the best.

3.5 Evaluate the effectiveness of different models

To verify the effectiveness of our proposed model, S\_TransNeXtM, several popular models, including ViT (Vision Transformer) (25), Swin Transformer (26), ConvNeXt (27), TransNeXt (20), EfficientViT (28), TransNeXtM and CNN-LSTM (11) are employed for comparison. The comparison results are shown in Table 7.

Table 7 demonstrates that our proposed S\_TransNeXtM model achieves the best performance across all evaluation criteria. Specifically, the S\_TransNeXtM achieves a top accuracy of 94.53%, surpassing ViT (83.21%), Swin Transformer (90.51%), ConvNeXt (92.52%), TransNeXt (92.88%), EfficientViT (90.51%), CNN-LSTM (90.00%) and TransNeXtM (93.20%) by margins of 11.32%, 4.02%, 2.01%, 1.65%, 4.02%, 4.53% and 1.33% respectively. Furthermore, the S\_TransNeXtM attains the lowest loss value of 0.335, showing 61.04%, 31.66%, 13.09%, 29.87%, 27.43%, 38.43% and 25.9% relative reductions compared to ViT (0.8597), Swin Transformer (0.4902), ConvNeXt (0.3991), TransNeXt (0.4777), EfficientViT (0.4616), CNN-LSTM (0.4291) and TransNeXtM (0.4521). In terms of Precision, Recall, and F1-Score, the S\_TransNeXtM achieved 94.52%, 94.54%, and 94.51%, respectively, which were 1.36% to

10.37%, 1.43% to 11.1% and 1.41% to 10.61% higher than the other models.

To further validate the effectiveness of the S\_TransNeXtM, Figure 8 shows the accuracy and loss curves of the comparison models under different epochs.

In Figure 8a, the accuracy of the S\_TransNeXtM surpasses that of the ViT, Swin Transformer, ConvNeXt, TransNeXt and EfficientViT models. Furthermore, in Figure 8b, the S\_TransNeXtM model achieves the lowest loss. These results further validate the effectiveness of the S\_TransNeXtM.

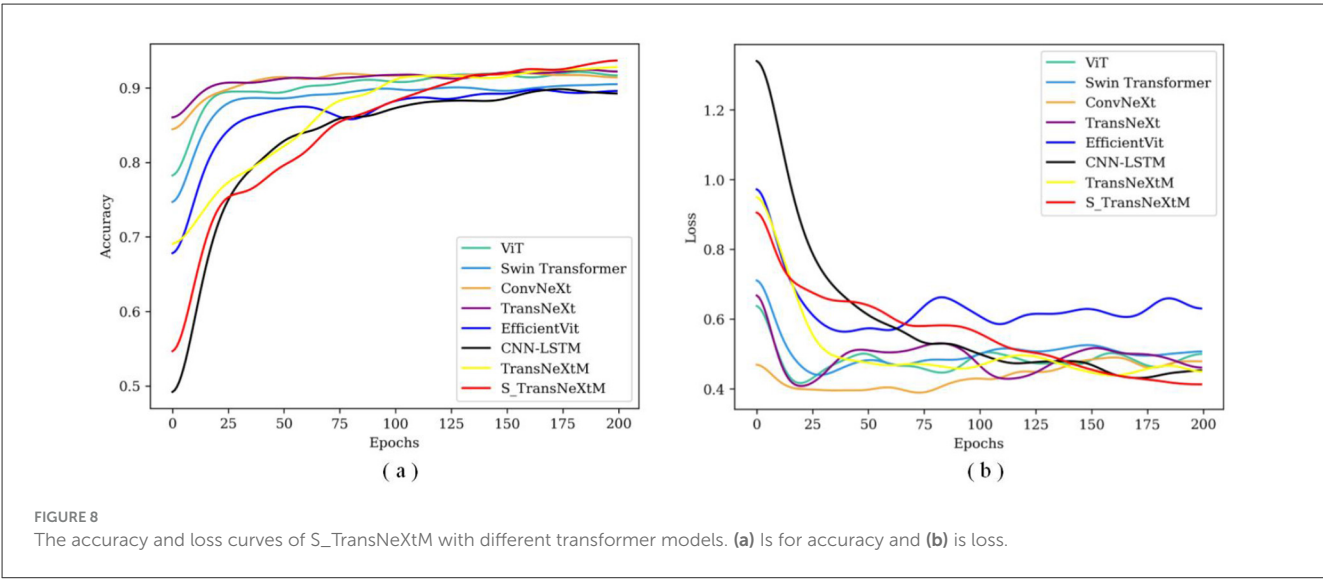
Table 8 demonstrates the superior performance of the S\_TransNeXtM model in six behavioral recognition tasks for pigs. Specifically, this model achieved the highest accuracy in recognizing the three behaviors of “Eating,” “Lying,” and “Walking.” For the recognition of “Drinking,” “Fighting,” and “Exploring” behavior, although the result was slightly inferior to the other models, it still reached a suboptimal level. The accuracy gap in these three categories arises from inherent challenges. Specifically, “Drinking” involves subtle spatiotemporal patterns with minimal head movement and short duration, leading to weaker feature distinguishability. “Fighting” recognition accuracy is constrained by limited data quality, as video blur weakens the model’s ability to capture key motion features of pigs. “Exploring” is difficult to distinguish from “Walking” due to their similar characteristics. Despite these category-specific limitations, the S\_TransNeXtM effectively captures dominant behavioral features, ensuring robust overall performance.

The superiority of the S\_TransNeXtM model for pig behavior recognition can be attributed to the following reasons. Firstly, the S\_TransNeXtM is with two modules: the TransNeXtM and the sLSTM. The bio-inspired Aggregate Attention Mechanism in TransNeXtM enables the model to capture global features effectively. Secondly, the Mamba unit in TransNeXtM allows the model to capture more discriminative features. Furthermore, the exponentially gated mechanism in sLSTM permits the model to establish long temporal sequence dependencies within the corresponding video. Consequently, the promising capabilities

TABLE 7 Comparison result of S\_TransNeXtM with different transformer models.

| Model               | Accuracy (%) | Loss          | Precision (%) | Recall (%)   | F1-Score (%) |
|---------------------|--------------|---------------|---------------|--------------|--------------|
| ViT                 | 83.21        | 0.8597        | 83.79         | 83.44        | 83.99        |
| Swin Transformer    | 90.51        | 0.4902        | 90.88         | 90.32        | 90.40        |
| ConvNeXt            | 92.52        | 0.3991        | 92.48         | 92.52        | 92.74        |
| TransNeXt           | 92.88        | 0.4777        | 92.87         | 92.87        | 92.86        |
| EfficientViT        | 90.51        | 0.4616        | 90.57         | 90.57        | 90.50        |
| CNN-LSTM            | 90.00        | 0.4291        | 90.14         | 89.92        | 89.82        |
| TransNeXtM          | 93.25        | 0.4521        | 93.25         | 93.21        | 93.20        |
| <b>S_TransNeXtM</b> | <b>94.53</b> | <b>0.3350</b> | <b>94.52</b>  | <b>94.54</b> | <b>94.51</b> |

The bold values indicate that the performance of this model is the best.



of global feature extraction and the establishment of long temporal sequence dependencies enhance the performance of pig behavior recognition.

### 3.6 Ablation experiments

In this section, a series of ablation experiments are conducted to systematically assess the effectiveness of the Mamba and sLSTM modules. Specifically, separate introduction of the Aggregation Attention Mechanism or Convolutional GLU module leads to a drawback. The model fails to simultaneously and effectively capture global and local image features. Consequently, the performance of the model declines markedly in the task of pig behavior recognition. Therefore, this experiment adopts the model introducing both of these two modules as the baseline (Index 1).

Based on Table 9, when both Mamba and the Aggregated Attention Mechanism are introduced (Index 2), shows no performance gain over baseline. While Mamba excels at capturing global information from sequential data, it struggles with extracting local features. Similarly, the Aggregated Attention Mechanism emphasizes global features. Due to the lack of the ability to precisely

capture local features, consequently impacts performance of the model in pig behavior recognition tasks.

The combination of the Mamba and Convolutional GLU modules (Index 3), it demonstrates a 0.16% increase in accuracy, and a 2.34% decrease in loss. These two modules are respectively adept at capturing global and local features, and the Mamba is more capable than the Aggregation Attention Mechanism in handling long temporal sequence dependencies. Therefore, compared with the baseline, the performance of the model has been improved.

When the Mamba unit is added (Index 4) in baseline, it exhibits a 0.32% increase in accuracy and a 5.36% decrease in loss compared to the baseline (Index 1). Furthermore, the introduction of Mamba facilitates the filtration of noise and redundant information, enabling the model to extract more discriminative pig behavioral features, which in turn improves the model's performance.

When the sLSTM modules is introduced (Index 5), it demonstrates a 1.1% increase in accuracy, a 19.2% decrease in loss. Due to the unique design of sLSTM featuring exponential gating activation functions and normalized state, the model can effectively capture longer temporal sequence dependencies in videos, improve the accuracy of its data processing, and thus achieve higher accuracy in pig behavior recognition.

TABLE 8 Comparison of the recognition performance of different models in six pig behavior categories.

| Model            | Metric    | Category      |              |              |              |              |              | Avg Prec     | Avg Rec      | Avg F1       |
|------------------|-----------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                  |           | Drinking      | Eating       | Fighting     | Exploring    | Lying        | Walking      |              |              |              |
| Vit              | Precision | 88.30         | 91.67        | 86.05        | 80.00        | 94.12        | 62.62        | 83.94        | 83.21        | 83.28        |
|                  | Recall    | 97.65         | 90.72        | 85.06        | 65.98        | 85.11        | 76.14        |              |              |              |
|                  | F1-Score  | 92.74         | 91.19        | 85.55        | 72.32        | 89.39        | 68.72        |              |              |              |
| Swin Transformer | Precision | 96.47         | 95.92        | 94.94        | 78.95        | 94.79        | 84.21        | 90.77        | 90.51        | 90.44        |
|                  | Recall    | 96.47         | 96.91        | 86.21        | <b>92.78</b> | 96.81        | 72.73        |              |              |              |
|                  | F1-Score  | 96.47         | 96.41        | 90.36        | 85.31        | 95.79        | 78.05        |              |              |              |
| ConvNeXt         | Precision | 93.41         | 95.96        | 96.43        | 90.72        | 94.44        | 83.91        | 92.52        | 92.52        | 92.49        |
|                  | Recall    | <b>100.00</b> | 97.93        | 93.10        | 90.72        | 90.43        | 82.95        |              |              |              |
|                  | F1-Score  | 96.59         | 96.94        | 94.74        | 90.72        | 92.39        | 83.43        |              |              |              |
| TransNeXt        | Precision | 96.43         | 97.87        | 88.89        | 90.63        | 95.79        | 87.64        | 92.94        | 92.88        | 92.90        |
|                  | Recall    | 95.29         | 94.85        | 91.95        | 89.69        | 96.81        | 88.64        |              |              |              |
|                  | F1-Score  | 95.86         | 96.34        | 90.40        | 90.16        | 96.30        | 88.14        |              |              |              |
| Efficient Vit    | Precision | 95.45         | 97.89        | 87.64        | 93.10        | 91.49        | 77.89        | 90.73        | 90.51        | 90.55        |
|                  | Recall    | 98.82         | 95.88        | 89.66        | 83.51        | 91.49        | 84.09        |              |              |              |
|                  | F1-Score  | 97.11         | 96.88        | 88.64        | 88.04        | 91.49        | 80.87        |              |              |              |
| CNN-LSTM         | Precision | 92.22         | 93.75        | 85.26        | 84.11        | 97.67        | 87.84        | 90.18        | 89.96        | 89.87        |
|                  | Recall    | 97.65         | 92.78        | 93.10        | 92.78        | 89.36        | 73.86        |              |              |              |
|                  | F1-Score  | 94.86         | 93.26        | 89.01        | 88.24        | 93.33        | 80.25        |              |              |              |
| TransNeXtM       | Precision | 97.65         | 93.00        | 92.22        | 94.44        | 95.79        | 86.36        | 93.27        | 93.25        | 93.23        |
|                  | Recall    | 97.65         | 95.88        | <b>95.40</b> | 87.63        | 96.81        | 86.36        |              |              |              |
|                  | F1-Score  | 97.65         | 94.42        | 93.79        | 90.91        | 96.30        | 86.36        |              |              |              |
| S_TransNeXtM     | Precision | 94.38         | 97.94        | 91.95        | 91.75        | 95.83        | 90.69        | <b>94.11</b> | <b>94.53</b> | <b>94.29</b> |
|                  | Recall    | 98.82         | <b>97.94</b> | 91.95        | 91.75        | <b>97.87</b> | <b>88.64</b> |              |              |              |
|                  | F1-Score  | 96.55         | 97.94        | 91.95        | 91.75        | 96.84        | 89.65        |              |              |              |
| Total Sample     |           | 85            | 97           | 87           | 97           | 94           | 88           |              |              |              |

The bold values indicate that the performance of this model is the best.

TABLE 9 Performance comparison of the models with different module combinations.

| Index | Block |                     |                   | sLSTM | Accuracy (%) | Loss          |
|-------|-------|---------------------|-------------------|-------|--------------|---------------|
|       | Mamba | Aggregate Attention | Convolutional GLU |       |              |               |
| 1     | ×     | ✓                   | ✓                 | ×     | 92.88        | 0.4777        |
| 2     | ✓     | ✓                   | ×                 | ×     | 92.29        | 0.4760        |
| 3     | ✓     | ×                   | ✓                 | ×     | 93.04        | 0.4665        |
| 4     | ✓     | ✓                   | ✓                 | ×     | 93.20        | 0.4521        |
| 5     | ×     | ✓                   | ✓                 | ✓     | 93.98        | 0.3860        |
| 6     | ✓     | ✓                   | ✓                 | ✓     | <b>94.53</b> | <b>0.3500</b> |

The bold values indicate that the performance of this model is the best.

Finally, when the Mamba and sLSTM are simultaneously added (Index 6), the S\_TransNeXtM achieves peak performance with a 1.65% increase in accuracy, a 26.7% decrease in loss compared to the baseline. This combination maximizes the utilization of their complementary strengths. Specifically,

the Mamba suppresses noise and effectively captures the characteristics of pigs. Additionally, sLSTM handles long temporal sequence dependence.

By leveraging the synergistic effects among these modules, the S\_TransNeXtM model addresses the limitations of traditional pig

behavior recognition models in effectively extracting image features and analyzing long temporal sequence dependencies.

## 4 Conclusions

This paper proposes a novel pig behavior recognition model, S\_TransNeXtM, which leverages both spatial and temporal information underlying the video. Specifically, for the spatial domain, the TransNeXtM is initially introduced, leveraging a Aggregated Attention Mechanism, a Convolutional GLU, and a Mamba unit to capture more discriminative global and local features. This allows the model to perceive more subtle differences in the pig's behavior. Additionally, in the temporal domain, sLSTM's exponential gating and stabilized states provide improved capability for processing long temporal sequence dependencies. Consequently, the S\_TransNeXtM enhances the performance of pig behavior recognition. We conducted numerous experiments to validate the effectiveness of the proposed model. Experimental validation demonstrates the state-of-the-art performance with 94.53% accuracy, surpassing the mentioned methods by 0.55%–11.32% and reducing loss by 52.47%.

In the future work, we will focus on processing longer temporal sequence data to further improve the performance of pig behavior recognition, via optimizing the architecture of the model. Additionally, we will work on developing an adaptive learning mechanism that enables the model to automatically adjust its parameters. Thus, it can cope with the diversity and dynamic changes of pig behavior under different environmental conditions. Furthermore, we will invest more resources into collecting high-quality, diverse, and representative datasets of pig behavior under various conditions.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The animal study was approved by Experimental Animal Ethics Committee of Shanxi Agricultural University. The study was conducted in accordance with the local legislation and institutional requirements.

## Author contributions

WH: Conceptualization, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. XH: Data curation, Visualization, Writing – original draft. YX: Investigation, Writing – original draft. HS: Software, Validation, Writing – review & editing. MH: Formal analysis, Project administration, Resources, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the GHfund D [ghfund202407042032]; the Shanxi Province Basic Research Program [202203021212444]; Shanxi Agricultural University Science and Technology Innovation Enhancement Project [CXGC2023045]; Shanxi Postgraduate Education and Teaching Reform Project Fund [2022YJJG094]; Shanxi Agricultural University doctoral research start-up project [2021BQ88]; Shanxi Agricultural University Academic Restoration Research Project [2020xshf38]; Young and Middle-aged Top-notch Innovative Talent Cultivation Program of the Software College, Shanxi Agricultural University [SXAUKY2024005]; Teaching Reform Project of Shanxi Agricultural University [JG-202523] and Zhejiang Province Key Research and Development Program [No. 2024C01212].

## Acknowledgments

The authors thank all the staff of the pig breeding farm of Nonglv Yuan Agricultural Co., Ltd. in Xiangfen County, Linfen City, Shanxi Province, as well as the Software College of Shanxi Agricultural University for their assistance in this research.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



## References

- Huang L, Xu L, Wang Y, Peng Y, Zou Z, Huang P. Efficient detection method of pig-posture behavior based on multiple attention mechanism. *Comput Intell Neurosci.* (2022) 2022:1759542. doi: 10.1155/2022/1759542
- Melfsen A, Lepsien A, Bosselmann J, Koschmider A, Hartung E. Describing behavior sequences of fattening pigs using process mining on video data and automated pig behavior recognition. *Agriculture.* (2023) 13:1639. doi: 10.3390/agriculture13081639
- Wei J, Tang X, Liu J, Zhang Z. Detection of pig movement and aggression using deep learning approaches. *Animals.* (2023) 13:3074. doi: 10.3390/ani13193074
- Wangli H, Kai Z, Li Z, Meng H, Wangbao H, Fuzhong L, et al. TSML: a new pig behavior recognition method based on two-stream mutual learning network. *Sensors.* (2023) 23:5092. doi: 10.3390/s23115092
- Qiumei Y, Deqin X. A review of video-based pig behavior recognition. *Appl Anim Behav Sci.* (2020) 233:105146. doi: 10.1016/j.applanim.2020.105146
- Jiangtao Q, Xiangnan L, Kai L, Farong X, Hui G, Xinliang T, et al. An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease. *Comput Electron Agric.* (2022) 194:106780. doi: 10.1016/j.compag.2022.106780
- Franziska H, Borge JD. Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods. *Front Vet Sci.* (2023) 9:1099347. doi: 10.3389/fvets.2022.1099347
- Hengyi J, Guanghui T, Jionghua Y, Yanbin W, Huixiang D, Yanrong Z. Efficient aggressive behavior recognition of pigs based on temporal shift module. *Animals.* (2023) 13:2078. doi: 10.3390/ani13132078
- Yue G, Kai Y, Baisheng D, Hongmin S, Yanling Y, Runze L, et al. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. *Animals.* (2023) 13:2078. doi: 10.1016/J.COMPAG.2022.107606
- Lili N, Bugao L, Fan J, JiaYuan S, Tingting Y, Zhenyu L. ASPP-YOLOv5: a study on constructing pig facial expression recognition for heat stress. *Comput Electron Agric.* (2023) 214:108346. doi: 10.1016/j.compag.2023.108346
- Junjie H, Janice S, Dirk C, Raymond L, Anna B, Chen C, et al. Evaluation of computer vision for detecting agonistic behavior of pigs in a single-space feeding stall through blocked cross-validation strategies. *Comput Electron Agric.* (2023) 204:107520. doi: 10.1016/j.compag.2022.107520
- Ma C, Deng M, Yin Y. Pig face recognition based on improved YOLOv4 lightweight neural network. *Inf Process Agric.* (2024) 11:356–71. doi: 10.1016/j.inpa.2023.03.004
- Zhang Y, Yang X, Liu Y, Zhou J, Huang Y, Li J, et al. A time-series neural network for pig feeding behavior recognition and dangerous detection from videos. *Comput Electron Agric.* (2024) 218:108710. doi: 10.1016/j.compag.2024.108710
- Vestbjerg Larsen ML, Jensen MB, Pedersen LJ. Increasing the number of wooden beams from two to four increases the exploratory behaviour of finisher pigs (2019). Available online at: <https://www.sciencedirect.com/science/article/pii/S0168159119300528> (Accessed March 16, 2025).
- Xu J, Ying Y, Wu D, Hu Y, Cui D. Recent advances in pig behavior detection based on information perception technology (2025). Available online at: <https://www.sciencedirect.com/science/article/pii/S0168169925004338> (Accessed March 16, 2025).
- Wang W, Xie E, Li X, Fan DP, Song K, Liang D, et al. PVT v2: improved baselines with pyramid vision transformer. *arXiv Preprint arXiv:2106.13797.* (2023). doi: 10.48550/arXiv.2106.13797
- Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. *arXiv Preprint arXiv:2312.00752.* (2024). doi: 10.48550/arXiv.2312.00752
- Chen L, Juntao Z, Wei S. Time series classification based on convolutional network with a Gated Linear Units kernel. *Eng Appl Artif Intell.* (2023) 123:106296. doi: 10.1016/j.engappai.2023.106296
- Zhao F, Feng F, Ye S, Mao Y, Chen X, Li Y, et al. Multi-head self-attention mechanism-based global feature learning model for ASD diagnosis. *Biomed Signal Process Control.* (2024) 91:106090. doi: 10.1016/j.bspc.2024.106090
- Shi D. TransNeXt: robust foveal visual perception for vision transformers. *arXiv Preprint arXiv:2311.17132.* (2023). doi: 10.1109/CVPR52733.2024.01683
- Qingguo Z, Xiangru L, Haitao L. Concat Convolutional Neural Network for pulsar candidate selection. *arXiv Preprint arXiv:2004.00257.* (2020). doi: 10.48550/arXiv.2004.00257
- Beck M, Pppel K, Spanring M, Auer A, Prudnikova O, Kopp M, et al. xLSTM: extended long short-term memory. *arXiv Preprint arXiv:2405.04517.* (2024). doi: 10.48550/arXiv.2405.04517
- Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models. *J Artif Intell.* (2024) 6:301–60. doi: 10.32604/jai.2024.054314
- Nwankpa C, Ijomah W, Gachagan A, Marshall S. *Activation functions: comparison of trends in practice and research for deep learning* (2020). Available online at: [https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1h1c0xm0sj1k0j80sj4a06c0vc596274&site=xueshu\\_se](https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1h1c0xm0sj1k0j80sj4a06c0vc596274&site=xueshu_se)
- Kai H, Yunhe W, Hanting C, Xinghao C, Jianyuan G, Zhenhua L, et al. *A Survey on Vision Transformer* (2022). doi: 10.1109/TPAMI.2022.3152247. Available online at: [https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1a0n04804y4n0r70s21k0ev0an160577&site=xueshu\\_se](https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1a0n04804y4n0r70s21k0ev0an160577&site=xueshu_se)
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows (2021). doi: 10.1109/ICCV48922.2021.00986. [https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1q730ru0se7n0480eg3c0jd0j9331460&site=xueshu\\_se](https://xueshu.baidu.com/ndscholar/browse/detail?paperid=1q730ru0se7n0480eg3c0jd0j9331460&site=xueshu_se)
- An Y, Yi Y, Han X, Wu L, Su C, Liu B, et al. A hybrid attention-guided ConvNeXt-GRU network for action recognition. *Eng Appl Artif Intell.* (2024) 133:108243. doi: 10.1016/j.engappai.2024.108243
- Liu X, Peng H, Zheng N, Yang Y, Hu H, Yuan Y. EfficientViT: Memory efficient vision transformer with cascaded group attention. *arXiv Preprint arXiv:2305.07027.* (2023). doi: 10.48550/arXiv.2305.07027