Expert Systems With Applications 237 (2024) 121462

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Analyzing the pregnancy status of giant pandas with hierarchical behavioral information

Xianggang Li [a], Jing Wu [b], Rong Hou [c], Zhangyu Zhou [d], Chang Duan [e], Peng Liu [c], Mengnan He [c], Yingjie Zhou [a,*], Peng Chen [c,*], Ce Zhu [d]

[a] College of Computer Science, Sichuan University, Chengdu, China
[b] Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia
[c] Chengdu Research Base of Giant Panda Breeding, Sichuan Key Laboratory of Conservation Biology for Endangered Wildlife, Chengdu, China
[d] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China
[e] School of Electrical Engineer and Information, Southwest Petroleum University, Chengdu, China

ARTICLE INFO

Keywords:
Giant pandas
Behavior analysis
Neural network
Action recognition
Intelligent video analysis

ABSTRACT

As giant pandas (*Ailuropoda melanoleuca*) are difficult to conceive and prone to abortion, humans have turned to artificial captive breeding to increase their population. Therefore, it is crucial for their reproduction to analyze their pregnancy status accurately and promptly in artificial captive breeding. To determine whether a giant panda is pregnant, with the current methods, experts must keep a close eye on them and frequently collect their urine or blood, which requires significant resources with a high misdiagnosis rate, and will adversely affect giant pandas' daily lives. Consequently, it is essential to rapidly advance the development of automated, precise methods that will not disrupt the pandas' lives to analyze giant pandas' behaviors and determine whether or not they are pregnant. In this paper, we propose an end-to-end intelligent system for predicting the pregnancy status of giant pandas and their Expected Date of Delivery (EDD). We first introduce expert knowledge to machine learning methods to solve this problem, which can significantly improve the accuracy of prediction. Experimental results show that this system achieves an accuracy of 91.5% for the pregnancy diagnosis and 0.579 days of mean average error for EDD prediction when the observation period is 5 days. Our automated system significantly reduces the need for human intervention, thus minimizing disruptions to the pandas' daily lives. It has the potential to contribute to the health and genetic diversity of the giant pandas, as well as aid in the panda's artificial reproduction and population growth.

1. Introduction

The charismatic giant panda is a conservation icon on a global scale and possesses exceptional scientific, economic, and cultural value. They were, however, on the verge of extinction due to factors such as low reproductive capacity and habitat destruction. The panda population is beginning to recover thanks to decades of successful conservation efforts. Panda's status on the IUCN Red List[1] was upgraded from "endangered" to "vulnerable" in 2016 (Swaisgood, Wang, & Wei, 2016). Chinese authorities also reclassified the giant panda as vulnerable in 2021 (Elías, de la Vega, et al., 2022). Nevertheless, they are still regarded as a rare species.

To increase the giant panda population, humans have taken a series of protective measures, including artificial captive breeding and habitat conservation (Kang, 2022). With the methods that are currently in use, however, experts need to keep a close eye on them and frequently collect their urine or blood in order to determine whether or not a giant panda is pregnant. We can categorize the studies in this field as either a hormone or behavior analysis. Hormone analysis can tell if a giant panda is pregnant by tracking fluctuations in the hormones found in its urine (Chen, Yuhua, Yuanzhi, & Yueming, 1985), which is based on the observation that the hormone levels (*e.g.,*Estradiol and Progesterone) in the urine of pregnant giant pandas differ significantly from those of non-pregnant giant pandas (Luo, Jian, Zhihe, & Rong, 2011). However, it requires sampling as soon as possible after urination to reduce contamination and relies heavily on giant panda excretion, which is not always stable (Liu, Yue, Liran, Bo, & Hui, 2005). The manual behavior analysis method determines whether a giant panda is pregnant by observing the behaviors of pregnant giant pandas. Pregnant giant pandas' external behaviors will differ depending on their gestational stage (Pan, Zuofu, Chao, & Peng, 2015). It is easier to implement than

* Corresponding authors.
E-mail addresses: yjzhou@scu.edu.cn (Y. Zhou), capricorncp@panda.org.cn (P. Chen).

[1] The International Union for Conservation of Nature (IUCN) Red List of Threatened Species, also known as the IUCN Red List or Red Data Book

https://doi.org/10.1016/j.eswa.2023.121462
Received 6 April 2023; Received in revised form 25 August 2023; Accepted 2 September 2023
Available online 7 September 2023
0957-4174/

the hormone analysis method, while it relies heavily on professional expertise and necessitates significant human effort. Besides, It has more effects on giant pandas than the former and will also disrupt the pandas' daily life, which could have serious consequences (*e.g.,* abortion) (Yang, Jingbo, & Yingjie, 2019). Overall, the methods currently used to predict pregnancy in giant pandas are unable to produce stable and accurate predictions with small manual effort and negligible impact on giant pandas (Huang, Jianqiu, Xuanzhen, et al., 1999).

Due to the giant panda's status as a rare animal with limited reproductive capabilities, there is a scarcity of literature related to giant panda pregnancy analysis, and a majority of the available literature is outdated. Information obtained from the Chengdu Research Base of Giant Panda Breeding, the world's largest giant panda breeding center, indicates that the existing analysis methods primarily involve frequent manual observations combined with urine analysis. However, these methods have the drawbacks mentioned earlier, including causing certain disturbances to the giant pandas' natural behaviors and heavy reliance on human intervention mentioned above. Our understanding of this issue is derived not solely from literature, but also from real-world observations.

In this study, we propose an end-to-end intelligent system for predicting the pregnancy status of giant pandas and their expected date of delivery. We first introduce expert knowledge to machine learning methods to solve this problem, which can significantly improve the accuracy of prediction and help save the lives of giant pandas in delivery. Precisely, the proposed method consists of three modules: **Hierarchical Behavioral Information Acquisition**, **Hierarchical Behavioral Time-series Construction**, and **Pregnancy Analysis**.

(1) The first module aims to construct an annotated giant pandas' behavior recognition dataset. At first, we divide the giant panda's behaviors into four levels, each with distinct categories of typical behaviors. After that, based on our definition, we annotate the video dataset for behavior recognition of giant panda, which contains position and hierarchical behavioral labels. (2) In the second module, to analyze the videos of the giant pandas produced through captive breeding, we construct a time series dataset with hierarchical behavioral information. We start by training a behavior recognition module, SlowFast networks (Feichtenhofer, Fan, Malik, & He, 2019), using the hierarchical behavior dataset that was generated in the first step. Next, we deploy this model to recognize videos of artificially bred giant pandas and count the recognition results to construct hierarchical behavior time series data. (3) The objective of the final module is to analyze the pregnancy status of giant pandas based on hierarchical behavior time series data, including two components: pregnancy diagnosis and EDD prediction. Each component uses multi-LSTM units (Hochreiter & Schmidhuber, 1997) to extract features of each behavior independently, then this module employs the Efficient Channel Attention (ECA) module (Feichtenhofer, Fan, Malik, & He, 2020) to learn the significance of each feature to the target task, and finally outputs the result by utilizing a classification head or prediction head. This module provides a comprehensive understanding of the potential spatial–temporal correlation between behavioral data and pregnancy status. We also develop a data augmentation technique based on scalable sliding window (SSW) to increase the number of training samples and improve the performance of our classification and regression models' performance, reducing the small sample size's negative effect. Experiments show that the system achieves an accuracy of 91.5% for pregnancy diagnosis and 0.579 days of mean average error (MAE) for EDD prediction when the observation time is 5 days.

The contributions of this paper are summarized as follows:

1. We design an automatic system to analyze the pregnancy status of giant pandas by employing intelligent video analysis technology. Four levels panda behaviors along with the typical categories at each level are built in the system, which enables the effective classification of the pregnancy status and the accurate prediction of the EDD while decreasing the workload of manual observation. To the best of our knowledge, this is the first work that analyzes the pregnancy status of giant pandas based on intelligent video analysis.
2. We design a pregnancy diagnosis and delivery prediction algorithm based on the Efficient Channel Attention (ECA) network, which can capture the relationship between different levels of behaviors and pregnancy. Moreover, for the limited number of observable giant pandas, we propose a data augmentation method named scalable sliding window (SSW) to increase the number of samples, significantly improving the method's performance.
3. The proposed system achieves high performance in terms of pregnancy diagnosis and EDD prediction, which has the potential to help the artificial breeding of pandas and increase their population.

This paper is structured as follows: In Section 2, we provided a summary of the current research on pregnancy diagnosis, action recognition, and classification/prediction using time series data. Section 3 describes the proposed system in detail. In Section 4, we conducted a series of experiments to demonstrate the effectiveness of our designed system. Finally, we conclude and discuss our work in Sections 5 and 6.

2. Related work

In this section, we present the related work in three parts. The first part summarizes existing methods for the pregnancy analysis of artificially bred giant pandas. Behavior recognition methods in computer vision are briefly introduced in the second part. Finally, we provide the relevant advancements for time series analysis.

2.1. Analysis of giant pandas' pregnancy

Currently, zoologists have two primary methods for analyzing the pregnancy status of giant pandas: hormone analysis in urine and behavior analysis based on manual observation.

During different stages of pregnancy, the content of certain hormones in giant pandas varies significantly (Luo et al., 2011). For instance, giant pandas' progesterone levels rapidly decrease before giving birth (Zhao, Zhang, et al., 2011). The method of hormone analysis in urine determines pregnancy by measuring changes in the concentration of particular hormones. To predict the EDD, a regression equation is constructed for the numbers of days between the absolute peak of estradiol concentration and the cutoff date of the first rising period of progesterone concentration (Luo et al., 2011). Alterations in Panda Chorionic Gonadotrophin (PCG) levels in the urine can also be used to diagnose pseudo-pregnancy (Chen et al., 1985). The method based on hormone concentrations analysis in urine has the advantages of a high degree of precision and robust interpretability. Nonetheless, this method depends on the time of the giant panda's exclusion, and requires a prompt response following the excretion, while necessitating manual and continuous tracking. The determination of hormones is frequently delayed, leading to a miscalculation of the due date and an inappropriate feeding management (Liu et al., 2005).

The feeding, resting, and other behaviors of pregnant and pseudo-pregnant giant pandas are significantly distinct. Zhou and Ji (2016). Gandia et al. conducted an assessment of the circadian rhythms (13 h) and annual behavioral patterns of 24 captive giant pandas in a zoo setting. They also observed variations in the patterns and intensity of behavioral cycles based on life stages and gender (Gandia, Herrelko, Kessler, & Buchanan-Smith, 2023). Based on manual observation, the behavior analysis method can tell if a panda is pregnant by looking at statistically significant differences in how pandas act during the reaction period (Pan et al., 2015). This method is simple to operate and does not need sampling, but it requires continuous manual observation and relies heavily on the personal experience of the breeder. In addition, frequent contact and observation will diminish the behavioral diversity of giant pandas (Yang et al., 2019).

Observing vaginal reproductive epithelial cells and immunology are additional methods for determining the pregnancy status of giant pandas. The former determines whether a giant panda is pregnant by measuring the keratinization rate of vaginal epithelial cells (Huang et al., 1999). The latter observes changes in the number of E-Rose (E-RFC) rings formed in the blood by T lymphocytes and sheep erythrocytes (Liu et al., 1986). These methods require routine sampling of the panda's vagina or blood, severely disrupting its daily life. In addition, because the giant panda is a rare protected animal, it is inappropriate to sample their vagina or blood frequently.

However, due to the rarity of giant pandas, obtaining research samples has proven challenging, leading to a limited number of studies on panda pregnancies. In recent years, numerous algorithms from the field of computer science have been applied to analyze pregnancies in other animals. For instance, by quantifying real-time ultrasound (RTU) imaging through numerical models, artificial neural networks (ANN) have been trained to predict the litter size of pregnant sows (Kousenidis, Kirtsanis, Karageorgiou, & Tsiokos, 2022). And utilizing transformer neural network to predict and interpret pregnancy loss from activity data in Holstein dairy cows (Lin, Kenéz, McArt, & Li, 2023). However, these methods require hands-on data collection, which could potentially disrupt the natural behaviors of giant pandas. Based on the increase in oxytocin levels in pregnant sows leading to nesting behavior and increased postural changes prior to parturition, the YOLOv5 algorithm is employed to detect postural changes in sows and this information is then used to predict the timing of parturition (Chen et al., 2023). And an approach employs a simple random walk model to predict the timing of calving events in pregnant cows (Zin et al., 2022). While these methods take into account the potential link between behavioral changes and pregnancy, they are limited to short-term parturition prediction just before delivery. They cannot facilitate pseudo pregnancy diagnosis nor fulfill the requirement for observing the complete pregnancy cycle of giant pandas.

2.2. Behavior recognition

Behavior recognition aims to identify the action categories of the observation target based on their behaviors. According to their different feature extraction methods, the recognition methods can be divided into manual and deep learning methods. Manual methods employ artificially designed features such as HOG (Dalal & Triggs, 2005), HOF (Chaudhry, Ravichandran, Hager, & Vidal, 2009), and MBH (Dalal, Triggs, & Schmid, 2006) to extract behavioral features from videos or images. Then they use machine learning techniques such as Hidden Markov Models (Baum & Petrie, 1966), SVMs (Cortes & Vapnik, 1995), and others for action recognition and classification. Such methods are fast and interpretable, but their precision is usually limited. Deep learning techniques typically employ neural networks to extract behavioral features in an end-to-end manner. 3D-CNN (Ji, Xu, Yang, & Yu, 2012) is the first method of presentative deep learning for action recognition that can simultaneously extracts information from the temporal and spatial domains. The two-stream method (Simonyan & Zisserman, 2014) extracts spatial and temporal information simultaneously. It uses a 2D-CNN network to extract spatial information from single-frame images and a 3D-CNN network to extract temporal information by processing multi-frame density optical flow fields. The SlowFast (Feichtenhofer et al., 2019) samples the video at two different rates and inputs the frames into "Slow" and "Fast" paths to effectively extract semantic and motion features. The parameter quantities in the two branches are comparable to the ratios of Magnocellular (responding to high frequencies) and Parvocellular (lower temporal resolution, sensitive to spatial detail and color) cells in the human retina. The rapid development of deep learning methods has made intelligent video analysis of giant pandas possible. The mentioned methods all employ an equidistant frame sampling strategy, which might overlook certain key frames and limit recognition performance. SMART (Gowda, Rohrbach, & Sevilla-Lara, 2021), on the other hand, considers these frames jointly rather than selecting them one by one, thereby enhancing the model's recognition accuracy. Inspired by the success of Vision-Language models (VLMs), the LIKE (Wu et al., 2023) model introduces a bi-directional cross-modal approach by pretraining on Video-to-Text data to enhance video representations. VideoMAE V2 (Wang et al., 2023), however, employs a self-supervised strategy for pretraining and presents a dual masking strategy to enhance efficiency in pre-training, thereby further reducing the overall computational cost.

2.3. Time-series analysis

Time series usually consists of a sequence of historical observations on particular attributes of the target object. It contains information about the object's inherent descriptions and trends. There have been numerous studies on time series analysis recently. Through analyzing the time series data of the water plant sensors, Li et al. (2021) determine whether the operation in the water plant was abnormal. Martis, Acharya, Mandana, Ray, and Chakraborty (2012) monitored the status of human health based on the ECG data.Ai, Pan, and Li (2017) predicted haze weather based on historical observations of weather time series data. Using the user's historical consumption data, Ghost et al. performed credit card fraud detection (Ghosh & Reilly, 1994). According to expert knowledge in zoology, there are differences between the activity patterns of actual giant pandas and those of pseudo pregnant ones across the period of gestation (Pan et al., 2015). In light of these studies, we conduct research on analyzing the pregnancy status of giant pandas using hierarchical behavioral time series data.

In the field of pregnancy analysis, Lin et al. proposed a Transformer-based network that predicts the probability of pregnancy loss in cows using continuous activity data (Lin et al., 2023). This work will be referred to as CowTrans in this paper, as the original author did not provide a specific name.

3. Methodology

In this section, we designed a system for analyzing the pregnancy status of giant pandas through the historical records of behavioral observations. Determining whether giant pandas are pregnant by analyzing their behavioral changes. The system has minimal impact on giant pandas and a low reliance on manpower, which can provide important reference information for artificial breeding. The architecture of the system is shown in Fig. 1, which consists of the following three modules: **Hierarchical Behavioral Information Acquisition**, **Hierarchical Behavioral Time-series Construction**, and **Pregnancy Analysis**. The first module defines four levels of panda behaviors based on zoological knowledge (*i.e.*, **Posture**, **Action**, **Activity**, and **Female Behaviors during Pregnancy**) and typical categories (*e.g.,*sitting, eating, exploring) at each level. Then, based on these definitions, it construct a video dataset with hierarchical behavioral labels. The second module employs the labeled dataset to train a behavior recognition model that will identify the behaviors from videos and generate hierarchical behavior time-series data containing high-level information, such as health and pregnancy status. From the obtained hierarchical behavior data, the third module learns the inherent rules (*e.g.,*abdominal contractions occur with different frequencies at different stages of pregnancy) among hierarchical behaviors and the potential connection between the rules and pregnancy, then determines the pregnancy status.

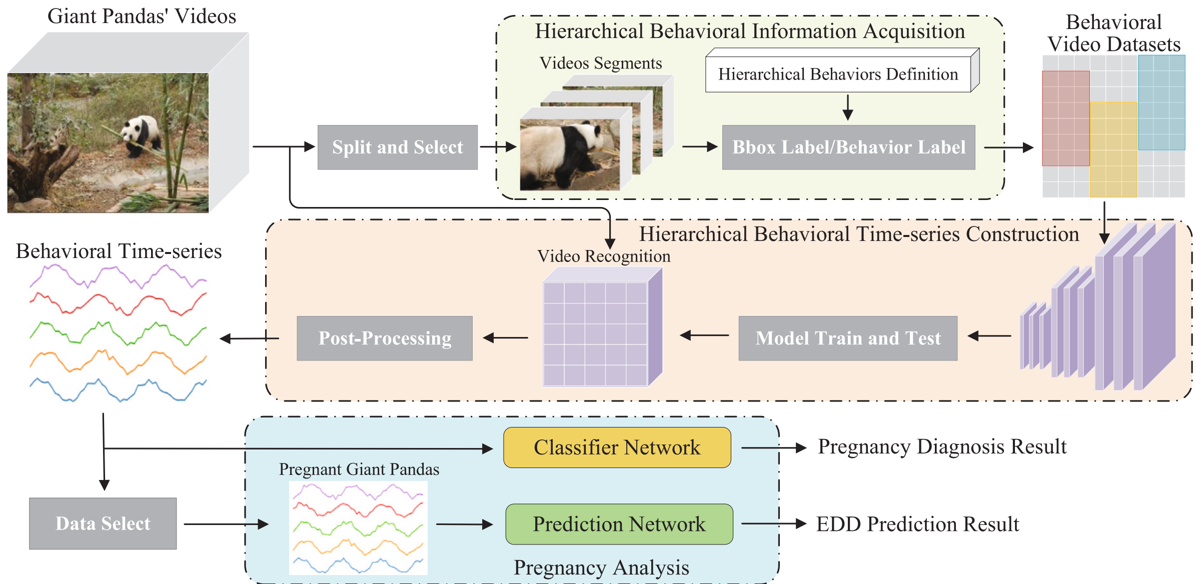

Fig. 1. The architecture of the designed system for analyzing the pregnancy status. The system contains three modules: hierarchical behavioral information acquisition module defines hierarchical behaviors for giant pandas and construct a video dataset with hierarchical behavioral labels; hierarchical behavioral time-series construction module trains a hierarchical behavior recognition model and count the recognition results to construct hierarchical behavioral time series data; pregnancy analysis module is used to diagnose pregnancy or predict EDD.

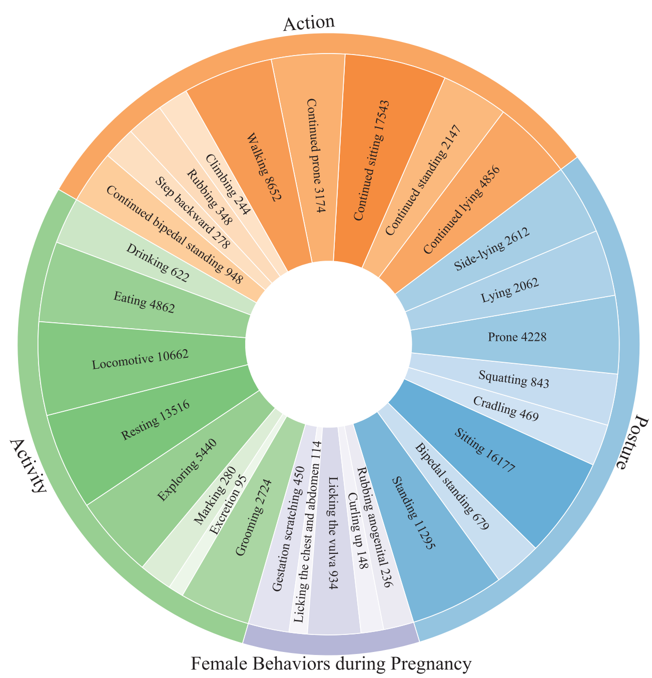

Fig. 2. Distribution of the behaviors from different levels. The dataset contains a total of 30 behaviors from 4 different levels, and the numbers of different behaviors are extremely unbalanced. There are 17543 frames labeled as continued sitting, while 95 frames are labeled as excretion.

3.1. Hierarchical behavioral information acquisition

This module aims to construct the dataset for training the hierarchical behavior recognition model from the collected videos of artificially inseminated giant pandas, including two tasks: behavior definition and dataset annotation. According to relevant zoological knowledge, we categorize the behaviors of giant pandas into different levels and define the characteristic categories at each level. Accordingly, the preprocessed video data is annotated with the behavior definitions.

The objective of defining the giant panda behaviors is to establish a scientific, accurate, and exhaustive classification standard for describing the behaviors of giant pandas, which can fully express the external exhibitions and high-level status of giant pandas to provide a reference for labeling the dataset. From a behavioral perspective, giant pandas can simultaneously exhibit multiple levels of behaviors, such as eating bamboo while being seated and exploring while standing. Sitting and standing are relatively low-level behavioral status, containing only simple posture information, whereas eating and exploring are higher-level status, containing information such as panda activities and intentions. These behaviors exist simultaneously and describe the status of giant pandas at various levels, whereas the existing giant panda knowledge graph lacks this information. According to zoological knowledge, we divide the behaviors of giant pandas into four levels: **Posture**, **Action**, **Activity**, and **Female Behaviors during Pregnancy**. We identify 46 typical behaviors from those four levels, including 9 at the posture level, 14 at the action level, 9 at the activity level, and 14 at the female behaviors during pregnancy.

In contrast to the existing giant panda behavior knowledge graph, our hierarchical behavior definition distinguishes giant panda behaviors with comprehensively considering their differences and interconnections. Our definition of behaviors fully accounts for the hierarchical nature of giant pandas' behaviors, allowing a detailed multidimensional behavioral description and facilitating the discovery of potential connections between behavioral exhibitions and pregnancy. Among the four levels, **Posture** is a kind of behaviors determined by a single frame image, representing an instantaneous (shallow) status, such as sitting or standing. **Action** indicates a short-term dependence, such as walking or climbing, which can be determined by a video clip. **Activity** is a long-term dependence, indicating deeper state information (*e.g.,*exercise and rest). **Female Behaviors during Pregnancy** has more occurrences during pregnancy and occurs less frequently at other periods, such as abdominal contraction and rubbing anogenital. These behaviors show different distribution characteristics at different stages of pregnancy. In addition, the behavior definition meets the two criteria listed below: (a) Representative: behavioral categories adequately represent

Table 1
The information of video data collected.

Name	View	Days	Duration (h)	Pregnancy	ADD
Zhaomei	2	7	66.1	True	2021–07–23
Meilun	2	6	40.5	True	2021–07–25
Yayun[a]	1	4	25.5	Pseudo	–
Nida	1	23	210.6	Pseudo	–
Chengji	1	26	231.3	Pseudo	–
Nini	1	30	246.4	Pseudo	–
Xingya	1	2	7.8	Pseudo	–
Fuya	1	21	195.3	Pseudo	–
Qifu	1	6	19.2	True	2021–07–17
Yayun[a]	1	5	18.1	Pseudo	–
Yazai	2	27	346.3	Pseudo	–
Jili	1	23	203.1	Pseudo	–
Yali	2	13	100.7	True	2021–07–31
Total	–	193	1711.9	–	–

[a] Two different pandas with the same English name.

the behavioral status of giant pandas at this level. (b) Atomicity: the behavior categories in each level are mutually exclusive. The categories and descriptions of the defined behaviors are provided in detail in Appendix A.

To validate the feasibility of our designed system, we collected the videos regarding giant panda pregnancy behaviors described in Table 1 at Chengdu Research Base of Giant Breeding from July to August, 2021. These videos include the data of 13 artificially inseminated giant pandas, of which 4 are pregnant and 9 are pseudo-pregnant. The video of each panda has been collected for 2 to 30 days with one or two cameras, approximately 9962 clips with a total duration of 1711.9 h in total. The data of 3 pseudo-pregnant and 1 pregnant giant pandas are excluded due to insufficient observation time or poor quality.

After the annotation of the video data, a dataset could be generated for training the hierarchical behavior recognition model. There are two types of labels: location and behavior. The former utilizes the BBox (Bounding Box) labels to identify the relative position of giant pandas in the video and is used to train an object detection model for finding the position of giant pandas in an unlabeled video. The latter is annotated according to the above definition of the giant panda's hierarchical behaviors. It will be used together with the former to train a giant panda hierarchical behavior recognition model. We utilize the keyframe annotation technique to strike a balance between accuracy and efficiency. In each video, we only select the first, last, and middle frames per second as keyframes. Each annotated video clip is thirty seconds long and contains thirty-two keyframes, which significantly reduces the number of labels and guarantees the accuracy of annotations.

In the end, we labeled 1200 30-second video clips containing 38 252 bounding boxes and 153 600 behavior labels. There are 30 types of behaviors in the scenario of our dataset. The numbers of behavior categories are 8, 9, 8, and 5 for posture, action, activity, and female behaviors during pregnancy, respectively. In addition, the distribution of labels for each behavior in the dataset is also unbalance, as shown in Fig. 2.

3.2. Hierarchical behavioral time-series construction

This module's objective is to construct hierarchical behavior time series on the base of intelligent identification for artificially inseminated giant pandas' behaviors. Behavioral information for the pregnancy analysis are contained in time series. The module consists of three parts: the object detector, the hierarchical behavior classifier, and the time series constructor. Specifically, the object detector is used to detect and locate giant pandas from the video. The hierarchical behavior classifier identifies hierarchical behaviors on these pandas located by the object detection model. The time series constructor counts the recognition results and constructs time series data containing behavior profiles.

The camera is fixed at a certain distance from the giant panda, thus the giant panda only occupies a small part of the videos. If we directly input the original video into the behavior classification model, the classification accuracy will be unacceptable due to the influence of irrelevant background. It is necessary to determine the position of the giant panda in the original video and then input the video containing only the pandas into the behavior recognition model. For identifying giant pandas and locating their relative positions in videos, existing works, such as YOLOv5 (Jocher, Chaurasia, Stoken, Borovec, NanoCode012, et al., 2022), Faster RCNN (Ren, He, Girshick, & Sun, 2015), and RetinaNet (Lin, Goyal, Girshick, He, & Dollár, 2017), achieve satisfactory performance in the object detection. We select Faster RCNN as the giant panda object detection model because it could seamlessly incorporate officially by SlowFast, the behavior classifier utilized in the following process.

The hierarchical behavior classifier recognizes the behavior categories of cropped pandas videos, which is located by the object detection model. The behavior classifier will provide the corresponding behavioral information and convert video data into behavioral data. Current behavior classifiers achieve remarkable performance (*e.g.,* Two Stream Simonyan & Zisserman, 2014, SlowFast Feichtenhofer et al., 2019, SMART Gowda et al., 2021). In this paper, we choose SlowFast as the hierarchical behavior recognition model, which consists of two branches, "Slow" and "Fast", to capture spatial details and temporal changes in the video separately.

The definition of the giant panda's hierarchical behavior divides the panda's behaviors into four levels. Note that multiple levels of behaviors may coincide. Thus, there are two strategies for training a behavior recognition model: **(a) single-level classification**: conducting model training for each behavioral level separately. **(b) multi-level classification**: training one hierarchical behavior recognition model that shares the same backbone but coops with diverse classifiers for different behavior levels. We selected the latter training strategy for SlowFast due to its high performance and low computational costs, and detailed experimental settings are introduced in Section 4.4. For a video clip $\boldsymbol{X}^{\text{clips}} \in \mathbb{R}^{W\times H\times T}$, the corresponding multi-label behavior label is $\boldsymbol{Y}^{\text{clips}} \in \mathbb{R}^{T}$. Input $\boldsymbol{X}^{\text{clips}}$ and $\boldsymbol{Y}^{\text{clips}}$ into the behavior recognition network for training, we can obtain the trained network weights $\boldsymbol{\Theta}_{sf}$. The corresponding procedures in Algorithm 1 are Step 1–6.

The time series constructor utilizes the trained hierarchical behavior recognition model to obtain the behavior profiles for the observed giant pandas and analyze the corresponding mathematical statistics to generate hierarchical behavior time series. In the experiments, we totally collected approximately 697-hour videos of 10 giant pandas for testing after removing those videos with poor quality. In order to obtain the hierarchical behavioral time series data of each giant panda, we firstly count the occurrence time of different behavior categories for each panda in a day, and calculate their proportions to the total observation time of the day as $v_i \in \mathbb{R}^{1\times C}$, which is a fixed-length vector. And vectors that continually change over time constitute our target time series $V = [v_1, v_2, \ldots, v_C] \in \mathbb{R}^{T\times C}$, which concentrates the behavior profiles in ascending order over time. Here, C is the number of behaviors and T is the number of observed days. In addition, we used linear interpolation to complete data for missing values. The corresponding procedures in Algorithm 1 are Step 7–12.

3.3. Pregnancy analysis

This module intends to determine the pregnancy status of giant pandas based on temporal deep learning models. As shown in Fig. 3, our system employs multi-LSTM and ECA based neural networks for feature extraction, and a classification or regression head to output analysis results.

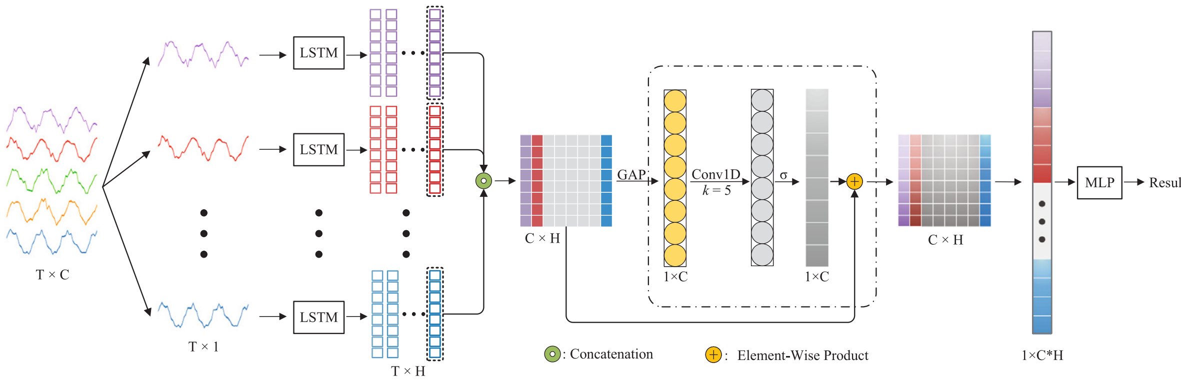

Fig. 3. The designed network structure. The pregnancy analysis module exploits multi-LSTM to extract features of different behaviors, followed the by the ECA based block to integrate various features, and finally utilizes a classification/prediction head to output the pregnancy analysis results.

Algorithm 1 Hierarchical Behavioral Time-series Construction

Stage 1: Behavior Recognition Network Training
Input: $\boldsymbol{X}^{\text{clips}} \in \mathbb{R}^{W\times H\times T}, \boldsymbol{Y}^{\text{clips}} \in \mathbb{R}^{T}$
Output: $\psi(\cdot;\boldsymbol{\Theta}_{sf})$
Randomly initialize $\boldsymbol{\Theta}_{sf}$
repeat
  Randomly sample one batch of labeled video clips.
  Compute the loss $\mathcal{L}_{sf}$ and update weight matrices of $\psi(\cdot;\boldsymbol{\Theta}_{sf})$
until converge
return $\psi(\cdot;\boldsymbol{\Theta}_{sf})$
Stage 2: Inference and Construction
Input: $\boldsymbol{X}^{\text{unlabeled}} \in \mathbb{R}^{W\times H\times T}$
Output: $V \in \mathbb{R}^{C\times T}$
Randomly initialize $\boldsymbol{\Theta}_{g}$ and load $\boldsymbol{\Theta}_{e}$ from the pre-trained network $\psi(\cdot;\boldsymbol{\Theta}_{e})$
repeat
  Sequentially extracting video clips for inputting to the trained network $\psi(\cdot;\boldsymbol{\Theta}_{e})$ yields hierarchical behavioral representation
until all video have been recognized
Calculate the ratio of the cumulative duration of each type of behavior per day to the total observation duration, resulting in $V \in \mathbb{R}^{C\times T}$.
return $V \in \mathbb{R}^{C\times T}$

Multi-LSTM. The multivariate behavioral time series V consists of multiple univariate behavioral time series, all with T time points. We use C LSTM units to extract features from each univariate time series, respectively. Benefiting from these units, the input time series could be with variable lengths. To augment the training data, we design a scalable sliding window mechanism to generate expanded inputs with variable lengths (detailedly introduced in Section 4.1). The LSTM units output multiple representations for the corresponding univariate behavioral time series, respectively. For each LSTM and univariate time series $v_j \in \mathbb{R}^{1\times T}$, we have

$$f_t = \sigma\left(W_f \cdot \left[h_{j_{t-1}}; v_{j_t}\right] + b_f\right), \tag{1}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{j_{t-1}}; v_{j_t}\right] + b_i\right), \tag{2}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{j_{t-1}}; v_{j_t}\right] + b_o\right), \tag{3}$$

$$cS_t = f_t * cs_{t-1} + i_t * \tanh\left(W_{cs} \cdot \left[h_{j_{t-1}}; v_{j_t}\right] + b_{cs}\right), \tag{4}$$

$$h_{j_t} = o_t * \tanh\left(cS_t\right), \tag{5}$$

where $h_{j_t} \in \mathbb{R}^{1\times H}$ represents the hidden features of jth LSTM network at the time point t. f_t, i_t and o_t are the forget, input and output gates. σ and $tanh$ are the sigmoid and hyperbolic tangent function. $\cdot$ and $*$ represent the Matmul product and Hadamard product, respectively. W and b are the weight matrix and bias. To get a fixed-length representation, we only leverage the last hidden feature of each LSTM as output. These outputs are then concentrated into a high-level behavioral representation $D = [d_1, \dots, d_j, \dots, d_C] \in \mathbb{R}^{\text{C}\times\text{H}}$. Here $d_j = h_{j_{T-1}}$ is the jth behavioral representation vector, and H is the length of the hidden unit in LSTM.

ECA network. Efficient Channel Attention (ECA) (Feichtenhofer et al., 2020) based network is designed to capture the interaction relationship among different behavioral time series. As shown in Fig. 3, the ECA based network firstly uses global average pooling (GAP) to obtain a sketch of diverse behaviors. Then it is leveraged to generate channel weights by performing an 1D convolution along with an activation function of sigmoid. Finally, the aforementioned output are multiplied the origin features D to generate $D' \in \mathbb{R}^{\text{C}\times\text{H}}$. And the calculation can be formulated as

$$D' = \sigma\left(Conv1D_k\left(Concat\left(g\left(d_j\right)\right)\right)\right) \cdot D, \tag{6}$$

where $g(d_j) = \frac{1}{H}\sum_{i=1}^{H} d_{j_i}$ is the GAP operation, $Concat()$ represents the concatenation operation, σ is a sigmoid function, and $Conv1D_k$ indicates $1D$ convolution with kernel size set to k. The module could efficiently learn the importance of characteristic behaviors with respect to the target classification or regression task while requiring much less computational costs compared to the global attention mechanisms.

Head. The classification or regression head is used to output the analysis results about the pregnancy status. The head is simply implemented by a two-layer fully connected network with an activation function of sigmoid activation in the hidden layer.

To sum up, the first part of our designed neural network uses multi-LSTM units to perform feature extraction on each of the univariate behavioral time series $V \in \mathbb{R}^{C\times T}$, respectively. It generate a high-level feature representation of each behavior. Then, the ECA based network, an efficient channel attention-based model is employed to adaptively learn the relationship among the feature representations of different behaviors as well as their corresponding importance. Finally, the output of the ECA based network is fed into the classification head (referred to as $\phi(\cdot;\boldsymbol{\Theta}_{dg})$ as a whole) or regression head (referred to as $\phi(\cdot;\boldsymbol{\Theta}_{pd})$ as a whole) towards the target task, performing pregnancy diagnosis or EDD prediction for giant pandas, respectively. The corresponding training procedures are in Algorithm 2.

Algorithm 2 Pregnancy Analysis

```
Training 1: Pregnancy Diagnosis Network
Input: V ∈ ℝ^{C×T}, Y^{dg} ∈ {0,1}
Output: φ(·; Θ_dg)
Randomly initialize Θ_dg
repeat
    Randomly sample one batch of time series data.
    Compute the loss L_dg and update weight matrices of φ(·; Θ_dg)
until converge
return φ(·; Θ_dg)
Training 2: EDD Prediction Network
Input: V ∈ ℝ^{C×T}, Y^{pd} ∈ ℝ
Output: φ(·; Θ_pd)
Randomly initialize Θ_pd
repeat
    Randomly sample one batch of time series data of pregnant giant
    pandas only.
    Compute the loss L_pd and update weight matrices of φ(·; Θ_pd)
until converge
return φ(·; Θ_pd)
```

4. Experiments

In this section, we conduct various experiments to evaluate the performance of the proposed system. Firstly, we introduce the experimental setup and the evaluated metrics used in the experiments. Then, we evaluate the performance of the proposed method, comparing to those of other existing approaches. Thirdly, we conduct ablative analysis to demonstrate the effectiveness of the designed strategies and components in the proposed system. Also, we analyze the sensitivity of parameters. Furthermore, we perform experiments on the four public datasets to validate the effectiveness of the proposed classification or regression model involved. Our code will be available at https://github.com/lxg199788/PAGP.

4.1. Setup

In this subsection, We will introduce the experiment setup, including the designed scalable sliding window, sample description, dataset preparation and competitive methods.

Scalable sliding window. Most existing methods analyze time series through sliding a fixed-length window to generate training samples for neural networks. However, our longest observation time in the dataset is no more than 30 days. Thus, generating enough samples for training is difficult by using a fixed-length window, even if employing a window of short time duration. To address this issue, we design the scalable sliding window for data augmentation. Specifically, for a time series $x \in R^{M\times N}$, where M is the number of metrics, and N is the number of observation length, we set a minimal window length l_{min} and a maximal window length l_{max} that equals to N in this paper, and then use windows of different lengths from l_{min} to l_{max} to slide in the time series from the beginning to the end for generating samples. For example, in an epoch with a window length W, for a window of length W, we could obtain $N-W+1$ samples as $[x_1,\dots,x_W]$, $[x_2,\dots,x_{W+1}]$, …, $[x_{N-W+1},\dots,X_N]$. This mechanism will enrich the samples for training greatly. From the point of the observer, the aforementioned strategy outputs samples with different lengths of historical observations, while the fixed window mechanism could only generate samples with historical memory of a certain time duration.

Sample description. We utilize the scalable sliding window to construct the input of our pregnancy analysis models with two kinds of labels for the target tasks, respectively. For pregnancy diagnosis, we label the real pregnant giant pandas' samples as positive and the pseudo-pregnant ones as negative. For the EDD prediction, we use the number of days before the actual delivery date (ADD) as hypothesis output.

Dataset preparation. For the pregnancy diagnosis experiments, after removing videos with poor quality, there are only ten giant pandas' data available, including three pregnant pandas (positive data) and seven pseudo-pregnant pandas (negative data). We randomly select from 3 times by stratified sampling as the training dataset and treat the rest to be the testing dataset while cross-validation is also performed. Besides, we use the over-sampling technology to balance the positive and negative samples in the dataset. For the experiments of the EDD prediction, we can only use the video data of three pregnant pandas. We also split 2/3 pandas' data into the training dataset and the remaining part as the testing dataset and performing for 3 times.

Baseline. we include six compared methods as baseline methods with the classification and regression head, respectively. They are based on different machine learning or deep learning models, including support vector machine (SVM) (Cortes & Vapnik, 1995), random forest (Breiman, 2001), multilayer perceptron (MLP) (Murtagh, 1991), long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997), Transformer (Vaswani et al., 2017) and CowTrans (Lin et al., 2023). CowTrans is a similar work in which it is utilized to predict the probability of pregnancy loss in cows. And it's head is also adapted to accommodate both pregnancy diagnosis and EDD prediction tasks. The Transformer, LSTM, CowTrans and the proposed method use the scalable sliding window mechanism for data augmentation. In contrast, the other methods, SVM, Random Forest, and MLP, use the fixed sliding window because they cannot deal with variable-length sequences.

Parameter selection and loss function. For SVM or SVR, we choose the kernel as "rbf" or "linear", and simultaneously perform a simple parameter search for penalty coefficients using 0.5, 1, 2, and 5. For Random Forest, we perform a search for the number of leaf nodes from 20, 40, 60, 80, 100. In the pregnancy diagnosis task, deep learning models select the cross-entropy loss function; for the EDD prediction task, the Mean Squared Error (MSE) loss function is used.

4.2. Performance metrics

We used seven metrics to evaluate the performance of the proposed methods in pregnancy diagnosis and the EDD prediction. Pregnancy diagnosis is a classification task, and four metrics are employed for evaluating it: including $Accuracy$, $Precision$, $Recall$, and $F1\_score$. $Accuracy$ represents the proportion of samples with accurate classifications in the entire instances, $Precision$ represents the proportion of actual pregnancy samples in the samples diagnosed as pregnancy by the system, while $Recall$ denotes the proportion of actual pregnancy samples judged as positive by the system in all actual pregnancy samples. $F1\_score$ is the harmonic mean of $Precision$ and $Recall$. The definitions of the four metrics are as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{7}$$

$$Precision = \frac{TP}{TP+FP}, \tag{8}$$

$$Recall = \frac{TP}{TP+FN}, \tag{9}$$

$$F1\_score = \frac{2Precision \times Recall}{Precision + Recall}, \tag{10}$$

where TP (True Positive) represents the correctly predicted pregnant panda samples, FP (False Positive) represents the incorrectly predicted samples that actually belongs to the pseudo-pregnant pandas, TN (True

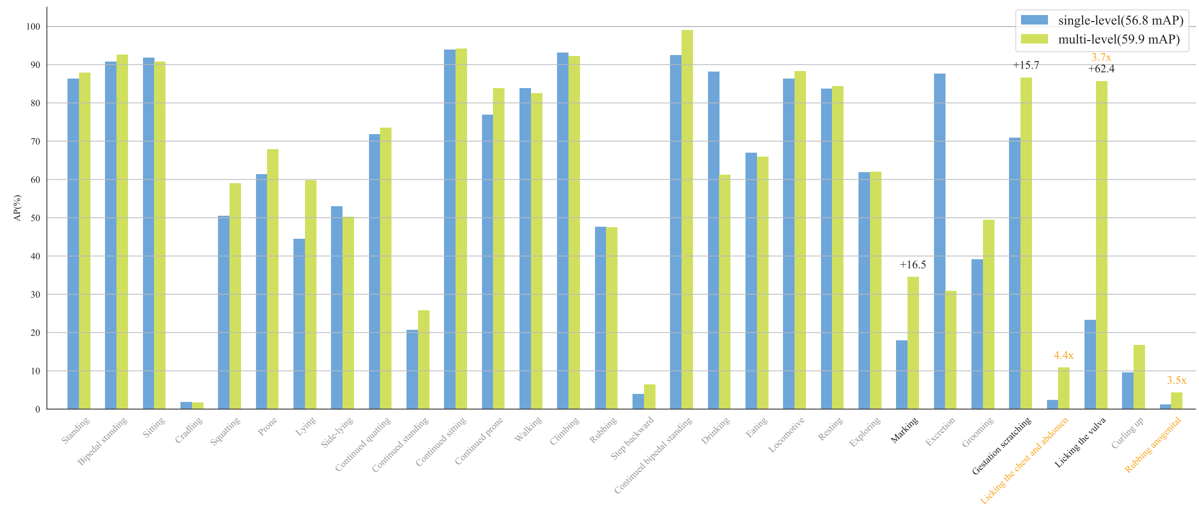

Fig. 4. Average precision for each category of behaviors. 'single-level': training one model for each level of behaviors, respectively. 'multi-level': training s single model for all levels of behaviors. The highlighted categories are the top 3 highest absolute increments (**black**) or the top 3 highest relative increments compared to the result of corresponding single-level strategy (**orange**). Categories are sorted by the number of examples, and both groups of experiments have the same hyper-parameters except the training strategy with single or multiple of behaviors simultaneously.

Table 2
Comparison with other methods on pregnancy diagnosis.

Methods	Accuracy	Precision	Recall	F1_score
SVM	0.707	0.764	0.605	0.673
Random Forest	0.743	0.879	0.565	0.685
MLP	0.737	0.743	0.733	0.732
LSTM	0.840	0.882	0.789	0.829
Transformer	0.874	0.861	0.896	0.877
CowTrans	0.882	0.832	0.909	0.868
Ours	**0.915**	**0.892**	**0.946**	**0.918**

Table 3
Comparison with other methods on the EDD prediction.

Methods	MAE	RMSE	R2
SVR	0.744	0.806	0.020
Random Forest	0.680	0.815	0
MLP	0.678	0.801	0.036
LSTM	**0.510**	**0.671**	**0.303**
Transformer	0.525	0.671	0.186
CowTrans	0.501	0.632	0.203
Ours	0.579	0.727	0.063

Negative) represents the correctly predicted pseudo-pregnant panda samples, FN (False Negative) represents the mispredicted samples that actually belongs to pregnant ones.

Three metrics including MAE (Mean Absolute Error), $RMSE$ (Root Mean Square Error), and $R2$ are employed for evaluating the EDD prediction in the experiments. MAE and $RMSE$ represent two kinds of differences between the model-predicted EDD and the actual delivery date while $R2$ represents the ratio of the sum of squares regression for the model-predicted EDDs to the sum of squares of the deviations from the average estimation. Their calculation formulas are as follows:

$$MAE = \frac{1}{n}\sum_{1}^{n}|\hat{y}_i - y_i|, \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}(\hat{y}_i - y_i)^2}, \tag{12}$$

$$R2 = 1 - \frac{\sum_{1}^{n}(\hat{y}_i - y_i)^2}{\sum_{1}^{n}(\overline{y} - y_i)^2}, \tag{13}$$

where $\hat{y}_i$ represents the predicted EDD for the ith sample by the system, y_i represents the actual delivery date of the corresponding giant panda for the ith sample, $\overline{y}$ represents the average value of actual EDDs for all the samples, and n is the number of the entire samples.

4.3. Experiment results with competitive methods

4.3.1. Pregnancy diagnosis

The comparison between the proposed method and the other baseline methods is shown in Table 2. It is apparent that the performance of our proposed method is better than those of the other six competitive methods on all the four metrics. In addition, we obtain two observations. (1) The deep learning based methods outperform the conventional based machine learning methods. Compared with machine learning based methods, deep learning models have excellent learning ability on complex tasks, which can learn the inherent connection between behavior features and pregnancy. (2) Three sequential models perform significantly better than the non-sequential models in terms of ACC and F1_score. It is because that sequential models can learn temporal dependencies and also be benefited from the scalable sliding window mechanism for data augmentation, both contributing to the learning performance.

4.3.2. EDD prediction

The results of our method and the other six methods on the EDD prediction task are shown in Table 3. The performance of our method is better than those of three methods, *ie.*, SVM, Random Forest and MLP, but lower than the performance of the other methods, *ie.*, LSTM , CowTrans and Transformer. The LSTM model achieves the best performance. According to the analysis, the model's performance is lower than those of LSTM and Transformer due to the small data set size. We will use four public datasets in Section 4.6 to evaluate the performance of our model and other models again.

4.4. Ablative analysis

4.4.1. Comparison of multi-level and single-level strategies

In this part, we conduct experiments to compare the performance of the hierarchical behavioral recognition model trained by two different strategies, *ie.*, multi-level classification and single-level classification described in Section 3.2. We randomly chose 1082 of the annotated 1200 videos as the training dataset and 118 as the testing dataset to evaluate both strategies.

Table 4
Ablative experiments on pregnancy diagnosis and the EDD prediction.

ECA	Multi-LSTM	SSW	Pregnancy diagnosis				EDD prediction		
			Accuracy	Precision	Recall	F1_score	MAE	RMSE	R2
✓	✓	✓	**0.915**	**0.892**	**0.946**	**0.918**	0.579	0.727	0.063
	✓	✓	0.908	0.885	0.940	0.911	0.570	0.719	0.055
		✓	0.840	0.882	0.789	0.829	**0.510**	**0.671**	**0.303**
✓	✓		0.780	0.732	0.891	0.804	0.684	0.808	0.017

The results are shown in Fig. 4. If we train one model for each behavioral level separately, the recognition system achieves 56.8 mAP (mean Average Precision) and over 50 AP (Average Precision) in 19 behavioral categories. In comparison, the trained single model for recognizing behaviors of all levels by multi-level classification strategy improves the recognition performance in 21 of the 30 behavior categories with an increasement of 3.1 mAP to 59.9 mAP, which shows that the recognition performance of the multi-level strategy is superior to that of single-level strategy. We think that the latter case achieves a better performance because the multi-level strategy can facilitate the recognition by sharing information and complementing each other. In addition, the multi-level classification strategy shares the backbone and only needs to train one model, while the single-level classification strategy conducts four training process separately, significantly reducing the computational costs.

4.4.2. Ablative experiments of three components

In this part, we conduct several experiments to verify the effects of different components on the performance of our proposed method. Specially, we exclude the ECA network, the scalable sliding window mechanism, and multi-LSTM (requiring along with ECA network together) separately to explore their effectiveness following the same setting as described in Section 4.3.

The results are shown in Table 4. For both pregnancy diagnosis and the EDD prediction, the model's performance decreases slightly after removing the ECA network, while dropping significantly without the scalable sliding window mechanism. It indicates that the ECA module is functional, and the scalable sliding window is valuable. We also noticed that after removing the multi-LSTM module, the method's performance decreases on pregnancy diagnosis but improves on the EDD prediction. Considering that we can only use the data of 3 pregnant giant pandas with a duration of 7 days in the EDD prediction test, the model may be over-fitted. Thus, our method and the Transformer model, which have a large number of parameters, are not as good as a simple LSTM model. It shows that multi-LSTM may not be effective on the dataset with extremely limited training samples.

4.5. Parametric analysis

In this subsection, we conduct experiments to explore the influence of the hyper-parameters, *ie.*, the length of scalable sliding window , on pregnancy diagnosis. The experimental results are shown in Table 5. However, there are tiny gaps in three of four indicators when the window length is six, comparing the results with respect to a window length of five. Generally, as the window length increases, the model's performance improves, which is also in line with our intuition that the more comprehensive behavioral time series data seen by the model, the more accurate the judgment will be. Thus, we think that the length of scalable sliding window is appropriate to be set to 5. Due to the small number of data samples that can be used for the date of birth prediction experiment, we will use four public datasets to conduct experiments in Section 4.6.

Table 5
Experimental results to evaluate the effectiveness with different length of the scalable sliding window.

Days	Accuracy	Precision	Recall	F1_score
1	0.739	0.763	0.691	0.725
2	0.766	0.790	0.724	0.756
3	0.810	0.832	0.776	0.803
4	0.885	0.876	0.896	0.886
5	**0.915**	0.892	**0.946**	**0.918**
6	0.914	**0.904**	0.927	0.915

Table 6
Description for the four public datasets used in the experiments.

Subset	Train	Test	Length	Classes	Dims
Cricket	108	72	1197	12	6
Epilepsy	137	138	206	4	3
AWR	275	268	144	25	9
MSL	12	12	2158	12	25

4.6. Evaluation on public datasets

In this subsection, we conduct experiments to compare the performance of our model with the compared models on public datasets since our datasets are relatively small, especially for the EDD prediction. The UEA datasets (Bagnall et al., 2018) are public multivariate time series datasets consisting of 30 different types of multivariate time series subsets. Considering the diversity with the length and the dimension of the time series, we selected three subsets: Cricket, Epilepsy and AWR (Articulary Word Recognition) from those. MSL (Mars Science Laboratory rover) (Hundman, Constantinou, Laporte, Colwell, & Soderstrom, 2018) is another well-known multivariate time series datasets, we cut a portion from it for experiments since the original datasets are large. Details are shown in Table 6.

Experiment design. To eliminate the negative impact of the extremely limited training samples in our real world dataset, we conduct experiments on the four public datasets against other competitive methods. Among them, we only compare our proposed method with MLP, LSTM, Transformer and CowTrans in the prediction experiments since SVM and Random Forest (denoted as RF) are inconvenient to predict vectors. We set the category to which each time series belongs as the label for the classification task and set the vector of the next time point as the predicted value for the prediction task. The training dataset and the testing dataset are split according to those datasets' default settings instead of the cross-validation above. Note that the experiments run without the scalable sliding window mechanism because the data sample is sufficient. When employing the fixed-length mechanism, the step is pinned to 20, and the length of the sliding window is set to one of the following values: 20, 40, 60, 80, and 100.

Result analysis. The results are shown in Fig. 5. In the classification experiments, the proposed model along with MLP, LSTM, Transformer and CowTrans, achieves 1.0 F1_score under different sliding window lengths on the two subsets, which is significantly higher than the performance of SVM and Random Forest. On the AWR subset, the superiority of the deep learning method is not obvious enough, due to the short length and insufficient sample diversity, but our method still performs well. And it shows satisfactory performance on the remaining

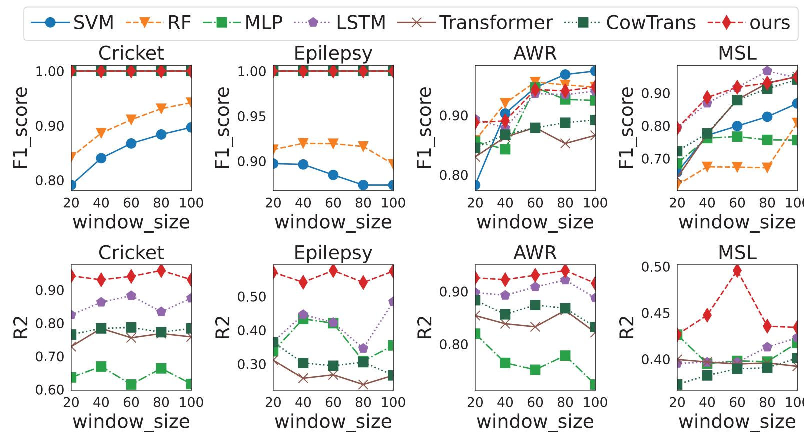

Fig. 5. Comparison with the competitive methods on four public datasets. The upper figures are the results of the classification experiments while the lower ones are about the prediction tasks. Our method performs well on both tasks.

Table 7
Time consumption for training and inference in a single iteration.

	Data construction		Pregnancy analysis	
	Training (bs = 6)	Inference (bs = 1)	Training (bs = 128)	Inference (bs = 128)
Time (s)	1.18	0.45	2.03	1.87

MSL dataset. And it is worth noting that our approach has consistently outperformed cowTrans. In the prediction experiments, the proposed model performs the best on the all four datasets. These experimental results indicate that our model exhibits favorable performance well in both classification and prediction tasks.

4.7. Time complexity

In this subsection, an evaluation of the time complexity is conducted. In the hierarchical behavioral time-series construction part, the behavior recognition network with the object detection network can be considered as a whole. The time complexity for both training and inference is directly proportional to the video's duration T_v, yielding a complexity of $O(T_v)$. In the pregnancy analysis part, the Multi-LSTM module introduces a computation complexity that scales linearly with the number of channels C in the time series. The designed data augmentation method SSW leads to a quadratic increase in time complexity with respect to the length T_t of the time series. Hence, the time complexity for this part is $O(CT_t^2)$. However, SSW is only employed when the time series length is insufficient. When the dataset is ample, and SSW is not used, the time complexity reduces to $O(CT_t)$, making it manageable. The training and inference times for both parts in a single iteration are summarized in Table 7. According to the table, the training or inference time for one iteration of those models is in the range of seconds. We completed the training of the behavior recognition network in about 5 h, and the training of the pregnancy diagnosis network in about 2 h, which are acceptable. And the experimental platform consists of an Intel i7-9700 CPU and two Nvidia 1080Ti GPUs.

5. Discussion

Due to the variability in the behavior of giant pandas, we have imposed the following three limitations to ensure the reliability of the experiments: (1) Only giant pandas that have undergone artificial insemination, thus ensuring their potential for pregnancy, were included in the pregnancy prediction model used in the experiments. (2) The sliding window time of the pregnancy prediction model should not be less than one day, as accurate prediction results require a longer observation period. (3) The data used in the experiment were manually annotated, which introduces the possibility of some labeling errors.

We also noticed problems such as the limited number of giant pandas and few observed days in the dataset. These problems are subject to experimental conditions. Shortly the giant panda is a rare animal with limited observable objects, which brings specific difficulties to data analysis. Thus, how to improve the scale and quality of data is a significant point in the next step. Apart from this, the recognition accuracy still needs to be higher in hierarchical behavior recognition, especially for some behavior categories with insufficient labels. If the recognition accuracy of the model can be further improved, these downstream tasks, **e.g.**, pregnancy diagnosis and the EDD prediction, will benefit from it.

6. Conclusion

In this paper, we analyze the giant pandas' pregnancy status from a new perspective, *ie.*, utilizing sequence models to perform pregnancy analysis via giant pandas' hierarchical behavioral information from video data. Specifically, we (1) classify the behaviors of giant pandas into different levels and define typical actions at each level to obtain a more detailed and accurate description; (2) collected more than 1700 h of videos from 13 giant pandas (4 pregnant) at the Chengdu Giant Panda Breeding Base, and labeled 1200 30-second video clips based on the aforementioned hierarchical behavioral definitions; (3) designed and implemented a giant pandas' pregnancy analysis system, which can automatically recognize hierarchical behaviors from videos to construct hierarchical behavioral time-series data, and judge whether the giant panda is pregnant or predict EDD. Besides, we have also shown the effectiveness of the system through experiments. The ACC, Precision, Recall and F1_score of pregnancy diagnosis is 0.915, 0.892, 0.946 and 0.918, respectively; and the RMSE, MAE and R2 for the EDD prediction is 0.579, 0.727 and 0.063, respectively. The experimental results demonstrates that our system has remarkable performance on the pregnancy diagnosis task and could perform well on the EDD prediction task.

In the future work, we will address the following considerations: (1) Manually annotating the data is costly and prone to labeling errors. Therefore, we will explore the use of unsupervised models for giant panda behavior recognition. (2) Achieving faster and real-time detection of giant panda behavior would significantly enhance our monitoring efforts. Therefore, our future goal is to reduce the time complexity of the model. (3) While our current model focuses on predicting pregnancy in individual panda videos, we will also investigate the problem of predicting pregnancy in multi-panda videos.

Table A.1
Hierarchical behavioral definition of the giant pandas.

Viewpoint	Label	Description
Posture	Standing	Animal bears the bodyweight with four upright limbs.
	Bipedal standing	Animal bears the bodyweight with hind limbs. Forelimbs leave the ground.
	Sitting	Hips contact the ground directly; hind limbs stretch forward, forelimbs prop up on the ground or off the ground, tree trunks, or other supports, in general, forelimbs hang naturally when the animal is resting.
	Cradling	In sitting posture, animal hold the target body in the arms by the forelimbs.
	Squatting	Animal half stand with hips tilted down, close to the ground. Forelimbs grasp the supports.
	Prone	Head, chest, and abdomen facing down, chest and abdomen touching the ground, tree poles, or other supports. Forelimbs stretch forward, curled under the head, chest, or hanging. Hind limbs stretch backward, curled or hanging.
	Handstand	Upright forelimbs propped up on the ground or the supports. The unilateral or bilateral hind limbs lift upward to contact the wall or trees, or other supports.
	Lying	Back or bodyside contact with the ground.
	Side-lying	A side of body contact with the ground and back does not contact, with or without supports.
Action	Continued lying	Back or bodyside contact with the ground for a period of time.
	Continued standing	Animal bears the bodyweight with four upright limbs for a period of time.
	Continued sitting	For a period of time, hips contacts the ground directly, hind limbs stretch forward, forelimbs prop up on the ground or off the ground, tree trunks, or other supports.
	Continued quatting	For a period of time, animal half stand with hips tilted down, close to the ground for a period of time. Forelimbs grasp the supports.
	Continued prone	For a period of time, head, chest, and abdomen facing down, chest and abdomen touching the ground, tree poles, or other supports. Forelimbs stretch forward, curled under the head, chest, or hanging. Hind limbs stretch backward, curled or hanging.
	Continued bipedal standing	Bear the bodyweight with hind limbs. Forelimbs leave the ground or other supports for a while.
	Licking	Animal constantly contacts the other part of the body or objects with the tongue.
	Running	Animal moves forward rapidly by staggering the fore and hind limbs from side to side.
	Walking	Animal moves forward slowly by staggering the fore and hind limbs from side to side.
	Climbing	Animal staggers up and down with the fore and hind limbs, moving on supports.
	Turning	Animal changes direction following the fore and hind limbs and hips.
	Rubbing	Animal rubs walls, protrusions, or the ground in a circular or straight line.
	Rolling	Animal lies on the ground and rolling around.
	Step backward	Animal moves backward by staggering the fore and hind limbs from side to side.
Activity	Drinking	Animal uses the mouth to suck water from a drinking basin (tank).
	Eating	Animal processes and swallows food with the mouth and paws.
	Locomotive	Animal moves to other sites, including walking, running, climbing.
	Resting	Animal lies or sits in one place, with occasional changes of position.
	Exploring	Animal checks object with mouth, nose or paws, standing and gazing.
	Playing	Animal acts for no apparent purpose. Such as rolling, games, and fiddling with various objects.
	Marking	Animal smears with perianal glands, vulva, feces, urine on objects.
	Excretion	Defecation of stool or urine.
	Grooming	Animal scratches or licks fur with the claws or mouth, or rubbing.
Female behaviors during pregnancy	Wood-biting	Animal gnaws on the wooden perches.
	Food refusal	Animal refuses to collect food and reduced food intake.
	Gestation scratching	Animal scratches the ground or other objects in a standing or other posture.
	Nesting	Animal articulate bamboo or pieces of wood that have fallen from wooden perches, placing the objects somewhere in the enclosure to make a nest.
	Licking the chest and abdomen	Animal sits on the ground and licks the chest and abdomen.
	Licking the vulva	Animal sits on the ground and licks the vulva.
	Abdominal contraction	Abdomen contract in sitting, standing, or side-lying posture.
	Delivery	Animal delivers the cub in sitting, standing, or side-lying posture.
	Licking the amniotic fluid	Animal licks the amniotic fluid after delivery.
	Lick the placenta	Animal licks the placenta after delivery.
	Lick the caul	Animal licks the caul after delivery.
	Curling up	Animal curl up the head and limbs.
	Lick the lochia	Animal licks the lochia after delivery.
	Rubbing anogenital	Animal rubs the perianal gland area on walls, protrusion, or the ground.

CRediT authorship contribution statement

Xianggang Li: Writing – original draft, Investigation, Methodology, Software, Data curation. **Jing Wu:** Writing – original draft, Methodology, Validation. **Rong Hou:** Methodology, Writing – review & editing, Project administration. **Zhangyu Zhou:** Investigation, Methodology, Writing – review & editing. **Chang Duan:** Formal analysis, Methodology, Writing – review & editing. **Peng Liu:** Data curation, Formal analysis. **Mengnan He:** Data curation, Investigation. **Yingjie Zhou:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Peng Chen:** Conceptualization, Validation, Funding acquisition, Writing – review & editing. **Ce Zhu:** Writing, Investigation, Review & editing, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Peng Chen, Yingjie Zhou, Rong Hou, Chang Duan, Ce Zhu, Zhangyu Zhou, Xianggang Li, Jing Wu, Peng Liu, Mengnan He has patent #CN202211653700.1 pending to Peng Chen, Yingjie Zhou, Rong Hou, Chang Duan, Ce Zhu, Zhangyu Zhou, Xianggang Li, Jing Wu, Peng Liu, Mengnan He, Pengcheng Wu, Qingyue Min.

Data availability

The authors do not have permission to share data.

Acknowledgments

We would like to thank Pengcheng Wu at Chengdu Research Base of Giant Panda Breeding, Anheng Xie and Xianda Pu at Southwest Petroleum University, and Wei Lai at Sichuan University for collecting and labeling datasets. This research is Supported by Sichuan Science and Technology Program (2023NSFSC1926; 2022NSFSC0020), Chengdu Science and Technology Program (2022-YF09-00019-SN), and the Chengdu Research Base of Giant Panda Breeding (NO. 2021KCPB-03).

Ethics approval and consent to participate

The methods, the use of materials and all experimental procedures involving animals were approved by the Institutional Animal Care and Use Committee of the Chengdu Research Base of Giant Panda Breeding protocol #2018017. All methods were performed in accordance with the relevant guidelines and regulations under the Law of the People's Republic of China.

Appendix A. Hierarchical behavior definition for the giant pandas

See Table A.1.

References

Ai, Hongfu, Pan, He, & Li, Yingbin (2017). Research on optimization of PM2.5 content prediction in air haze. *Computer Integrated Manufacturing Systems*, *34*, 392–396.

Bagnall, Anthony, Dau, Hoang Anh, Lines, Jason, Flynn, Michael, Large, James, Bostrom, Aaron, et al. (2018). The UEA multivariate time series classification archive, 2018. arXiv preprint arXiv:1811.00075.

Baum, Leonard E., & Petrie, Ted (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, *37*(6), 1554–1563.

Breiman, Leo (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Chaudhry, Rizwan, Ravichandran, Avinash, Hager, Gregory, & Vidal, René (2009). Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 1932–1939). IEEE.

Chen, Yushu, Yuhua, Chen, Yuanzhi, Chen, & Yueming, Lin (1985). Study on the reproductive biology of the giant panda. *Acta Theriologica Sinica*, *5*, 233–238.

Chen, Jinxin, Zhou, Jie, Liu, Longshen, Shu, Cuini, Shen, Mingxia, & Yao, Wen (2023). Sow farrowing early warning and supervision for embedded board implementations. *Sensors*, *23*(2), 727.

Cortes, Corinna, & Vapnik, Vladimir (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Dalal, Navneet, & Triggs, Bill (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Vol. 1* (pp. 886–893). Ieee.

Dalal, Navneet, Triggs, Bill, & Schmid, Cordelia (2006). Human detection using oriented histograms of flow and appearance. In *European conference on computer vision* (pp. 428–441). Springer.

Elías, García Ramírez, de la Vega, Ana Dolores Díaz, et al. (2022). Comparación de la actividad en cautiverio en una cohorte bianual del panda gigante (Ailuropoda melanoleuca), México. *Brazilian Journal of Animal and Environmental Research*, *5*(1), 797–812.

Feichtenhofer, Christoph, Fan, Haoqi, Malik, Jitendra, & He, Kaiming (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6202–6211).

Feichtenhofer, Christoph, Fan, Haoqi, Malik, Jitendra, & He, Kaiming (2020). ECA-Net: efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 11534–11542).

Gandia, Kristine M., Herrelko, Elizabeth S., Kessler, Sharon E., & Buchanan-Smith, Hannah M. (2023). Understanding circadian and circannual behavioral cycles of captive giant pandas (ailuropoda melanoleuca) can help to promote good welfare. *Animals*, *13*(15), 2401.

Ghosh, Sushmito, & Reilly, Douglas L. (1994). Credit card fraud detection with a neural-network. In *System sciences, 1994. Proceedings of the twenty-seventh hawaii international conference on, Vol. 3* (pp. 621–630). IEEE.

Gowda, Shreyank N., Rohrbach, Marcus, & Sevilla-Lara, Laura (2021). Smart frame selection for action recognition. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 35* (pp. 1451–1459).

Hochreiter, Sepp, & Schmidhuber, Jürgen (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, Xiangming, Jianqiu, Yu, Xuanzhen, Liu, et al. (1999). Observation on behaviors associated with gestation and parturition in giant pandas. *Sichuan Journal of Zoology*, *18*, 179–181.

Hundman, Kyle, Constantinou, Valentino, Laporte, Christopher, Colwell, Ian, & Soderstrom, Tom (2018). Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 387–395).

Ji, Shuiwang, Xu, Wei, Yang, Ming, & Yu, Kai (2012). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 221–231.

Jocher, Glenn, Chaurasia, Ayush, Stoken, Alex, Borovec, Jirka, NanoCode012, et al. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA realtime instance segmentation.

Kang, Dongwei (2022). A review of the habitat restoration of giant pandas from 2012 to 2021: Research topics and advances. *Science of the Total Environment*, *852*, Article 158207.

Kousenidis, Konstantinos, Kirtsanis, Georgios, Karageorgiou, Efstathia, & Tsiokos, Dimitrios (2022). Evaluation of a numerical, real-time ultrasound imaging model for the prediction of litter size in pregnant sows, with machine learning. *Animals*, *12*(15), 1948.

Li, Zhihan, Zhao, Youjian, Han, Jiaqi, Su, Ya, Jiao, Rui, Wen, Xidao, et al. (2021). Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 3220–3230).

Lin, Tsung-Yi, Goyal, Priya, Girshick, Ross, He, Kaiming, & Dollár, Piotr (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).

Lin, Dan, Kenéz, Ákos, McArt, Jessica A. A., & Li, Jun (2023). Transformer neural network to predict and interpret pregnancy loss from activity data in Holstein dairy cows. *Computers and Electronics in Agriculture*, *205*, Article 107638.

Liu, Juan, Yue, Chen, Liran, Guo, Bo, Gu, & Hui, Liu (2005). Stereotypic behavior in relation to fecal cortisol level in captive giant pandas. *Journal of Beijing Normal University (Natural Science)*, *41*, 75–78.

Liu, Weixin, et al. (1986). Pregnancy diagnosis of giant panda. *Acta Theriologica Sinica*, *4*, 317–318.

Luo, Li, Jian, Li, Zhihe, Zhang, & Rong, Hou (2011). The progesterrone change in parturition giant pandas. *Journal of Southwest University for Nationalities - Natural Science Edition*, *37*, 752–756.

Martis, Roshan Joy, Acharya, U. Rajendra, Mandana, K. M., Ray, Ajoy Kumar, & Chakraborty, Chandan (2012). Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. *Expert Systems with Applications*, *39*(14), 11792–11800.

Murtagh, Fionn (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5–6), 183–197.

Pan, Zaiyang, Zuofu, Xiang, Chao, Chen, & Peng, Chen (2015). Behavioural changes and activity rhythms of giant pandas during the gestation period. *Sichuan Journal of Zoology*, *34*, 817–823.

Ren, Shaoqing, He, Kaiming, Girshick, Ross, & Sun, Jian (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*.

Simonyan, Karen, & Zisserman, Andrew (2014). Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems*, *27*.

Swaisgood, R., Wang, D., & Wei, F. (2016). Ailuropoda melanoleuca. *The IUCN Red List of Threatened Species*, *2016*, e–T712A102080907.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Wang, Limin, Huang, Bingkun, Zhao, Zhiyu, Tong, Zhan, He, Yinan, Wang, Yi, et al. (2023). Videomae v2: Scaling video masked autoencoders with dual masking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14549–14560).

Wu, Wenhao, Wang, Xiaohan, Luo, Haipeng, Wang, Jingdong, Yang, Yi, & Ouyang, Wanli (2023). Bidirectional cross-modal knowledge exploration for video recognition with pre-trained vision-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6620–6630).

Yang, Yong, Jingbo, Jia, & Yingjie, Zhao (2019). Effects of human factors on panda behavior in captivity. *Journal of Northeast Forestry University*, *47*, 104–106.

Zhao, Pengpeng, Zhang, Qiang, et al. (2011). Variation of estradiol and progesterone levels in urine of female giant pandas during the breeding season. *Progress in Veterinary Medicine*, *32*, 45–48.

Zhou, Xingyu, & Ji, Lihong (2016). The judgement of actual-pregnant and pseudo-pregnant Ailuropoda melanoleuca. *Journal of Sichuan Forestry Science and Technology*, *37*, 91–93.

Zin, Thi Thi, Sumi, Kosuke, Seint, Pann Thinzar, Tin, Pyke, Kobayashi, Ikuo, & Horii, Yoichiro (2022). A stochastic modeling procedure for predicting the time of calving in cattle. *ICIC Express Letters, Part B: Applications*, *13*(1), 49–56.