Cattle Action Recognition with Multi-Viewpoint Cameras based on Deep Learning

Muhammad Fahad Nasir[1,2], Alvaro Fuentes[1,2][0000-0001-8847-1541], Shujie Han[1,2], Jongbin Park[1,2], Sook Yoon[3], and Dong Sun Park[1,2]

[1] Department of Electronics Engineering, Jeonbuk National University, Jeonju, 54896 South Korea
[2] Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju, 54896 South Korea
[3] Department of Computer Engineering, Mokpo National University, Muan, 58554 South Korea
dspark@jbnu.ac.kr; syoon@mokpo.ac.kr

Abstract. Cattle activity recognition is an essential element in monitoring cattle health. Cattle behavior is a primary indicator of its well-being, any anomalous behavior displayed by cattle is the earliest indication of illness, and if treated immediately it can prevent aggravation and development of diseases, which is detrimental to cattle's health. To study cattle activity, embedded devices have been used but they can be a cause of discomfort, and stress. This paper focuses on training a deep learning model to detect and localize multiple cattle in video frames, captured from multiple cameras at multiple angles during day and night with an overall precision of 95.3%. Our system provides information about cattle mobility, which can detect any inconsistent behavior exhibited by cattle, leading to early detection and prevention of disease. Furthermore, we share an in-depth analysis of the model's performance on our raw dataset and its effectiveness in recognizing individual, group, and part cattle behavior, and facilitating cattle health monitoring.

Keywords: deep learning, animal welfare, action recognition, video, camera.

1 Introduction

The growing population of the human race is challenging the cattle industry, with sheer intensity of demand for cattle products [3]. The agriculture industry needs to optimize itself to keep up, and this opened the doors for innovation, technology, and automation [5,12]. Hence, cattle welfare is of the utmost importance for maintaining ideal cattle produce. Cattle are susceptible to disease and illness, in general, diseases are detected

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2019R1A6A1A09031717); Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm R&D Foundation(KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration(RDA)(421044-04); and National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (2020R1A2C2013060).

too late when the spread has continued for too long causing severe illness, and often death [4]. However, the intensity of disease is preventable by monitoring cattle behavior meticulously [2]. Cattle tend to depict anomalous behavior when they are under duress, anxiety, or the influence of disease [6]. There are key behavioral indicators depicted which can help in the early detection and prevention of diseases [7].

Traditionally, cattle welfare has been carried out via manual checkups conducted by a farmer or physician [2]. This approach is laborious and time-consuming, and the quality of results is mainly dependent on the intuition, and expertise of the observer [3]. To improve the monitoring and observations of cattle behavior, a better alternative has been using the assistance of technology [12]. In this regard, the term Precision Livestock Farming (PLF) has been coined, for monitoring of every aspect of cattle activity using technology [1]. For instance, portable and wearable devices have been extensively adapted in cattle farming, but these technologies have been also called into question as they tend to be intrusive, stressful, and uncomfortable for cattle, causing cattle to deviate from normal behavior [11]. To overcome these challenges, our approach is to utilize a non-intrusive, highly efficient deep learning model, using RGB cameras to detect animal behaviors and contribute to animal welfare.

Our work extends the work of [1], in categorizing cattle activities into 15 respective hierarchical behaviors. The behaviors are sub-categorized into three categories, namely: i) Individual cow actions, ii) Group cow actions, and iii) Part cow actions. These respective categories are illustrated in Fig 1.

For this research work, we installed RGB cameras in the farms to gather videos for our dataset. In our specific application, we faced significant challenges such as 1) Camera viewpoint, when the entire cowshed is not visible, and some cows may move to blind spots [16]; 2) Illumination: at night time, the cow shed is not as well-lit and affects the visibility [16]; 3) Deformation: cow poses from some viewpoints are deformed and it is difficult to interpret the specific action [8]; 4) Occlusion: this is one of the biggest challenges, as either cattle itself, cow shed structure or blind-spots can hide considerable part of cattle's body, stressing the visibility for camera and challenging the model robustness [17]; 5) Background Clutter: combined with occlusion, a model can overlook some cattle and generate false negatives [8]. Also, during nighttime cattle standing far away cannot be differentiated from the background.

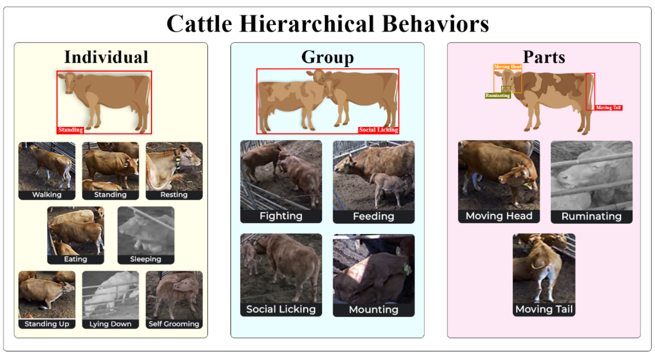


Fig. 1. Hierarchical behaviors depicted by Hanwoo cattle.

Our proposed solution is to use a deep learning model, on our dataset collected from multiple cowsheds, inhabiting 18 cows and some calves, from multi-viewpoint cameras, during day and night. The video recordings collected are converted to image frames and annotated using bounding boxes and classified into 15 different behavioral categories. For the robustness of our model, our dataset contains real-world challenges as stated above, this allows us to widen the applicability of our work in different farms. Our model performance for 90 percent of actions is more than 97% and above, and an overall accuracy of 95.3%.

Our main contributions in this paper are: i) A highly accurate deep learning-based model to detect cattle behavioral characteristics; ii) A diverse dataset, collected from different farms, with multiple cattle, from different RGB camera viewpoints and during the day and nighttime; and iii) An analysis and discussion on how these actions help detect a pattern for early detection of illness.

The remainder of the paper is structured as follows. We present the dataset and methodology in Section 2. Details of our experimentation and analysis are described in Section 3. We then conclude our findings in Section 4.

2 Methodology

Our work on cattle action recognition is aimed at studying Hanwoo, a cattle native to Korea, raised for its meat. Due to the mountainous geography of Korea, open cattle farming is not an option, so Hanwoo are raised indoors in cow sheds. Most research on Hanwoo are specific to just a few hierarchical behaviors such as ruminating [3], eating, and drinking [14], some are disease-oriented [15] or research conducted via labor work manually [2]. We propose a system that performs a comprehensive study of Hanwoo actions and discusses their impacts. Our system is composed of four stages namely, data collection, data preparation, annotations, and deep learning model as shown in Fig 2.

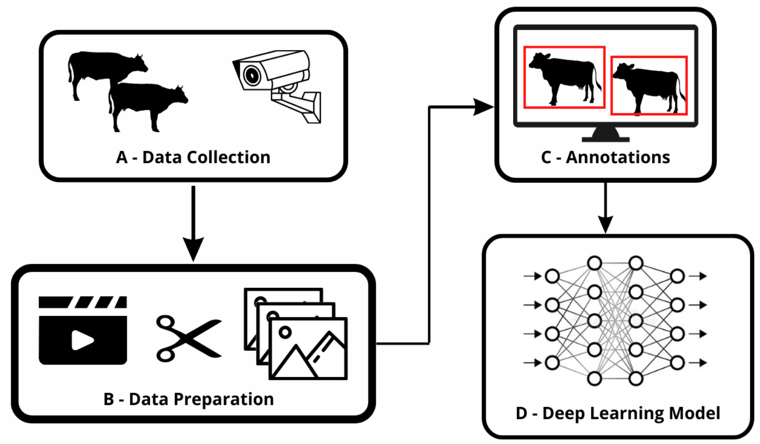


Fig. 2. Overview of the proposed system for Hanwoo cattle behavioral recognition.

2.1 Model Pipeline

For this study, the objective was to develop a framework that would aid in cattle welfare and thereby early detection and prevention of disease; another priority was non-intrusiveness. For these purposes, we installed RGB cameras in a cowshed to collect a diverse set of data, annotated it, and then proceeded to utilize that data in our model for training it. We share each step in detail below.

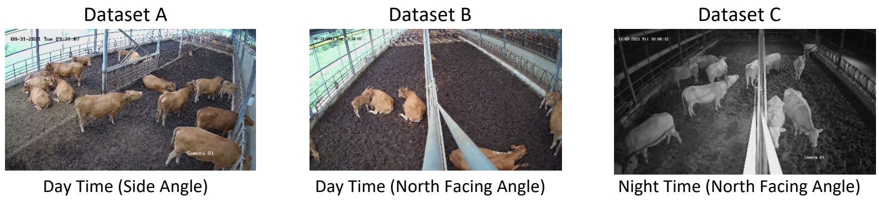


Fig. 3. Cattle sheds camera viewpoint for the dataset, of Hanwoo cattle.

A) Data Collection. To compose a dataset, we visited indoor Hanwoo cattle farms in South Korea. To get a realistic idea of the environment in Hanwoo farms, we shortlisted a cow shed with 18 cows and 3 calves of varying sizes and growth stages. We proceeded to install RGB cameras in the shed, at multiple angles. The dataset consists of RGB video frames of the Hanwoo cattle enclosure. We choose recording footage from three different cameras, some details are available in Fig 3.

B) Data Preparation. After the collection of data from the three cameras, we chose recording clips and form a unique dataset. The goal was to incorporate almost all types of actions performed by cattle as suggested in Fig 1. We compiled snippets of recordings from all three cameras and converted these recordings into image-level frames. We then proceeded to send the prepared data to the annotation stage.

C) Data Annotation. Image frames obtained were then labeled with each cow's actions. The categorization of these actions is 1) Individual Actions: it specifies action being performed by a single cow. It considers the entire cow and classifies it to a certain action as shown in Fig 1. 2) Group Actions: it specifies actions being performed by more than one cow. These actions represent cattle social interactivity with neighboring cattle, as shown in Fig 1. 3) Part Actions: they refer to the activity and movement of specific body parts of an individual cattle, including head movement, tail waggling, and ruminating, as shown in Fig 1.

The ground truth annotations were done manually by labeling every cow in the visible area. Table 1 showcases the instances of each of the 15 hierarchical actions being performed in our dataset.

D) Deep Learning Model. Our proposed approach is to perform action recognition of cattle farms. We faced numerous challenges as mentioned in Section 1, and multiple cattle (objects) to be detected. Since we aimed to create a system capable of detecting cattle activity in real time of multiple objects, we required a model operating in real-time, with high accuracy. We achieved this by using YOLOv5 [9], a single-stage object detector, capable of operating in real-time.

Table 1. Hanwoo cattle action instances. The values reflect the annotations.

Hierarchical category	Actions	Dataset A	Dataset B	Dataset C	Total
Individual	Walking	3,305	1,560	1,562	6,427
	Standing	25,585	10,853	25,044	61,482
	Resting	18,768	7,080	456	26,304
	Eating	-	-	-	-
	Sleeping	-	-	-	-
	Standing Up	243	284	-	527
	Lying Down	-	-	-	-
	Self Grooming	247	224	-	471
Group	Fighting	-	-	-	-
	Feeding	228	-	-	228
	Social Licking	-	-	-	-
	Mounting	-	-	-	-
Part	Moving Head	308	231	660	1,199
	Ruminating	5,430	1,241	4,161	10,832
	Moving Tail	3,505	1,998	16	5,519

A detailed overview of the proposed framework is depicted in Fig 4. We provide our annotated dataset to the model, it passes through the model backbone, where features are extracted from the image and a reduced spatial resolution of the respective feature maps is obtained. The feature maps are forwarded to the model neck which performs further processing by increasing the depth and reducing the spatial resolution. The results are sent to the model head where the final predictions for cattle actions are conducted.

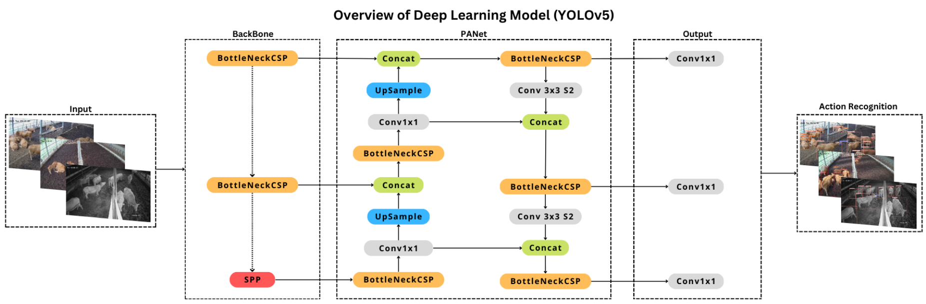


Fig. 4. The overall structure of the model and detailed insight into YOLOv5 architecture.

Our model provides us with the class of action detected and the confidence score of the accuracy of that prediction. During training, the loss function continues to optimize the model to overcome the discrepancy between predicted output and ground truth. In object detection, another factor of vital importance is the Intersection over

Union (IoU) of bounding boxes, as the model aims to find the exact coordinates of the bounding boxes, so the loss function equation can be written as:

$$Loss = L_{class} + L_{conf} + L_{CIoU} \tag{1}$$

where L_{class}, L_{conf}, and L_{CIoU} are the classification loss function, confidence loss function, and bounding box regression loss function, respectively. Details about each function are provided in [9].

3 Experimental Results

This section shares our model specifications, performance, and experiments on the cattle farm during different times of the day and from multiple viewpoints. We share detailed insights about the overall performance of our model, performance with independent actions, and how our system aids in cattle welfare.

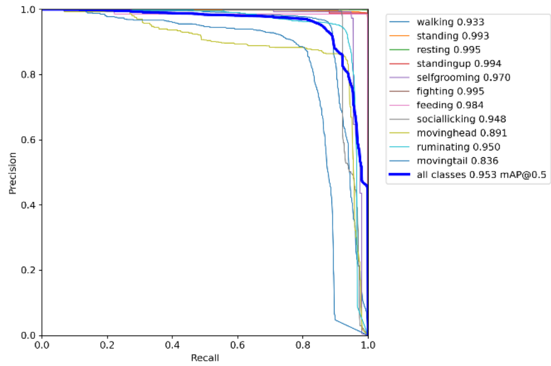
3.1 Implementation

We utilized the YOLOv5x model and train it on our dataset. To challenge our model, we made sure not to select temporally adjacent frames but rather built our test dataset by selecting random frames. These frames sometimes introduce part actions at slightly occlusive situations as well, this confronted the robustness of our model. We achieved the best results using a 70/30 train-test split of our dataset, and we trained the model for 300 epochs. Our server consisted of 2 NVIDIA TITAN V GPUs with 12 GBs each. The training time for the model was 34.7 hours with a batch size of 16. We also utilized several data augmentation techniques such as mosaic, copy-paste, albumentations, and random horizontal flip. Allowing our model to be able to deal with viewpoint, occlusion, and deformation challenges to increase our dataset for better performance of the model. These allow us to surge the dataset size and generate a robust model.

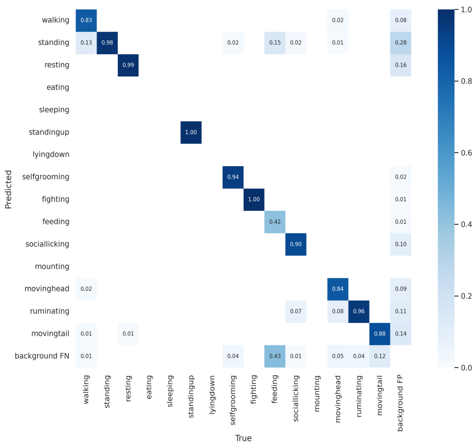
3.2 Quantitative Results

The metric for measuring the performance of object detectors is the mean average precision (mAP). For our model, we evaluated the performance with the mAP at an IoU threshold of 0.5, meaning the mapping of prediction and ground truth is computed at a 0.5 threshold. We adjusted the learning rate of the model at 0.01, whilst setting the momentum at 0.937 and a weight decay of $5\text{x}10^{-4}$. We achieved an overall mAP of 95.3%. Since our model was trained on the detection of 15 behavioral categories of cattle actions, individual accuracies can be seen observed in Fig 5a.

The confusion matrix provides an in-depth understanding of where the model makes incorrect detections, these are generally due to the vision challenges mentioned in the introduction section, and mainly due to transitionary actions. As shown in Fig 5b, an example of that is walking and standing, in video, these actions are quite clear due to the visible temporal information, but in images, when a cow starts walking there are instances where all of its limbs are touching the ground, but the cow is still in a walking instance and the model interprets to have stopped and incorrectly states it to be standing. Similarly, part actions such as moving head, ruminating, and moving tail have a few False Negatives (FN) and False Positives (FP) because of the non-availability of temporal information.



a) PR curves of our model



b) Confusion matrix of validation set

Fig. 5. Detailed analysis of model performance on our dataset.

3.3 Qualitative Results

We implemented our trained model on the video recordings of cattle farms and the results are showcased in Fig 6, along with the respective time duration. Our model provides bounding boxes along with action classification and a confidence score, with a threshold of 45% confidence during inference.

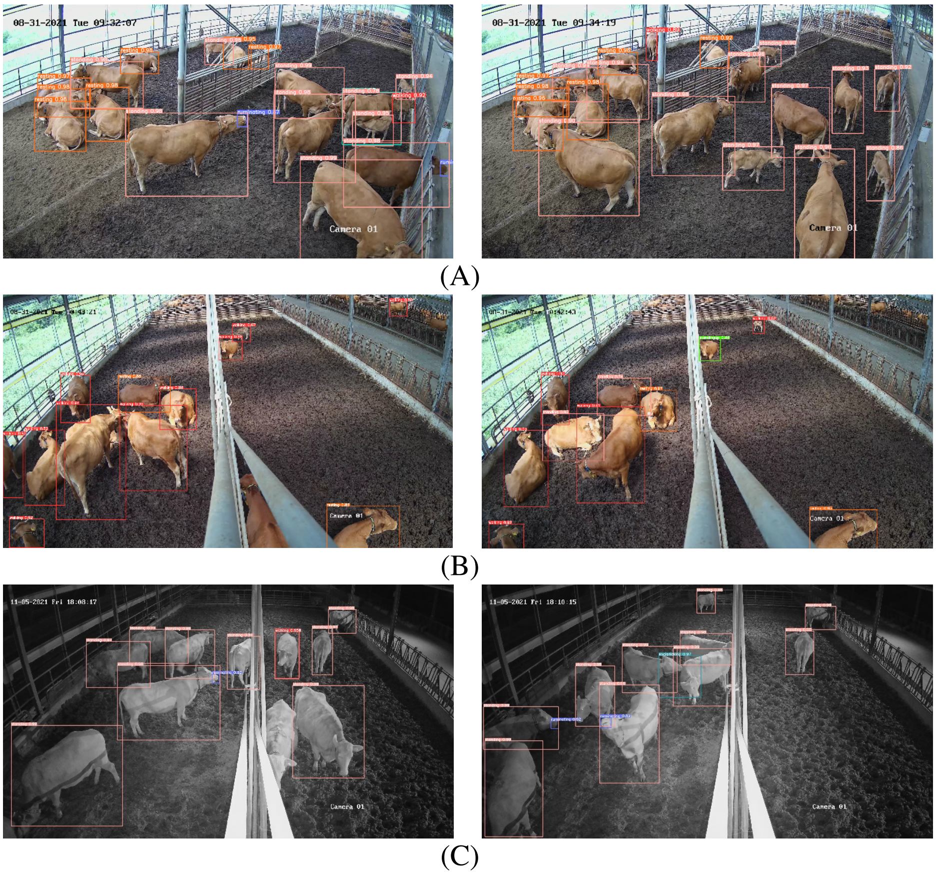


Fig. 6. The implementation of our deep learning model on video recordings from cattle farms during the day and nighttime. (A) Detection at a time interval of 09:00 – 10:00; (B) Detection during 10:00 – 11:00; (C) Results during nighttime 18:00 – 19:00. The bounding boxes inform us of the action being performed with a confidence score.

3.4 Cattle Welfare Analysis

For the results of our model, we observed that actions were performed with high accuracy. Using this, we can monitor and analyze individual cow actions and detect any anomalous behavioral patterns. Cow actions reflect animal health such as rumination time, as farmers monitor the rumination time to predict the calving period and oversee the process to avoid any complications [3]. Similarly, early indications of lameness can be observed by walking patterns and head movements [6,10]. Cows socializing less, spending more time resting or lying down, spending less time walking, and reduced eating and ruminating behavior are indications of a cow under stress or suffering from a disease [13]. In general, healthy cows ruminate more than cows suffering from disease [14]. If suddenly a cow's rumination time decreases, it is a sign of ailment [3]. Any significant change in the daily activity routine of cows provides information about their health. These indicators are incredibly important as early detection of any disease can lead to timely medical treatment and a higher likelihood of recovery of the animal and less cost to the cattle owner [15].

4 Conclusion

In this paper, we proposed an effective cattle monitoring technique that utilizes cattle farm video recordings and recognizes cattle behaviors. For the execution of this study, we collected a diverse dataset from indoor cattle farms using RGB cameras placed at multiple viewpoints and collecting data during the day and nighttime. Our system is accurate and robust as it is trained on a raw dataset, with numerous visibility challenges, multiple objects for detection, 15 classes, and real-time rendering capabilities. We also provided an analysis on how using our research on cattle activity can help in detecting a pattern for cattle stress and ideally inform cattle owners of early disease development in cattle, thereby leading to early treatment or prevention of disease.

References

1. Fuentes, A., Yoon, S., Park, J., & Park, D.S. (2020). Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Comput. Electron. Agric., 177*, 105627.
2. Kim, N.Y., Kim, S.J., Jang, S.Y., Oh, M.R., Tang, Y., Seong, H.J., Yun, Y.S., & Moon, S. (2017). Behavioral characteristics of Hanwoo (Bos taurus coreanae) steers at different growth stages and seasons. *Asian-Australasian Journal of Animal Sciences, 30*, 1486 - 1494.
3. Ayadi, S., Said, A.B., Jabbar, R., Aloulou, C., Chabbouh, A., & Achballah, A.B. (2021). Dairy Cow rumination detection: A deep learning approach. *ArXiv, abs/2101.10445*.
4. Ryan, C., G'ueret, C., Berry, D.P., Corcoran, M., Keane, M.T., & Mac Namee, B. (2021). Predicting Illness for a Sustainable Dairy Agriculture: Predicting and Explaining the Onset of Mastitis in Dairy Cows. *ArXiv, abs/2101.02188*.
5. Halachmi, I., Guarino, M., Bewley, J.M., & Pastell, M. (2019). Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. *Annual review of animal biosciences, 7*, 403-425.
6. Arazo, E., Aly, R., & McGuinness, K. (2022). Segmentation Enhanced Lameness Detection in Dairy Cows from RGB and Depth Video. *ArXiv, abs/2206.04449*.
7. Ryan, C., Gu'eret, C., Berry, D.P., & Mac Namee, B. (2020). Can We Detect Mastitis earlier than Farmers? *ArXiv, abs/2011.03344*.
8. Nguyen, C.H., Wang, D., Richter, K.V., Valencia, P., Alvarenga, F.A., & Bishop-Hurley, G. (2021). Video-based cattle identification and action recognition. *2021 Digital Image Computing: Techniques and Applications (DICTA)*, 01-05.
9. Jocher, G.R., Stoken, A., Borovec, J., NanoCode, Chaurasia, A., TaoXie, Changyu, L., Abhiram, Laughing, tkianai, yxNONG, Hogan, A., lorenzomammana, AlexWang, Hájek, J., Diaconu, L., Marc, Kwon, Y., Oleg, wanghaoyang, Defretin, Y., Lohia, A., ah, M., Milanko, B., Fineran, B., Khromov, D.P., Yiwei, D., Doug, Durgesh, & Ingham, F. (2021). ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations.

10. Helwatkar, A., Riordan, D., & Walsh, J. (2014). Sensor Technology For Animal Health Monitoring. *International Journal on Smart Sensing and Intelligent Systems, 7*, 1 - 6.
11. Tuyttens, F.A., Molento, C.F., & Benaissa, S. (2022). Twelve Threats of Precision Livestock Farming (PLF) for Animal Welfare. *Frontiers in Veterinary Science, 9*.
12. Mahmud, M.S., Zahid, A., Das, A.K., Muzammil, M., & Khan, M.U. (2021). A systematic literature review on deep learning applications for precision cattle farming. *Comput. Electron. Agric., 187*, 106313.
13. Zheng, Z., Zhang, X., Qin, L., Yue, S., & Zeng, P. (2023). Cows' legs tracking and lameness detection in dairy cattle using video analysis and Siamese neural networks. *Computers and Electronics in Agriculture*.
14. Borchers, M., Chang, Y.M., Tsai, I.C., Wadsworth, B.A., & Bewley, J.M. (2016). A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. *Journal of dairy science, 99 9*, 7458-7466.
15. Liang, D., Arnold, L.M., Stowe, C.J., Harmon, R.J., & Bewley, J.M. (2017). Estimating US dairy clinical disease costs with a stochastic simulation model. *Journal of dairy science, 100 2*, 1472-1486.
16. Han, S., Fuentes, A., Yoon, S., Park, J., & Park, D.S. (2022). Multi-Cattle tracking with appearance and motion models in closed barns using deep learning. *Korean Institute of Smart Media*.
17. Singh, C. (2021). Applications and Challenges of Deep Learning in Computer Vision. *Health Information Science*.