E3S Web of Conferences 664, 01002 (2025) https://doi.org/10.1051/e3sconf/202566401002
ISSAT 2025

Real-time gender classification of Garut sheep within a small flock using deep learning

Rida Hudaya[1*], *Noor Cholis* Basjaruddin[1], *Hepi* Ludiyati[1], *Eril* Mozef[1], *Farrah* Vauzia[1], *Bima Gerry* Pratama[1], *Anding* Nyuak[2], *Tze Ching* Ong[3], and *Mike Joe* Juing[4]

[1]Department of Electrical Engineering, Politeknik Negeri Bandung, Bandung, Indonesia
[2]Department of Information Technology and Communication, Politeknik Kuching Sarawak, Sarawak, Malaysia
[3]Department of Mechanical Engineering, Politeknik Kuching Sarawak, Sarawak, Malaysia
[4]Department of Electrical Engineering, Politeknik Kuching Sarawak, Sarawak, Malaysia

Abstract. Garut sheep are a unique Indonesian breed with both agricultural and cultural importance. Accurate gender identification is crucial for breeding management, health monitoring, and behavioral analysis. This study proposes a hybrid deep learning approach to automatically classify the gender of Garut sheep using video data. The method integrates You Only Look Once version 8 (YOLOv8) for real-time sheep detection with a Convolutional Neural Network (CNN) for gender recognition. Publicly available videos were collected, annotated, and preprocessed to construct a custom dataset. YOLOv8 was applied to detect sheep in frames, and cropped regions were then passed to the CNN for classification. Experimental results show that the combined YOLOv8–CNN framework performs effectively in distinguishing male and female Garut sheep under varying lighting and motion conditions. The system achieved an accuracy of 89.6% while maintaining real-time performance at 25–30 frames per second (FPS) from CCTV video input. These results demonstrate the promise of computer vision for intelligent livestock monitoring in precision agriculture. Nevertheless, the study remains limited by the small dataset size, which may affect generalizability. Future work will focus on expanding the dataset, capturing more diverse conditions, and validating the system's robustness on larger flocks.

1 Introduction

Garut sheep (Ovis Aries), a native Indonesian breed, are widely recognized for their economic and cultural significance, particularly in West Java. These sheep are not only valuable as a source of meat and income for local farmers, but also hold a prominent role in traditional contests such as adu domba, which are deeply rooted in Sundanese cultural heritage [1, 2]. The distinct physical characteristics and behavior of Garut sheep make them an important subject in animal husbandry and livestock development programs. Morphologically, male Garut sheep typically exhibit a more robust body frame, pronounced

* Corresponding author: rida_hudaya@polban.ac.id

 https://doi.org/10.1051/e3sconf/202566401002

muscle mass, and prominent curved horns, which differentiate them from females that generally possess a smaller body size and less aggressive appearance [3]. Currently, gender identification of Garut sheep is still commonly performed through manual visual observation, relying on the experience and subjective judgement of farmers. This traditional method is inherently time-consuming and prone to human error, particularly when multiple sheep are moving simultaneously or viewed from suboptimal angles [4]. These limitations underscore the need for an automated and reliable gender classification system specifically designed for Garut sheep, capable of operating in real-time using visual data obtained from videos or surveillance cameras.

To address these limitations, computer vision and deep learning technologies offer promising solutions for automating visual-based tasks in livestock monitoring [5]. Among these, Convolutional Neural Network (CNN) have become a foundational architecture for image classification tasks due to their ability to learn spatial hierarchies of features from raw visual data automatically [6, 7]. CNN has been widely adopted in agricultural applications, including livestock monitoring, because they are able to accurately distinguish small visual characteristics that differentiate between categories, such as gender or breed [8, 9]. In addition to CNN-based classification, object detection plays a crucial role in identifying sheep within video frames. One of the most widely used real-time object detection algorithms is You Only Look Once (YOLO), which performs detection in a single forward pass, allowing high-speed and accurate analysis of visual data [10]. The latest version, YOLOv8, introduces significant improvements in detection precision, model efficiency, and adaptability to various input resolutions, making it highly suitable for real-time livestock applications [11-13]. In the context of Garut sheep monitoring, YOLOv8 enables the system to automatically detect individual Garut sheep from video input, even in dynamic or crowded environments. The combination of CNN and YOLOv8 forms a powerful deep learning pipeline capable of detecting Garut sheep in video frames and classifying their gender with high reliability, even under challenging environmental conditions.

Several studies combine CNN and YOLO to develop integrated frameworks capable of both detecting and classifying objects in a single pipeline. Singh et al. are combining CNN and YOLO to leverage the strengths of both approaches to improve detection speed and accuracy for applications such as surveillance, autonomous vehicles, and smart environments [14]. Aroujo et al. are combining CNN and YOLOv8I for cow detection in Complex Farm Environments. The paper integrates the CBAM attention module into the YOLOv8l backbone for cow detection [15]. Bello et al. combined (CNN) and the YOLO model to enhance object detection and monitoring in animal husbandry applications [16]. However, most of these systems do not specifically address local breeds such as Garut sheep, and have not focused on real-time performance using CCTV in small flock settings.

Unlike previous studies that applied CNN and YOLO for general livestock monitoring, this study specifically targets the Garut sheep, a local Indonesian breed, and demonstrates a real-time CCTV-based pipeline for small flock settings. To our knowledge, this is the first work that combines YOLOv8 with CNN for gender classification of Garut sheep, thereby advancing the applicability of computer vision in precision livestock farming for local breeds. We hypothesize that combining YOLOv8 and CNNs offers a reliable and efficient approach for gender classification in Garut sheep, supporting more precise and effective livestock management for this specific breed.

2 Research method

This study implements a two-stage deep learning pipeline consisting of object detection and gender classification. The overall architecture integrates YOLOv8 for detecting and localizing Garut sheep in video frames, and a CNN for classifying the gender of the detected

Garut sheep. The research method includes five main stages: data collection and annotation, data augmentation, object detection using YOLOv8, CNN-based gender classification, and performance evaluation. Each stage is carefully designed to ensure robustness, accuracy, and generalizability across different video conditions, such as variations in lighting, camera angle, and Garut sheep movement. A visual overview of the research method is presented in Fig. 1.

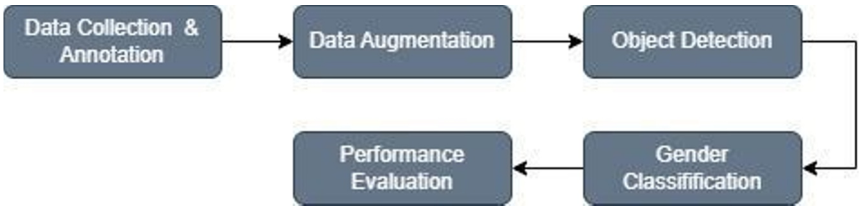

Fig. 1. Research method.

2.1 Data collection and annotation

The raw data used in this study consists of both static images and video recordings of Garut sheep. The videos were obtained from publicly available footage on the internet and real-time recordings captured via CCTV cameras installed in local sheep pens. Selected frames were extracted from these videos and combined with static images to form a unified visual dataset. The data were then manually reviewed and categorized into two classes, *Male* and *Female*, based on visible morphological characteristics such as horn curvature, body size, and posture. Each class contained a total of 75 images, resulting in a balanced dataset of 150 labeled samples. The images were stored in separate subdirectories named *Male* and *Female* under a common root folder labeled *Dataset*, allowing for an organized structure to support subsequent processing and model training as shown in Fig. 2.

(a) (b)

Fig. 2. (a) Male garut sheep and (b) female garut sheep

Once the dataset structure was established, Python (version 3.11) was utilized to support further preprocessing. This included resizing all images to 128 × 128 pixels to ensure consistent input dimensions for the classification model. Due to the limited dataset size (150 images), the data were divided into an 80:20 training–validation split. A separate test set could not be created, and the validation set was therefore used as a proxy for estimating the model’s generalization performance. This limitation is acknowledged and will be addressed in future work with larger-scale data collection.

2.2 Data augmentation

To increase the diversity of the training data and enhance the generalization capability of the model, data augmentation techniques were applied to the training set. The augmentation process included random rotations, zooming, shear transformations, horizontal flipping, and spatial shifts. These transformations were selected to simulate visual variations commonly encountered in real-world farm environments, such as different camera angles, lighting changes, and animal movement. All augmentation techniques were applied only to the training data to prevent data leakage into the validation set. This ensures that the evaluation process remains unbiased and allows the model to be trained on a more varied representation of the Garut sheep images.

2.3 Object detection

In this study, object detection served as the first stage in the real-time pipeline to locate individual Garut sheep within video frames. The YOLOv8 (You Only Look Once, version 8) model was selected for its strong real-time detection performance and low computational cost. The lightweight YOLOv8n (nano) variant was used to optimize speed and efficiency, making it practical for resource-constrained environments such as CCTV-based livestock monitoring. The system was configured to process RTSP video streams from multiple CCTV cameras around the sheep pen, enabling simultaneous monitoring of different enclosure areas.

Each incoming video stream was processed frame by frame. YOLOv8 was applied to detect the presence of Garut sheep and to generate bounding boxes around each detected object. This detection phase enabled the system to isolate regions of interest (ROIs), effectively separating individual Garut sheep from the surrounding environment, which is often cluttered or dynamic in nature. The cropped regions corresponding to the detected Garut sheep were then forwarded to the next stage for gender classification.

2.4 Gender classification

Following the detection stage, each cropped image of a Garut sheep was resized to 128 × 128 pixels and normalized. The preprocessed image was then fed into a pre-trained CNN consisting of three convolutional layers (3×3 kernel size, ReLU activation), each followed by 2×2 max-pooling, then passed through a fully connected layer with 128 neurons, and a final output layer with 2 neurons for gender classification. The CNN employed a softmax activation function in its output layer to produce probabilistic predictions for two categories: *Male* and *Female*. In simplified terms, a CNN model transforms the input image x into a prediction vector ŷ through a series of convolutional, non-linear activation, and pooling operations followed by fully connected layers. This process can be represented as:

$$\hat{y} = softmax(W.x + b) \quad (1)$$

where:

- x is the extracted feature vector from the final convolutional or pooling layer,
- W and b are the weight matrix and bias vector of the final dense layer,
- and ŷ is the output probability vector for each class, calculated using the softmax function:

$$softmax(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{C} \left(e^{Z_j}\right)} \quad (2)$$

Here, C is the number of classes (2 in this case: Male and Female), and Z_i is the score for class i before normalization. The model assigns the final gender label based on the class with the highest predicted probability. This label, along with its confidence score, is then overlaid on the corresponding Garut sheep in the video frame using a color-coded scheme, where green indicates male sheep and blue indicates female sheep. This classification step enhances the system's utility by not only detecting Garut sheep but also providing real-time demographic insights, which are beneficial for automated livestock monitoring and breeding decision systems.

2.5 Performance evaluation

The performance of the proposed system was evaluated through two main components: gender classification and real-time object detection. For the classification component, evaluation was conducted on a dataset of 150 labeled images, split into training and validation sets with an 80:20 ratio. The CNN model was assessed using standard performance metrics, including accuracy, precision, recall, and F1-score for both male and female classes. These metrics were calculated using a confusion matrix generated from the validation set to measure the model's generalization capability.

For the object detection component, the YOLOv8 model's performance was assessed based on detection confidence scores, frame processing speed (frames per second), and bounding box consistency under varying conditions. The system received video input through multiple RTSP camera streams to simulate real-time monitoring conditions. Evaluation focused on measuring processing latency per frame, average FPS, and detection robustness against variations in lighting and camera angle.

3 Result and discussion

3.1 Classification performance

Table 1 illustrates the classification performance metrics.

Table 1. Classification performance.

Metric	Male (%)	Female (%)	Average (%)
Accuracy	89.6	89.6	89.6
Precision	88.4	91.1	89.7
Recall	92.4	86.4	89.4
F1-score	90.4	88.7	89.5

The gender classification model developed using CNN demonstrated solid performance on the validation dataset, achieving an overall accuracy of 89.6% with precision, recall, and F1-score values indicating balanced performance across male and female classes. The classification accuracy of 89.6% corresponds to 27 correctly classified samples out of 30 in the validation set, with a 95% confidence interval of 74.4%–96.5%. These results suggest that although the system performs well, the limited dataset size constrains the reliability of accuracy estimates. The close alignment between training and validation accuracy indicates minimal overfitting, supported by data augmentation strategies (rotation, flipping, zooming) that enhanced model generalization. However, some classification errors were observed. Specifically, female sheep were occasionally misclassified as male under conditions of low lighting or partial occlusion, where distinguishing morphological features such as horn curvature were less visible. Quantitatively, this contributed to a 6% drop in recall for the female class compared to male. Compared to related CNN–YOLO frameworks in cattle and goat monitoring [15, 16], our model demonstrated comparable robustness in cluttered environments, but still requires improvements for better handling of occlusion and motion blur.

These classification errors underline the system's sensitivity to environmental factors such as lighting and occlusion, which reduce the visibility of key morphological cues. While the overall performance is promising, these limitations highlight the need for improvements to increase robustness. Future research should therefore consider multi-angle camera setups to mitigate occlusion, deeper CNN architectures or attention mechanisms to capture finer features, and the integration of behavioral data such as mounting or locomotion patterns. Such enhancements would help the model generalize better across diverse farm conditions and reduce misclassification risks in practical deployments.

3.2 Real-time object detection performance

The real-time detection component using YOLOv8 was integrated with live video feeds from three RTSP camera streams. The model was able to detect and localize Garut sheep in each frame with an average confidence score of over 89%, maintaining an average frame rate of 25–30 FPS per stream. Real-time inference showed latency below 200 ms per frame, indicating that the system is suitable for practical deployment in live monitoring scenarios. The results of YOLOv8 evaluation on RTSP streams are shown in Table 2.

Table 2. YOLOv8 performance on rtsp streams.

Parameter	**Value**
Number of active RTSP streams	3
Average FPS per stream	25–30 fps
Average detection confidence	>89%
Classification latency	< 200 ms/frame
Input frame resolution	640 × 480 px
Number of Garut sheep per frame	1–3 (average: 2)

A summary of system-level detection performance is presented in Table 2, which confirms that the detection module consistently performed within acceptable latency and accuracy ranges. The average number of Garut sheep per frame ranged between 1 to 3, with the system reliably handling multiple animals without significant drop in accuracy or processing speed.

3.3 Integrated system behavior

The integrated pipeline, which combines YOLOv8 for detection and CNN for classification, performed effectively during live testing. Detected Garut sheep were cropped and classified in real-time, with the results visualized through bounding boxes and gender labels that were directly overlaid on the video streams. The application of distinct color codes (such as green for male and blue for female) improved visual clarity during observation. This visual feedback allowed operators to monitor multiple Garut sheep simultaneously with minimal effort, reducing the likelihood of human error in manual observation. As depicted in Figure 3, multiple Garut sheep are classified by gender in real time.

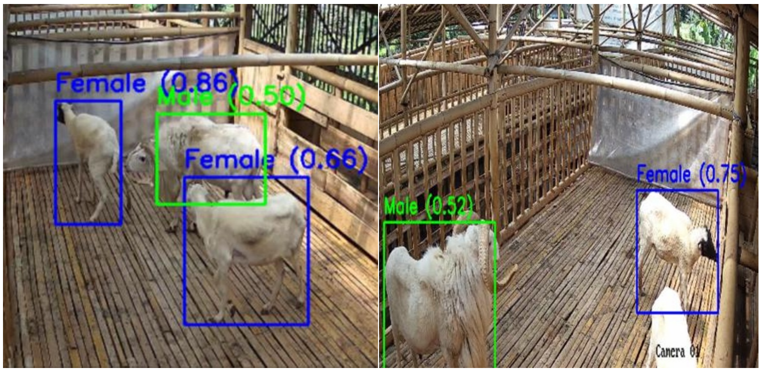

Fig. 3. Real-time gender classification of multiple Garut sheep.

As can be seen in Figure 3, Figure 3 demonstrates the capability of the integrated system to simultaneously detecting and classifying multiple Garut sheep in real time. This frame, taken from a live CCTV stream inside a local sheep pen, shows three Garut sheep being successfully identified and labeled according to their gender. The bounding boxes indicate two female Garut sheep with confidence scores of 0.86 and 0.66, and one male with 0.50. Meanwhile, in the other figure, taken from a different CCTV angle, the system also successfully identified one female Garut sheep with a confidence score of 0.75, and one male with 0.52. Although the confidence score for the male sheep is relatively low (0.50), this can be attributed to the limited visibility of distinguishing features, such as partially occluded horns or suboptimal viewing angle. Such conditions reduce the clarity of morphological cues used by the CNN model, especially when animals are close together or in motion. This highlights a known challenge in real-world livestock environments, where lighting, occlusion, and movement are highly dynamic.

Moreover, the system showed resilience in handling occlusion and complex spatial arrangements, highlighting its utility for intelligent livestock management. It was validated in a live setting at a Garut sheep pen in West Java, using three CCTV cameras connected via RTSP streams over several days. During field trials, the system consistently detected and classified multiple sheep under varied lighting and behaviors. Farm staff reported that it enabled quick gender identification without manual inspection, especially during feeding and health checks. This feedback confirms its practical value and ease of integration into livestock workflows.

4 Conclusion

In conclusion, this research successfully developed a real-time gender classification system for Garut sheep by integrating YOLOv8 for object detection and CNN for gender classification. The system achieved an overall accuracy of 89.6% while maintaining stable real-time performance with low latency and high detection confidence across multiple video streams. These results demonstrate that the proposed approach is both effective and practical for automated livestock monitoring, providing a reliable tool to support more precise and efficient sheep farming management in real-world environments. However, the dataset used in this study was relatively small (150 images) and collected under limited recording conditions. Future work should focus on expanding the dataset, incorporating more diverse environmental conditions (e.g., lighting variations, multi-angle recordings, and different farm environments), and validating the system in larger-scale trials with more sheep to ensure robustness and generalizability.

For future research, further improvements can be made by optimizing the classification model's performance, such as exploring more advanced CNN architectures or incorporating additional features like age or body condition, to enhance the system's accuracy and applicability in diverse farming conditions. In addition, this research serves as an important initial step toward further studies aimed at enhancing intelligent livestock systems. Future work will focus on developing a system for detecting mounting behavior in Garut sheep, which is a key indicator in reproductive management. By integrating behavioral detection capabilities, the system can provide farmers with deeper insights into breeding cycles and animal health, supporting more effective and automated livestock management strategies.

This research was financially supported by the World Bank. The funding provided played a crucial role in enabling data collection, system development, and experimental validation. The authors sincerely acknowledge and appreciate the World Bank's contribution to advancing research in intelligent livestock monitoring systems.

References

1. S. S. Mansjoer, T. Kertanugraha, C. Sumantri. Estimation of Genetic Distance between Agile and Meat Type Garut Sheep. *Media Peternak*. **30**, 129–138 (2007)
2. D. Wijayanti, F. Ardigurnita, N. Frasiska, N. F. F. Ihsan, A. Dudung. Effect of Parity and Type of Birth on Reproduction Performance of Garut Ewes in Indonesia. *Bull. Anim. Sci*. **49**, 313–318 (2025). https://doi.org/10.21059/buletinpeternak.v%vi%i.100020
3. I. Inounu, Erfan, R. H. Mulyono. Characteristics of Body Measurement and Shape of Garut Sheep and Its Crosses with Other Breeds. *Indonesian J. Anim. Vet. Sci*. **14**, 295–306 (2009). https://doi.org/10.14334/jitv.v14i4.330
4. I. Devi, N. Singh, K. Dudi, R. Ranjan, S. S. Lathwal, D. S. Tomar, H. Nagar. Deep learning aided computer vision system for automated linear type trait evaluation in dairy cows. *Smart Agric. Technol*. **8**, 100509 (2024)
5. M. Jorquera-Chavez, S. Fuentes, F. R. Dunshea, E. C. Jongman, R. D. Warner. Computer vision and remote sensing to assess physiological responses of cattle to pre-slaughter stress, and its impact on beef quality: A review. *Meat Sci*. **156**, 11–22 (2019)
6. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, L. Farhan. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **8** (2021)
7. Y. Lecun, Y. Bengio, G. Hinton. Deep Learning. *Nature* **521**, 436–444 (2015). https://doi.org/10.1038/nature14539
8. M. H. Kang, S. H. Oh. Research trends in livestock facial identification: a review. *J. Anim. Sci. Technol*. **67**, 43–55 (2025)
9. Z. Wan, F. Tian, C. Zhang. Sheep Face Recognition Model Based on Deep Learning and Bilinear Feature Fusion. *Animals* **13** (2023)
10. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You Only Look Once: Unified, Real-time Object Detection. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. 779–788 (2016). https://doi.org/10.1109/CVPR.2016.91

11. C. Medjahed, N. W. Krolkral, F. Mezzoudj, A. Slimani. YOLO Accuracy Enhancement for Dense and Dynamic Scenes. *J. Inf. Syst. Eng. Manag*. **10**, 462–470 (2025)
12. H. Dou, S. Chen, F. Xu, Y. Liu, H. Zhao. Analysis of vehicle and pedestrian detection effects of improved YOLOv8 model in drone-assisted urban traffic monitoring system. *PLoS One* **20**, 1–22 (2025)
13. M. Safaldin, N. Zaghden, M. Mejdoub. An Improved YOLOv8 to Detect Moving Objects. IEEE Access 12, 59782–59806 (2024). https://doi.org/10.1109/ACCESS.2024.3393835
14. K. Singh, A. Dubey, A. K. Verma, A. Pathak. Combining YOLO and R-CNN for Robust Object Detection. *Int. J. Res. Appl. Sci. Eng. Technol*. **13**, 5684–5688 (2025)
15. V. M. Araújo, I. Rili, T. Gisiger, S. Gambs, E. Vasseur, M. Cellier, A. B. Diallo. AI-powered cow detection in complex farm environments. *Smart Agric. Technol*. **10**, 1–16 (2025)
16. R.-W. Bello, R. O. Ogundokun, P. A. Owolawi, E. A. van Wyk, C. Tu. Application of Convolutional Neural Networks in Animal Husbandry: A Review. *Mathematics* **13**, 1906 (2025)