Journal of Electronics and Information Science (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2026.110110
ISSN 2371-9524 Vol. 11 Num. 1

Intelligent Poultry Health Recognition from Fecal Images Using YOLO11n

Yajie Liu[1,a], Zhi Liu[1,b], Shuaichen Yuan[1,c,*], Yuqin Wang[1,d], Yiyi Long[1,e], Lu Gao[1,f]

[1]School of Information Engineering, Henan Vocational University of Science and Technology, Zhoukou, China

[a]012021070737@sgs.msu.edu.my, [b]402097323@qq.com, [c]yuancn2022@163.com, [d]1911505141@qq.com, [e]2845352025@qq.com, [f]gl20060407@qq.com

**Corresponding author*

Keywords: Poultry Healthy; Fecal Images; YOLO11n; Object Detection; Transfering Learning

Abstract: Poultry industry plays a vital role in the livestock industry. In poultry production, disease breakouts can cause a fall in production performance and, at times, can cause significant mortality which could result in tremendous economic loss. To realize the low-cost, contactless poultry health status detection for large-scale poultry houses, this study explores the method of detecting the four types of poultry health status according to the images of poultry feces. A total of 6812 photographs were taken on farms, of four types: Healthy, Coccidiosis, Newcastle Disease and Salmonellosis. The dataset has been divided into training, test and validation sets in a ratio of 7:2:1. In this study YOLO11n models were trained from scratch and with pretrained weights. The results show that the pretrained model achieved Precision, Recall, mAP@0.5 and mAP@0.5:0.95 of 85.1%, 73.1%, 83.6%, and 63.1%, respectively; the model trained from scratch achieved 81.5%, 76.0%, 82.8%, and 59.9%, respectively. In addition, pretraining boosted the overall Precision by 3.6%, mAP@0.5 by 0.8%, and mAP@0.5:0.95 by 3.2%. This procedure can be used as an automatic preliminary screening procedure for abnormal feces in chicken houses.

1. Introduction

Poultry production provides the world's population with efficient and affordable animal protein, but intensive farming also increases the risk of infectious disease spread, production losses, and animal welfare issues [1]. Traditional manual inspections rely on the observer's experience, making it difficult to conduct continuous, objective monitoring of large-scale flocks; while laboratory testing offers high diagnostic value, it requires sampling, specialized equipment, and time [2]. Artificial intelligence, the Internet of Things, and computer vision are therefore being progressively deployed to convert signals such as images, sounds, body temperature, and behavior into actionable health alerts [3].

Poultry health monitoring primarily includes thermal imaging, sound analysis, behavioral analysis, postural analysis, and visual analysis of feces. Thermal infrared methods can reflect changes in body surface temperature but rely on specialized sensors and environmental compensation [4]. Sound analysis is suitable for continuous flock monitoring but are susceptible to mechanical noise and overlapping sounds [5]. Behavioral and postural methods can detect lameness, heat stress, or abnormal activity, but dense crowding, individual tracking, and differences in age can compromise cross-scenario stability [6]. Fecal images are characterized by low acquisition costs, minimal disturbance to the chickens, and ease of integration with edge-detection visual systems; their color, moisture content, morphology, and texture can indicate risks associated with certain gastrointestinal or infectious diseases [7]. However, these phenotypic traits are also influenced by feed, age, lighting, and house environment, and therefore can only support risk screening but cannot constitute a pathogen diagnosis on their own.

Previous studies have employed convolutional neural networks, VGG16, MobileNetV2, Xception, and models combining YOLO with classification networks to identify poultry feces images [8-10]. While these studies have demonstrated the computability of feces phenotypes, three shortcomings remain: First, whole-image classification cannot clearly identify abnormal regions; second, inconsistencies in the data domains, classification systems, and evaluation metrics across different studies make it difficult to directly compare the high accuracy rates reported; third, the reliability of detection for minor classes, such as Newcastle disease, is easily masked by overall metrics. YOLO11n achieves end-to-end localization and classification with a relatively small number of parameters, providing a viable foundation for edge deployment [11].

2. Related Works

Non-contact poultry monitoring has evolved into a multimodal technical system. Wang et al. fused RGB, thermal infrared, and depth information, and used YOLOv8n-mvc to localize chicken heads and perform distance calibration, reporting a mAP@0.5 of 96.0% [12]. Ansarimovahed et al. used YOLOv8 to extract thermal regions of the head and legs, then employed support vector machines to distinguish between avian influenza and Newcastle disease [13]. In the acoustic domain, Soster et al. utilized over 2,000 manually annotated clips to distinguish four types of broiler chicken vocalizations, achieving a balanced accuracy of 91.1% [14]. Sun et al. combined time-frequency domain, MFCC, and sparse features, using a random forest to statistically analyze abnormal vocal patterns and cough rates [15]. These two categories of methods provide physiological and population-level acoustic evidence, respectively; however, equipment costs, ambient temperature, background noise, and individual localization continue to limit their universal applicability.

Behavior and posture studies primarily estimate health status based on video trajectories, heatmaps, or keypoints. Merenda et al. combined YOLOv5l, individual tracking, and a gradient-boosted classifier to identify feeding and drinking behaviors, but the system’s ability to maintain identity declined during long-term tracking [16]. Yan et al. used a lightweight YOLO-HGP with multi-object tracking to assess heat stress behaviors [17]. Fodor et al. used DeepLabCut to extract eight keypoints and found that stride height and the tarsus-to-foot distance ratio were correlated with broiler gait scores [18]. Garg and Goel employed YOLOv5n and a convolutional network to detect lameness in young broilers [19]. These methods can characterize dynamic or postural abnormalities but are susceptible to occlusion, camera angles, and growth stages.

Visual analysis of feces is most relevant to this paper. Machuve et al. established a dataset of four types of feces labeled by both laboratory researchers and farm experts, and compared various transfer learning classifiers [8]. Degu and Simegn used YOLOv3 to localize feces and ResNet50 for classification, and implemented a mobile prototype [20]. Qin et al. compared YOLOv3 through YOLOv9 on a dataset of five types of abnormal feces collected from commercial chicken farms, demonstrating that lightweight models offer a good speed-accuracy trade-off [21].

Overall, different perception pathways reflect different physiological mechanisms, and high performance in a single study does not directly prove diagnostic capability across different chicken houses. This study selected fecal images because of their low cost, ease of collection, and suitability for edge-based visual processing; the objective of the study is to locate and screen for abnormal phenotypes, rather than to claim that fecal images are disease-specific.

3. Materials and Methods

3.1. Dataset Preparation

The data for this study were derived from farm-labeled images collected by Machuve et al. at poultry farms in Tanzania [8]. The source data categories were provided by agricultural technicians and included "Healthy," "Coccidiosis," "Newcastle disease," and "Salmonellosis." The local dataset used in this paper consists of a total of 6812 JPEG images, including 2057 images labeled as "Healthy," 2103 as "Coccidiosis," 376 as "Newcastle disease," and 2276 as "Salmonellosis."

To transform the whole-image classification task into an object detection task, LabelImg was used to draw tight bounding boxes around recognizable feces regions in each image, applying a fixed set of four labels. Upon file-level verification, all 6,812 images were labeled with corresponding YOLO text annotations manually. The total number of object instances was 8,829.

The dataset was split into training, validation, and testing dataset into 7:2:1 as shown in Table 1. 4768 images in the training set, 1362 in the validation set, and 682 in the test set. The testing dataset is kept separate during training and model selection and is used solely for the final independent evaluation. The training, validation, and testing sets contain 6154, 1774, and 901 annotated instances, respectively.

Table 1: Dataset split.

	Training	Validation	Testing	Total
Healthy	1,440	411	206	2,057
Coccidiosis	1,472	421	210	2,103
Newcastle disease	263	75	38	376
Salmonellosis	1,593	455	228	2,276
Images	4,768	1,362	682	6,812
Instances	6,154	1,774	901	8,829

3.2. YOLO11n and Training Strategy

YOLO11n adopts a three-level object detection architecture consisting of a backbone, neck, and head, as shown in Figure 1. Compared to the C2f module used in YOLOv8, YOLO11 employs C3k2 as its primary feature extraction unit. C3k2 inherits the concept of cross-stage partial connectivity and, based on the c3k parameter, internally selects either a Bottleneck or a C3k unit to strike a balance between computational overhead and feature representation. C2PSA divides the input features into a bypass branch and an attention branch, performing position-sensitive attention on a subset of channels before fusion, thereby avoiding the additional computational overhead associated with applying attention to all channels.

For a 640×640 input, the backbone network performs stepwise downsampling through stride convolutions and the C3k2 module to generate multiscale features at multiple levels. The SPPF and C2PSA modules are located at the end of the backbone and are used to expand the receptive field and enhance high-level spatial feature modeling, respectively.

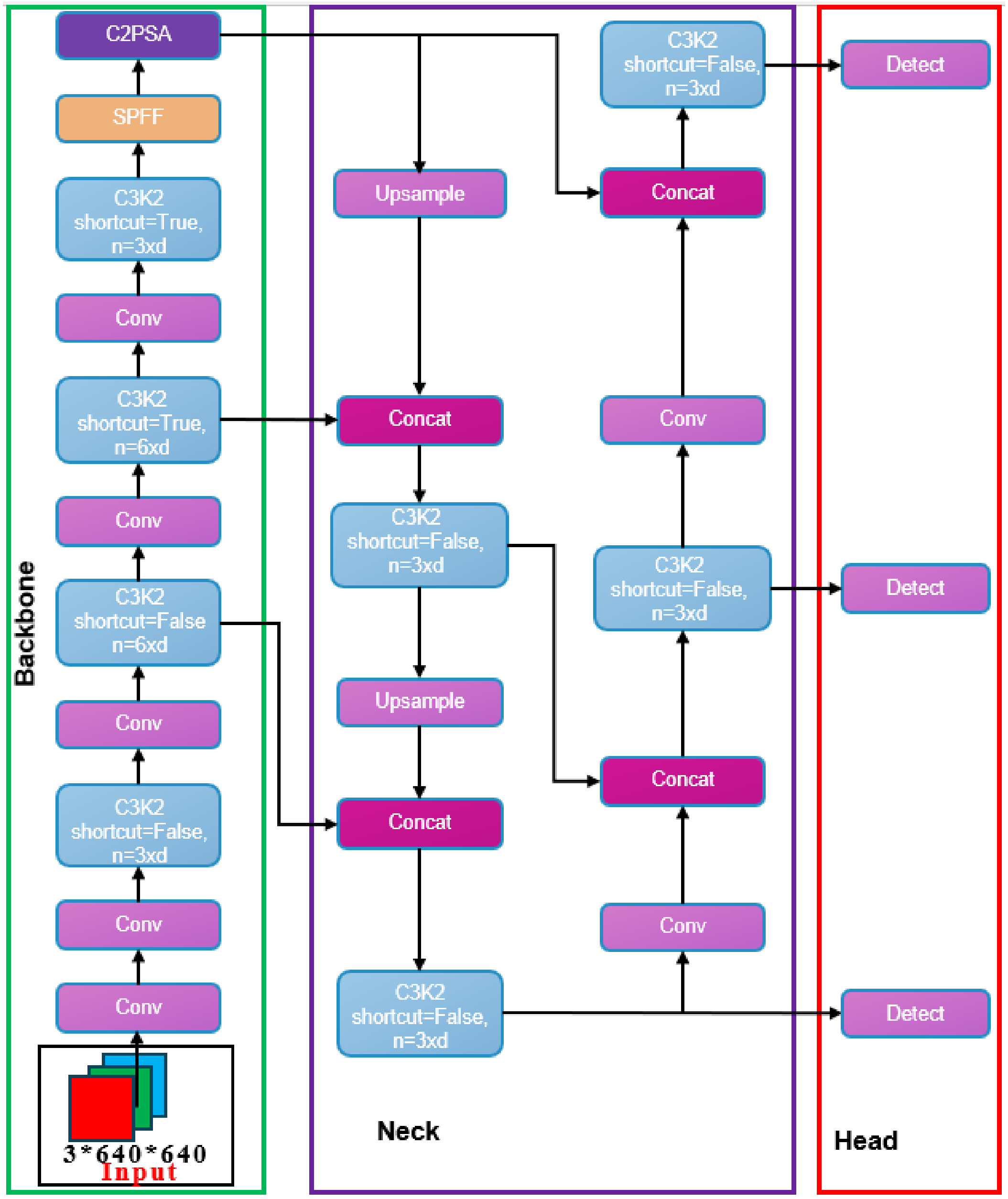


Figure 1: Architecture of YOLO11n.

The Neck consists of a top-down feature pyramid path and a bottom-up path aggregation structure, which fuses shallow localization information with high-level semantic information through upsampling, concatenation, stride convolution, and C3k2. Finally, an anchor-free, decoupled Detect module simultaneously receives the three fused scales and performs bounding box distribution regression and class prediction, respectively.

To isolate the effects of transfer learning, we conducted two experiments that differed only in their initialization methods: (1) Pretrained: Loaded the pretrained weights yolo11n.pt; (2) Scratch: Used the same YOLO11n architecture with random initialization. Both sets of models were trained for 100 epochs, with an input size of 640×640, a batch size of 16. The uniform training environemnts are described as Table 2. The best weights were automatically saved based on the overall fitness on the validation set.

Table 2: Training environments.

Training Platfrom Configuration	Specification	Hyperparameters	Value
CPU	Intel® Xeon® Silver 4214R CPU (2.40 GHz)	Pretrained	yolo11n.pt
GPU	NVIDIA - GeForce RTX 4080 16GB	Epoch	100
RAM	16 GB	Training Size	640
CUDA	11.3	Batch Size	16
Python	3.10	Learning Rate (lr0)	0.01
PyTorch	2.7.0	Optimizer	AdamW

3.3. Evaluation Metrics

To evaluate the detection performance of YOLO11n on four types of poultry manure targets, we used Precision, Recall, Average Precision (AP), and Mean Average Precision (mAP). And Precision (P) and Recall (R) are defined as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

Here, TP refers to the number of predicted bounding boxes that are correct in class and match the ground-truth bounding box; FP refers to the number of predicted bounding boxes that are incorrect in class, fail to meet the IoU requirement, or are incorrectly matched to the background; and FN refers to the number of true objects not detected by the model. Precision reflects the proportion of correct detections in the prediction results, while Recall reflects the proportion of true objects that are detected.

Since a single confidence threshold cannot fully describe a detector's performance, this study calculates the AP for each class based on the Precision-Recall curves at different confidence thresholds:

$$AP = \int_0^1 P(R)\, dR \tag{3}$$

For detection tasks involving multiple classes, mAP is defined as the arithmetic mean of the AP_s for each category:

$$mAP = \frac{1}{N}\sum_{1}^{N} AP_i \tag{4}$$

Here, N represents the number of classes included in the evaluation; in this study, N = 4; AP_i represents the average precision for the ith class.

4. Related Works

Both sets of models completed 100 training iterations. The pretrained model's combined validation set fitness peaked at iteration 66, while the scratch model peaked at iteration 95, indicating that the pretrained weights formed effective features suitable for this task earlier. As shown in Figure 2, the training and validation losses for Box, Classification, and DFL of the pretrained model were lower than those of the Scratch model throughout the whole training phase; The Scratch model showed a rapid decline in the early stages, followed by gradual convergence, but its overall loss level remained high. No sustained divergence between the training and validation curves was observed.

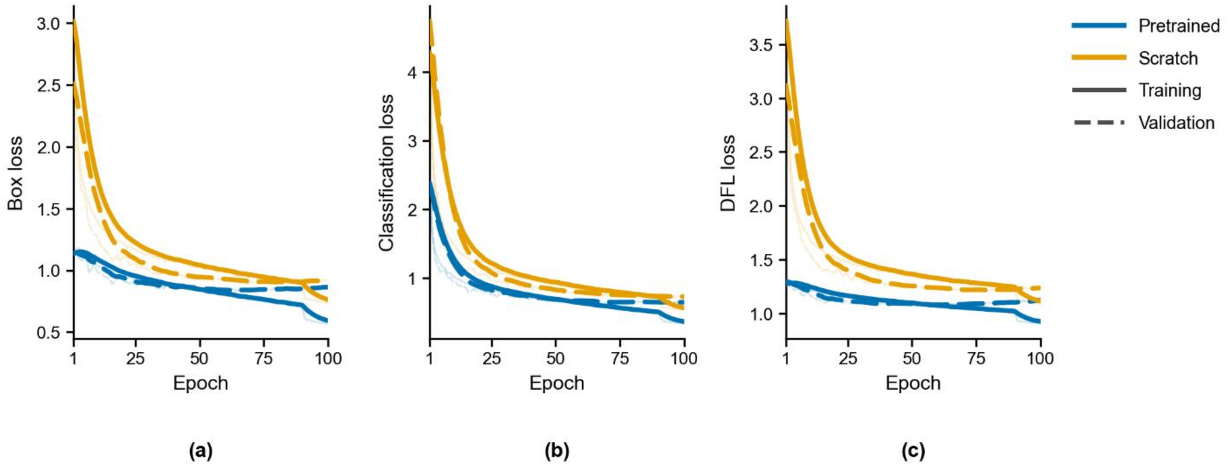


Figure 2: Loss curves: (a) Box loss; (b) Classification loss; (c) DFL loss.

Figure 3 summarizes the overall performance of the optimal checkpoint across 1,362 validation images and 1,774 objects. For the pretrained model, under a threshold of 0.5:0.95, the precision, recall, mAP@0.5, and mAP were 86.9%, 81.3%, 86.7%, and 64.0%. Whereas the model trained from scratch (Scratch model) achieved corresponding metrics of 84.9%, 76.3%, 86.45, and 61.8%. Compared to the model trained from scratch, pretraining increased precision by 2%, recall by 5.0%, mAP@0.5 by 0.3%, and mAP@0.5:0.95 by 2.2%. This indicates that pretraining improves detection coverage and localization quality but also introduces more false positives. Since the validation set is used to optimize weight selection.

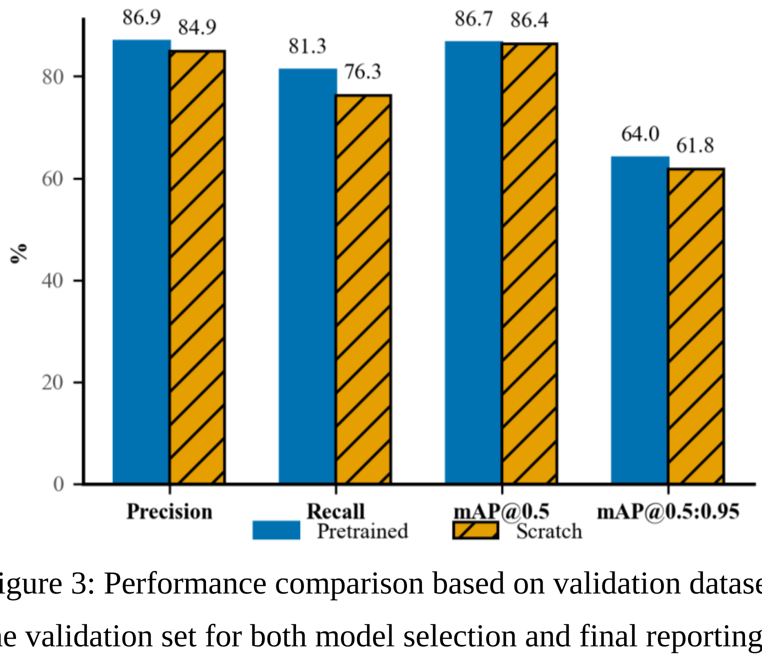


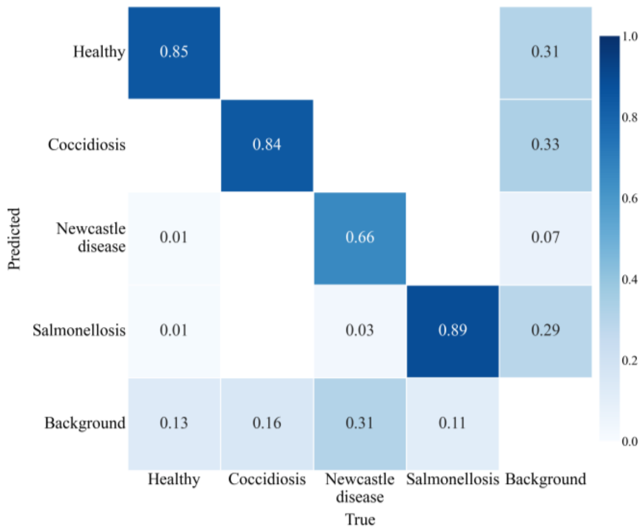
Figure 3: Performance comparison based on validation dataset.

To avoid using the validation set for both model selection and final reporting, this study conducts an independent evaluation using testing dataset. As shown in Table 3, the pretrained model achieved Precision, Recall, mAP@0.5, and mAP@0.5:0.95 of 85.1%, 73.1%, 83.6%, and 63.1%, respectively, while the Scratch model achieved 81.5%, 76.0%, 82.8%, and 59.9%, respectively. Compared to the Scratch model, pretraining increased Precision by 3.6%, mAP@0.5 by 0.8%, and mAP@0.5:0.95 by 3.2%, but decreased Recall by 2.9%.

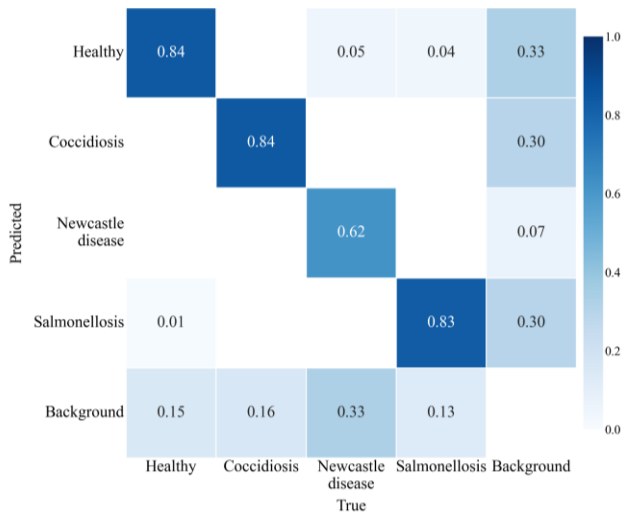
Table 3: Performance comparison based on testing dataset.

Schemes	Class	P (%)	R(%)	mAP@0.5(%)	mAP@0.5:0.95 (%)
Pretrained	All	85.1	73.1	83.6	63.1
Pretrained	Healthy	87.7	78.0	89.0	69.3
Pretrained	Coccidiosis	88.9	76.3	87.8	70.3
Pretrained	Newcastle disease	75.7	55.2	67.2	42.7
Pretrained	Salmonellosis	87.6	82.6	90.4	70.2
Scratch	All	81.5	76.0	82.8	59.9
Scratch	Healthy	81.6	82.8	87.6	64.8
Scratch	Coccidiosis	85.7	80.5	87.7	67.7
Scratch	Newcastle disease	77.4	58.6	69.6	41.3
Scratch	Salmonellosis	81.5	81.9	86.4	65.6

The results further illustrate that the pretrained model achieved the highest mAP@0.5 for Salmonellosis (90.4%), while the mAP@0.5:0.95 for Coccidiosis, Salmonellosis, and Healthy were 70.3%, 70.2%, and 69.3%, respectively. Compared to the Scratch model, the pre-trained model improved the mAP@0.5:0.95 for "Healthy," "Coccidiosis," "Newcastle disease," and "Salmonellosis" by 4.5%, 2.6%, 1.4%, and 4.6%, respectively. However, the Recall for "Healthy," "Coccidiosis," and "Newcastle disease" decreased by 4.8%, 4.2%, and 3.4%, respectively. Newcastle disease was the weakest class in both models: the pretrained model achieved Recall and mAP@0.5 of 55.2% and 67.2%, respectively, while the Scratch model achieved 58.6% and 69.6%, respectively. This indicates that pretraining consistently improves localization of target objects.



(a) Pretrained



(b) Scratch

Figure 4: Confusion matrix comparison: (a) Pretrained; (b) Scratch.

Figure 4 shows the normalized confusion matrix on the independent testing set, with the true class on the x-axis and the predicted class on the y-axis. At the confidence level used for plotting, the diagonal ratios of the pretrained model for Healthy, Coccidiosis, Newcastle disease, and Salmonellosis are 0.85, 0.84, 0.66, and 0.89, respectively, while those for the Scratch model were 0.84, 0.84, 0.62, and 0.83, respectively. This indicates that pretraining improved the correct classification rates for "Healthy," "Newcastle disease," and "Salmonellosis" at this working point, while the rate for "Coccidiosis" remained largely unchanged. The primary errors for both model groups were not due to cross-confusion between disease categories, but rather the misclassification of true targets as "background." The false negative rates for the "background" class in the pre-trained models were 0.13, 0.16, 0.31, and 0.11, respectively, while those for the Scratch models were 0.15, 0.16, 0.33, and 0.13. Newcastle disease had the highest false negative rate, further highlighting the issues of insufficient samples in minority classes and phenotypic heterogeneity.

5. Conclusions

This paper establishes a YOLO11n-based object detection workflow for four categories of poultry manure images and compares COCO pre-training with random initialization on 6,812 images and 8,829 objects. Independent testing showed that the pre-trained model achieved a mAP@0.5 of 83.6% and a mAP@0.5:0.95 of 63.1%, representing improvements of 0.8% and 3.2%, respectively, over the scratch model. Meanwhile, Precision increased by 3.6%, but Recall decreased by 2.9%. Class-by-class results show that pretraining improved the mAP@0.5:0.95 for four classes; however, for Newcastle disease, Recall remained below 60% under both training methods, and this class exhibited the highest rate of false negatives in the confusion matrix. In summary, the benefits of transfer learning are primarily reflected in localization quality and false positive control, rather than a uniform increase in detection rates. YOLO11n supports low-cost, non-contact preliminary screening for abnormal feces; however, model outputs should not replace veterinary examinations or laboratory confirmation until external validation across different farms, expansion to minor classes, and diagnostic consistency studies have been completed.

Acknowledgements

This work was supported by Henan Vocational University of Science and Technology, China (Grant No. HZDKYL2025-043), and the Henan Provincial Education Department, China (Grant No. 27BQ520010).

References

[1] X. Yang et al., "Computer Vision-Based cybernetics systems for promoting modern poultry Farming: A critical review," Computers and Electronics in Agriculture, vol. 225, p. 109339, 2024, doi: 10.1016/j.compag.2024.109339.

[2] R. B. Bist et al., "Sustainable poultry farming practices: a critical review of current strategies and future prospects," Poultry Science, vol. 103, no. 12, p. 104295, 2024, doi: 10.1016/j.psj.2024.104295.

[3] H. M. Taleb et al., "Using Artificial Intelligence to Improve Poultry Productivity – A Review," Annals of Animal Science, vol. 25, no. 1, pp. 23-33, 2025, doi: 10.2478/aoas-2024-0039.

[4] M. H. Natsir, W. F. Mahmudy, M. Tono, and Y. F. Nuningtyas, "Advancements in artificial intelligence and machine learning for poultry farming: Applications, challenges, and future prospects," Smart Agricultural Technology, vol. 12, p. 101307, 2025, doi: 10.1016/j.atech.2025.101307.

[5] E. Hassan, S. Elbedwehy, M. Y. Shams, T. Abd El-Hafeez, and N. El-Rashidy, "Optimizing poultry audio signal classification with deep learning and burn layer fusion," Journal of Big Data, vol. 11, no. 1, p. 135, 2024, doi: 10.1186/s40537-024-00985-8.

[6] L. Yajie, L. Zhi, M. Qingxun, H. Yuxi, and J. Md Gapar Md, "Intelligent Poultry Health Recognition Using an Improved YOLOv8," International Journal of Engineering and Technology Innovation, vol. 16, no. 3, pp. 381-395, 05/22 2026, doi: 10.46604/ijeti.2026.16307.

[7] S. A. S. Mola, E. L. Tade, T. Widiastuti, O. C. Yanda, M. Isnan, and B. Pardamean, "Poultry Diseases Classification Based on Fecal Images Using Pre-Trained Models," in 2025 Tenth International Conference on Informatics and Computing (ICIC), 2025: IEEE, pp. 1-6.

[8] D. Machuve, E. Nwankwo, N. Mduma, and J. Mbelwa, "Poultry diseases diagnostics models using deep learning," Frontiers in Artificial Intelligence, vol. 5, p. 733345, 2022.

[9] S. Harini, K. Aakash, S. Godwin Joe, V. K. Kaliappan, G. B. Hiremath, and D. Jaganathan, "Classification and Detection of Poultry Disease from Chicken Fecal Images Using Deep Learning Techniques," presented at the 2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), 2025.

[10] R. Sridhar, J. Sharma, R. N. K. Dhenia, I. J. Kanani, D. Banerjee, and G. Sharma, "Early Detection of Chicken

Diseases Using EfficientNetB3: A Deep Learning Approach for Poultry Health Monitoring," in 2025 Global Conference on Information Technology and Communication Networks (GITCON), 2025: IEEE, pp. 1-5.
[11] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," arXiv preprint arXiv:2410.17725, 2024.
[12] P. Wang et al., "Chicken body temperature monitoring method in complex environment based on multi-source image fusion and deep learning," Computers and Electronics in Agriculture, vol. 228, p. 109689, 2025, doi: 10.1016/j.compag.2024.109689.
[13] A. Ansarimovahed, A. Banakar, G. Li, and S. M. Javidan, "Separating Chickens' Heads and Legs in Thermal Images via Object Detection and Machine Learning Models to Predict Avian Influenza and Newcastle Disease," Animals, vol. 15, no. 8, p. 1114, 2025, doi: 10.3390/ani15081114.
[14] P. de Carvalho Soster et al., "Automated detection of broiler vocalizations a machine learning approach for broiler chicken vocalization monitoring," Poultry Science, vol. 104, no. 5, p. 104962, 2025, doi: 10.1016/j.psj.2025.104962.
[15] Z. Sun, W. Tao, M. Gao, M. Zhang, S. Song, and G. Wang, "Broiler health monitoring technology based on sound features and random forest," Engineering Applications of Artificial Intelligence, vol. 135, p. 108849, 2024, doi: 10.1016/j.engappai.2024.108849.
[16] V. R. Merenda, V. U. C. Bodempudi, M. D. Pairis-Garcia, and G. Li, "Development and validation of machine-learning models for monitoring individual behaviors in group-housed broiler chickens," Poultry Science, vol. 103, no. 12, p. 104374, 2024, doi: 10.1016/j.psj.2024.104374.
[17] Y. Yan, Z. Sheng, Y. Gu, Y. Heng, H. Zhou, and S. Wang, "Research note: A method for recognizing and evaluating typical behaviors of laying hens in a thermal environment," Poultry Science, vol. 103, no. 11, p. 104122, 2024, doi: 10.1016/j.psj.2024.104122.
[18] I. Fodor, M. van der Sluis, M. Jacobs, B. de Klerk, A. C. Bouwman, and E. D. Ellen, "Automated pose estimation reveals walking characteristics associated with lameness in broilers," Poultry Science, vol. 102, no. 8, p. 102787, 2023, doi: 10.1016/j.psj.2023.102787.
[19] D. Garg and N. Goel, "Early detection model for lameness disease in small broilers using YOLOv5n and Convolution neural network," Procedia Computer Science, vol. 258, pp. 4000-4007, 2025, doi: 10.1016/j.procs.2025.04.651.
[20] M. Z. Degu and G. L. Simegn, "Smartphone based detection and classification of poultry diseases from chicken fecal images using deep learning techniques," Smart Agricultural Technology, vol. 4, p. 100221, 2023, doi: 10.1016/j.atech.2023.100221.
[21] W. Qin, X. Yang, Y. Wang, Y. Wei, Y. Zhou, and W. Zheng, "YOLOPoul: Performance evaluation of a novel YOLO object detectors benchmark for multi-class manure identification to warn about poultry digestive diseases," Smart Agricultural Technology, vol. 12, p. 101145, 2025, doi: 10.1016/j.atech.2025.101145.