



VentVision: A multimodal vision-based system for automated vent-based chick sexing

Marta Vezanzones-Rodriguez ^{a,*}, Thanh Phan ^a, Arthur F.A. Fernandes ^b, Vivian Breen ^b, Jesus Arango ^b, Michael T. Kidd ^c, Ngan Le ^a

^a Department of Computer Science and Computer Engineering, 1 University of Arkansas, Fayetteville, 72701, Arkansas, USA

^b Cobb Vantress, Inc, 4703 US HWY 412 E, Siloam Springs, 72761, Arkansas, USA

^c Department of Poultry Science, 1260 W. Maple, POSC O-114, Fayetteville, 72701, Arkansas, USA

ARTICLE INFO

Keywords:

Chick sexing

Poultry

Vent

Computer vision (CV)

Gender classification

Deep learning (DL)

ABSTRACT

Chick sexing has been a crucial practice for many years due to the need for early gender identification in poultry production. Traditional methods, such as vent sexing, are highly effective but labor-intensive, requiring expert knowledge and years of training, making it a challenging and time-consuming process. Moreover, skilled professionals are increasingly scarce, making the process difficult to scale. To address this challenge, we proposed **VentVision**, an automated vision-based classification system that leverages computer vision and deep learning techniques to improve the efficiency and scalability of chick gender classification. Our system follows a four-step pipeline: multi-modality video capture, temporal segment selection, vent region detection, and gender classification, automating vent-based sexing. We evaluated our approach using RGB imaging, infrared (IR) imaging, and a multimodal RGB-IR feature-level fusion framework based on a dual-stream architecture. The multimodal model achieved an accuracy of 98.09%, outperforming single-modality approaches. These results demonstrate that VentVision achieves expert-level chick sexing performance while eliminating the need for years of specialized training, with manual handling limited to vent exposure, thereby simplifying the sexing workflow.

1. Introduction

Chick sexing, the process of determining the gender of chicks, is a well-established practice in the poultry industry. However, there is an increasing demand for innovative and automated techniques, as hatcheries seek less invasive and more scalable alternatives [1–3]. Early gender identification is crucial, particularly in layer breeds where only female chicks are required for egg production and male chicks are typically culled [4]. In contrast, in broiler parental lines, both sexes are normally raised for meat production [5]. Identifying sex at early stages offers significant operational advantages and helps reduce commercial production costs [6].

Among the most widely implemented traditional methods for day-old chicks are feather sexing and vent sexing, which present complementary advantages and limitations. Feather sexing relies on analyzing the length of primary and secondary wing feathers and can achieve high accuracy, but its applicability is limited to specific strains of chickens [7,8]. In contrast, vent sexing is applicable across virtually all chicken breeds and offers high accuracy [9], but it requires highly trained experts who undergo years of training, making labor availability a bot-

tleneck [10,11]. The method relies on identifying a very small whitish feature in males after cloacal eversion, which is difficult to detect due to its low contrast and subtle appearance. Skilled personnel are increasingly scarce [12] and the procedure itself is invasive, as it involves manually evert and inspecting the chick's cloaca, which can cause discomfort, physical injury, or even mortality if not performed properly [13,14]. Despite these drawbacks, vent sexing remains widely used due to its reliability and general applicability [15].

Recent advances in machine learning (ML) have demonstrated strong potential in automating visual inspection tasks, including applications in poultry production [16–18]. In particular, ML-based approaches have successfully automated traditional feather sexing, improving efficiency and scalability through image-based analysis [11,19–21]. These systems leverage clearly visible morphological features, making them well-suited for computer vision techniques. However, vent sexing presents a substantially more challenging problem, as it relies on subtle, low-contrast morphological cues that are often difficult to observe even for trained experts [22]. Variability in appearance, deformation during handling, and occlusions further increase the difficulty of reliable classification [23]. This highlights the need for approaches capable of

* Corresponding author.

E-mail address: mveganzo@uark.edu (M. Vezanzones-Rodriguez).

<https://doi.org/10.1016/j.atech.2026.102236>

Received 7 April 2026; Received in revised form 12 May 2026; Accepted 20 May 2026

Available online 22 May 2026

2772-3755/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

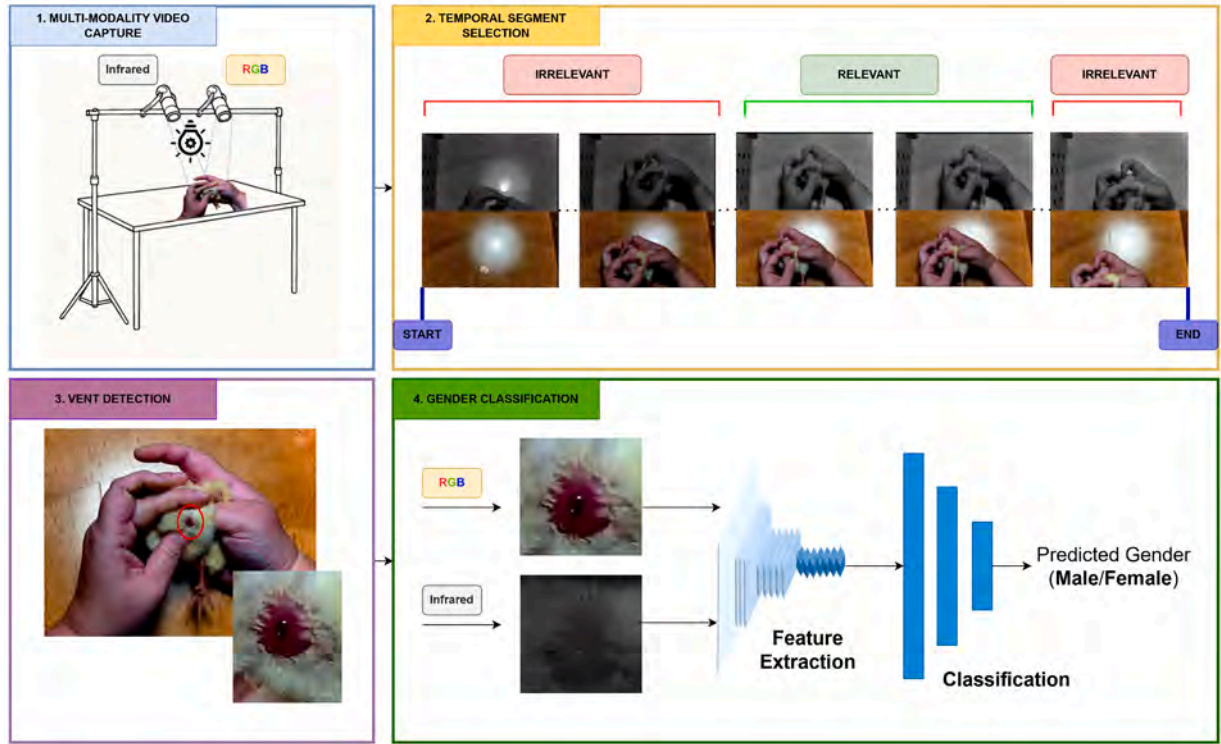


Fig. 1. Overall pipeline of the proposed **VentVision** system, consisting of four main components: multi-modality video capture, temporal segment selection, vent region localization, and gender classification.

capturing subtle and variable visual cues beyond what can be reliably interpreted through manual inspection or single-modality imaging.

Despite the operational importance of vent sexing, its automation remains largely underexplored in the research literature. While commercial systems have been deployed in industrial settings, their methodologies are not publicly documented, limiting reproducibility and scientific evaluation [24]. This lack of publicly documented approaches creates a gap between industrial practice and reproducible, research-driven solutions. As a result, vent sexing continues to rely heavily on manual inspection by trained experts in most hatchery operations. To the best of our knowledge, publicly documented and reproducible computer vision approaches for automated vent sexing remain extremely limited.

To address this gap, we propose a reproducible RGB-Infrared computer vision framework for automated vent sexing. The framework combines multimodal image acquisition, vent-region localization, temporal segment selection, and RGB-IR feature-level fusion to analyze the everted vent region without relying on more invasive imaging approaches. By integrating visible structural information from RGB images with infrared reflectance contrast, the proposed method aims to better capture subtle sex-related vent cues while reducing reliance on expert manual interpretation. The practical positioning of the proposed framework relative to conventional feather and vent sexing methods is summarized in Table 1.

Table 1

Comparison of strain identification methods.

Method	Strain	Expert Required	Training Time	Scalability
Manual feather sexing	Limited	Visual classification	6 months	No
Manual vent sexing	General	Everting vent & Classification	2 years	No
Our VentVision	General	Everting vent	6 months	Yes

Our contributions are summarized as follows:

- **Image-Based Dataset:** We obtained a dataset of vent images of one-day-old chicks from 176 females and 139 males through a semi-automated process. The images were captured in two modalities RGB

and infrared, comprising a total of 20,560 RGB images and 20,412 infrared images. The dataset has been carefully annotated by two poultry experts and two computer vision experts.

- **Methodology:** We introduce, *VentVision*, an automated chick sexing system capable of accurately identifying the gender of day-old chicks. This contribution includes:
 - **Automated Vent Sexing Pipeline:** A end-to-end pipeline that automates the analysis and classification stages of vent sexing through: multi-modality video capture, temporal segment selection, vent region localization, and deep-learning-based gender classification.
 - **Multimodal Fusion Framework:** A reproducible multimodal RGB-Infrared learning framework that integrates complementary visual cues for vent-based chick sexing, demonstrating that *feature-level fusion* can achieve expert-level accuracy in a task traditionally reliant on human expertise.

2. Materials and methods

2.1. Data acquisition

2.1.1. Multi-modality video capture

To enable model training (Fig. 1), multi-modality videos (RGB and infrared) were recorded from pre-sexed chicks provided by Cobb-Vantress. The chicks were processed and sexed at a commercial hatchery by an expert sexer. The designed setup, illustrated in Fig. 2, involved placing two Reolink 842A CCTV cameras (4K resolution, 25 fps, 5× optical zoom, infrared support) mounted side by side with no gap between them. One camera streaming in infrared mode and the other in RGB mode, and both were positioned approximately 50 cm from the chick to ensure detailed imaging of the vent area. A standard light bulb was positioned above the cameras, aligned with their optical axis, to provide uniform illumination of the vent region during data collection.

This dual-camera approach leverages the strengths of each modality. RGB images provide color cues, texture, and structural information

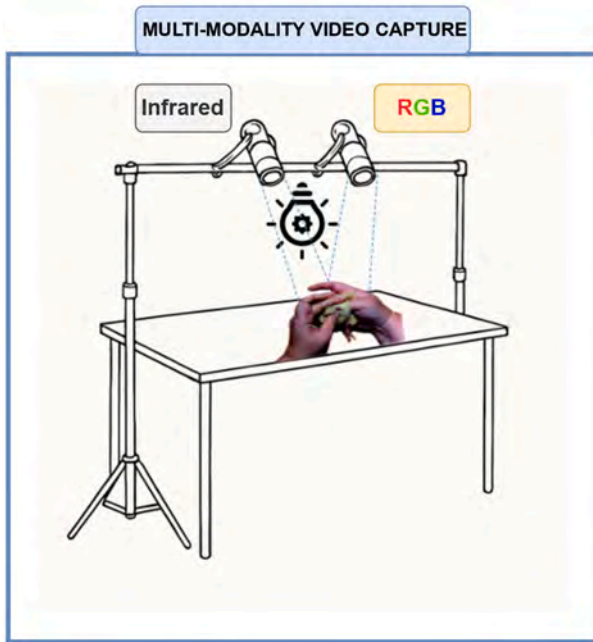


Fig. 2. Data acquisition process. Our acquisition process involves setting up two modality cameras for data acquisition and preprocessing.

from the visible spectrum, which is useful for characterizing the overall appearance of the everted vent region. However, the identification of male chicks relies on detecting very small protruding structure that can appear as a faint whitish feature once the cloaca is manually everted. In practice, even when the expected feature is known, it can remain difficult to identify because of its small size, low contrast with surrounding tissue, making it challenging to distinguish reliably. Infrared imaging provides complementary reflectance information beyond the visible spectrum and is less dependent on visible-light illumination, offering additional contrast that may help disambiguate regions that are difficult to distinguish in RGB images alone. By combining RGB and infrared information, the model can integrate visible structural cues with complementary reflectance contrast, resulting in a more robust representation than either modality alone.

During recording, the chick is held upside down and positioned within the frame. The abdomen is gently squeezed to remove any fecal matter from the vent area, and the cloaca is manually everted to expose the region of interest. This process reveals a small whitish feature in males, while it is absent in females, as illustrated in Fig. 3; however, this feature is subtle and not consistently easy to identify. For each chick, our goal was to capture the vent area clearly for approximately 5 to 10 seconds to ensure that enough frames contained the anatomical features needed for sexing.

2.2. Data annotation

2.2.1. Temporal segment selection and data preprocessing

After capturing the videos, we needed to preprocess them to focus on relevant content. Initially, our recordings included periods before the chick entered the camera frame and after it exited, resulting in irrelevant content. Therefore, we selected and extracted the desired segment by assigning a start and end frame. This segment begins when the cloacal area of the chick, after manual eversion, starts to exhibit subtle anatomical features used in vent sexing, which are visible but require expert knowledge to interpret reliably, such as a small white bump indicating a male or its absence indicating a female. It ends just before these features are no longer visible. The selection process was performed manually by reviewing each video and annotating the start and end frames. An example



Fig. 3. Comparison of male (left) and female (right) chick cloacas. The green circle indicates the small white bump used to identify males. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of this selection process is shown in Fig. 4, where the relevant segment to keep for further analysis is highlighted in green.

Multiple frames were retained from each selected segment because the appearance of the everted vent region can vary across time due to small changes in eversion, tissue exposure, hand positioning, focus, illumination, and partial occlusion. Therefore, these frames provide multiple views of the same chick under slightly different visual conditions.

Consequently, after selecting this temporal segment, we cropped the original videos to retain only the targeted data. During this data preprocessing phase, we curated the dataset to ensure data quality and reliability. Specifically, we discarded certain videos due to data contamination. The resulting distribution of our study is presented in Table 2. The final dataset for both modalities for further analysis consists of 315 chicks.

Table 2

Summary of the distribution of chick IDs by gender.

	# chicks of both modalities
Female	176
Male	139
Total	315

2.2.2. Chick vent detection

Since our primary objective is to examine the vent area of the chick, it is essential to discard background elements that could introduce noise into the dataset. To achieve this, we focus on localizing only the genitals within the frame. For this purpose, we utilized the TrackMe annotation tool [25], which facilitates interpolation between frames in extended video segments that lack predefined detection boxes. This tool helped automate part of the annotation process by allowing us to manually add bounding boxes to a subset of frames and then estimate their positions for the remaining frames through interpolation. All interpolated results were subsequently reviewed and adjusted as necessary to ensure accuracy. This approach enhanced the precision of our annotations by ensuring consistency across the entire set of frames. Finally, once the vent area is detected, we crop the original image based on the predicted bounding box, retaining the cloacal region of interest for further analysis.

2.2.3. Data preparation

After a thorough data processing and curation, the resulting datasets for RGB and infrared image-based gender classification were generated. These datasets reflect the distribution of chicks, as previously mentioned in Table 2, and are presented below with a more detailed specification of their contents.

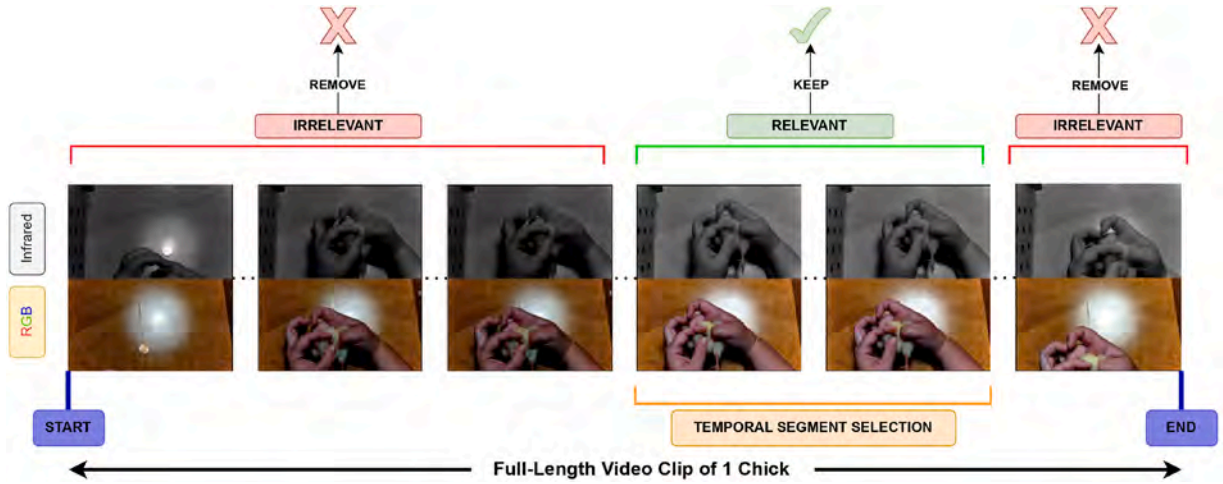


Fig. 4. Temporal segment selection in RGB and infrared frames, highlighting the relevant portion where the genital area is visible for classification while removing all the irrelevant frames.

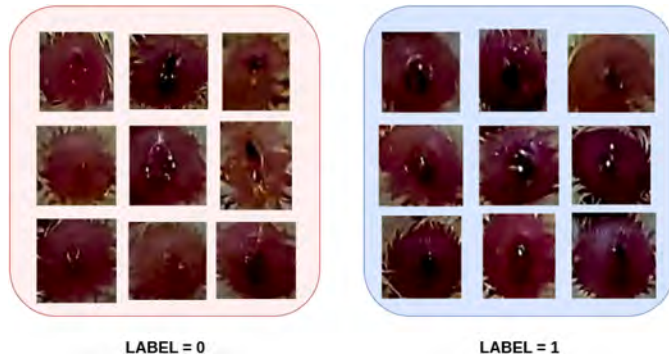


Fig. 5. Representative RGB samples from the dataset. Each image shows the vent region of a chick together with its annotated label, where 0 corresponds to Female and 1 to Male.

i. Image-Based Distribution set

The original videos were recorded at 25 frames per second (FPS). To optimize storage and increase processing speed, the frame rate was reduced to half of the original. As a result, the RGB dataset consists of a total of 20,560 images, of which 7,569 correspond to female chicks and 12,991 to male chicks. Similarly, the infrared dataset contains 20,412 images, with 7,525 corresponding to female chicks and 12,887 to male chicks. The slight difference in frame counts is due to imperfect synchronization between the RGB and infrared streams, with temporal drift occurring during longer recording sessions. This distribution is summarized in Table 3.

Table 3
Summary of the distribution by modality and gender.

Gender	# IDs	# RGB Images	# Infrared Images
Female	176	7,569	7,525
Male	139	12,991	12,887
Total	315	20,560	20,412

In addition to the numerical distribution, Fig. 5 illustrates representative outcomes of the annotated data. Each image corresponds to the vent region of a chick captured in the RGB modality. The assigned label indicates the ground truth gender: 0 = Female and 1 = Male.

2.3. VentVision

Building upon the curated dataset described above, we propose a system for automatic chick sex classification from vent images. In our setting, the input corresponds to the vent region already annotated and extracted during the preprocessing stage. We introduce three complementary frameworks: (i) using RGB images as input, (ii) using infrared (IR) images as input, and (iii) a dual-stream RGB-IR fusion model. All three frameworks adopt ResNet-50 [26] as the backbone for feature extraction, selected for its robustness, strong performance on fine-grained visual tasks, and suitability for moderate-sized datasets, and are followed by a custom classification head. The overall processing pipeline is illustrated in Figs. 6–8.

RGB-Only Framework. We extracted meaningful visual features from the vent region in the preprocessed RGB frames, which serve as input for the chick sex classification task. For this purpose, we employ a ResNet-50 [26] as the feature extraction backbone, due to its strong performance in visual recognition tasks.

For image-based classification, the input image I has the dimension $I \in \mathbb{R}^{H \times W \times C}$, with $C = 3$ channels for the RGB modality. The ResNet-50 network, consisting of 50 layers organized in residual blocks, first captures low-level features such as edges, textures, and corners, and progressively extracts higher-level representations through deeper convolutional stages. The resulting 2048-dimensional embedding is passed through a custom fully connected classification head to produce a binary output indicating the chick's gender (Fig. 6).

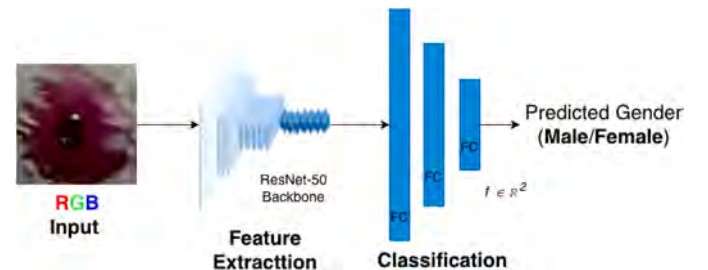


Fig. 6. Pipeline of the RGB-based framework: vent images are processed through a ResNet-50 to extract features, followed by a classification head that predicts gender (0 = Female, 1 = Male).

Infrared-Only Framework. The infrared (IR) framework follows the same design as the RGB one, differing only in the input modality. Although the IR images were originally grayscale, they were processed in

three-channel format by replicating the single channel across all three, in order to match the input requirements of the pretrained ResNet-50 backbone.

As in the RGB framework, ResNet-50 extracts hierarchical features from the vent images, ranging from low-level structures in early layers to more abstract representations in deeper layers. The resulting 2048-dimensional feature embedding is then fed to the same fully connected classification head to produce the binary prediction of gender (Fig. 7).

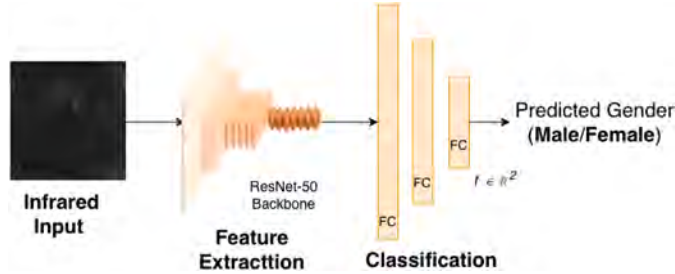


Fig. 7. Pipeline of the IR-based framework: vent images are processed through a ResNet-50 to extract features, followed by a classification head that predicts gender (0 = Female, 1 = Male).

RGB-IR Fusion Framework. In addition to the single-modality models, we implemented a dual-stream architecture for feature level fusion of RGB and infrared (IR) images, allowing each modality to be processed independently in order to preserve modality-specific information. RGB images capture texture and appearance cues under visible light, while infrared images provide complementary contrast patterns that are less sensitive to illumination variability. Two ResNet-50 encoders pretrained on ImageNet were used, one per modality, with the final fully connected layers removed, yielding 2048-dimensional embeddings per stream. The embeddings are concatenated into a joint 4096-dimensional representation and passed through a multi-layer perceptron that produced the final binary prediction via a sigmoid activation. This feature-level fusion strategy enables the model to learn cross-modal interactions between complementary visual cues within a unified representation. Unlike decision-level fusion approaches that combine independent modality predictions, the proposed design allows joint representation learning rather than statistical combination of outputs (Fig. 8).

2.3.1. Gender classification

Once the most relevant features have been extracted using Resnet-50, the network outputs a value called *logits*. For binary classification, this value is transformed into a probability value using the sigmoid function, defined as:

$$p = \sigma(\text{logit}) = \frac{1}{1 + e^{-\text{logit}}} \quad (2.1)$$

where $\sigma(x)$ is the standard logits function, which is a sigmoid function that provides the probability of the output values between 0 and 1. To determine the gender, a threshold of 0.5 was used. The model classifies the image or video as male if p is greater than 0.5 and female if p is equal to or less than 0.5.

$$\text{Gender} = \begin{cases} \text{Male} & \text{if } p > 0.5 \\ \text{Female} & \text{if } p \leq 0.5 \end{cases}$$

2.4. Evaluation metrics

The classification capability of our vent chick sexing model is evaluated through a set of widely used performance metrics. These include Accuracy, Recall, Precision, F1-score, and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). Each metric captures a different aspect of prediction quality and is calculated using the components of the confusion matrix: True Positive (TP), True Negative (TN), False

Positive (FP), and False Negative (FN) values. Performance metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.3)$$

$$F1 - \text{Score} = \frac{TP}{TP + 0.5 * (FP + FN)} \quad (2.4)$$

In our evaluation, all performance metrics are assessed at the chick ID level rather than at the individual image level. Per-ID values are obtained by aggregating predictions across multiple images of the same chick. For instance, per-ID Accuracy is calculated as follows: for a given chick ID represented by N images, the accuracy for that ID is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^N TP_i}{N} \quad (2.5)$$

where TP_i represents correctly classified images of the chick and N is the total number of images associated with that chick.

The Area Under the Curve (AUC) quantifies this performance, representing the percentage of times a randomly chosen female receives a higher predicted score than a randomly chosen male. It evaluates the ability of the model to distinguish between genders without a fixed threshold. The closer the AUC is to 1.0, the closer the model is to being a perfect classifier [27].

2.5. Implementation details

To evaluate model generalization and reduce the risk of overfitting we implemented κ -fold cross-validation with κ set to 5. This approach ensures that accuracy is measured on unseen data. The RGB and infrared datasets were divided into five folds, four folds for training and one fold for validation. The data was split at the chick ID level to ensure all frames from a given chick were assigned to a single fold. To maintain class balance across folds, male and female IDs were first separated, sorted alphabetically within each class, and then evenly distributed into five subsets. Corresponding male and female folds were combined to form the final folds. As a result, the distribution of images per fold for training and evaluation for RGB and infrared dataset are presented in Tables 4 and 5 respectively.

Table 4

Distribution of IDs and RGB images per fold.

Fold	Gender	Training (IDs / Images)	Evaluation (IDs / Images)
Fold 0	Female	140 / 5,651	36 / 1918
	Male	111 / 9819	28 / 3,172
Fold 1	Female	141 / 5,783	35 / 1,786
	Male	111 / 10,258	28 / 2,733
Fold 2	Female	141 / 6,231	35 / 1,338
	Male	111 / 9,999	28 / 2,992
Fold 3	Female	141 / 6,459	35 / 1,110
	Male	111 / 11,086	28 / 1,905
Fold 4	Female	141 / 6,152	35 / 1,417
	Male	112 / 10,802	27 / 2,189

As observed in Tables 4 and 5, the greater number of male images is a consequence of the longer duration of male chick videos. This reflects the difficulty of detecting the small white spot inside the vent, which is the key feature used to identify males and requires more time to confirm visually. In contrast, female classification does not involve identifying such feature, making the recording process comparatively simpler.

Likewise, Table 6 represents the data distribution across the five folds for the RGB-IR fusion setting. In this configuration, all available paired RGB-IR images are used both for training and evaluation. For each ID,

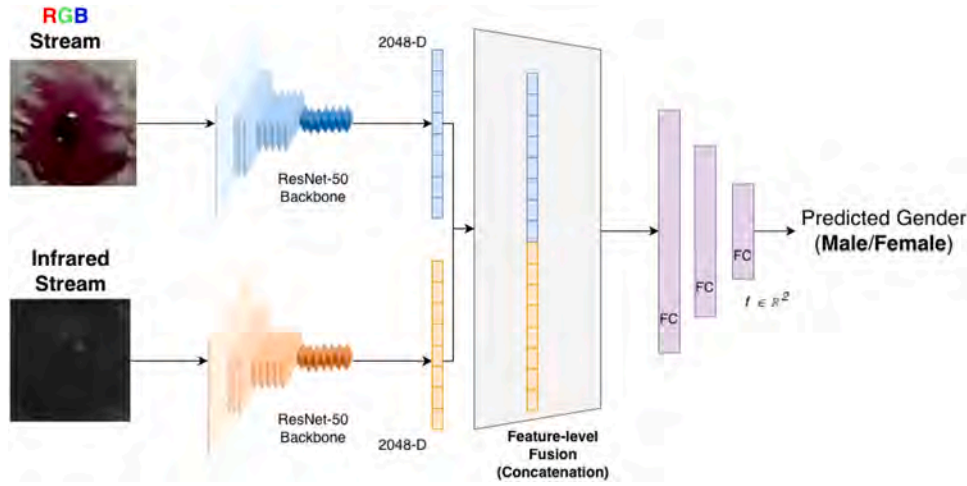


Fig. 8. Pipeline of the proposed multimodal RGB-IR fusion framework with dual ResNet-50 encoders and feature-level fusion via embedding concatenation for automated chick gender classification.

Table 5

Distribution of IDs and **infrared (IR)** images per fold.

Fold	Gender	Training (IDs / Images)	Evaluation (IDs / Images)
Fold 0	Female	140 / 5,597	36 / 1,928
	Male	111 / 9,716	28 / 3,171
Fold 1	Female	141 / 5,739	35 / 1,786
	Male	111 / 10,172	28 / 2,715
Fold 2	Female	141 / 6,241	35 / 1,284
	Male	111 / 9,935	28 / 2,952
Fold 3	Female	141 / 6,415	35 / 1,110
	Male	111 / 11,027	28 / 1,860
Fold 4	Female	141 / 6,108	35 / 1,417
	Male	112 / 10,698	27 / 2,189

the number of paired samples corresponds to the minimum between the number of RGB and IR frames available. Unpaired images are discarded. Consequently, the total number of paired samples is always less than or equal to the counts reported in the single-modality [Tables 4 and 5](#). Importantly, the number of IDs remains unchanged, since only identities with images in both modalities are included.

Table 6

Distribution of IDs and **paired RGB-IR** images per fold.

Fold	Gender	Training (IDs, Images)	Evaluation (IDs, Images)
Fold 0	Female	140 / 5,597	36 / 1,918
	Male	111 / 9,716	28 / 3,171
Fold 1	Female	141 / 5,729	35 / 1,786
	Male	111 / 10,172	28 / 2,715
Fold 2	Female	141 / 6,231	35 / 1,284
	Male	111 / 9,935	28 / 2,952
Fold 3	Female	141 / 6,405	35 / 1,110
	Male	111 / 11,027	28 / 1,860
Fold 4	Female	141 / 6,098	35 / 1,417
	Male	112 / 10,698	27 / 2,189

Within each validation fold, the accuracy parameter was optimized by selecting the best-performing model over 10 epochs given by the multilayer perceptron (MLP). The final performance accuracy was then determined by averaging the accuracies obtained across all folds. For image-based classification, training relied on binary cross-entropy loss to optimize classification performance. Both image modalities were also trained using the Adam optimizer [28], and a learning rate of 1×10^{-5} .

2.6. Results and discussion

2.6.1. Quantitative results

This section presents a comparative analysis of the gender classification model performance obtained from single-modality models (RGB and infrared) and the multimodal fusion framework. The evaluation is conducted across five folds, utilizing the performance metrics detailed earlier in [Section 2.4](#).

i. Image-Based Gender Classification

[Table 7](#) summarizes the chick gender classification results on the RGB image-based dataset. The average ID accuracy achieved by the ResNet-50 model is 97.17%, with recall, precision, and F1-score also consistently exceeding 97%, demonstrating consistent and robust performance across the five folds. Notably, fold 3 reached 100% accuracy, underscoring the model's capability to extract features that provide distinctive information for gender differentiation. The slight variation observed in fold 0 (90.62% accuracy) suggests differences in data complexity or variations in training conditions. This could be attributed to variations in lighting, as the chick was repositioned around the camera to capture the genital area. Overall, these results confirm the effectiveness and stability of ResNet-50 for chick gender classification task using RGB images, as the cross-validation process ensures that performance is not biased toward specific classes.

Table 7

Performance comparison on RGB image set with five folds.

Folds	RGB Image set				
	Accuracy	Recall	Precision	F1-score	AUC
Fold 0	90.62	90.87	90.39	90.54	94.64
Fold 1	98.41	98.57	98.28	98.40	100
Fold 2	98.41	98.21	98.61	98.39	99.49
Fold 3	100	100	100	100	100
Fold 4	98.39	98.15	98.61	98.35	99.37
Average	97.17	97.16	97.18	97.14	98.70

Similarly, [Table 8](#) outlines the results for the infrared image-based dataset. The performance metrics exhibit higher values across all evaluation metrics compared to the RGB image set, with an average accuracy of 97.46%, indicating a small increase in classification effectiveness. Folds 1, 2, and 3 achieved 100% accuracy. This can be attributed to the observation that accuracy is computed at the ID level, meaning multiple images contribute to the final classification decision, ensuring robust evaluation, reducing the likelihood of overfitting. Consequently, it reflects the model's ability to generalize well on unseen data.

Table 8
Performance comparison on infrared image set with five folds.

Folds	Infrared Image set				
	Accuracy	Recall	Precision	F1-score	AUC
Fold 0	93.75	94.05	93.53	93.69	95.54
Fold 1	100	100	100	100	100
Fold 2	100	100	100	100	100
Fold 3	100	100	100	100	100
Fold 4	93.55	93.86	93.31	93.49	98.52
Average	97.46	97.58	97.37	97.44	98.81

The following Table 9 presents the classification results obtained using the RGB-Infrared multimodal fusion. In this approach, features from each modality are extracted independently using two ResNet-50 backbones, with their classification layers removed and replaced by a custom fully connected network. Those features are concatenated into a single vector and passed through three fully connected layers to predict the gender probability. During training, all available paired RGB-IR samples belonging to the training IDs of each fold are used, ensuring one-to-one correspondence between RGB and infrared images for every chick. In evaluation, the model is tested exclusively on the held-out validation IDs, using all paired RGB-IR images from those individuals. Predictions from all paired RGB-IR samples belonging to the same ID are averaged to obtain a single score per individual, from which the final performance metrics are computed at the ID level. This represents feature-level fusion, where paired RGB-IR features are learned jointly within a single network rather than combining predictions from separate models.

This strategy yields strong and reliable performance across all five folds, with an average accuracy of 98.09%. Recall, precision, and F1-score remain high across folds, averaging 97.07%, 98.60%, and 97.80%, respectively. Notably, no fold scored below 96% accuracy, indicating consistent generalization. When compared with models trained on individual modalities, the fused approach outperforms both RGB-only (97.17%) and infrared-only (97.46%) average accuracy. While the improvement in accuracy may seem marginal, the fusion model consistently leads across all major metrics, demonstrating enhanced robustness in classification. Most notably, the fused method achieves the best AUC (99.18%), reflecting a higher ability to correctly rank female chicks higher than male chicks based on probability scores, without relying on a fixed threshold. These findings indicate that RGB and infrared cues provide complementary information, enabling the fusion model to capture patterns that single-modality models overlook.

Table 9
Performance comparison on fused RGB-IR image set with five folds.

Backbones	Fusion of RGB and Infrared Image set				
	Accuracy	Recall	Precision	F1-score	AUC
Fold 0	96.88	96.43	96.43	96.43	96.53
Fold 1	98.41	100	96.55	98.25	100
Fold 2	98.41	96.43	100	98.18	100
Fold 3	100	100	100	100	100
Fold 4	96.77	92.59	100	96.15	99.37
Average	98.09	97.07	98.60	97.80	99.18

Further insights are provided by the confusion matrices in Fig. 9, which illustrate the classification performance, showing, from left to right, the results for the RGB and infrared single-modality models, and for the multimodal RGB-IR fusion framework. Each matrix illustrates the average number of correctly and incorrectly predicted IDs per gender, computed across the five cross-validation folds. The fused model achieves the most balanced outcome overall, with an average of 0.4 misclassified female and 0.8 misclassified male IDs. While infrared-only slightly outperforms the fusion model in male classification (0.4 misclassified males), it shows a higher error for females (1.2). This results in a

more balanced overall error distribution, although the improvement is primarily driven by gains in female classification.

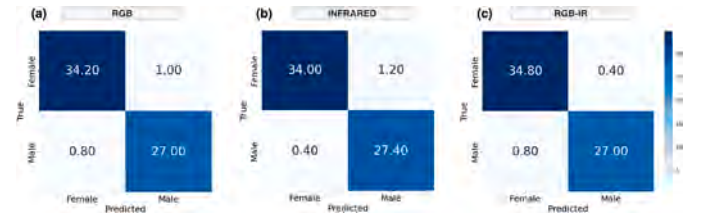


Fig. 9. Confusion matrices showing the number of correctly and incorrectly classified IDs for (a) RGB, (b) infrared, and (c) RGB-IR models.

Additionally, another graphical representation of the binary results is the ROC curve [27]. The ROC curve illustrates the trade-off between sensitivity (true positive rate) and specificity (false positive rate) in different classification thresholds. A true positive rate of 1.0 and a false positive rate of 0.0 will lead to a perfect classifier. As shown in Fig. 10, both single-modality (RGB and infrared) and multimodal fusion models demonstrate high sensitivity and specificity, with AUC values close to the perfect classifier. The RGB-IR fusion framework shows the most consistent behavior, achieving perfect AUC in three out of five folds. The infrared-only modality closely follows, with slightly more variability but still excellent AUC values. The RGB-only performs similarly, though it shows a reduced sensitivity in fold 0. Overall, the fusion approach confirms its strong generalization capability.

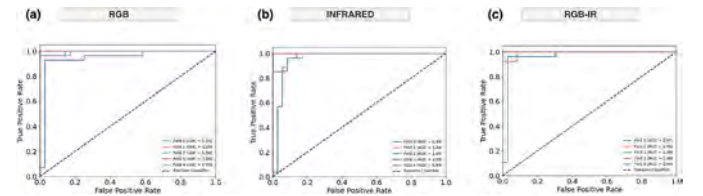


Fig. 10. Receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) scores for the ID classification task using (a) RGB, (b) infrared, and (c) RGB-IR multimodal fusion.

To contextualize the achieved performance, Table 10 summarizes representative results from related chick sexing approaches.

The proposed method achieves an accuracy of 98.09%, which is comparable to expert-level manual vent sexing. Feather-based methods report high accuracy under specific genetic conditions, while vent-based approaches maintain broader applicability across breeds.

2.6.2. Qualitative results

Once the quantitative results have been analyzed, another way to understand what the model is learning is by visual inspection. For this purpose, in this section, we explore the use of the GRAD-CAM ++ (Gradient weighted Class Activation Mapping plus) technique [29]. This technique will provide a visual understanding of the regions within the input images that contribute most to the model's classification decisions. Warmer areas represented with brighter colors in the resulting heatmaps indicate where the model's attention is most concentrated.

The resulting heatmaps reveal that the model consistently focuses on the vent region, which corresponds anatomically to where gender-specific characteristics are located. This observation aligns well with expert practices, as the cloacal area contains subtle morphological cues that experts rely on for manual classification. Specifically, in male samples, the model tends to highlight the small white bump located at the bottom of the cloaca. For female samples, the model also focuses on the inner vent region, suggesting it is also learning from internal patterns as well. Each image in Fig. 11 displays the original RGB on the left and the corresponding Grad-CAM ++ heatmap on the right, enabling a clear visual comparison of input and model focus. Additionally, male images in

Table 10
Contextual comparison of chick sexing approaches in terms of accuracy and operational characteristics.

Study / Method	Sexing Type	Modality	Reported Accuracy	Operational Characteristics
Manual vent sexing [13]	Vent	Human inspection	>98%	Invasive Expert-dependent Labor-intensive
Otsuka et al. [7]	Vent	Endoscopic imaging	~98%	Invasive Specialized equipment Operator handling required
Escamilla-Garcia et al. [19]	Feather	RGB imaging	94.4%	Non-invasive Breed-limited Genetic line dependent
TARGAN [20]	Feather	RGB imaging	>97%	Industrial system Breed-limited Proprietary equipment
VentVision (Ours)	Vent	RGB + IR imaging	98.09%	Image-based Automated classification Reduced expert dependency

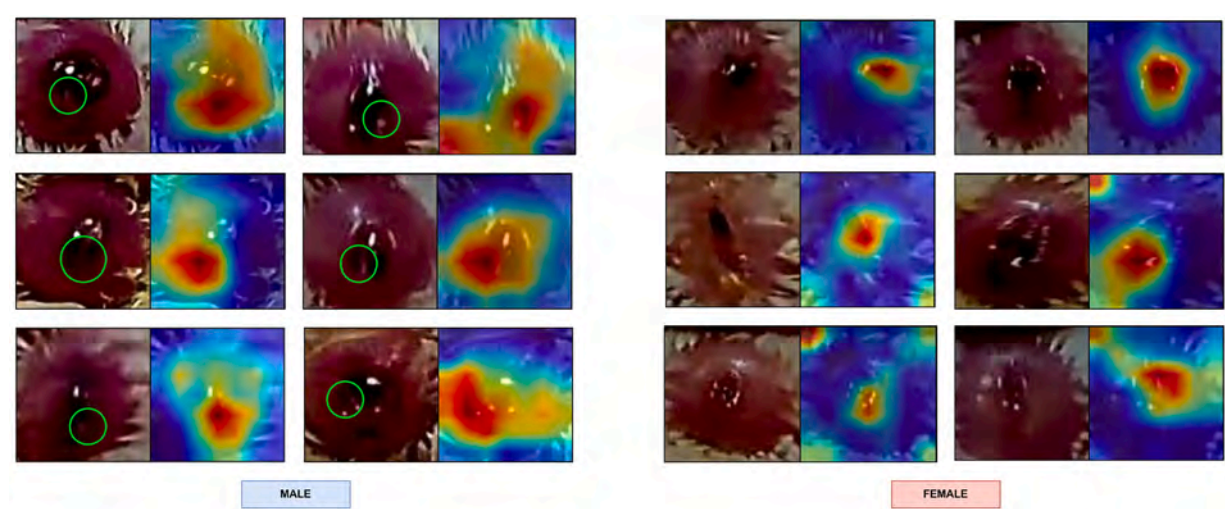


Fig. 11. Representation of correct gender predictions over the RGB modality.

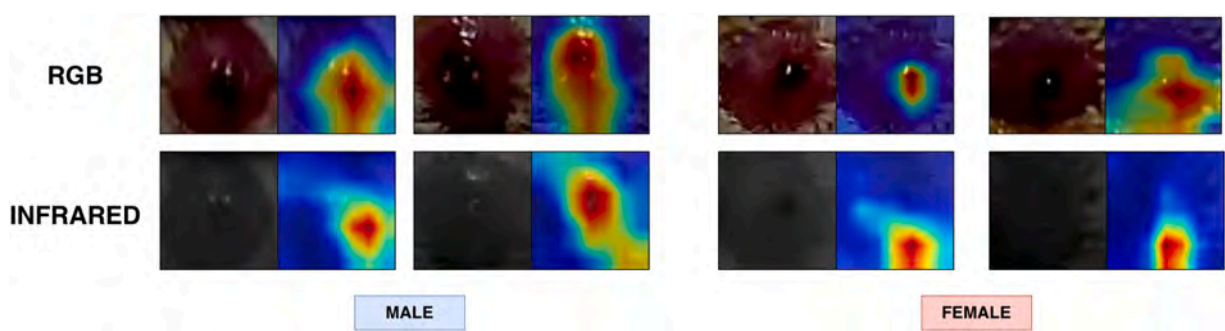


Fig. 12. Infrared Grad-CAM visualization with RGB reference for correctly classified samples. The top row shows RGB images and Grad-CAM overlays for anatomical context, while the bottom row shows the corresponding infrared images and Grad-CAM overlays, showing activation over the cloacal region.

Fig. 11 show the small white feature visually emphasized with a green circle to facilitate interpretation.

Similarly, the Grad-CAM visualizations for the infrared model are shown in Fig. 12. Because anatomical structures are more difficult to interpret directly in infrared images, the corresponding RGB frames are included as visual references. The top row shows the RGB images and RGB Grad-CAM overlays, while the bottom row shows the corresponding infrared images and infrared Grad-CAM overlays. The infrared Grad-

CAM maps show activation over the cloacal region, consistent with the localization observed in the RGB visualizations.

Additionally, Fig. 13 shows an example in which the infrared model correctly classified the sample and produced a more localized Grad-CAM activation within the cloacal region than the corresponding RGB model. This qualitative result is consistent with the quantitative improvement observed for the infrared-only model compared with the RGB-only model.

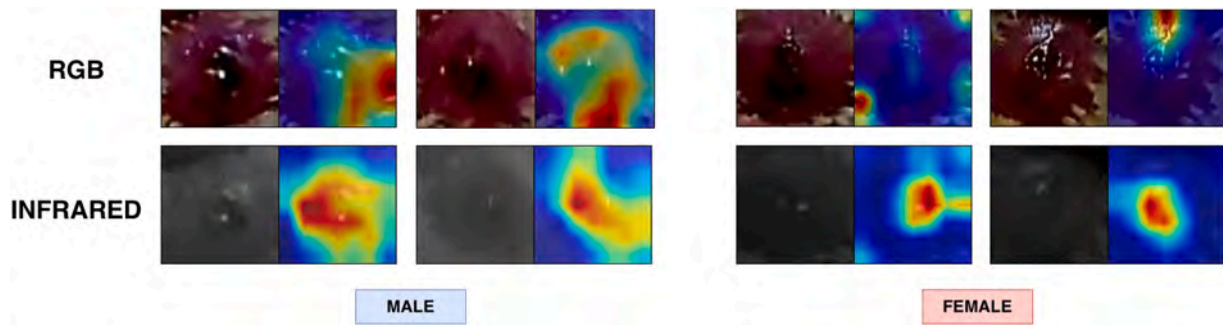


Fig. 13. Qualitative comparison of RGB and infrared Grad-CAM localization for the same chick ID. The infrared Grad-CAM shows a more localized activation within the cloacal region, whereas the RGB activation is more diffuse. The RGB image is included to provide anatomical context.

For the RGB-IR fusion model, feature integration occurs at the representation level after modality-specific feature extraction. Since the RGB and infrared embeddings are concatenated into a non-spatial feature vector before classification, a single fused Grad-CAM attention map cannot be directly generated for the final multimodal prediction.

3. Discussion

This study demonstrates the feasibility of automating the traditional vent chick sexing technique using a multimodal computer vision approach. The results show that the multimodal RGB-IR model achieves higher overall classification performance and more stable behavior across cross-validation folds compared to single-modality approaches. In addition, the model reaches performance levels comparable to expert-level manual vent sexing. Qualitative analysis further indicates that the model consistently focuses on the genital region, suggesting that predictions are driven by anatomically relevant features rather than spurious correlations.

The effectiveness of the proposed approach can be attributed to the complementary nature of RGB and infrared imaging. While RGB captures structural and textural details, infrared imaging enhances contrast in regions that are difficult to distinguish visually, particularly in scenarios where visible features are difficult to distinguish due to poor illumination or low visual contrast [30].

Although the fused model achieves the strongest overall performance, the confusion matrices indicate that the benefit of multimodal fusion is not uniform across both classes. RGB produces fewer female errors than IR (1.00 vs 1.20), whereas IR produces fewer male errors (0.40 vs 0.80). The fusion model further reduces female errors (0.40), but does not retain the lower male error observed with IR alone (0.80 vs 0.40). This suggests that the two modalities capture class-relevant anatomical cues differently. Despite this class-dependent behavior, the fusion model maintains the best overall performance while producing a more balanced distribution of errors across classes.

This class-dependent behavior may be explained by differences in the visual cues required for female and male classification. Female classification relies more on the overall visual appearance of the vent region, including its shape, texture, and continuity, whereas male classification depends more on identifying a small, localized protruding structure. In practice, this structure can vary in size, may be weakly expressed, and is highly dependent on proper eversion and imaging conditions, making it difficult to detect consistently [10,31]. Because RGB images depend on visible-light illumination, they can be affected by reflections, shadows, and illumination variability. In our task, such artifacts may resemble male-specific structures and contribute to false positive male predictions. The use of infrared imaging may help mitigate these effects by providing more stable contrast, thereby reducing female-to-male misclassifications. Conversely, errors in male classification are more likely to occur when the relevant structure is partially occluded, poorly exposed, or not clearly visible in the captured image.

The combination of RGB and infrared information therefore appears beneficial for female samples, where integrating structural cues from RGB with contrast enhancement from infrared helps resolve ambiguous cases. However, male classification depends on reliably capturing localized contrast cues associated with this protruding structure. In these situations, the addition of broader structural information from RGB may reduce the influence of these localized contrast cues captured by infrared imaging, making the detection of the relevant feature less consistent. This provides a possible explanation for why fusion improves female classification while not preserving the lower male error observed with infrared alone.

From an industrial perspective, current automation efforts in chick sexing have largely focused on feather-based systems, such as WingScan or Genesys [20,21], which rely on clearly visible morphological traits. In contrast, automated vent-based solutions remain limited and are typically not publicly documented [24], restricting reproducibility and broader adoption. Alternative approaches, such as in-ovo sexing technologies, have been proposed; however, their deployment is constrained by high infrastructure costs and operational complexity [32]. As a result, manual vent sexing continues to be widely used in commercial hatcheries, highlighting the need for accessible and scalable solutions.

The proposed system addresses this need by reducing dependence on highly trained experts. While some level of operator involvement is still required to handle and position the chick, the task no longer demands years of specialized training. This facilitates deployment across a wider range of hatchery environments, particularly in settings where access to skilled personnel is limited, and supports improved operational efficiency.

Despite these promising results, the approach still requires manual handling and is sensitive to imaging conditions. Future work will focus on improving robustness under less controlled conditions, enabling real-time in hatchery environments, and evaluating broader architectural benchmarking.

4. Conclusion

In conclusion, this study demonstrates the feasibility of automating vent sexing in day-old chicks using computer vision techniques across RGB, infrared, and multimodal imaging. The proposed approach achieved high classification performance, exceeding 97% accuracy for individual modalities and 98% with multimodal fusion, indicating that RGB and infrared imaging provide complementary information for this task. The multimodal model achieved the strongest overall performance and a more balanced error distribution, although its benefit was class-dependent. These results highlight the potential of reducing reliance on expert-level interpretation in vent sexing. However, the method still depends on manual cloacal eversion and controlled data acquisition conditions, which may limit its direct applicability in fully automated industrial environments. Future work will focus on improving robustness

under less controlled conditions, further reducing manual intervention, and evaluating broader model architectures and fusion strategies.

CRedit authorship contribution statement

Marta Veganzones-Rodriguez: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Thinh Phan:** Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Conceptualization; **Arthur F.A. Fernandes:** Writing – review & editing, Validation, Resources, Data curation; **Vivian Breen:** Writing – review & editing, Validation, Resources, Data curation; **Jesus Arango:** Writing – review & editing, Validation, Resources, Data curation; **Michael T. Kidd:** Writing – review & editing, Resources, Data curation; **Ngan Le:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Data availability

The data that has been used is confidential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is based upon work supported by University of Arkansas.

References

- [1] N. Jia, B. Li, J. Zhu, H. Wang, Y. Zhao, W. Zhao, A review of key techniques for in ovo sexing of chicken eggs, *Agricultural* 13 (3) (2023) 677.
- [2] M.-E. Krautwald-Junghanns, K. Cramer, B. Fischer, A. Förster, R. Galli, F. Kremer, E.U. Mapesa, S. Meissner, R. Preisinger, G. Preusse, et al., Current approaches to avoid the culling of day-old male chicks in the layer industry, with special reference to spectroscopic methods, *Poult. Sci.* 97 (3) (2018) 749–757.
- [3] J.N. Sreedhara, L. Yesappa, K.V. Prakash, V. Raghavendra, A review on poultry chick sex identification: traditional, molecular and advanced methods, *Uttar Pradesh J. Zool.* 46 (20) (2025) 8–15.
- [4] R. Preisinger, Sex determination in poultry-breeder's view, *Avian Sex Determ. Sex Diagn.* 59 (2003) 54.
- [5] S. Aerts, R. Boonen, V.E. Bruggeman, J. De Tavernier, E. Decuyper, Culling of day-old chicks: opening the debates of moria?, in: *Ethical Futures: Bioscience and Food Horizons*, Wageningen Academic, 2009, pp. 115–122.
- [6] U. Asil, E. Nasibov, Sex detection in the early stage of fertilized chicken eggs via image recognition, (2023). [arXiv:2305.02325](https://arxiv.org/abs/2305.02325)
- [7] M. Otsuka, O. Miyashita, M. Shibata, F. Sato, M. Naito, A novel method for sexing day-old chicks using endoscope system, *Poult. Sci.* 95 (11) (2016) 2685–2689.
- [8] R. Feltwell, *Small-Scale Poultry Keeping: A Guide to Free-Range Poultry Production*, Faber & Faber, 2011.
- [9] H. Cerit, K. Avanus, Sex identification in avian species using DNA typing methods, *World's Poul. Sci. J.* 63 (1) (2007) 91–100.
- [10] I. Biederman, M.M. Shiffrar, Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task, *J. Exp. Psychol. Learn. Mem. Cogn.* 13 (4) (1987) 640.
- [11] S.D. Lu, J.I. Padilla, L. Asis, J.C. Basa, H.D. Menguito, C. Lazana, E. Versoza, A. Bandala, L. Asis, Chick-check: design and development of a feather sexing machine for chicks using YOLOv10 and UV optical system, in: *2024 IEEE 16th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, IEEE, 2024, pp. 1–6.
- [12] C.S. Gibbs, *A Guide to Sexing Chicks*, Read Books Ltd, 2016.
- [13] P. Phelps, A. Bhutata, S. Bryan, A. Chalk, B. Ferrell, S. Neuman, C. Ricks, H. Tran, T. Butt, Automated identification of male layer chicks prior to hatch, *Worlds Poul. Sci. J* 59 (1) (2003) 33–38.
- [14] W.J. Miller, F.H. Wagner, Sexing mature columbiformes by cloacal characters, *Auk* 72 (3) (1955) 279–285.
- [15] F. Ellendorf, S. Klein, S. Nandi, D. McBride, R. Blanco, M. Clinton, I. Hardy, Avian sex determination and sex diagnosis, *World's Poul. Sci. J.* 59 (1) (2003) 5–6.
- [16] P. Tripathi, N. Kumar, M. Rai, P.K. Shukla, K.N. Verma, Applications of machine learning in agriculture, in: *Smart Village Infrastructure and Sustainable Rural Communities*, IGI Global, 2023, pp. 99–118.
- [17] M. Tran, S. Truong, A.F.A. Fernandes, M.T. Kidd, N. Le, CarcassFormer: an end-to-end transformer-based framework for simultaneous localization, segmentation and classification of poultry carcass defect, *Poult. Sci.* 103 (8) (2024) 103765.
- [18] L. Yang, J. Song, Real-time detection of dead broilers in high-density poultry houses using high-resolution video and semantic segmentation, *Poult. Sci.* (2026) 106969.
- [19] A. Escamilla-Garcia, G.M. Soto-Zarazúa, M. Toledano-Ayala, A. Gastélum-Barrios, A new application of morphometric variables and image processing to determine day-old chicken sex, *J. Appl. Res. Technol.* 20 (5) (2022) 564–569.
- [20] Targan, Automated feather sexing enhancing poultry production performance, 2023, (Accessed February 2025). <https://www.targan.com/wingscan>.
- [21] Ceva, GeneSys: gender sorting system, (Accessed September 2025), (2024) <https://poultrycontent.ceva.com/ceva-genesys-gender-sorting-system.com>.
- [22] R.D. Reynnells, C.J. Flegal, Method and apparatus for avian pre-hatch sex determination, 2000. US Patent 6,029,080.
- [23] W. Blount, *Sexing Day-Old Chicks-A Treatise on Sex Detection in Pure and Cross-Breed Chicks*, Read Books Ltd, 2012.
- [24] Xiashu Technology, Automatic chicken sexing system, (Accessed 7 April 2026), (2024). <http://www.xiashutech.com/>.
- [25] T. Phan, I. Phillips, A. Lockett, M.T. Kidd, N. Le, TrackMe: A Simple and Effective Multiple Object Tracking Annotation Tool, (2024). [arXiv:2410.15518](https://arxiv.org/abs/2410.15518)
- [26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [27] G. Developers, Classification: ROC and AUC, 2024, (Accessed March 2025). <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>.
- [28] Z. Zhang, Improved adam optimizer for deep neural networks, in: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, IEEE, 2018, pp. 1–2.
- [29] A. Chattopadhyay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-cam ++: generalized gradient-based visual explanations for deep convolutional networks, in: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018, pp. 839–847.
- [30] J. Ma, Y. Ma, C. Li, Infrared and visible image fusion methods and applications: a survey, *Inf. Fusion* 45 (2019) 153–178.
- [31] T.H. Canfield, Sex determination of day-old chicks, *Poult. Sci.* 19 (4) (1940) 235–238.
- [32] M. Lapididou, Genus focus non-invasive in-ovo sexing fully aligned with animal welfare standards, 2023, (Accessed January 2026). <https://www.orbem.ai/newsroom/genus-focus-aligned-to-animal-welfare>.