Poultry-CViTNet. A Hybrid ViT--CNN Network for Non-Invasive Poultry Health Monitoring Based on Fecal Images

Syed Mohammad Ali Shah[1], Shams ur Rahman[1], Khalid Khan[2], Abid Elahi[3] and Syed Jaffar Ali Shah[4]

[1]Department of Computer Software Engineering, University of Engineering and Technology Mardan, Pakistan.

[2]Institute of Computer Science and Information Technology, University of Agriculture, Peshawar, Pakistan.

[3]Department of Computer Sciences and Information Technology, Sarhad university of science and information technology Peshawar.

[4]Department of Computer Science Brains postgraduate college Peshawar, Khyber Pakhtunkhwa, Pakistan.

*__Correspondence__: shams@uetmardan.edu.pk

Citation | Shah. S. M. A, Rahman. S. U, Khan. K, Elahi. A, Shah. S. J. A, "Poultry-CViTNet. A Hybrid ViT--CNN Network for Non-Invasive Poultry Health Monitoring Based on Fecal Images", IJIST, Vol. 8 Issue 5 pp 1995-2011, August 2026

DOI | https://doi.org/10.33411/IJIST/1969

Received | July 22, 2026 **Revised |** Aug 12, 2026 **Accepted |** Aug 15, 2026 **Published |** Aug 18,2026.

Poultry diseases remain a major challenge for sustainable poultry production because delayed diagnosis can increase mortality, treatment costs, and the risk of disease transmission across farms. Conventional diagnostic methods based on manual observation, veterinary inspection, and laboratory testing are often time-consuming, labor-intensive, and difficult to scale in low-resource farming environments. To address these limitations, this study proposes a hybrid Convolutional Neural Network (CNN) and Vision Transformer (ViT) framework for automated poultry health monitoring using fecal images. The proposed model combines the local texture extraction capability of CNNs with the global contextual modeling ability of ViTs, enabling effective discrimination between healthy and unhealthy fecal samples. The proposed CNN–ViT model was evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis, with additional ablation experiments. The experimental results show that the hybrid model achieved an accuracy of 97.47 ± 0.36%, precision of 97.50 ± 0.38%, recall of 97.47 ± 0.41%, and F1-score of 97.47 ± 0.39%, outperforming the CNN-only and ViT-only baseline models. The ablation study confirms that integrating local convolutional features with transformer-based global attention improves classification reliability. These results demonstrate the potential of the proposed framework as a non-invasive, low-cost, and scalable decision-support tool for early poultry disease screening in smart farming environments. Additionally, the model contained only 3.62 million parameters, required 2.60 GFLOPs, and processed 194.95 images per second. Furthermore, Grad-CAM-based interpretability was applied to visualize the model's decision-making process.

Keywords: Disease Prediction, Machine Learning, Rough Set Theory, Maximum Value Attribute, Extra Trees, XG Boost, Light GBM, Feature Selection, Hybrid MVA, Clinical Symptoms

Introduction:

Poultry production is one of the most important sources of animal protein, contributing substantially to the global supply of meat and eggs. It represents a rapidly expanding sector of agriculture because of its relatively low production cost, efficient feed conversion ratio, short production cycle, and high nutritional value. Poultry products provide essential proteins, vitamins, and minerals, making them an important component of the human diet. In addition, poultry farming contributes to rural employment, livelihood generation, and food security, particularly in developing economies.

With the continuous growth of the global population and increasing consumer demand for affordable animal protein, poultry production is expected to play an even greater role in meeting future food requirements. The world population is projected to reach approximately 8.6 billion by 2030, creating significant pressure on agricultural systems to produce sufficient and sustainable food resources [1]. However, the increasing demand for poultry products also requires more efficient production systems capable of addressing major challenges such as disease outbreaks, environmental stressors, biosecurity risks, and the maintenance of overall poultry health [2]. Therefore, reliable and timely poultry health monitoring has become essential for improving productivity, reducing economic losses, and supporting sustainable poultry farming.

In recent years, deep learning has emerged as a powerful tool for automated disease detection and diagnosis [3][4][5][6][7][8][9][10]. Unlike conventional diagnostic approaches that depend heavily on manual observation, laboratory testing, or expert veterinary inspection, deep learning models can automatically learn complex visual patterns from large-scale image datasets. In particular, computer vision techniques based on convolutional neural networks (CNNs) have shown strong potential for analyzing poultry-related images by identifying disease-indicative changes in texture, color, shape, and surface appearance.

These capabilities make deep learning highly suitable for non-invasive poultry health assessment, especially when visual symptoms appear in fecal samples, body posture, comb color, or behavioral patterns. By enabling fast, consistent, and scalable analysis, deep learning-based systems can support early disease detection, reduce diagnostic delays, minimize economic losses, and assist farmers in making timely management decisions.

Although existing poultry disease detection studies have reported promising results using machine learning, CNN-based architectures, object detection models, acoustic analysis, Internet of Things (IoT)-based monitoring, and hybrid CNN–Transformer networks, several research gaps remain. First, many existing approaches depend mainly on either local visual features or global image-level representations. Such representations may be insufficient for fecal-image analysis because disease-related cues can simultaneously appear as fine-grained texture changes, color variations, shape irregularities, and spatially distributed appearance patterns.

Second, several reported models have been evaluated under limited experimental settings and do not sufficiently discuss model robustness, interpretability, computational efficiency, or practical deployment feasibility. Third, previous poultry health-monitoring studies often focus primarily on overall classification accuracy, while limited attention has been given to designing a fecal-image-specific local–global feature-learning framework for non-invasive poultry health assessment.

To address these limitations, this study proposes Poultry-CViTNet, a hybrid CNN–ViT framework specifically designed for poultry fecal-image-based health monitoring. Unlike conventional CNN models that primarily emphasize local feature extraction, the proposed framework integrates convolutional texture encoding with transformer-based global contextual modeling. The CNN branch captures local disease-related visual patterns, such as fecal texture, color changes, boundary irregularities, and morphological variations, whereas the

ViT branch models long-range dependencies among different image regions through self-attention.

By fusing these complementary representations, Poultry-CViTNet provides a more discriminative feature representation for classifying healthy and unhealthy poultry samples.

The main contributions of this work are as follows:

A lightweight dual-branch CNN–ViT model is proposed to capture both local visual features and global contextual representations for poultry health classification using fecal images.

A dual-stream fusion network is introduced to adaptively combine local and global features according to their learned feature weights.

A comprehensive experimental evaluation is conducted using five-fold cross-validation, ablation studies, and comparative analysis with state-of-the-art methods.

The computational efficiency of the proposed model is also evaluated in terms of parameters, floating-point operations, and throughput, demonstrating that it achieves high classification accuracy with a low computational cost suitable for real-world deployment.

The remainder of this paper is organized as follows. Section 2 presents the relevant literature. Section 3 describes the experimental setup, dataset, and proposed model architecture. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper and outlines future research directions.

Novelty of the Study:

The novelty of Poultry-CViTNet lies in its fecal-image-specific integration of local convolutional feature learning and global transformer-based attention for non-invasive poultry health monitoring. Unlike previously published generic hybrid CNN–ViT models, the proposed lightweight dual-branch architecture separately captures fine-grained fecal texture, color, boundary, and morphological patterns through the CNN branch, while the ViT branch models long-range contextual relationships among spatially distributed image regions. These complementary representations are adaptively fused to improve the discrimination of healthy and unhealthy poultry samples while maintaining low computational complexity suitable for practical deployment.

Related Work:

Image-based disease detection is rapidly advancing in agriculture because of recent developments in computer vision and deep learning. In poultry farming, visual symptoms provide important indicators for early disease detection. Among these indicators, fecal appearance is commonly used by veterinary experts to assess bird health. Changes in fecal color, texture, shape, and consistency may indicate infections or abnormal health conditions. Therefore, automated fecal-image analysis can support faster, non-invasive, and scalable poultry disease diagnosis.

Several machine-learning and deep-learning approaches have been proposed for poultry disease detection and health monitoring. [11] developed a solar-powered system integrating a decision tree model to classify six chicken symptoms using audio and visual data. The system achieved 84.6% accuracy using visual data and 86.1% accuracy using combined visual and audio data. [12] proposed a three-way comb-color analysis method based on the CIE XYZ color space and logistic regression, achieving 100% sensitivity and 83% specificity. [13] employed the ChicNetV6 classifier to diagnose Salmonella, coccidiosis, and Newcastle disease, achieving an accuracy of 94.49%.

Deep-learning-based image-classification models have also demonstrated promising performance in poultry disease recognition. [14] employed a Convolutional Neural Network (CNN) model and achieved an accuracy of 93.23%. [15] used MobileNetV2 and reported an accuracy of 82%. Similarly, [16] proposed a posture-estimation model based on different deep

neural network and object-detection architectures, including YOLOv3, RetinaNet, and Faster R-CNN. The model achieved a precision of 92.18% and a recall of 89.96%.

[2] employed XceptionNet and achieved an accuracy of 94%, while a fully trained CNN achieved 93.67% accuracy. [2] proposed a YOLOv5-based detection model and achieved an accuracy of 89.2%. More recently, [17] introduced ResTFG, a hybrid CNN–Vision Transformer model, and achieved an accuracy of 96.9%.

In addition to image-based methods, several studies have explored audio-, sensor-, and Internet of Things-based approaches for poultry health monitoring. [18] developed a sound-based model using Mel-frequency cepstral coefficients (MFCCs) and a CNN, achieving accuracies of 97.43% and 95.84% under different experimental settings. [19] proposed a poultry disease-classification system based on DenseNet and MobileNet architectures, achieving an accuracy of 97%.

[20] developed a CNN-based system using visual features such as edges and Harris corners, achieving an accuracy of 94.14%. [21] used auditory biomarkers to identify bronchitis and Newcastle disease through wavelet entropy, MFCC features, and a Support Vector Machine classifier. Their approach achieved accuracies of 83% for bronchitis and 80% for Newcastle disease. [22] introduced a remote poultry monitoring system based on ResNet-FPN and achieved an accuracy of 93.7%. [23] combined generative adversarial networks, machine-learning methods, wearable sensors, and Internet of Things (IoT) technologies in a real-time predictive model, achieving an accuracy of 97%.

A summary of the existing studies on poultry disease detection and health monitoring is presented in Table 1.

Table 1. Summary of related studies on poultry disease detection and health monitoring.

Reference	Model Used	Accuracy/Performance
[11]	Decision Tree	84.6% using visual data; 86.1% using visual and audio data
[12]	Logistic Regression and CIE XYZ color space	83% specificity
[13]	ChicNetV6, CNN, and Transfer Learning	94.49% accuracy
[14]	CNN	93.23% accuracy
[15]	MobileNetV2	82% accuracy
[16]	DNN, YOLOv3, RetinaNet, and Faster R-CNN	92.18% precision and 89.96% recall
[2]	CNN and XceptionNet	94% using XceptionNet; 93.67% using a fully trained CNN
[17]	YOLOv5	89.2% accuracy
[18]	ResTFG, CNN, and Transformer	96.9% accuracy
[19]	MFCC and CNN	97.43% and 95.84% accuracy
[20]	DenseNet and MobileNet	97% accuracy
[21]	CNN and Harris Corner Detection	94.14% accuracy
[22]	MFCC, Wavelet Entropy, and SVM	83% for bronchitis and 80% for Newcastle disease
[23]	ResNet-FPN	93.7% accuracy
[24]	Internet of Things, GAN, and Machine Learning	97% accuracy

Although previous studies have achieved promising results in poultry disease detection, several limitations remain. Many existing approaches depend on controlled imaging

conditions, limited datasets, single-feature analysis, or conventional CNN-based models that may not fully capture both local and global disease-related patterns. In practical poultry-farm environments, variations in lighting, background, image quality, and fecal appearance can reduce the reliability of automated diagnostic systems.

Furthermore, many existing studies primarily report classification accuracy without sufficiently evaluating model robustness, interpretability, computational efficiency, or deployment feasibility. Conventional CNN-based approaches are effective in capturing fine-grained local features; however, they may have limited capability to model long-range relationships among spatially separated disease-related regions. In contrast, transformer-based models can capture global contextual relationships but may overlook important local texture information, particularly when the available dataset is limited.

Therefore, there remains a need for a robust, non-invasive, interpretable, and scalable poultry health-monitoring system that can accurately analyze fecal images and support early disease detection. To address these limitations, the present study proposes a hybrid CNN–ViT framework that combines local texture extraction with global attention-based feature learning. The proposed framework is designed to improve the classification of healthy and unhealthy poultry fecal samples by integrating complementary local and global visual representations.

Proposed Methodology:

The proposed network provides an automated and non-invasive solution for poultry health assessment using fecal-image analysis, as illustrated in Figure 1. The system first standardizes the input images through preprocessing operations, including resizing, normalization, noise reduction, contrast enhancement, and data augmentation.

A hybrid CNN–ViT architecture is then employed to learn discriminative visual representations. The CNN branch captures local texture and morphological variations, whereas the ViT branch models global contextual relationships among different image regions. The processed input 224×224×3 is passed to both branches. The CNN generates feature maps of 32×112×112, 64×56×56, 128×28×28, and 256×14×14, followed by global pooling to obtain a 256-dimensional local feature vector. The ViT divides the image into 3,136 patches, represents each patch with a 64-dimensional embedding (3,136×64), and projects the transformer output to a 256-dimensional global feature vector. The adaptive fusion module therefore receives two aligned vectors, $F_{cnn}, F_{ViT} \in R^{256\times14\times14}$, and learns their relative contributions.

Finally, the effectiveness of the proposed model is evaluated using standard performance measures, including accuracy, precision, recall, F1-score, and confusion-matrix analysis. The following sections explain the complete working mechanism of the proposed Poultry-CViTNet.

Experimental Setup:

All experiments were conducted in the Kaggle environment using an NVIDIA Tesla P100 GPU to accelerate the training and testing processes. Before being passed to the proposed model, several preprocessing operations were performed. These operations included Gaussian smoothing for noise reduction, resizing all images to a fixed resolution of 224 × 224 pixels, and normalizing the pixel values.

Data-augmentation techniques were also applied during training to reduce overfitting and improve the generalization capability of the model. The experimental setup and training configuration used for the proposed CNN–ViT model are summarized in Table 2.

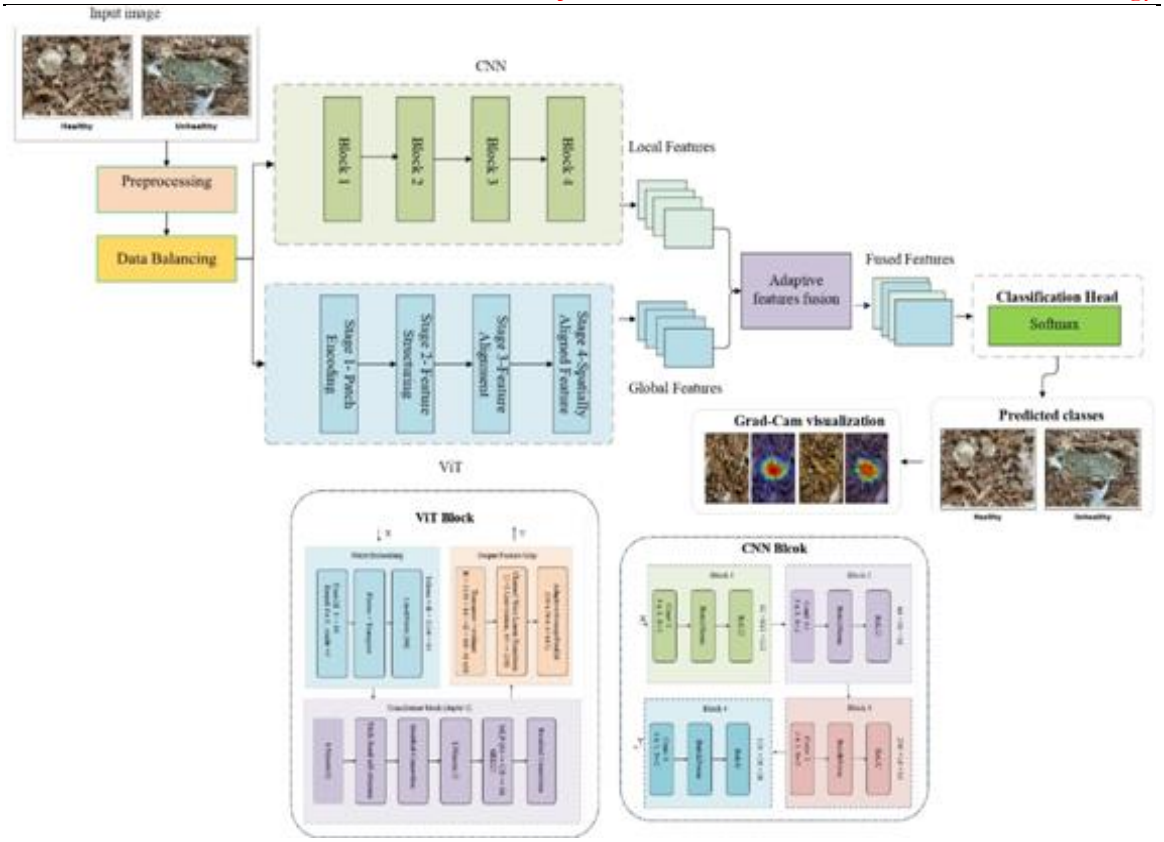


Figure 1. Proposed overall architecture of Poultry-CViTNet.

Table 2. Experimental setup and training configuration.

Parameter	Value
Experimental Environment	Experimental Environment
Validation Protocol	Validation Protocol
Number of Epochs	Number of Epochs
Learning Rate	Learning Rate
Batch Size	Batch Size
Optimizer	Optimizer
Early-Stopping Patience	Early-Stopping Patience
Loss Function	Loss Function
Input Image Size	Input Image Size
Total Training Time	Total Training Time

Dataset Description:

This study uses the publicly available Poultry Birds Poo Imagery Dataset for Health Status Prediction obtained from the Mendeley Data repository. The dataset contains fecal images of broiler and layer chickens collected from poultry farms in South-West Nigeria.

The dataset consists of two classes: healthy and unhealthy. It contains a total of 14,618 validated images stored in JPEG format. Representative samples from the dataset are shown in Figure 2. These images demonstrate considerable variations in fecal color, texture, shape, size, and background conditions.

Such variations make poultry health classification challenging and highlight the importance of effective preprocessing and robust feature extraction. The proposed Poultry-CViTNet is therefore designed to learn both fine-grained local patterns and global contextual information from the fecal images.



Figure 2. Sample images from the poultry fecal-image dataset.

Image Preprocessing:

All images are passed through a preprocessing pipeline that includes image resizing, filtering, contrast enhancement, data augmentation, normalization, and dataset balancing [23][24][25][26][27][28][29][30][31].

First, each image is resized to a fixed resolution of 224 × 224 pixels to ensure a consistent input size for the proposed model. Gaussian smoothing and bilateral filtering are then applied to reduce image noise while preserving important boundaries and texture information.

Contrast-Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance local contrast and improve the visibility of disease-related characteristics. Pixel values are subsequently normalized to provide a consistent input range and improve model convergence.

Data-augmentation techniques are applied to increase the diversity of the training samples and help address class imbalance. These techniques included image rotation, horizontal flipping, translation, scaling, and brightness variation. After preprocessing and augmentation, the dataset is divided according to a five-fold cross-validation protocol. The processed images are then passed to Poultry-CViTNet for feature extraction and classification.

CNN Branch of Poultry-CViTNet:

The CNN branch of the proposed Poultry-CViTNet is designed to extract local discriminative features from poultry fecal images. It consists of four convolutional blocks.

Each block contains a two-dimensional convolutional layer, batch-normalization layer, Rectified Linear Unit activation function, and max-pooling operation.

The input image is represented as:

$$X \in \mathbb{R}^{224 \times 224 \times 3}$$

where 224 × 224 represents the spatial resolution and 3 represents the red, green, and blue color channels.

The input image is progressively transformed into compact feature maps that capture disease-related visual characteristics, including fecal texture, color variations, boundary irregularities, and morphological patterns. The convolutional operation can be expressed as:

$$F_CNN = ReLU[BN(Conv2D(X, W) + b)] \quad (1)$$

Where X represents the input image or the feature map from the preceding layer; W represents the learnable convolutional filters; b represents the learnable bias; BN represents batch normalization; ReLU represents the Rectified Linear Unit activation function; and F_CNN represents the output feature map generated by the CNN branch.

Max pooling is applied after each convolutional block to reduce the spatial dimensions of the feature maps, suppress redundant information, and gradually increase the effective receptive field. The final CNN representation is forwarded to the feature-fusion stage, where it is combined with the global representation learned by the ViT branch.

ViT Branch of Poultry-CViTNet:

The Vision Transformer branch is designed to extract global contextual features from poultry fecal images. The input image is represented as:

$$X \in \mathbb{R}^{224 \times 224 \times 3}.$$

The image is divided into fixed-size, non-overlapping patches. Each image patch is flattened into a one-dimensional vector and projected into a higher-dimensional embedding space through a learnable linear-projection layer.

For an image with dimensions H × W × C and a patch size of P × P, the total number of patches is calculated as:

$$N = (H \times W) / P^2 \quad (2)$$

where N represents the number of image patches. Each flattened patch is transformed into a patch embedding. A learnable classification token and positional embeddings are added to the patch sequence to preserve spatial information. The embedded tokens are passed through a series of transformer encoder blocks. Each transformer encoder consists of multi-head self-attention, layer normalization, a multilayer perceptron, and residual connections.

The multi-head self-attention mechanism enables the model to capture long-range relationships among different fecal-image regions. It can therefore model global patterns involving color distribution, texture consistency, structural appearance, and spatial relationships.

The self-attention operation is expressed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}(QK^T / \sqrt{d_k})V \quad (4)$$

where: Q denotes the query matrix; K denotes the key matrix; V denotes the value matrix; and d_k denotes the dimensionality of the key vectors.

The final ViT representation contains global contextual information and is forwarded to the feature-fusion module, where it is combined with the local features extracted by the CNN branch.

Feature Fusion and Classification:

The CNN and ViT branches generate complementary feature representations. The CNN branch focuses on local texture and morphological characteristics, whereas the ViT branch captures global contextual relationships among different image regions.

Let FCNN and FViT represent the feature vectors generated by the CNN and ViT branches, with dimensions dc and dv, respectively. Since these vectors may have different dimensions, they are first projected into a common d-dimensional feature space as follows:

$$Zc = \varphi(Wc\ FCNN + bc) \quad (5)$$

$$Zv = \varphi(Wv\ FViT + bv) \quad (6)$$

Here, Wc and Wv are learnable projection matrices, b_c and b_v are bias terms, and φ represents a nonlinear activation function. The projected feature vectors Z_c and Z_v both have dimension d. The projected features are then concatenated and passed through a fully connected layer to estimate the adaptive importance weights:

$$[\alpha c, \alpha v] = \text{Softmax}(Wa\ [Zc \,||\, Zv] + ba) \quad (7)$$

where αc and αv represent the learned weights assigned to the CNN and ViT features, respectively. The Softmax function ensures that: $\alpha c + \alpha v = 1$. The final fused feature representation is obtained by calculating the weighted sum of the two projected feature vectors:

$$Ffused = \alpha c\ Zc + \alpha v\ Zv \quad (5e)$$

The resulting Ffused vector has dimension d. This adaptive fusion mechanism allows the model to dynamically control the contributions of local CNN features and global ViT features for each input image.

Training Process of Poultry-CViTNet:

The proposed Poultry-CViTNet is trained using the AdamW optimizer and cross-entropy loss. Early stopping is employed to prevent overfitting and retain the model that provides the best validation performance.

Algorithm 1. Training process of Poultry-CViTNet

Input: Dataset D, learning rate η, maximum number of epochs E, and early-stopping patience p
Output: Best-performing model parameters θ^*
1. Initialize the model parameters θ.
2. Set the best validation score to negative infinity.
3. Set the waiting counter to zero.
4. Divide the dataset into training, validation, and testing subsets according to the five-fold cross-validation protocol.
5. For each epoch from 1 to E:
6. For each mini-batch (x_b, y_b) in the training dataset:
7. Apply the preprocessing and augmentation function Π to the input images.
8. Extract local features using the CNN branch $F_C = \Phi_{CNN}(x)$
9. Extract global features using the ViT branch: $F_V = \Phi_{ViT}(x)$
10. Fuse the CNN and ViT feature representations: $F = Concat(F_C, F_V)$
11. Generate the predicted class scores: $\hat{y} = f_{cls}(F)$
12. Compute the cross-entropy loss: $L = CrossEntropy(\hat{y}, y_b)$
13. Update the model parameters using the AdamW optimizer.
14. Evaluate the model on the validation dataset and compute the validation performance score m_val.
15. When m_val is greater than the best recorded validation score:
16. Update the best validation score.
17. Save the current model parameters as θ^*.
18. Reset the waiting counter to zero.
19. Otherwise, increase the waiting counter by one.
20. Stop the training process when the waiting counter becomes equal to or greater than the predefined patience value p.
21. Load the best model parameters θ^*.
22. Evaluate the selected model on the independent test dataset.

Result and discussion:

To evaluate the performance of the proposed Poultry-CViTNet, the training and validation accuracy and loss were analyzed across the five folds. Figure 3 presents the training and validation accuracy and loss curves over the training epochs.

The loss curves show a consistent reduction as the number of epochs increases, indicating stable convergence of the proposed model. Similarly, the accuracy curves demonstrate continuous improvement during training. The training and validation curves remain close to each other throughout the training process, with no substantial divergence. This behavior indicates that the model generalizes effectively and shows no clear evidence of overfitting.

Moreover, the training and validation losses remain low after convergence, suggesting that the model successfully learns the discriminative characteristics of the poultry fecal images. The continuous decrease in loss and improvement in accuracy suggest that the model does not exhibit clear signs of underfitting.

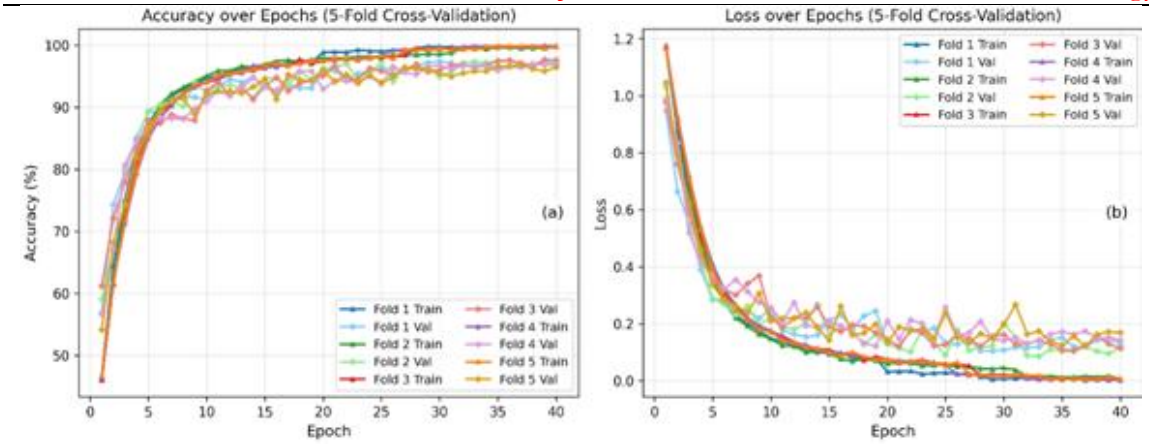


Figure 3. Show the training and validation accuracy of the proposed scheme.

To provide a comprehensive evaluation, multiple performance metrics were calculated, including accuracy, precision, recall, and F1-score. A detailed summary of these metrics is presented in Figure 4.

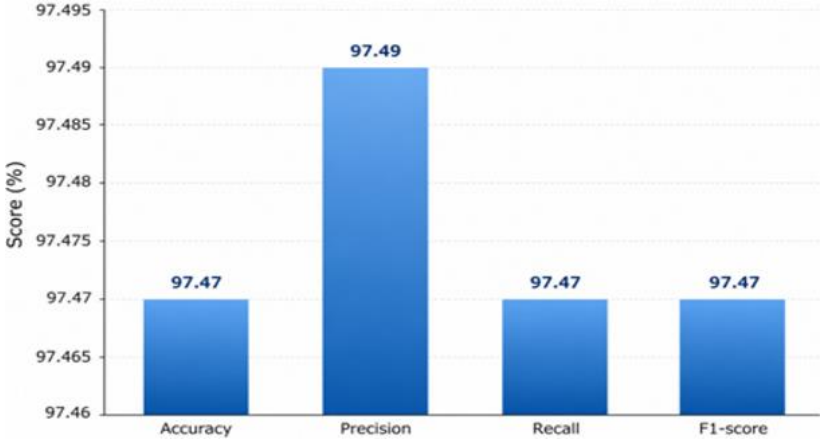


Figure 4. Show the overall performance of the proposed scheme.

The proposed Poultry-CViTNet achieved an overall accuracy of 97.47 ± 0.36%, precision of 97.50 ± 0.38%, recall of 97.47 ± 0.39%, and F1-score of 97.47 ± 0.41%, as summarized in Table 3.

These results demonstrate that the proposed model provides accurate and consistent classification performance across the five folds. The closely balanced precision, recall, and F1-score further indicate that Poultry-CViTNet effectively distinguishes between healthy and unhealthy poultry fecal samples without exhibiting a strong bias toward either class.

Table 3. Five-fold cross-validation performance of Poultry-CViTNet on the poultry fecal-image dataset.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	97.82	97.85	97.82	97.82
Fold 2	96.93	96.97	96.93	96.93
Fold 3	97.58	97.61	97.58	97.58
Fold 4	97.14	97.17	97.14	97.14
Fold 5	97.88	97.91	97.88	97.88
Mean ± Standard Deviation	97.47 ± 0.36	97.50 ± 0.38	97.47 ± 0.41	97.47 ± 0.39

The confusion matrix presented in Figure 5 illustrates the classification performance of the proposed Poultry-CViTNet across the two classes: healthy and unhealthy.

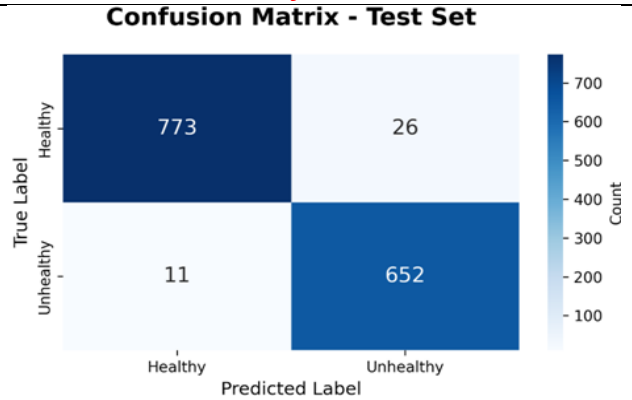


Figure 5. Show the confusion matric of the proposed scheme.

The matrix demonstrates strong diagonal dominance, indicating that most samples were correctly classified.

Only 26 healthy samples were incorrectly classified as unhealthy, while 11 unhealthy samples were misclassified as healthy. The relatively small number of off-diagonal predictions confirms that the proposed model can effectively distinguish between healthy and unhealthy poultry fecal samples. Overall, the confusion-matrix results demonstrate the strong and reliable performance of Poultry-CViTNet in binary poultry health classification.

Computational Complexity Analysis:

To evaluate the computational efficiency of the proposed Poultry-CViTNet, its performance was compared with several strong baseline models. All experiments were conducted under the same experimental protocol and hardware configuration to ensure a fair comparison. Multiple computational metrics were considered, including the number of trainable parameters, floating-point operations (FLOPs), inference throughput, classification accuracy, and Macro-F1 score.

As shown in Table 4, Poultry-CViTNet contains only 3.62 million trainable parameters, making it considerably more lightweight than the baseline models. It also requires 2.60 GFLOPs and achieves the highest throughput of 194.95 images per second. Despite its low computational cost, the proposed model achieves the highest classification accuracy of 97.47 ± 0.36% and a Macro-F1 score of 97.47 ± 0.39%.

Compared with DenseNet121, EfficientNet-B0, Swin-T, Swin-S, and MaxViT-T, Poultry-CViTNet provides a better balance between predictive performance and computational efficiency. These results indicate that the proposed model is suitable for real-time and resource-constrained poultry health-monitoring applications.

Table 4. Computational complexity comparison of Poultry-CViTNet and baseline models on the poultry fecal-image dataset.

Model	Parameters (M)	FLOPs (G)	Throughput (images/s)	Accuracy (%)	Macro-F1 (%)
DenseNet121	7.12	3.85	88.42	92.46 ± 0.06	92.17 ± 0.06
EfficientNet-B0	5.33	1.92	165.23	93.13 ± 0.07	93.22 ± 0.08
Swin-T	26.78	4.21	108.76	92.60 ± 0.16	92.95 ± 0.18
Swin-S	47.35	9.33	62.43	93.14 ± 0.08	93.85 ± 0.07
MaxViT-T	29.88	6.28	49.87	94.88 ± 0.05	94.87 ± 0.06
Poultry-CViTNet	3.62	2.60	194.95	97.47 ± 0.36	97.47 ± 0.39

Ablation Study:

To evaluate the contribution of each component of Poultry-CViTNet, an ablation study was conducted using three model configurations: the complete hybrid model, a CNN-only variant, and a ViT-only variant. All variants were trained and evaluated under the same experimental conditions to ensure a fair comparison. The results are presented in Table 5.

The CNN-only model achieved an accuracy of 95.70 ± 0.21%, precision of 95.20 ± 0.42%, recall of 95.40 ± 0.35%, and F1-score of 95.40 ± 0.23%. The ViT-only model performed better than the CNN-only variant, achieving an accuracy of 96.20 ± 0.26%, precision of 95.30 ± 0.32%, recall of 96.10 ± 0.35%, and F1-score of 95.80 ± 0.25%. Simple concatenation of CNN–ViT features achieved 96.37 ± 0.45% accuracy and 96.45 ± 0.56% F1-score.

The complete Poultry-CViTNet achieved the best performance, with an accuracy of 97.47 ± 0.36%, precision of 97.50 ± 0.38%, recall of 97.47 ± 0.41%, and F1-score of 97.47 ± 0.39%. Compared with the CNN-only model, the hybrid model improved accuracy by 1.77 percentage points. It also improved accuracy by 1.27 percentage points compared with the ViT-only model.

These results confirm that the CNN and ViT branches provide complementary information. The CNN branch captures local texture, color, boundary, and morphological characteristics, whereas the ViT branch models long-range contextual relationships among different image regions. Their integration produces a more discriminative feature representation and improves the reliability of healthy versus unhealthy poultry classification.

Table 5. Ablation study of the proposed Poultry-CViTNet model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN Only	95.70 ± 0.21	95.20 ± 0.42	95.40 ± 0.35	95.40 ± 0.23
ViT Only	96.20 ± 0.26	95.30 ± 0.32	96.10 ± 0.35	95.80 ± 0.25
ViT-CNN	96.37±0.45	95.68±0.25%	96.75±0.65%	96.45 ± 0.56
Poultry-CViTNet	97.47 ± 0.36	97.50 ± 0.38	97.47 ± 0.41	97.47 ± 0.39

Statistical Significance Analysis:

Using the reported mean and standard deviation values over five folds, two-sided Welch's t-tests were conducted to compare Poultry-CViTNet with the baseline models. Poultry-CViTNet achieved significantly higher accuracy than DenseNet121 (mean difference = 5.01 percentage points, $t = 30.70$, $p < 0.00001$), EfficientNet-B0 (difference = 4.34, $t = 26.46$, $p < 0.00001$), Swin-T (difference = 4.87, $t = 27.64$, $p < 0.00001$), Swin-S (difference = 4.33, $t = 26.25$, $p < 0.00001$), and MaxViT-T (difference = 2.59, $t = 15.93$, $p = 0.00007$).

Similarly, the Macro-F1 score of Poultry-CViTNet was significantly higher than that of DenseNet121 (difference = 5.30 percentage points, $t = 30.03$, $p < 0.00001$), EfficientNet-B0 (difference = 4.25, $t = 23.87$, $p < 0.00001$), Swin-T (difference = 4.52, $t = 23.53$, $p < 0.00001$), Swin-S (difference = 3.62, $t = 20.43$, $p = 0.00002$), and MaxViT-T (difference = 2.60, $t = 14.73$, $p = 0.00009$). All comparisons remained statistically significant after Holm correction for multiple testing, with adjusted p-values below 0.001. These findings provide statistical evidence that the observed improvements are unlikely to be attributable to random variation under the assumptions of the applied tests. However, because only summary statistics were available, these results represent approximate unpaired Welch's t-tests. A final paired analysis should use the fold-wise results of all models obtained from identical cross-validation splits.

Visual Interpretability:

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to interpret the predictions of the proposed Poultry-CViTNet, as illustrated in Figure 6. Grad-CAM computes gradient-based importance scores from the final convolutional layer and generates heatmaps that highlight the image regions contributing most strongly to the model's classification decision.

The resulting visualizations indicate that Poultry-CViTNet primarily focuses on meaningful characteristics of the fecal samples, including texture variations, color changes, boundary patterns, and structural irregularities, rather than irrelevant background regions. for

unhealthy samples, the model tends to concentrate on localized abnormal regions. In contrast, for unhealthy samples, the model tends to concentrate on localized abnormal regions exhibiting irregular texture, color, or morphology.

These findings suggest that the model learns discriminative visual features associated with poultry health conditions. The Grad-CAM analysis therefore improves the transparency of the proposed framework and provides visual evidence supporting its classification decisions.

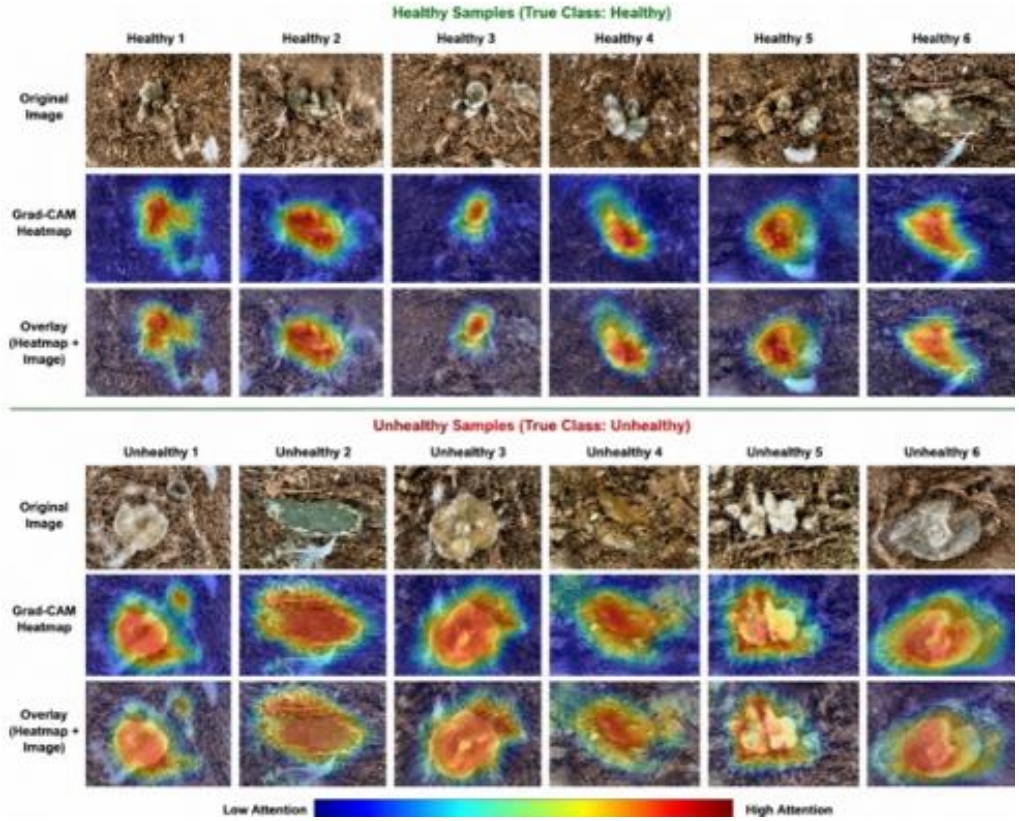


Figure 6. Grad-CAM visualization of predictions generated by the proposed Poultry-CViTNet.

Comparison with State-of-the-Art Methods:

The performance of the proposed Poultry-CViTNet was compared with several existing poultry disease-detection and health-monitoring methods. The comparative results are presented in Table 6.

Among the evaluated approaches, traditional machine-learning and CNN-based models achieved accuracies ranging from 82.00% to 94.49%. More advanced architectures, including hybrid CNN–Transformer models, audio-based CNN models, and Internet of Things-based predictive systems, achieved accuracies between 96.90% and 97.43%.

The proposed Poultry-CViTNet achieved the highest reported accuracy of 97.47%. It exceeded the performance of the ResTFG hybrid CNN–Transformer model by 0.57 percentage points and the IoT-, GAN-, and machine-learning-based model by 0.47 percentage points. Compared with the MFCC–CNN method, Poultry-CViTNet achieved a marginal improvement of 0.04 percentage points.

The improved performance can be attributed to the complementary integration of local convolutional features and global transformer-based representations. The CNN branch captures fine-grained texture, color, boundary, and morphological patterns, whereas the ViT branch models long-range contextual relationships among spatially separated image regions.

The adaptive fusion of these representations enables the proposed model to distinguish healthy and unhealthy fecal samples more effectively.

Table 6. Accuracy comparison of the proposed Poultry-CViTNet with existing studies.

Reference	Model Used	Accuracy (%)
[3]	Decision Tree	86.10
[5]	ChicNetV6, CNN, and Transfer Learning	94.49
[6]	CNN	93.23
[7]	MobileNetV2	82.00
[9]	CNN and XceptionNet	94.00
[11]	ResTFG, CNN, and Transformer	96.90
[12]	MFCC and CNN	97.43
[14]	CNN and Harris Corner Detection	94.14
[16]	ResNet-FPN	93.70
[2]	Internet of Things, GAN, and Machine Learning	97.00
Proposed	Poultry-CViTNet	97.47

All comparative models were evaluated using the same poultry fecal-image dataset. Identical preprocessing, five-fold cross-validation, training settings, and evaluation metrics were applied to ensure a fair comparison. Therefore, the reported results represent a direct experimental comparison rather than values collected from different studies.

Conclusion:

This study presented Poultry-CViTNet, a hybrid deep-learning framework for automated and non-invasive poultry health classification using fecal images. The proposed model integrates a Convolutional Neural Network branch for extracting local texture, color, boundary, and morphological features with a Vision Transformer branch for capturing global contextual relationships among different image regions. The extracted local and global representations are fused and passed to a classification head for predicting healthy and unhealthy poultry samples.

The proposed Poultry-CViTNet achieved an accuracy of 97.47%, demonstrating strong performance in binary poultry health classification. The experimental results, including five-fold cross-validation, ablation analysis, computational complexity evaluation, confusion-matrix analysis, and Grad-CAM visualization, confirmed the effectiveness of combining convolutional feature extraction with transformer-based global attention. The model also achieved high throughput with a relatively low number of trainable parameters, indicating its potential for deployment in practical and resource-constrained poultry-farming environments. Poultry-CViTNet can support non-invasive health monitoring by analyzing fecal images and supporting early identification of abnormal health conditions. Its lightweight design of 3.62 million parameters, 2.60 GFLOPs, and 194.95 images/s indicates potential for integration with edge devices and smart-farming platforms. Despite these promising results, the study has several limitations. First, the proposed model was evaluated using a single dataset and was not validated across independent datasets collected under different farm, camera, lighting, breed, and environmental conditions. This may limit its generalizability to diverse real-world settings. Second, Grad-CAM was used as the primary interpretability technique, which provides visual localization but may not fully explain the reasoning behind individual predictions. Third, the current framework performs binary classification and does not identify specific poultry diseases or disease severity levels. Future work will therefore focus on evaluating Poultry-CViTNet across multiple datasets and real poultry-farm environments. The framework will also be extended to multiclass disease classification and severity assessment. In addition, future studies will investigate complementary explainable artificial intelligence techniques and large

language model-assisted explanations to provide more comprehensive and understandable interpretations of the model's predictions.

References:

[1] X. Fang and A. L. Katchova, "Evaluating the OECD–FAO and USDA agricultural baseline projections," *Q Open*, vol. 3, no. 2, Jul. 2023, doi: 10.1093/QOPEN/QOAD029.

[2] H. Mbelwa, D. Machuve, and J. Mbelwa, "Deep Convolutional Neural Network for Chicken Diseases Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 759–765, 2021, doi: 10.14569/IJACSA.2021.0120295.

[3] G. Z. Khan and I. Ullah, "Efficient Technique for Monkeypox Skin Disease Classification with Clinical Data using Pre-Trained Models," *J. Innov. Image Process.*, vol. 5, no. 2, pp. 192–213, Jun. 2023, doi: 10.36548/JIIP.2023.2.009.

[4] G. Zaman Khan, I. Ali Shah, Farhatullah, M. A. Hassan, H. Junaid, and F. Sardar, "Intelligent Systems for Early Malaria Disease Detection in Patient Cells Using Transfer Learning Approaches," *2023 4th Int. Conf. Comput. Math. Eng. Technol. Sustain. Technol. Socio-Economic Dev. iCoMET 2023*, 2023, doi: 10.1109/ICOMET57998.2023.10099260.

[5] T. A. Serbessa, Y. G. Geleta, and I. O. Terfa, "Review on diseases and health management of poultry and swine," *Int. J. Avian Wildl. Biol.*, vol. 7, no. 1, pp. 62–65, Apr. 2023, doi: 10.15406/IJAWB.2023.07.00187.

[6] E. Abayneh, H. Goba, and M. Shurbe, "Salmonellosis prevalence and risk factors in chicken breeding farms in and around Arba Minch town, Gamo Zone, Ethiopia," *J. Infect. Dev. Ctries.*, vol. 17, no. 2, pp. 226–235, Feb. 2023, doi: 10.3855/JIDC.17553.

[7] G. Zaman Khan *et al.*, "An Efficient Deep Learning Model based Diagnosis System for Lung Cancer Disease," *2023 4th Int. Conf. Comput. Math. Eng. Technol. Sustain. Technol. Socio-Economic Dev. iCoMET 2023*, 2023, doi: 10.1109/ICOMET57998.2023.10099357.

[8] U. Ahmad, U. Bin Ahmad, G. Z. Khan, S. Shahbaz, and I. Ullah, "Automated Detection and Classification of Plasmodium Parasites in Blood Smears Using the RFCN Deep Learning Model," *2025 4th Int. Conf. Commun. Comput. Digit. Syst. C-CODE 2025*, 2025, doi: 10.1109/C-CODE67372.2025.11204078.

[9] I. A. Khan *et al.*, "A Comprehensive Review on 3D Volumetric CT Liver Segmentation: Techniques, Challenges, Trends, and Future Research Directions," *ICCK Trans. Appl. Intell. Cybern.*, vol. 1, no. 1, pp. 5–35, 2026, doi: 10.62762/TAIC.2025.965486.

[10] R. Zaman, I. A. Shah, N. Ullah, and G. Z. Khan, "A Robust Deep Learning-Based Approach for Detection of Breast Cancer from Histopathological Images," *Eng. Proc. 2023, Vol. 56, Page 313*, vol. 56, no. 1, p. 313, Dec. 2023, doi: 10.3390/ASEC2023-16598.

[11] M. M. D. Quintana, R. R. D. Infante, J. C. S. Torrano, and M. C. Pacis, "A Hybrid Solar Powered Chicken Disease Monitoring System using Decision Tree Models with Visual and Acoustic Imagery," *2022 14th Int. Conf. Comput. Autom. Eng. ICCAE 2022*, pp. 65–69, 2022, doi: 10.1109/ICCAE55086.2022.9762418.

[12] M. A. A. A. Bakar, P. J. Ker, S. G. H. Tang, H. J. Lee, and B. S. Zainal, "Classification of Unhealthy Chicken based on Chromaticity of the Comb," *2022 IEEE Int. Conf. Comput. ICOCO 2022*, pp. 1–5, 2022, doi: 10.1109/ICOCO56118.2022.10031812.

[13] M. K. Gourisaria, A. Arora, S. Bilgaiyan, and M. Sahni, "Chicken Disease Multiclass Classification Using Deep Learning," *Lect. Notes Networks Syst.*, vol. 614 LNNS, pp. 225–238, 2023, doi: 10.1007/978-981-19-9331-2_19/SAVE-RESEARCH.

[14] A. K. Srivastava and B. P. Pandey, "Deep Learning Based Classification of Poultry

Disease," *Int. J. Autom. Smart Technol.*, vol. 13, no. 1, 2023, doi: 10.5875/AUSMT.V13I1.2439.

[15] Vandana, K. K. Yogi, and S. P. Yadav, "Chicken Diseases Detection and Classification Based on Fecal Images Using EfficientNetB7 Model," *Evergreen*, vol. 11, no. 1, pp. 314–330, Mar. 2024, doi: 10.5109/7172288.

[16] H. ; ; Zheng *et al.*, "Study on Poultry Pose Estimation Based on Multi-Parts Detection," *Anim. 2022, Vol. 12, Page 1322*, vol. 12, no. 10, p. 1322, May 2022, doi: 10.3390/ANI12101322.

[17] W. Widyawati and W. Gunawan, "Early detection of sick chicken using artificial intelligence," *Tek. J. Sains dan Teknol.*, vol. 18, no. 2, p. 136, Nov. 2022, doi: 10.36055/TJST.V18I2.17337.

[18] P. He *et al.*, "A reliable and low-cost deep learning model integrating convolutional neural network and transformer structure for fine-grained classification of chicken Eimeria species," *Poult. Sci.*, vol. 102, no. 3, p. 102459, Mar. 2023, doi: 10.1016/J.PSJ.2022.102459.

[19] K. Cuan, T. Zhang, J. Huang, C. Fang, and Y. Guan, "Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network," *Comput. Electron. Agric.*, vol. 178, p. 105688, Nov. 2020, doi: 10.1016/J.COMPAG.2020.105688.

[20] S. Suthagar, G. Mageshkumar, M. Ayyadurai, C. Snegha, M. Sureka, and S. Velmurugan, "Faecal Image-Based Chicken Disease Classification Using Deep Learning Techniques," *Lect. Notes Networks Syst.*, vol. 563, pp. 903–917, 2023, doi: 10.1007/978-981-19-7402-1_64/SAVE-RESEARCH.

[21] "(PDF) A computer Vision System for Early Detection of Sick Birds in a Poultry Farm Using Convolution Neural Network on Shape and Edge Information." Accessed: Jul. 25, 2026. [Online]. Available: https://www.researchgate.net/publication/369040188_A_computer_Vision_System_for_Early_Detection_of_Sick_Birds_in_a_Poultry_Farm_Using_Convolution_Neural_Network_on_Shape_and_Edge_Information

[22] A. Mahdavian, S. Minaei, P. M. Marchetto, F. Almasganj, S. Rahimi, and C. Yang, "Acoustic features of vocalization signal in poultry health monitoring," *Appl. Acoust.*, vol. 175, p. 107756, Apr. 2021, doi: 10.1016/J.APACOUST.2020.107756.

[23] H. Zhang and C. Chen, "Design of Sick Chicken Automatic Detection System Based on Improved Residual Network," *Proc. 2020 IEEE 4th Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2020*, pp. 2480–2485, Jun. 2020, doi: 10.1109/ITNEC48623.2020.9084666.

[24] G. Ahmed, R. A. S. Malick, A. Akhunzada, S. Zahid, M. R. Sagri, and A. Gani, "An Approach towards IoT-Based Predictive Service for Early Detection of Diseases in Poultry Chickens," *Sustain. 2021, Vol. 13, Page 13396*, vol. 13, no. 23, p. 13396, Dec. 2021, doi: 10.3390/SU132313396.

[25] S. H. Badshah *et al.*, "An Efficient Classification Techniques for Brain Tumor Using Features Extraction and Statistic Methods, with Machine Learning Algorithms," *Lect. Notes Data Eng. Commun. Technol.*, vol. 158, pp. 757–775, 2023, doi: 10.1007/978-3-031-24475-9_62/SAVE-RESEARCH.

[26] "Improved Neural Network-Based System for Early and Accurate Diagnosis of Alzheimer Disease | Request PDF." Accessed: Aug. 17, 2026. [Online]. Available: https://www.researchgate.net/publication/387755906_Improved_Neural_Network-Based_System_for_Early_and_Accurate_Diagnosis_of_Alzheimer_Disease

[27] Sajad Ulhaq, G. Z. Khan, Imran Ulhaq, I. Ullah, and F. Rabbi, "Epilepsy Seizures Classification with EEG Signals: A Machine Learning Approach," *J. Comput. Sci.*

Technol. Stud., vol. 5, no. 3, pp. 25–34, Aug. 2023, doi: 10.32996/JCSTS.2023.5.3.3.

[28] Z. Cai, J. Cui, H. Yuan, and M. Cheng, "Application and research progress of infrared thermography in temperature measurement of livestock and poultry animals: A review," *Comput. Electron. Agric.*, vol. 205, Feb. 2023, doi: 10.1016/j.compag.2022.107586.

[29] G. Z. Khan, I. Ulhaq, I. Adil, S. Ulhaq, and I. Ullah, "A Privacy-Preserving Based Technique for Customer Churn Prediction in Telecom Industry," *VFAST Trans. Softw. Eng.*, vol. 11, no. 3, pp. 73–80, Sep. 2023, doi: 10.21015/VTSE.V11I3.1642.

[30] J. Y. Noh *et al.*, "Thermal Image Scanning for the Early Detection of Fever Induced by Highly Pathogenic Avian Influenza Virus Infection in Chickens and Ducks and Its Application in Farms," *Front. Vet. Sci.*, vol. 8, p. 616755, May 2021, doi: 10.3389/FVETS.2021.616755/TEXT.

[31] C. H. Chuang, C. Y. Chiang, Y. C. Chen, C. Y. Lin, and Y. C. Tsai, "Goose Surface Temperature Monitoring System Based on Deep Learning Using Visible and Infrared Thermal Image Integration," *IEEE Access*, vol. 9, pp. 131203–131213, 2021, doi: 10.1109/ACCESS.2021.3113509.

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.