JWPR
Journal of World's Poultry Research
2026, Scienceline Publication
J. World Poult. Res. 16(1): 31-41, 2026
Research Paper
DOI: https://dx.doi.org/10.36380/jwpr.2026.4
PII: S2322455X2600004-16
ISSN: 2322-455X

Assessment of Chicken Meat Quality Through Categorized Image Analysis Using Artificial Intelligence

Arash Biabanian[1*], Seyed Vahab Shojaedini[2], and Nima Misaghi[3]

[1]*Department of Computer Engineering, Faculty of Engineering, Electronic Branch, Islamic Azad University, Tehran, Iran*
[2]*Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran*
[3]*Department of Computer Engineering, Senior software developer, Aviv-Group GmbH, Berlin, Germany*

*Corresponding author's E-mail: arashbiabanian@iauec.ac.ir

Received: December 14, 2025, Revised: January 19, 2026, Accepted: February 15, 2026, Published: March 05, 2026

ABSTRACT

As consumer demands for high food quality have risen, the swift and non-destructive assessment of chicken meat freshness has become crucial for public health. The present study developed an artificial-intelligence-based digital imaging approach to estimate chicken meat quality by analyzing images categorized as healthy, defrosted, or rotten. A total of seven pre-trained convolutional neural network architectures, including Xception, Inception V3, MobileNet, DenseNet121, VGG16, VGG19, and a baseline convolutional neural network (CNN), were considered for the current study. Chicken meat samples were photographed under controlled lighting, and their quality categories were confirmed by expert inspections. Model performance was evaluated by overall classification accuracy. Among the tested architectures, the VGG19 network achieved the highest mean accuracy of 94.3%, outperforming compared to the baseline CNN by approximately 9%, while other networks exceeded 85% accuracy, indicating reliable recognition capability. This study confirmed that AI-based digital image analysis can accurately classify chicken meat freshness in a non-destructive manner. Among the tested models, VGG19 achieved the highest performance and demonstrated its strong capability for reliable meat quality assessment.

Keywords: Accuracy, Artificial intelligence, Chicken meat quality, Deep learning, Transfer learning

INTRODUCTION

The increasing global consumption of animal products since the mid-twentieth century has intensified the demand for efficient meat-quality evaluation systems. Among these products, chicken meat holds a prominent place in the human diet owing to its affordability, accessibility, and favorable nutritional profile (Sans and Combris, 2015). As consumer expectations for safety and freshness rise, ensuring the quality of chicken meat has become a central issue in the poultry industry and public-health sectors. Chicken meat is particularly vulnerable to microbial spoilage and contamination, which can lead to severe foodborne illnesses if undetected (Botta et al. 2024). Therefore, reliable and rapid quality-assessment tools are vital to protect consumers and reduce economic losses.

Food spoilage generally refers to any deterioration that makes food unacceptable for consumption. In meat, such degradation manifests through physical and sensory changes, including color fading, slimy surface texture, off-odor, or abnormal taste (Nychas et al. 2008). These alterations often arise from the metabolic activity of spoilage microorganisms such as *Clostridium*, *Bacillus*, and *Pseudomonas* species, which generate volatile compounds during protein breakdown (Zhu et al. 2022). Conventional freshness evaluation depends heavily on visual inspection and odor detection by a human operator. However, such subjective methods suffer from inconsistency among evaluators and cannot ensure reliable discrimination between fresh and spoiled meat.

To achieve higher precision, laboratory-based analytical techniques such as microbial enumeration, chemical composition analysis, and spectroscopy have traditionally been employed. Nondestructive approaches—including near-infrared spectroscopy, Fourier-transform near-infrared spectroscopy, mid-infrared spectroscopy, ultraviolet–visible, Raman spectroscopy, laser-induced breakdown spectroscopy, hyperspectral imaging, and thermal imaginghave shown promising accuracy in meat-quality detection. Nonetheless, these methods typically require expensive instrumentation, complex calibration procedures, and expert operators, which limits their

routine industrial application (Markiewicz-Keszycka et al., 2017).

In recent years, digital imaging combined with artificial intelligence (AI) has emerged as an effective, nondestructive, and low-cost alternative for assessing meat freshness (Shi et al., 2021). Computer-vision systems can capture subtle visual features, such as color gradients, surface texture, and structural uniformity, that correlate strongly with quality deterioration. Deep learning, a subset of AI based on convolutional neural networks (CNNs), enables automated feature extraction from large image datasets without manual preprocessing. Such models have demonstrated remarkable performance in classifying and grading agricultural and food products, including fruits, fish, and red meat (Kaushal et al., 2024).

A previous study has applied deep-learning or machine-vision methods to poultry-meat evaluation, yet limitations remain. Many earlier works relied on hyperspectral or thermal imaging, which, while accurate, increase system cost and complexity. Others used relatively small datasets or single architectures, restricting model generalization (Jian et al., 2025). In contrast, the present study introduced a comprehensive framework for visual quality estimation of sliced chicken meat using standard grayscale adaptation for pretrained red-green-blue networks (RGB) digital images. This detection approach provided a practical and affordable solution that can be implemented with readily available imaging devices, such as smartphones or laboratory cameras (Zou et al. 2024).

Beyond technical innovation, applying deep learning to poultry quality evaluation offers broader benefits. It can assist producers and regulatory agencies in monitoring freshness along the supply chain, reduce food waste through early spoilage detection, and increase consumer confidence in marketed products. Additionally, developing reliable digital image datasets for poultry contributes to establishing open resources that facilitate future AI research in food safety and quality control (Cai et al., 2024).

This study aimed to develop and evaluate a nondestructive framework based on digital imaging and deep-learning algorithms for classifying chicken meat quality into healthy, defrosted, and rotten categories. Specifically, this study compared the accuracy of seven convolutional architectures using transfer learning to determine the most suitable model for automated chicken-meat quality prediction.

MATERIALS AND METHODS

Ethical approval

All procedures involving post-slaughter, sampling, transport, and imaging were followed by biosafety regulations and national food-safety standards according to the guidelines of Islamic Azad University, Tehran, Iran.

Study design and sampling

This experimental, laboratory-based cross-sectional study evaluated chicken meat quality using artificial-intelligence-based digital image analysis. Sixty broiler carcasses (Ross 308, approximately 42 days old) were randomly selected from the processing line at Sahar Roshan Slaughterhouse, Shahriyar, Tehran Province, Iran, handled aseptically, and transported under cold-chain conditions (4 °C) to the food quality laboratory, Pooya Protein, Shahriyar, Iran. Boneless chicken breast muscle was selected due to its uniform texture and relevance to industrial processing, enhancing visual consistency for image-based analysis. Samples were included if they were intact, hygienically handled, and free from visible defects, while damaged, contaminated, or abnormally discolored samples were excluded. Fresh meat (healthy class) was defined as meat stored at ≤ 4 °C for ≤ 72 h post-slaughter in accordance with FAO guidelines (FAO, 2003). Each breast was portioned into approximately 5 × 5 cm subsamples and imaged under controlled lighting conditions. The same subsamples were subsequently frozen at –18 °C for 48 h, thawed at 25 °C, and re-imaged to represent the defrosted class, then stored at 25 °C for five days until standard spoilage indicators (color darkening, surface slime, and foul odor) were observed and imaged as the rotten class. The final dataset comprised 1,800 RGB images (600 per class), and the primary evaluated parameter was meat quality classification accuracy across healthy, defrosted, and rotten categories.

Image acquisition

Images were acquired in a light-controlled, scanner-style photo box using a Samsung Galaxy A50 smartphone camera (South Korea) mounted on a fixed holder, with the terms “photo box” and “cellphone” referring to the same standardized acquisition setup. All images were saved as RGB JPEG files at a resolution of 3000 × 4000 pixels using the 25-MP sensor. To enhance reproducibility and reduce illumination-related variability, constant imaging conditions were maintained, including automatic exposure (1/125 s), automatic ISO (150), automatic white balance, and a fixed camera-to-sample distance of 30 cm. The imaging chamber was equipped with two symmetrically positioned LED light sources to assess robustness across color temperatures while minimizing directional shadows: a white light source (~ 6000 K, approximately 800 lm) and a warm yellow light source (~3000 K, approximately 800 lm). This standardized geometry and lighting configuration ensured that extracted image features reflected true meat quality rather than acquisition artifacts. Ground-truth labeling was performed independently by three experienced poultry meat inspectors, who classified each image as healthy, defrosted, or rotten based on visible surface color and texture characteristics, supported by contemporaneous odor observations recorded at the time of image capture. Any discrepancies were resolved

through consensus, and inter-rater reliability analysis demonstrated strong agreement among evaluators (Cohen's kappa, $\kappa = 0.91$).

Data splits and cross-validation

For model development and evaluation, a stratified hold-out sampling strategy was employed to ensure balanced representation of each meat quality class. The dataset was divided into 60% for training, 20% for validation, and 20% for testing on a per-class basis. Accordingly, each class (healthy, defrosted, and rotten) contributed 360 images to the training set, 117 images to the validation set, and 123 images to the test set, resulting in a total of 1,080 training images, 351 validation images, and 369 test images across all classes. The fixed test set, comprising 20% of the data, was reserved exclusively for final model evaluation, which accounts for the 123 images per class displayed in the confusion matrices. To optimize model hyperparameters and enhance generalizability, five-fold stratified cross-validation was additionally performed on the training subset, while the test set remained completely intact until the final performance assessment.

Preprocessing

All JPEG images were first resized to 224 × 224 pixels, normalized to the [0, 1] range, and augmented with random horizontal and vertical flips, rotations of ±15°, and brightness adjustments of ±10% to improve model generalization while maintaining a realistic appearance. Guided by established deep-learning workflows in poultry meat quality assessment, seven convolutional neural network (CNN) architectures, including Xception, InceptionV3, MobileNet, DenseNet121, VGG16, VGG19, and a baseline CNN, were compared, using ImageNet-pretrained convolutional bases combined with task-specific classification heads to create a cost-effective visible-RGB imaging pipeline aligned with best practices in more instrument-intensive modalities. Transfer learning was applied, initially freezing the convolutional base and later fine-tuning higher layers, with the classifier consisting of global average pooling, a ReLU-activated dense layer, dropout (0.3-0.5), and a Softmax output layer for three-class classification. Models were trained using the Adam optimizer (learning rate 1×10^{-4}), categorical cross-entropy loss, a batch size of 32, and early stopping based on validation loss. Performance was evaluated primarily via overall accuracy on a fixed held-out test set (20%; 369 images, 123 per class), supplemented by precision, recall, F1-score, and confusion matrices to ensure transparency and comparability across architectures. Standardized camera settings and controlled lighting with dual LEDs (6000 K and 3000 K) minimized variance due to illumination and geometry, allowing the models to focus on intrinsic meat-quality features such as color gradients and surface micro-texture while maintaining robustness to realistic color-temperature variations.

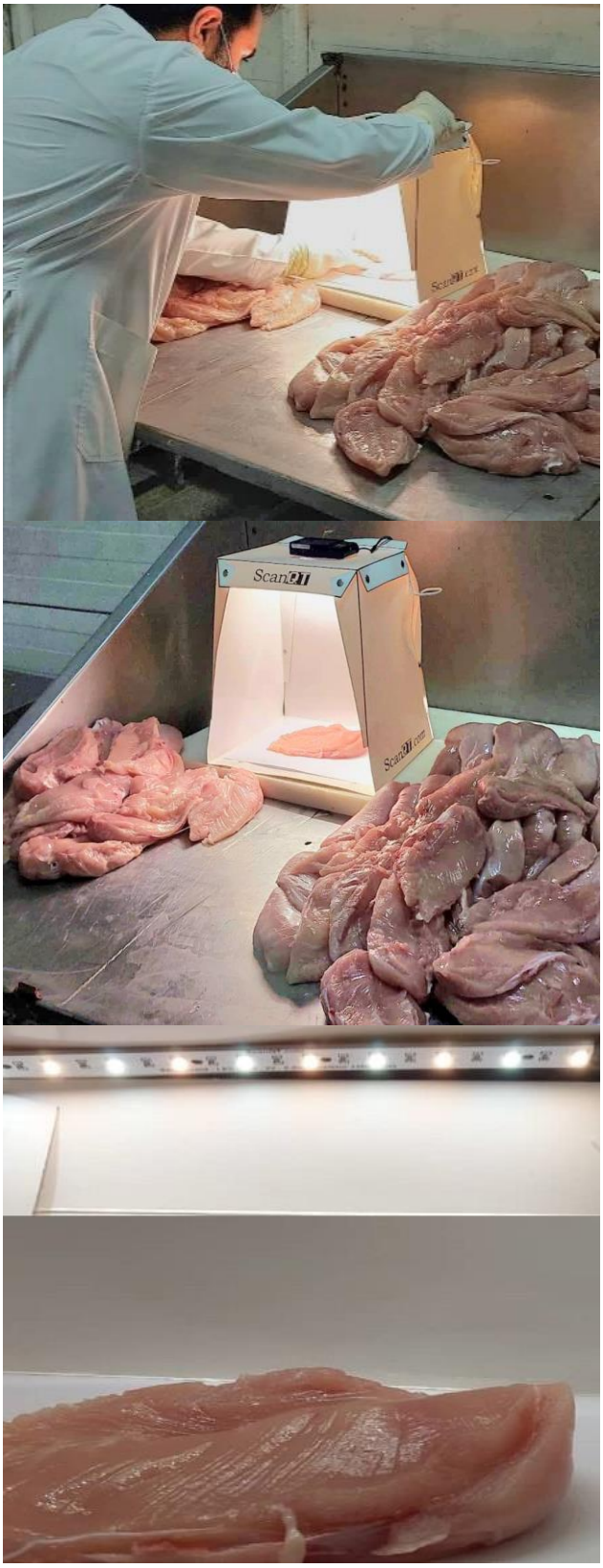
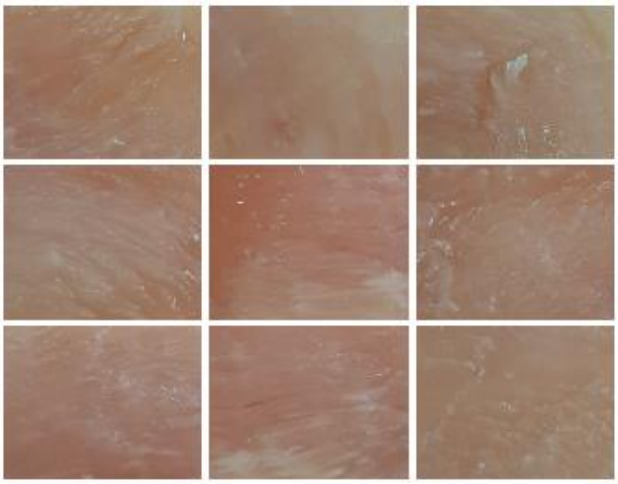


Figure 1. The process of acquiring images from chicken meat (captured by the authors)

(A) Fresh chicken meat



(B) Defrosted chicken meat

(C) Rotten chicken meat

Figure 2. Three sample images **(A-C)** are based on the defrosting stage, the freshness stage, and the rotting stage of chicken meat (captured by the authors)

The procedure of the proposed method

The pipeline performs standardized preparation of sliced chicken-breast images, grayscale adaptation compatible with ImageNet-pretrained RGB backbones, data augmentation, transfer-learning-based classification, and held-out testing. The goal was robust recognition of three quality classes (healthy, defrosted, rotten) under controlled yet practical imaging conditions.

Preparation and recognition workflow

The preparation and recognition process proceeds as follows. First, acquisition inside a light-controlled photo box (Section 3.2 Image Acquisition). Second, expert labeling into three classes with consensus and κ agreement is reported in Section 3.3. Third, preprocessing (crop → grayscale → resize → normalization). Fourth, augmentation during training only. Fifth, deep models trained with transfer learning and evaluated on a fixed 20% hold-out test set. This workflow was designed to minimize illumination-related variance while retaining surface-texture cues most relevant to freshness discrimination.

Grayscale adaptation for pretrained RGB networks

Images are converted from RGB to grayscale using the ITU-R BT.601 luma transform $Y = 0.299R + 0.587G + 0.114B$, then replicated to three channels $[Y, Y, Y]$ so that input tensors retain the 3-channel shape expected by ImageNet-pretrained models. This preserves the original convolutional weights without architectural changes and yields stable convergence across backbones. An alternative one-channel adaptation (collapsing first-layer RGB kernels by averaging) was tested and produced comparable results; three-channel replication was selected for simplicity and consistency. Grayscale reduces sensitivity to color-temperature shifts between warm (≈3000 K) and white (≈6000 K) LEDs and emphasizes micro-texture and structure (exudate, surface slime, fine marbling, discontinuities) that differentiate defrosted from healthy and early rotten tissues.

Preprocessing and cropping

All JPEGs are decoded to RGB, centrally cropped to remove borders and the photo-box frame (regions not informative for classification), converted to grayscale, replicated to three channels, resized to 224×224 px, and normalized to 0 and 1. For models that apply ImageNet normalization, the standard channel-wise mean ± SD are used (applied identically to the replicated channels) to maintain numerical compatibility with pretrained weights (Krizhevsky et al., 2012a).

Data augmentation

Data augmentations were applied exclusively to the training images and were implemented identically across all models. Geometric augmentations included random horizontal and vertical flips (probability = 0.5 each), random in-plane rotations of ± 15° (p = 0.8), and random resized-crop transformations (scale 0.90-1.00, aspect ratio 0.95-1.05) followed by bilinear resizing back to 224 × 224 pixels (p = 0.7). Photometric augmentations, adapted for grayscale images, consisted of brightness jitter of ±10% (p = 0.7) and contrast jitter of ±10% (p = 0.7), with hue and saturation jitter disabled after grayscale conversion. Noise and regularization techniques included additive Gaussian noise with σ ≈ 0.01 of the dynamic range (p = 0.5) and Gaussian blur with σ = 0.5 pixels (p = 0.2). Images were processed in batches of 32 with a deterministic seed to ensure reproducibility (Krizhevsky et al., 2012b).

Transfer learning

Seven convolutional neural network (CNN, Figure 3) architectures, including VGG16, VGG19, InceptionV3, Xception, MobileNet, DenseNet121, and a baseline CNN, were compared in this study. These model selections follow established deep-learning practices in food-quality sensing and poultry-related applications (Ametefe et al. 2024). While prioritizing low-cost visible-RGB deployment over more instrument-intensive spectroscopy methods. A fine-tuning policy was applied in which the convolutional bases were initially frozen for 10 epochs and then partially unfrozen for further training: for VGG16 and VGG19, Blocks 4-5 (the last 8 convolutional layers) were unfrozen; for InceptionV3, layers from Mixed 7 onward were made trainable; for Xception, the middle- and exit-flow layers (last 36 convolutional layers) were unfrozen; for MobileNet (v1), the last 20 layers (depthwise and pointwise pairs) were unfrozen; and for DenseNet121, Dense Block 4 along with its preceding transition layers were unfrozen. Each model concluded with a global average pooling layer, followed by a fully connected dense layer with 256 ReLU units, a dropout layer (0.5), and a final Softmax output layer for three-class classification.

Optimization, hyperparameters, and schedules

All models were trained using the Adam optimizer with an initial learning rate of 1×10^{-4}, coupled with a ReduceLROnPlateau learning-rate scheduler (factor 0.5, patience 5 epochs, minimum LR 1×10^{-6}). The categorical cross-entropy loss function was used, with a batch size of 32 and training for up to 100 epochs, incorporating early stopping based on validation loss (patience 12 epochs, restoring the best weights). Weight decay (L2 regularization) of 1×10^{-5} was applied to all trainable convolutional and dense layers. Convolutional bases were initialized with ImageNet pretrained weights, while newly added dense layers used He-normal initialization. To ensure a fair comparison, all models employed the same data splits, augmentation procedures, optimizer settings, and learning-rate schedules (He et al., 2015).

Cross-validation and held-out testing

A 5-fold stratified cross-validation is performed on the training set only (folds are drawn within the 60% training portion). The 20% validation split supports model selection within each fold. The fixed 20% test set (369 images; 123 per class) remains untouched until final evaluation. Reported model selection is based on mean validation accuracy across folds; final test results are computed once using the best configuration chosen by CV. Per-fold metrics (accuracy, precision, recall, F1) and mean ± SD are provided in supplementary Table S1, while the main text reports mean ± SD and the final test performance for transparency (Varma and Simon, 2006).

Runtime environment and software

Training and evaluation were conducted on Google Colab with NVIDIA T4 GPU (16 GB); no Tensor Processing Unit (TPU)was used. Dataset curation and figure rendering occurred on a local workstation (Intel Core i7, 16 GB RAM, GeForce GT-710) but did not perform model training. Python 3.10, TensorFlow 2.15 / Keras 2.15, CUDA 11.8, and cuDNN 8.9 were used. Random seeds were fixed to enhance reproducibility.

Mathematical formulation

Let $I(x,y)$ denote the input image and $f(u,v)$a convolution kernel. The 2-D convolution producing feature map $g(x,y)$ according to the following formula (Fukushima 1980).

$$g(x,y) = (I*f)(x,y) = \sum_{u} \sum_{v} I(x-u, y-v)\, f(u,v),$$

where u, v index the kernel's spatial support. For a generic layer L, the activation of the unit j was according to the following formula (Goodfellow et al. 2016).

$$X_j^{(L)} = \phi\left(\sum_{i \in M_j} X_i^{(L-1)} * W_{ij}^{(L)} + b_j^{(L)}\right),$$

with nonlinearity $\phi(\cdot)$, weights $W_{ij}^{(L)}$, bias $b_j^{(L)}$, and $\mathcal{M}_j$ set of input maps connected to j. The output spatial size (o_w, o_h) for a convolution with input (W, H), filter size (F_w, F_h), padding p, and stride s was according to the following formula (Dumoulin and Visin 2018).

$$o_w = \left\lfloor \frac{W - F_w + 2p}{s} \right\rfloor + 1, o_h = \left\lfloor \frac{H - F_h + 2p}{s} \right\rfloor + 1.$$

Freshness classification standards

Class labels follow operational standards consistent with FAO guidance and poultry-meat spoilage indicators: healthy (fresh) stored at ≤ 4 °C for ≤ 72 h post-slaughter with normal color/texture; defrosted previously frozen at –18 °C for 48 h and thawed at 25 °C; rotten stored at 25 °C for five days with off-odor, surface slime, and color darkening documented at capture (Sujiwo et al., 2018).

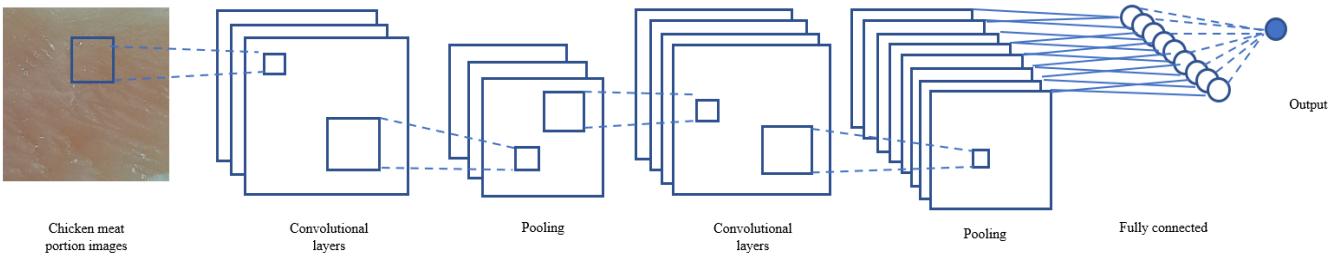


Figure 3. The structure illustration of a convolutional neural network (CNN) network (designed by the authors)

Visual geometry group networks (VGG16 and VGG19)

The VGG family comprises deep convolutional networks that stack multiple small (3×3) filters to enhance representational capacity while maintaining a simple layer design. VGG16 consists of five convolutional blocks (13 convolutional layers) followed by three fully connected layers (FC6-FC8), whereas VGG19 extends this depth to 16 convolutional layers with the same fully connected head; both were originally trained on ImageNet with 1,000 classes and were widely reused in transfer-learning settings for downstream image-classification tasks (Simonyan and Zisserman, 2015). InceptionV3 architectures factorize computation with parallel filters of varying sizes (1×1, 3×3, 5×5) to capture multi-scale features efficiently, while bottleneck 1×1 convolutions reduce dimensionality before costlier convolutions, improving computational efficiency and mitigating overfitting. Concatenation of parallel branches aggregates multi-scale features, and InceptionV3 further refines this approach with factorized convolutions and label smoothing, achieving an optimal accuracy-efficiency balance (Szegedy et al., 2015).

Xception replaces standard Inception modules with depthwise separable convolutions applying depthwise spatial filtering per channel followed by pointwise 1×1 mixing organized in a linear stack of modules with residual connections, comprising 36 convolutional layers in entry, middle, and exit flows for efficient feature extraction (Chollet, 2017). MobileNet (v1) employs depthwise separable convolutions throughout the network, combining depthwise spatial filters per channel with pointwise 1×1 convolutions, and uses batch normalization and ReLU activations to produce a lightweight backbone suitable for embedded or resource-constrained applications while remaining effective in transfer learning. Finally, DenseNet121 connects each layer to all subsequent layers within dense blocks, enabling feature reuse and improved gradient flow; transition layers with 1×1 convolutions and pooling control dimensionality between blocks. DenseNet121 provides a balance of depth and efficiency, making it a popular choice as a transferable backbone for image-classification tasks (Howard et al., 2017).

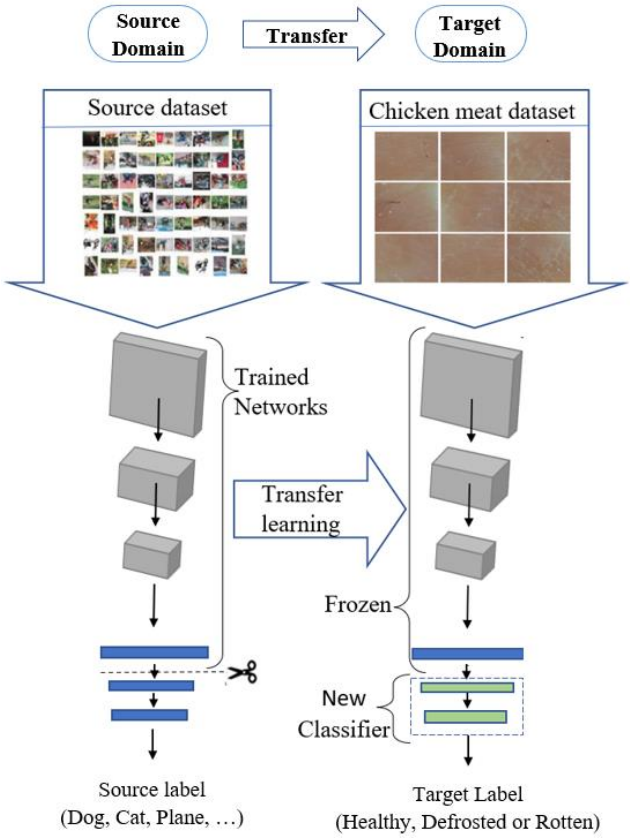


Figure 4. A schematic of the transfer learning technique (designed by the authors)

Statistical analysis

All analyses were performed in Python 3.10 using TensorFlow/Keras 2.15 for model training, scikit-learn 1.4 for performance metrics and confusion matrices, SciPy 1.11 and statsmodels 0.14 for statistical testing, and Matplotlib 3.8 for plotting. Primary outcomes included overall accuracy, per-class precision, recall (sensitivity), specificity, F1-score, and macro/micro averages. Confidence intervals for proportions were computed using the Wilson method where applicable.

Model-to-model comparisons on the same held-out test set used McNemar's test on paired prediction outcomes to evaluate whether differences in accuracy were statistically significant. For cross-validated performance, per-fold accuracies (and F1-scores) were compared with paired, non-parametric tests (Wilcoxon signed-rank), and effect sizes were reported. Statistical significance was set at $\alpha = 0.05$ with Holm correction for multiple pairwise comparisons. Runtime and size. For deployability, per-image inference time (ms) was measured on an NVIDIA T4 (16 GB) GPU with batch = 1 after a 20-step warm-up; the median of 100 timed runs is reported. Model size was taken as the serialized weight file size.

Environment, training, and evaluation were executed on Google Colab with NVIDIA T4 GPU; no TPU was used. Dataset curation (no training) occurred on a local workstation. Random seeds were fixed for reproducibility.

RESULTS

All images were first converted to grayscale using the ITU-R BT.601 standard and then replicated across three channels to maintain compatibility with ImageNet-pretrained backbones. They were resized to 224 × 224 pixels and normalized to the 0-1 range. Seven convolutional neural network architectures including VGG16, VGG19, InceptionV3, Xception, MobileNet, DenseNet121, and a baseline CNN were evaluated using a consistent transfer-learning protocol (initially frozen convolutional base followed by partial fine-tuning) and identical optimization settings, including the Adam optimizer (initial learning rate 1×10^{-4}), categorical cross-entropy loss, batch size 32, ReduceLROnPlateau scheduling, and early stopping. Model performance was evaluated not only by overall accuracy but also by per-class precision, recall (sensitivity), specificity, F1-score, and macro/micro averages (Tables 2-4), with confusion matrices provided in Figure 4. Mean ± SD values from 5-fold cross-validation on the training set were reported for validation metrics, while final metrics on the held-out test set were reported after selecting the best model configuration based on cross-validated performance. Learning curves of training and validation loss and accuracy for each architecture are presented in Figure S1 to illustrate convergence and assess overfitting. Statistical comparisons between top-performing models on the same test set were conducted using McNemar's test on paired predictions (Table 5), and cross-validated accuracies were compared using Wilcoxon signed-rank tests with Holm-adjusted p-values and effect sizes (Table S1). Additionally, practical deployment metrics, including per-image inference time on a T4 GPU and model size in megabytes, are reported in Table 6 to evaluate real-time feasibility and memory requirements for production settings.

Table 2. Summary of dataset composition, imaging/acquisition settings, preprocessing steps, and training hyperparameters used in this study

Network	Structure	Number of non-trainable parameters	Number of trainable parameters
VGG16	13 convolution, 5 max pooling	7,635,264	8,392,449
VGG19	16 convolution, 5 max pooling	12,944,960	7,473,665
Xception	71 layers deep	8,693,240	12,758,482
MobileNet	53 layer deep	3,210,051	21,888
DenseNet121	121 layers deep	1,149,440	7,626,817
Inception	48 layer deep	9,544,288	1,2848,546

Figure 6. Chicken breast (RGB visible imaging), healthy class (correct versus incorrect) model predictions. **T** (green) indicates correct prediction; **F** (red) indicates incorrect prediction. (Captured by the authors)

To contextualize the quantitative results, representative images from the three quality classes, including healthy, defrosted, and rotten were analyzed to relate model predictions to visible appearance. Figure 6 illustrates nine expert-labeled healthy chicken breast images captured under the study's controlled visible-light conditions using the standardized photo-box with dual LEDs. Ground-truth labels for the healthy class were assigned by three poultry-meat inspectors based on operational criteria, including storage at ≤4 °C within 72 hours post-slaughter, normal color and texture, and absence of off-odor or slime, with consensus achieved across inspectors (Cohen's $\kappa = 0.91$). In Figure 6, six images are marked T (green) where the neural network prediction matched the expert label, while three images are marked F (red) where the prediction disagreed and was counted as an error. This visual comparison directly illustrates the alignment between model decisions and expert assessment under controlled imaging conditions.

Figure 7 shows nine expert-labeled images of defrosted chicken breast captured under the study's controlled visible light conditions. Each image is annotated with the model's prediction. T (green) indicates agreement with the expert label (correct), while F (red) denotes disagreement (error).

Figure 8 similarly displays nine expert-labeled images of rotten chicken breast, using the same T/F notation to indicate correctness. Ground-truth labels were assigned by three inspectors according to the operational criteria described in the Methods section, with T (green) indicating correct model predictions and F (red) indicating errors.

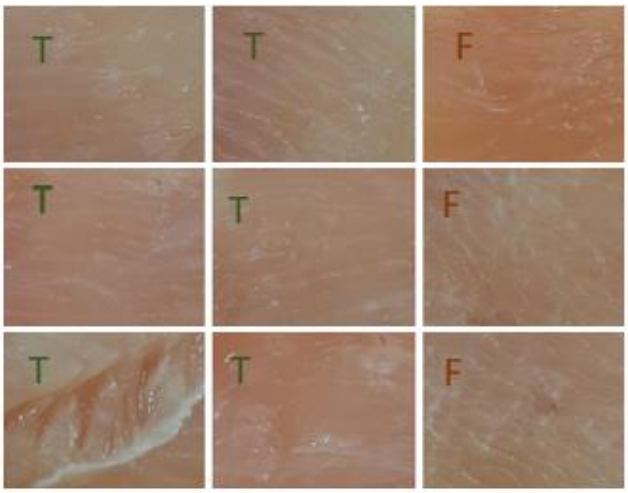


Figure 7. Chicken breast (RGB visible imaging), defrosted class (correct versus incorrect) model predictions. **T** (green) indicates correct prediction; **F** (red) indicates incorrect prediction. (Captured by the authors)

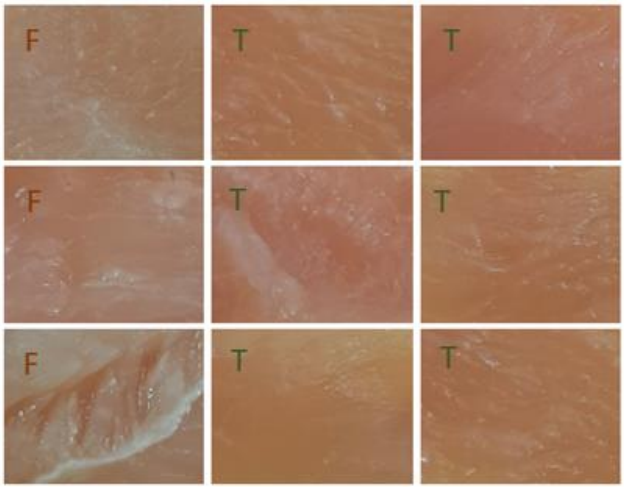


Figure 8. Chicken breast (RGB visible imaging), rotten class (correct versus incorrect) model predictions. **T** (green) indicates correct prediction; **F** (red) indicates incorrect prediction. (Captured by the authors)

In the classification of poultry meat images, particularly for samples in borderline states such as early defrosting or initial decay, visual characteristics may change minimally, making accurate differentiation challenging. This finding highlighted the need for more abstract and complex feature extraction in future studies to better distinguish such near cases. Conversely, completely rotten meat exhibits more pronounced visual changes, allowing artificial intelligence models to recognize it with higher accuracy and lower ambiguity. Among the tested

models, the baseline CNN achieved an average accuracy of 72.35%, primarily limited by lower defrosted meat detection, though it performed well in identifying rotten meat (97.56%) and moderately in healthy meat (85.36%). The VGG16 network detected rotten meat with 100% accuracy, performed moderately for healthy meat, and ranked second for defrosted detection (95.93%) after Xception. VGG19 improved healthy meat detection (95.22%) and achieved 89.43% accuracy for defrosted meat, with only two false positives for rotten samples (98.37%), yielding an overall mean accuracy of 94.3%. The Xception network, while less effective in identifying healthy meat (82.92%), performed well for defrosted (95.93%) and rotten meat (97.56%), highlighting its strength in detecting abnormal quality states. InceptionV3 excelled at healthy meat detection (91.05%), with slightly lower accuracy for defrosted meat (89.43%) and 97.56% for rotten meat. MobileNet achieved an overall accuracy of 92.82%, with 86.17% for healthy meat and 98.37% for rotten meat, while DenseNet121 reached 91.86% overall, detecting healthy meat at 86.17% and defrosted meat at 91.86%. Across all models, detection of rotten meat consistently outperformed the other two classes, whereas errors were primarily distributed between healthy and defrosted categories. In some architectures for example VGG16, Xception, MobileNet, and DenseNet, improving defrost detection led to reduced accuracy for healthy meat, while in others (CNN, VGG19, and Inception), higher healthy meat detection came at the expense of defrost accuracy. This pattern reflects the visual similarity between healthy and defrosted meat, making these borderline cases prone to misclassification, whereas rotten meat's distinct appearance minimizes confusion.

DISCUSSION

This study evaluated seven convolutional neural network architectures, including six ImageNet-pretrained backbones (Xception, InceptionV3, MobileNet, DenseNet121, VGG16, VGG19) and a baseline CNN, under a consistent protocol for classifying chicken breast images into healthy, defrosted, and rotten categories using standardized visible-light imaging. Two key observations emerged; rotten samples were consistently detected with the highest accuracy across models, while healthy and defrosted samples exhibited greater mutual confusion, reflecting their overlapping visual characteristics even under controlled lighting. This finding aligned with prior studies showing that pronounced spoilage cues such as color darkening, surface slime, and textural breakdown facilitate discrimination, whereas early or thaw-related changes are subtler (Frelka et al., 2019). The superior detection of rotten meat likely arises from distinctive surface features that remain informative even after grayscale conversion, whereas healthy versus defrosted meat differs primarily by subtle texture and reflectance cues, which can be attenuated by minor pose or brightness variations. Although several architectures achieved high overall accuracy, class-wise metrics and confusion matrices revealed trade-offs between sensitivity to defrosted meat and specificity for healthy meat, consistent with previous reports (Hamdy et al., 2020; Jian et al., 2025). Transfer learning proved advantageous under these controlled conditions, consistent with general food quality sensing practice (Li et al., 2022; Hidalgo et al., 2025), but results may not generalize beyond this dataset, which was acquired at a single slaughterhouse with one camera and photo-box setup, converted to grayscale, and lacking external validation. For practical deployment on processing lines, considerations include camera and lighting calibration, safety-critical thresholds to minimize false negatives for rotten meat, and throughput requirements, as per-image inference times and model sizes support near-real-time screening. Errors predominantly occurred between healthy and defrosted classes, suggesting that operating points can be tuned depending on risk priorities (Gallitto et al., 2025). This study is novel in three ways; first, it introduces a new, expert-verified dataset collected under controlled conditions across three freshness classes; second, it benchmarks seven ImageNet describing pre-trained CNNs using transfer learning to identify the most efficient model; and third, it shows that standard camera images without hyperspectral imaging or chemical tests can still deliver high accuracy, supporting a practical approach to real-time meat freshness screening.

CONCLUSION

This study evaluated seven convolutional neural network architectures on standardized visible-light images of chicken breast to classify samples as healthy, defrosted, or rotten. Across all models, two patterns were consistent, including rotten meat was identified most reliably, and most errors occurred between healthy and defrosted samples, where the visual differences are much more subtle, even under controlled lighting. Among the tested backbones, VGG19 achieved the highest overall test accuracy (~94.3%), while VGG16, InceptionV3, Xception,

MobileNet, and DenseNet121 also performed strongly, with the baseline CNN ranking lower. Class-wise results showed very high accuracy for rotten meat (~97-100%, depending on the model) and moderate trade-offs between sensitivity to defrosted and specificity for healthy. These outcomes demonstrated that transfer-learning-based models are effective for nondestructive screening of poultry meat quality when images are acquired under controlled geometry and lighting. Practically, this approach is suitable as a screening aid in quality-control workflows, particularly for flagging rotten samples, provided that sites implement camera and light calibration, maintain consistent acquisition protocols, and set class-specific operating thresholds, such as prioritizing low false-negative rates for safety-critical "rotten" samples. Further studies should be conducted on more meat samples from other animal species to evaluate the effectiveness of AI methods for assessing meat quality parameters.

DECLARATIONS

Funding

This study received no specific grant or financial support from any funding agency.

Availability of data and materials

The dataset and code used in this study are available upon reasonable request from the corresponding author.

Competing interests

The authors declare no competing interests.

Authors' contributions

Arash Biabanian coordinated data acquisition and labeling, drafted the manuscript, and led the critical revision process. Seyed Vahab Shojaedini conceived the study, designed the experiments, and developed the methodology and model implementation. Nima Misaghi conducted the analysis and contributed to the interpretation of the results. All authors reviewed the manuscript and approved the final version to be published.

Ethical considerations

All procedures involving post-slaughter poultry tissues complied with institutional biosafety regulations. Sampling permission was obtained from Sahar Roshan Slaughterhouse (Shahriyar, Iran). The authors did not use AI tools in preparation for the present study.

REFERENCES

Ametefe DS, Sarnin SS, Ali DM, Ametefe GD, John D, Aliu AA, and Zoreno Z (2024). Automatic classification and segmentation of blast cells using deep transfer learning and active contours. International Journal of Laboratory Hematology, 46(5): 837-849. DOI: https://www.doi.org/10.1111/ijlh.14305

Botta C, Buzzanca D, Chiarini E, Chiesa F, Rubiola S, Ferrocino I, Fontanella E, Rantsiou K, Houf K, and Alessandria V (2024). Microbial contamination pathways in a poultry abattoir provided clues on the distribution and persistence of *Arcobacter* spp. Applied and Environmental Microbiology, 90(5): e00296-24. DOI: https://www.doi.org/10.1128/aem.00296-24

Cai M, Li X, Liang J, Liao M, and Han Y (2024). An effective deep learning fusion method for predicting the TVB-N and TVC contents of chicken breasts using dual hyperspectral imaging systems. Food Chemistry, 456: 139847. DOI: https://www.doi.org/10.1016/j.foodchem.2024.139847

Chollet F (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 1800-1807. DOI: https://www.doi.org/10.1109/CVPR.2017.195

Dumoulin V and Visin F (2018). A guide to convolution arithmetic for deep learning. arXiv:1603.07285. Preprint, arXiv. DOI: https://www.doi.org/10.48550/arXiv.1603.07285

Food and agriculture organization (FAO) (2003). Pesticide residues in food. Report 2003. Joint FAO/WHO meeting on pesticide residues, pp. 1-280. Available at: https://openknowledge.fao.org/server/api/core/bitstreams/5b40fa1c-ca3f-4ad6-925b-0f4a40f436b8/content

Frelka JC, Phinney DM, Yang X, Knopp MV, Heldman DR, Wick MP, and Vodovotz Y (2019). Assessment of chicken breast meat quality after freeze/thaw abuse using magnetic resonance imaging techniques. Journal of the Science of Food and Agriculture, 99(2): 844-853. DOI: https://www.doi.org/10.1002/jsfa.9254

Fukushima K (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4): 193-202. DOI: https://www.doi.org/10.1007/BF00344251

Gallitto G, Englert R, Kincses B, Kotikalapudi R, Li J, Hoffschlag K, Bingel U, and Spisak T (2025). External validation of machine learning models—registered models and adaptive sample splitting. GigaScience, 14: giaf036. DOI: https://www.doi.org/10.1093/gigascience/giaf036

Goodfellow I, Bengio Y, and Courville A (2016). Deep learning. MIT Press., Cambridge, pp. 351-354. Available at: https://e-hir.org/upload/pdf/hir-22-351.pdf

Hamdy O, Abdel-Salam Z, and Abdel-Harith M (2020). Discrimination between fresh, chilled, and frozen/thawed chicken based on its skin's spectrochemical and optical properties. Analytical Methods, 12(16): 2093-3101. DOI: https://www.doi.org/10.1039/D0AY00324G

He K, Zhang X, Ren S, and Sun J (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. Proceedings of the IEEE International Conference on Computer Vision 2015 (ICCV), pp. 1026-1034. DOI: https://www.doi.org/10.1109/ICCV.2015.123

Hidalgo MM, Lima RC, Fernandes EA, Bacchi MA, and Sarriés GA (2025). Leveraging pre-trained computer vision models for accurate classification of meat freshness. Food Chemistry, 495(Pt 3): 146430. DOI: https://www.doi.org/10.1016/j.foodchem.2025.146430

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, and Adam H (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv

preprint arXiv:1704.04861. Available at: https://arxiv.org/abs/1704.04861

Jian R, Li G, Jun X, and Shi G (2025). Nondestructive freshness recognition of chicken breast meat based on deep learning. Scientific Reports, 15: 27538. Available at: https://www.nature.com/articles/s41598-025-13576-1

Kaushal S, Tammineni DK, Rana P, Sharma M, Sridhar K, and Chen HH (2024). Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances. Trends in Food Science & Technology, 146: 104408. DOI: https://www.doi.org/10.1016/j.tifs.2024.104408

Krizhevsky A, Sutskever I, and Hinton GE (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90. DOI: https://www.doi.org/10.1145/3065386

Li Y, Xue J, Wang K, Zhang M, and Li Z (2022). Surface defect detection of fresh-cut cauliflowers based on convolutional neural network with transfer learning. Foods, 11(18): 2915. DOI: https://www.doi.org/10.3390/foods11182915

Markiewicz-Keszycka M, Cama-Moncunill X, Casado-Gavalda MP, Dixit Y, Cama-Moncunill R, Cullen PJ, and Sullivan C (2017). Laser-Induced breakdown spectroscopy (LIBS) for food analysis: A review. Trends in Food Science & Technology, 65: 80-93. DOI: https://www.doi.org/10.1016/j.tifs.2017.05.005

Nychas GJ, Skandamis PN, Tassou CC, and Koutsoumanis KP (2008). Meat spoilage during distribution. Meat Science, 78(1-2): 77-89. DOI: https://www.doi.org/10.1016/j.meatsci.2007.06.020

Sans P and Combris P (2015). World meat consumption patterns: An overview of the last fifty years (1961-2011). Meat Science, 109: 106-111. DOI: https://www.doi.org/10.1016/j.meatsci.2015.05.012

Shi Y, Wang X, Borhan MS, Young J, Newman D, Berg E, and Sun X (2021). A review on meat quality evaluation methods based on non-destructive computer vision and artificial intelligence technologies. Food Science of Animal Resources, 41(4): 563-588. DOI: https://www.doi.org/10.5851/kosfa.2021.e25

Simonyan K and Zisserman A (2015). Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. Preprint, arXiv. DOI: https://www.doi.org/10.48550/arXiv.1409.1556

Sujiwo J, Kim D, and Jang A (2018). Relation among quality traits of chicken breast meat during cold storage: Correlations between freshness traits and torrymeter values. Poultry Science, 97(8): 2887-2894. DOI: https://www.doi.org/10.3382/ps/pey138

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A (2015). Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 2015, pp. 1-9. DOI: https://www.doi.org/10.1109/CVPR.2015.7298594

Varma S and Simon R (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7(1): 91. DOI: https://www.doi.org/10.1186/1471-2105-7-91

Zhu Y, Wang W, Li M, Zhang J, Ji L, Zhao Z, Zhang R, Cai D, and Chen L (2022). Microbial diversity of meat products under spoilage and its controlling approaches. Frontiers in Nutrition, 9: 1078201. DOI: https://www.doi.org/10.3389/fnut.2022.1078201

Zou X, Xin C, Wang C, Li Y, Wang S, Zhang W, Li JJ, Su S, and Xiao M (2024). Non-destructive detection of chicken freshness based on multiple features image fusion and support vector machine. International Journal of Agricultural and Biological Engineering, 17(6): 264-272. DOI: https://www.doi.org/10.25165/ijabe.v17i6.8783