

ELMF4EggQ: Ensemble Learning with Multimodal Feature Fusion for Non-Destructive Egg Quality Assessment

Md Zahim Hassan^{a,b}, Md. Osama^{a,b}, Muhammad Ashad Kabir^{c,d,*}, Md. Saiful Islam^{e,*}, Zannatul Naim^e

^aDepartment of Computer Science and Engineering, Bangladesh Army University of Science and Technology, Saidpur, 5310, Bangladesh

^bDepartment of Computer Science and Engineering, Rajshahi University of Engineering & Technology, Rajshahi, 6204, Bangladesh

^cSchool of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW, 2795, Australia

^dGulbali Institute for Agriculture, Water and Environment, Charles Sturt University, Wagga Wagga, NSW, 2678, Australia

^eDepartment of Animal Production and Management, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh

Abstract

Accurate, non-destructive assessment of egg quality is critical for ensuring food safety, maintaining product standards, and operational efficiency in commercial poultry production. This paper introduces *ELMF4EggQ*, an ensemble learning framework that employs multimodal feature fusion to classify egg grade and freshness using only external attributes – image, shape, and weight. A novel, publicly available dataset of 186 brown-shelled eggs was constructed, with egg grade and freshness levels determined through laboratory-based expert assessments involving internal quality measurements, such as yolk index and Haugh unit. To the best of our knowledge, this is the first study to apply machine learning methods for internal egg quality assessment using only external, non-invasive features, and the first to release a corresponding labeled dataset. The proposed framework integrates deep features extracted from external egg images with structural characteristics such as egg shape and weight, enabling a comprehensive representation of each egg. Image feature extraction is performed using top-performing pre-trained CNN models (ResNet152, DenseNet169, and ResNet152V2), followed by principal component analysis (PCA)-based dimensionality reduction, synthetic minority oversampling technique (SMOTE) augmentation, and classification using multiple machine learning algorithms. An ensemble voting mechanism combines predictions from the best-performing classifiers to enhance overall accuracy. Experimental results demonstrate that the multimodal approach significantly outperforms image-only and tabular (shape and weight) only baselines, with the multimodal ensemble approach achieving 86.57% accuracy in grade classification and 70.83% in freshness prediction. The framework demonstrates strong potential for real-time, low-cost deployment in commercial egg processing environments. It highlights the feasibility of using computer vision and lightweight structural inputs for scalable, non-invasive egg quality evaluation. All code and data are publicly available at https://github.com/Kenshin-Keeps/Egg_Quality_Prediction_ELMF4EggQ, promoting transparency, reproducibility, and further research in this domain.

Keywords: Egg quality, freshness, grade, imaging, machine learning, multimodal fusion, ensemble framework, classification.

1. Introduction

With the ongoing growth of the global population, there is a steadily increasing demand for food sources that are both affordable and nutritionally rich. Among the available options, eggs have emerged as a particularly important staple due to their high-quality protein content, relatively low cost, and well-balanced nutritional profile [8]. As a widely consumed food item, eggs contribute significantly to human nutrition by providing essential macronutrients and micronutrients, including proteins, fats, vitamins, and minerals. Their biological value, especially in terms of protein quality, is considered one of the highest among natural food sources, comparable to meat, dairy products, and legumes [59]. The significance of eggs as a global food source is underscored by the World Egg Organization (WEO), which recognizes the second Friday of October each year as World Egg Day to promote awareness of their nutritional and economic value [77]. Eggs are not only an affordable source of high-quality protein and relatively low in calories (about 140 kcal per 100 gm) but also provide a rich supply of essential fats, vitamins, and minerals [50].

However, nutritional value alone does not ensure consumer acceptance. Consistent quality assurance is critical for both producers and consumers to maintain market standards and ensure food safety. The term “quality” refers to the level of excellence of a product or its suitability for a specific purpose [64]. Quality assessment plays a crucial role in ensuring product consistency, extending shelf life, and safeguarding consumer health [44]. To maintain the nutritional and commercial value of eggs throughout the supply chain, effective and reliable quality assessment methods are essential. Egg quality encompasses external characteristics, such as size, and internal attributes, primarily yolk freshness and albumen (egg white) thickness, which degrade over time [15, 3]. External features such as egg size and weight significantly influence consumer perception of quality [40]. Traditionally, egg quality has been assessed using manual techniques such as candling, thermal imaging [84], hyperspectral imaging [12], and destructive methods that require breaking the egg to observe internal attributes [1, 2, 57, 40]. While effective, these techniques are labor-intensive, time-consuming, subjective, and often unsuitable for automated systems [39]. As a result, non-destructive testing methods, including shell texture analysis, have gained prominence for their ability to evaluate egg quality rapidly, consistently, and without damaging the product [40], forming the basis of our study’s innovative framework.

Regarding the physical characteristics of an egg, the texture of the shell is a key indicator of quality, reflecting its structural strength, freshness, and ability to resist microbial contamination [16, 54, 84, 33]. A smooth, uniform texture generally reflects optimal calcification during shell formation, whereas roughness, cracks, or abnormal ridges may signal physiological stress, aging, or nutritional deficiencies in laying hens. Such texture anomalies, including repaired cracks and porous surfaces, can compromise the eggshell’s barrier properties, leading to increased susceptibility to bacterial penetration and moisture loss [7, 75]. Notably, eggshell translucency and irregular surface patterns have been associated with weakened shell membranes and heightened microbial risk, particularly in older flocks and

*Corresponding authors.

Email addresses: zahimhassan@baust.edu.bd, zahim951xtgmovi@gmail.com (Md Zahim Hassan), osama_cse@baust.edu.bd, osamasohag39@gmail.com (Md. Osama), akabir@csu.edu.au (Muhammad Ashad Kabir), saiful.apma@sau.edu.bd (Md. Saiful Islam), znaim.apma@sau.edu.bd (Zannatul Naim)

under fluctuating environmental conditions [53]. Comparative studies further reveal that eggshell surface morphology, including roughness and pigmentation, is shaped by ecological and evolutionary factors, reinforcing its biological and functional significance [6]. While external features, such as shell texture, are critical for consumer appeal and safety, internal quality metrics, such as yolk freshness and albumen thickness, are equally vital for assessing nutritional value.

Among the significant parameters, the yolk index (YI) is a significant indicator for assessing egg freshness and overall internal quality. The theory behind the low yolk index method is rooted in the physical changes an egg undergoes as it ages. As an egg loses its freshness, its internal quality deteriorates, leading to a loss of structural integrity. This is specifically manifested as the yolk becoming more flattened and spreading out, while the egg white (albumen) also thins and loses viscosity. These changes result in a decrease in the ratio of yolk height to yolk diameter. A lower yolk index value, therefore, indicates an older and less fresh egg [50]. The Haugh unit (HU) is another well-established metric, widely used in commercial and research settings to evaluate the grade of eggs [82]. It is calculated based on the height of the thick albumen relative to the weight of the egg when broken onto a flat surface. Higher HU values indicate better egg quality. As eggs age during storage, the albumen becomes thinner and the HU value decreases, providing a quantitative means of grading egg quality. In this study, we employ the Haugh unit as the reference standard for egg grading, while the yolk index serves as the primary indicator for freshness detection.

Recent advancements in artificial intelligence and computer vision, particularly machine learning and deep learning, have enabled a shift toward fully automated, image-based methods for evaluating egg quality. However, many existing approaches [12, 24, 42] focus exclusively on either internal imaging modalities (e.g., hyperspectral or X-ray) or external visual features, without leveraging the correlation between the two. To address the limitations of traditional methods and the disconnect between internal and external quality assessments, this study presents a comprehensive approach that bridges internal and external data sources to improve the accuracy and reliability of egg quality assessment. The main contributions of this work are as follows:

- We construct and introduce a novel egg dataset with labels for both egg grade and freshness level. This dataset supports supervised learning for quality evaluation tasks and promotes further advancements in the field.
- We propose an ensemble-based multimodal feature fusion framework that combines egg external image features with egg shape and weight to improve egg grade and freshness classification.
- We evaluate the effectiveness of using external egg images as a low-cost, non-destructive method for predicting both grade and freshness. This approach is particularly suitable for real-time applications in commercial egg processing facilities.

2. Related Work

A diverse body of research has emerged focusing on egg quality measurement. Various studies have proposed different features as indicators of egg quality, including internal quality grading using egg-specific features, external

assessments based on conventional or processed images (such as thermal imaging), grading based on egg shape indices, evaluation of volume, weight, and size, detection of defects such as cracks, dark spots or double yolks, and the assessment of egg freshness.

A substantial body of research has been dedicated to various aspects of egg quality assessment. Many studies have focused on external appearance analysis, including the detection of surface defects such as cracks, dirt, and blood spots [27, 72, 37, 37, 17, 48]. Others have investigated size, weight, and volume estimation to classify eggs into different grades [41, 28, 70, 4, 63, 76]. Internal quality evaluation has also been explored, with methods addressing dark spot detection and double yolk identification [29, 35]. In addition, several works have targeted freshness assessment through non-destructive techniques, including infrared thermal imaging and spectral analysis of eggshell properties [49, 32, 65, 12, 20]. Overall, these studies demonstrate the diversity of features and modalities considered for automated egg quality assessment.

While external physical features such as shell thickness and egg shape provide preliminary information about an egg's quality, they do not always reliably reflect its internal condition. In practice, internal quality indicators, such as the Haugh Unit and Yolk Index, are more robust measures of egg grading and freshness. The Haugh unit [31], derived from the relationship between egg weight and albumen height, is a widely accepted international standard for grading eggs. Similarly, the yolk index serves as a critical parameter for evaluating the deterioration of yolk quality over time. Accurate and non-destructive evaluation of these internal parameters is crucial, especially for meeting consumer expectations and regulatory requirements regarding egg freshness and grading. Thus, methods that can directly estimate internal quality hold particular significance over approaches relying solely on external characteristics.

Table 1 summarizes several notable works in the field of egg quality assessment. Yang et al. [81] developed a system for categorizing eggs based on external defects and predicting weight using major and minor axes, without addressing internal freshness. Similarly, Omid et al. [44] proposed an expert system for grading eggs into five quality categories based on surface conditions and size-related features. Dai et al. [12] and Zhang et al. [84] explored freshness estimation; however, their methods relied on processed image features, such as spectral characteristics or thermal imaging, rather than directly using biological indicators. To the best of our knowledge, only our proposed method directly integrates internal quality grading based on the Haugh Unit and freshness detection using the Yolk index. By combining multiple feature modalities, including external image, shape index, and weight measurement, our approach uniquely addresses the internal quality and freshness evaluation of eggs within a single, unified framework.

Building upon these insights, our work proposes a novel system that takes advantage of easily available egg characteristics to achieve both internal quality grading and freshness classification. Unlike prior studies that often focused on isolated aspects of quality or employed complex and expensive imaging techniques, our approach emphasizes practical applicability by utilizing standard features such as external images, shape indices, and weight measurements.

Table 1: Summary of related work on egg quality assessment

Study	Objective			Feature Modality				Methodology
	Internal Grading	Freshness	Others	Image (External)	Shape	Weight	Others	
[81]	✗	✗	✓ ^a	✓	✗	✗	✗	Feature fusion using image diffraction patterns and hierarchical clustering; classified using RMTDet + RF.
[1]	✗	✗	✓ ^b	✓	✓	✗	✓	PCA and IGR used for feature selection from preprocessed images; classifiers include KNN.
[44]	✗	✗	✓ ^c	✗	✓	✗	✓	Color segmentation in HSV space and fuzzy logic-based inference for egg grading.
[57]	✗	✗	✓ ^d	✗	✗	✓	✓	Normalization and 10-fold cross-validation of extracted statistical features; classification with logistic regression.
[20]	✗	✗	✓ ^e	✓	✗	✗	✓	Grayscale conversion, background removal, and weight ratio estimation; classification using logistic regression.
[12]	✗	✗	✓ ^f	✓	✗	✗	✓	Spectral calibration and noise filtering of hyperspectral images; ensemble method used for classification.
[84]	✗	✗	✓ ^g	✗	✗	✗	✓	Thermal video processing with frame extraction, edge detection, and feature selection; classification using SegNet+SVM.
Ours	✓	✓	–	✓	✓	✓	–	Ensemble learning with multimodal feature fusion

^a Classifies eggs into five categories (intact, cracked, bloody, floor, and non-standard) and predicts weight using geometric features.

^b Grades eggs (AA, A, B, C, D, E) based on egg shape parameters.

^c Grades eggs (Excellent, Good, Medium, Bad, Wastage) using characteristics such as cracks, blood spots, breakage, and size.

^d Uses 20 features for binary classification into acceptable or unacceptable quality.

^e Estimates freshness using egg density based on height and width.

^f Applies Haugh unit-based criteria for shell quality assessment, but not yolk index.

^g Proposes a non-destructive method for freshness estimation using thermal imaging.

3. Dataset

3.1. Egg Collection

A total of 288 brown-shelled chicken eggs were collected from Dhaka, Bangladesh. To capture variability in origin, handling practices, and storage conditions, eggs were sourced equally ($n = 72$) from four distinct market categories: wholesale market (WM), super shop (SS), grocery shop (GS), and open shop (OS). After collection, eggs were transported to the laboratory on the same day of collection to preserve freshness and quality. The eggs originated from a diverse range of farms, sources both within and outside Dhaka. Each market category reflects a unique distribution channel. Wholesale markets (WM) serve as major distributors supplying large volumes of eggs to other retailers. Super shops (SS) obtain eggs either from wholesale markets or through corporate supply chains.

Grocery shops (GS) typically sell eggs alongside various daily necessities, while open shops (OS) operate through small independent stalls or mobile vendors that sell eggs directly to consumers.

This stratified and balanced sampling design was implemented to capture real-world market variability in egg supply chains. By including diverse market types, the dataset reflects the variation in sourcing, storage, and handling practices that consumers are exposed to. This diversity is crucial for evaluating how different retail conditions influence the external and internal quality of eggs, as well as their safety for consumption. All eggs were handled under controlled laboratory conditions upon arrival to maintain the integrity of quality.

3.2. Dataset Preparation

Following collection from the four distinct market categories, all eggs were stored under standard refrigeration conditions to ensure consistency in post-collection handling. At regular intervals of seven days, a random subset of eggs from each market group was selected for data acquisition. This process involved capturing external images and conducting internal assessments by breaking the eggs. This staggered sampling protocol was designed to introduce variability in egg freshness and quality, thereby allowing the dataset to reflect a broader spectrum of egg grades over time. By simulating real-world storage durations, the resulting dataset incorporates natural degradation patterns and quality transitions, which are critical for training and evaluating models aimed at freshness prediction and quality classification.

Each egg underwent standardized image capturing to support downstream analysis. All illustrated in Figure 1, all images were taken using a fixed digital camera setup positioned 15 cm above the egg surface, under controlled and uniform lighting conditions. This imaging process ensured visual uniformity across the dataset and supports future applications in computer vision-based egg quality assessment. Following image capture, the eggs were promptly taken to the controlled environment to preserve freshness and internal quality. To evaluate the physical characteristics of the collected eggs, two primary external attributes were measured: egg weight and shape index. Egg weight was measured using a precision digital weighing scale with an accuracy of ± 0.01 grams, with consistent procedures [56].

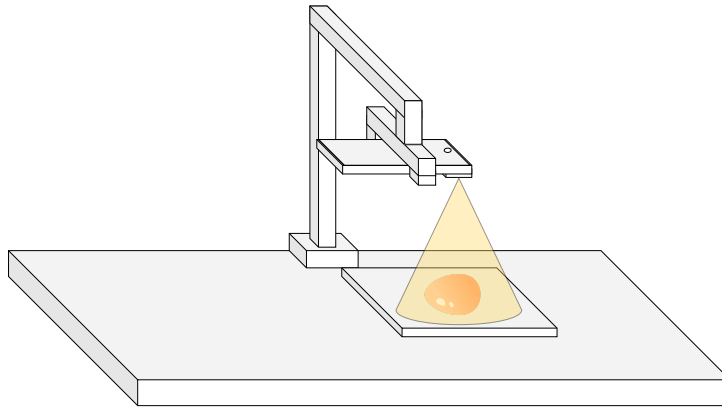


Figure 1: Image capturing setup for individual egg sample.

It plays a central role in egg classification and quality grading, particularly in retail and food safety contexts.

Shape index (SI) was used to describe the proportional relationship between the egg's width and length, offering insights into egg morphology and consumer preference. The shape index was calculated according to the formula [52] presented in Equation 1.

$$\text{Shape index (SI)} = \left(\frac{\text{Width}}{\text{Length}} \right) \times 100 \quad (1)$$

where *Width* represents the maximum breadth (width) of the egg and *Length* is the maximum longitudinal length, illustrated in Figure 2. Higher shape index values indicate more spherical eggs.

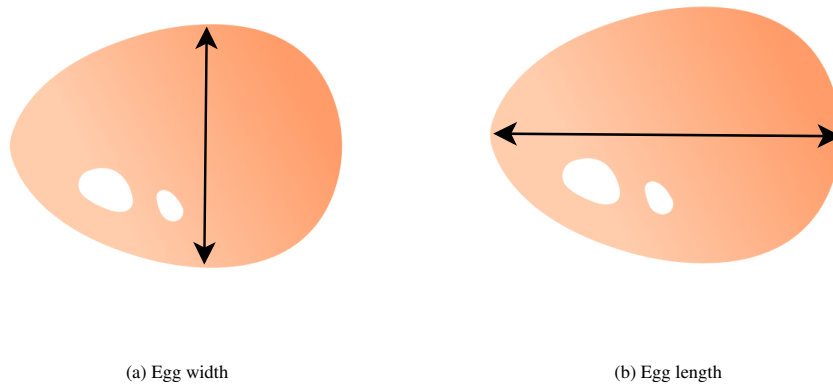


Figure 2: Measurement of Egg Shape Index using width and length.

Among internal quality parameters, the yolk index (YI) is widely recognized as a reliable indicator of egg freshness. Yolk diameter was measured using a digital caliper shown in Figure 3a, while yolk height was determined with a trivet micrometer as shown in Figure 3b. The yolk index was then computed using Equation 2, which expresses the ratio of yolk height to yolk diameter:

$$\text{Yolk index (YI)} = \left(\frac{\text{Yolk height}}{\text{Yolk diameter}} \right) \times 100 \quad (2)$$

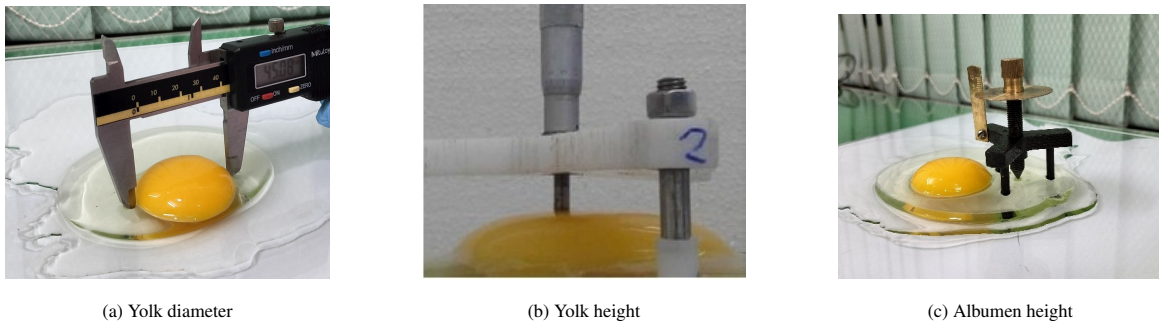


Figure 3: Measurement of an egg's internal attributes.

Table 2 provides clear thresholds for categorizing eggs into different freshness levels based on their yolk index (YI) [26]. Eggs with a YI greater than 38 are considered fresh, reflecting a rounder, firmer yolk and superior internal quality. Eggs with a YI between 34.5 and 38 are classified as moderately fresh, indicating slightly reduced yolk firmness, while values below 34.5 correspond to old eggs with a flatter yolk and diminished quality. This classification system allows for an objective assessment of egg freshness, providing a standardized approach that is useful for both research studies and industrial quality control processes.

Table 2: Yolk Index Classification

Freshness category	Yolk index (YI) Range
Fresh	$YI > 38$
Moderately fresh	$34.5 \leq YI \leq 38$
Old	$YI < 34.5$

The haugh unit (HU) is a widely accepted quantitative measure of internal egg quality. It incorporates both the egg's weight and the height of its thick albumen. HU values were calculated using the formula [26] that is given in Equation 3 :

$$HU = 100 \times \log_{10} (H + 7.6 - 1.7 \times W^{0.37}) \quad (3)$$

where H denotes the albumen height, which is measured as depicted in Figure 3c, and W is the egg weight in grams. Based on the resulting HU scores, each egg was labeled into quality grades (AA, A, B, or C) [30].

Table 3 summarizes egg quality grades based on the haugh unit, which incorporates both albumen height and egg weight. Grade AA represents the highest quality eggs with firm albumen, while Grade C denotes eggs of low internal quality. Grades A and B indicate intermediate quality levels based on albumen firmness and egg weight. Together with the yolk index, HU provides a comprehensive framework for internal egg quality grading. These measurements are essential for accurately assessing egg freshness and consistency and support reliable monitoring in quality control systems.

Table 3: Haugh unit-based grading

Grade	Haugh unit (HU) range
AA	$HU \geq 72$
A	$60 \leq HU < 72$
B	$31 \leq HU < 60$
C	$HU < 31$

3.3. Dataset Description

During the data collection process, a number of eggs were excluded due to various quality and integrity issues. Specifically, some eggs were accidentally broken during handling, found to be rotten upon inspection, or had damaged yolks that made internal assessment unreliable. In addition, samples with missing measurements, data inconsistencies, or extreme outlier values were systematically identified and removed during preprocessing. After applying these quality control steps, a total of 186 valid samples remained from the initial 288 collected eggs. These cleaned and verified samples form the final dataset used for analysis and model development in this study.

Categorical simplification was applied to ensure a more balanced dataset, as there was an imbalance in the number of samples within the initial classes. The original egg grading scheme, which included four distinct levels (AA, A, B, and C), was restructured into two broader categories: samples classified as AA or A were grouped under the “High” grade class, while those falling into grades B or C were assigned to the “Low” grade class. Similar to grading, the initial freshness classification, which distinguished among *Fresh*, *Moderately fresh* and *Old* eggs, was consolidated into two categories. Eggs labeled as either *Fresh* or *Moderately fresh* were grouped under the unified “Fresh” class, whereas eggs originally labeled as *Old* were retained in a separate “Old” category. This binary categorization improved class balance and enabled more robust model training and evaluation. Table 4 presents a summary of the descriptive statistics for the 186 finalized samples. The features include egg weight, shape index, yolk index, and Haugh unit, categorized by the newly defined binary grade (High, Low) and freshness labels (Fresh, Old). Statistics are reported in terms of minimum, maximum, mean, and standard deviation (SD).

Table 4: Descriptive Statistics of Egg Features by Grade and Freshness (n=186)

Attribute	Measure	Grade (n=186)		Freshness (n=186)	
		High (n=78)	Low (n=108)	Fresh (n=90)	Old (n=96)
Weight (g)	Min	47.69	44.78	47.69	44.78
	Max	77.15	70.30	77.15	70.30
	Mean (SD)	60.13 (4.69)	58.92 (4.57)	60.35 (4.62)	58.57 (4.53)
Shape index	Min	69.54	71.91	69.54	71.91
	Max	96.00	96.21	96.00	96.21
	Mean (SD)	78.84 (3.82)	77.47 (3.15)	78.50 (3.71)	77.62 (3.27)
Yolk index	Min	34.79	1.91	34.54	1.91
	Max	50.12	37.25	50.12	34.33
	Mean (SD)	38.66 (2.57)	25.96 (7.80)	38.27 (2.62)	24.74 (7.41)
Haugh unit	Min	61.10	1.80	32.91	1.80
	Max	104.50	59.70	104.50	59.70
	Mean (SD)	77.91 (9.18)	41.35 (13.96)	73.99 (13.52)	40.45 (14.28)
Market (count)	GS	17	28	22	23
	OS	27	26	28	25
	SS	11	33	13	31
	WM	23	21	27	17

As seen in the table, High-grade and Fresh eggs exhibited better values across multiple features. The yolk index

and haugh unit in particular showed strong contrast between High vs. Low grade, and Fresh vs. Old groups, reinforcing their effectiveness as egg quality and freshness indicators. The distribution across market categories remained well represented, enabling reliable comparative and predictive analyses in subsequent modeling tasks.

4. Methodology

Figure 4 illustrates the proposed ensemble-based multimodal feature fusion framework designed for automated egg grading and freshness classification. The framework leverages multimodal feature fusion by combining visual features extracted from egg images with egg structural attributes – specifically shape and weight – to capture comprehensive information about each egg. It employs multiple feature extractors to derive diverse feature representations from the egg image, which are subsequently fused and passed through a series of modules for augmentation, transformation, and classification. The ensemble mechanism at the final stage leverages majority voting across multiple classifiers to produce robust predictions for both egg grade (high vs. low) and freshness (fresh vs. old).

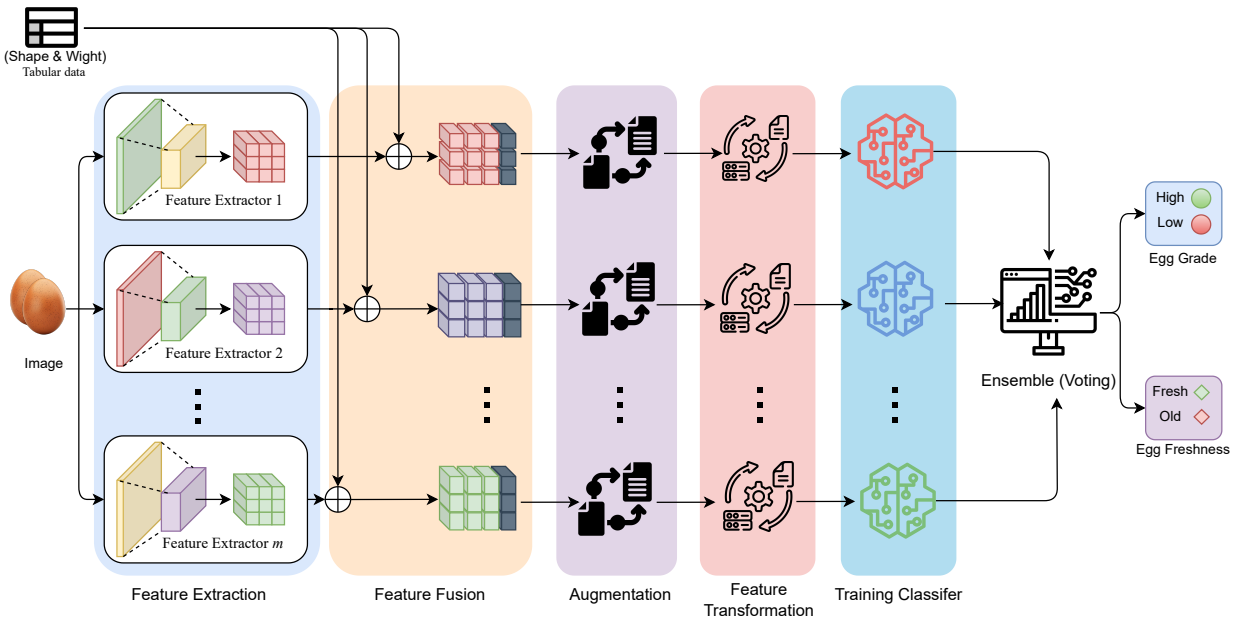


Figure 4: Proposed ensemble-based multimodal feature fusion framework for egg grading and freshness classification

4.1. Feature Extraction and Fusion

Feature extraction is often used in image processing for simplifying image representation, thus enhancing computational efficiency and improving performance [38]. Images contain various types of features, including statistical, color, and texture features, among others. However, extracting each feature type individually from every image is computationally intensive and inefficient.

To overcome these limitations, especially in scenarios with limited data resources, this study employs transfer learning using pre-trained convolutional neural networks (CNNs) for image feature extraction [14, 47, 11]. In such settings, models pre-trained on large-scale datasets (e.g., ImageNet) are leveraged by removing their final classification layers, thereby utilizing the learned representations as feature extractors. When an input image is passed through such a model, the output of the final global average pooling layer serves as a high-level feature vector representing the image. We evaluate the effectiveness of multiple state-of-the-art pre-trained CNN architectures for this purpose, including InceptionResNetV2 [66], Xception [10], ResNet101, ResNet152, ResNet12V2 [21], MobileNetV2 [55], DenseNet169, DenseNet201 [25], InceptionV3 [67], EfficientNetB7 [68], ConvNeXtTiny, ConvNeXtLarge [34] pre-trained models as the feature extractor.

Prior to feature extraction, all images were normalized and resized to a resolution of 224×224 pixels. For each CNN model, the final prediction layer was excluded, and the output of the last global average pooling layer was retained as the image feature vector. For instance, DenseNet169 yields a feature matrix of size $n \times 1664$ for n input images, while ResNet152 and ResNet152V2 produce matrices of size $n \times 2048$.

In addition to image-based features, tabular data comprising structural attributes such as egg shape and weight were also incorporated. A simple sample-wise concatenation strategy was employed to fuse these numerical features with the extracted image features, resulting in a comprehensive multimodal feature representation for each egg.

4.2. Augmentation

To address the issue of class imbalance in the dataset, an augmentation technique called Synthetic Minority Over-sampling Technique (SMOTE) is used to increase the samples to eliminate the class imbalance problem [9]. This method involves selecting an instance from the minority class and identifying its nearest neighbors. A new instance is generated by interpolating within the feature space between the original instance and its neighbors. Through this process, SMOTE produces additional instances of the minority class that closely resemble existing samples, thus increasing the sample size of the minority class and mitigating class imbalance. This method increases the number of minority class samples in a way that preserves the underlying data distribution while avoiding exact duplication of existing instances. As a result, SMOTE enhances the model's ability to generalize across underrepresented classes and mitigates the risk of biased learning due to class imbalance.

In this study, SMOTE was applied to the merged multimodal feature set comprising both image-derived and tabular features to generate a balanced dataset across all target classes. Table 5 reports the class distributions before and after the SMOTE operation, highlighting that after performing augmentation, the dataset becomes perfectly balanced.

4.3. Feature Transformation

High-dimensional feature spaces often introduce computational overhead, particularly in real-time applications, and may include redundant or non-informative variables that do not significantly contribute to the predictive performance of the model. To address these challenges, this study applies Principal Component Analysis (PCA) [36] as a

Table 5: Sample distribution of target classes (before and after augmentation)

Task	Category	Before Augmentation	After Augmentation
Egg Grade Prediction	High	78	108
	Low	108	108
Egg Freshness Prediction	Fresh	90	96
	Old	96	96

dimensionality reduction technique aimed at transforming the feature space while preserving 99% of the original data variance.

This transformation reduces the dimensionality of the feature vectors, resulting in more compact representations that retain the essential patterns and information necessary for classification tasks. The method begins by computing the covariance matrix of the original features to understand the relationships between variables. The covariance between any two features, w and q , is calculated using Equation 4:

$$\text{cov}(w, q) = \frac{\sum_{i=0}^n (w_i - \mu_w)(q_i - \mu_q)}{n - 1} \quad (4)$$

where n represents the number of samples, and μ_w and μ_q denote the mean values of features w and q , respectively.

Once the covariance matrix is obtained, PCA identifies new orthogonal axes, known as principal components, that maximize the variance in the data. These components are derived from the eigenvectors and eigenvalues of the covariance matrix, with eigenvalues indicating the amount of variance captured by each principal component. The components are ranked in descending order based on the variance they explain, and the top- k components that cumulatively account for at least 99% of the total variance are retained. The original high-dimensional data is then projected onto this reduced feature space, resulting in a transformed dataset that is computationally more efficient and less prone to overfitting.

4.4. Training Classifier

The transformed feature sets obtained after dimensionality reduction were used to train a variety of machine learning classifiers to evaluate performance across diverse algorithmic paradigms. By incorporating a diverse set of classifiers, this study aims to comprehensively evaluate the effectiveness of the proposed feature sets in classifying egg quality and freshness. The classifiers include: Regression, Decision Tree, Random Forest, Support Vector Classifier (SVC), Gradient Boosting, Multi-Layer Perceptron (MLP), XGBoost, LightGBM, and Adaptive Boosting. Logistic Regression is a statistical method that utilizes the sigmoid function for classification tasks. Decision Tree adopts a decision-based approach, which is to recursively partition a dataset into smaller subsets based on feature values, creating a tree-like structure where each node represents a decision or outcome to classify instances into target classes.

Random Forest, Gradient Boosting, XGBoost, and LightGBM are ensemble-based models. Random Forest leverages an ensemble of decision trees, each trained on random subsets of data and features, to generate predictions.

Gradient Boosting, another ensemble-based technique, constructs decision trees sequentially, with each tree designed to correct the errors of its predecessors. XGBoost uses an extreme gradient boosting algorithm with an ensemble approach for categorization tasks. Similarly, LightGBM employs a light-gradient boosting algorithm, focusing on minimizing the loss function through gradient descent. All gradient boosting methods operate by combining weak learners to form a robust predictive model. Support Vector Classifier (SVC) identifies an optimal hyperplane to separate the data so that prediction can be done based on this boundary. The Multi-Layer Perceptron (MLP), a type of multilayer feedforward neural network, is designed to capture complex, non-linear relationships for classification tasks. Adaptive Boosting, also an ensemble-based method, aggregates the outputs of multiple simpler algorithms to produce a final prediction.

4.5. Ensemble

In the final stage of the proposed framework, predictions from multiple classifiers are aggregated to produce a more robust and accurate final prediction. To achieve this, we employed the majority voting technique, one of the most widely used ensemble methods for classification tasks. In this approach, each individual classifier contributes a vote for the predicted class label, and the class that receives the majority of votes is selected as the final prediction for a given input. The ensemble decision mechanism can be formally expressed as:

$$\hat{y} = \text{mode}[C_1(x), C_2(x), \dots, C_m(x)] \quad (5)$$

where $\hat{y}$ denotes the final ensemble prediction, x represents the input sample, and $C_m(x)$ is the predicted label produced by the m^{th} classifier.

This majority voting strategy enhances the generalization capability of the model by leveraging the complementary strengths of individual classifiers and mitigating the impact of any single model's misclassification. By aggregating diverse decision boundaries, the ensemble model achieves improved stability and predictive performance compared to any single classifier in isolation.

5. Experimental Evaluation

5.1. Settings

To rigorously evaluate model performance for both egg grading and freshness classification tasks, a 10-fold cross-validation strategy was employed. This method ensures a robust and unbiased estimate of each model's generalization capability by partitioning the dataset into 10 equal subsets. In each fold, 9 subsets were used for training and 1 for testing, with the process repeated across all folds. The average classification accuracy across the 10 folds was computed and reported as the final performance metric for each experiment.

To ensure fair comparisons and optimal performance, Grid Search was applied during the cross-validation process to identify the best-performing hyperparameters for each classifier. The optimal hyperparameter configurations discovered through this process are summarized in Table 6.

Table 6: Hyperparameter table

Model	Hyperparameter	Search space	Best value
Logistic Regression	C	0.1, 1, 10, 100	100
	penalty	l1, l2	l2
	class weight	None, balanced	None
Decision Tree	max_depth	None, 2, 5, 10, 20	2
	min_samples_split	2, 5, 10	10
	criterion	gini, entropy, log_loss	entropy
	class weight	None, balanced	balanced
Random Forest	n_estimators	50, 100, 150	100
	max_depth	None, 2, 5, 10, 20	10
	min_samples_split	2, 5, 10	2
	criterion	gini, entropy, log_loss	entropy
	class_weight	None, balanced	balanced
SVC	C	0.1, 1, 10	10
	kernel	linear, poly, rbf, sigmoid	rbf
	gamma	scale, auto	scale
	class_weight	None, balanced	None
Gradient Boosting	n_estimators	50, 100, 150	100
	learning_rate	0.001, 0.01, 0.1	0.1
	max_depth	3, 5	3
Multi-Layer Perceptron	hidden_layer_sizes	(8,), (16,), (8,16), (16,32)	(16, 32)
	activation	relu, tanh	relu
	learning_rate_init	0.001, 0.01, 0.1	0.1
	alpha	0.0001, 0.001	0.0001
	solver	sgd, adam	adam
XGBoost	n_estimators	50, 100, 150	50
	learning_rate	0.001, 0.01, 0.1, 0.5	0.5
	max_depth	2, 3, 5, 7	2
	subsample	0.8, 1.0	0.8
	colsample_bytree	0.8, 1.0	1
	reg_lambda	0, 1	0
	min_child_weight	1, 3, 5	3

Model	Hyperparameter	Search space	Best value
LigthGBM	n_estimators	50, 100, 150	150
	learning_rate	0.01, 0.1	0.1
	max_depth	2, 3, 5	3
	num_leaves	5, 10, 15, 20	10
	subsample	0.8, 1.0	0.8
Adaptive Boosting	n_estimators	50, 100, 200	100
	learning_rate	0.001, 0.01, 0.1, 0.5, 1.0	0.01
	estimator	Decision Tree, Random Forest	Random Forest

Table 7: Feature transformation results

Feature Extractor	Feature count	
	Extracted	PCA
InceptionResNetV2	1536	125
Xception	2048	156
ResNet101	2048	129
ResNet152	2048	131
MobileNetV2	1280	163
DenseNet169	1664	166
InceptionV3	2048	160
ResNet152V2	2048	74
EfficientNetB7	2560	104
ConvNeXtTiny	768	138
ConvNeXtLarge	1536	132
DenseNet201	1920	157

Table 7 shows the number of features generated by different feature extractors. However, the raw feature dimensions are quite large, which can make model training computationally expensive and less efficient. To address this, we applied Principal Component Analysis (PCA) to reduce the dimensionality of the features. This reduction helps create a more efficient training while preserving the most important information from the original features.

5.2. Selection of Feature Extractors

The effectiveness of pre-trained convolutional neural networks (CNNs) for feature extraction varies significantly due to architectural differences, which influence how well models capture either object-specific or domain-specific features [22, 18]. To identify the most suitable feature extractors for egg quality classification, we evaluated a range of state-of-the-art pre-trained models on their ability to generate informative representations from egg images. Each model was used to extract features from the image dataset, followed by dimensionality reduction using PCA, and then classified using a Random Forest classifier. Random Forest was selected for this comparison due to its consistent and strong performance on PCA-transformed feature spaces within our preliminary experiments.

The comparative performance of the pre-trained models in the context of egg grade classification is presented in Table 8. This analysis guided the selection of the most effective image-based feature extractor for inclusion in the final ensemble framework.

Table 8: Performance comparison of pre-trained CNN models as feature extractors for egg grade classification using the Random Forest classifier.

Feature Extractor	Accuracy (%)
InceptionResNetV2	75.00
Xception	72.22
ResNet101	72.69
ResNet152	79.63
MobileNetV2	75.46
DenseNet169	79.17
InceptionV3	72.22
ResNet152V2	77.31
EfficientNetB7	76.39
ConvNeXtTiny	73.15
ConvNeXtLarge	71.76
DenseNet201	75.93

The results indicate that ResNet152 achieved the highest accuracy of 79.63%, followed closely by DenseNet169 at 79.17%, and ResNet152V2 at 77.31%. These top three models significantly outperformed others, such as ConvNeXtLarge (71.76%) and Xception (72.22%), with ResNet152’s accuracy being approximately 4% higher than EfficientNetB7 (76.39%) and 4.63% higher than InceptionResNetV2 (75.00%). The strong performance of ResNet-based models (ResNet152 and ResNet152V2) suggests that the residual learning framework may be particularly effective for this task. The range of accuracies across all models, from 71.76% to 79.63%, highlights the variability in feature extraction effectiveness.

Based on their superior performance, **ResNet152**, **DenseNet169**, and **ResNet152V2** were selected as the final fea-

ture extractors for the framework. Feature vectors from each model were independently extracted for every image, and their prediction outputs were subsequently integrated during the ensemble stage to determine the final classification result.

5.3. Egg Grading Classification - Tabular vs Image vs Multimodal

To evaluate the effectiveness of different data modalities in predicting egg grading, we conducted comparative experiments using three distinct input types: (i) tabular features (i.e., egg shape and weight), (ii) image-based features, and (iii) a multimodal fusion of both image and tabular features presented in Table 9. For both the image-based and multimodal approaches, features were extracted using the top-performing pre-trained CNN models identified in earlier experiments: ResNet152, DenseNet169, and ResNet152V2. These models were selected based on their superior performance in extracting discriminative visual representations relevant to egg quality.

The image-based approach leverages deep feature embeddings derived from these pre-trained CNN models to capture external egg characteristics such as shell color, texture, and contour. In contrast, the tabular approach relies solely on numerical descriptors of structural properties specifically, egg shape and weight. The multimodal framework integrates both visual and structural features through concatenation, enabling the model to learn richer and more holistic representations.

Table 9: Classification accuracy (%) comparison across different input modalities (Tabular, Image only, and Multimodal) and classifiers for egg grading. Image-based and multimodal approaches used features extracted from ResNet152, DenseNet169, and ResNet152V2. The highest accuracy for each modality is highlighted in bold.

Classifier	Tabular	Image only			Multimodal		
		ResNet152	DenseNet169	ResNet152V2	ResNet152	DenseNet169	ResNet152V2
Logistic Regression	62.96	70.83	78.70	72.69	68.98	77.31	71.30
Decision Tree	62.04	67.59	57.87	62.96	75.46	66.67	62.04
Random Forest	62.50	79.63	79.17	77.31	77.78	76.85	70.83
SVC	62.50	75.00	78.70	71.76	75.00	79.17	75.93
Gradient Boosting	60.65	72.69	67.59	72.69	76.39	68.52	70.83
Multi-Layer Perceptron	65.74	73.61	79.63	74.54	71.76	81.48	75.00
XGBoost	66.67	77.31	73.15	74.54	80.56	74.07	71.30
LightGBM	65.74	76.85	67.13	71.76	73.61	71.30	68.52
Adaptive Boosting	61.11	74.07	75.93	75.93	75.46	77.31	74.07

In the tabular-only approach, XGBoost came out on top with an accuracy of 66.67%. Its performance was just slightly better than Multi-Layer Perceptron and LightGBM, which both scored 65.74%. This shows that while all three models performed fairly close to each other, XGBoost had a small edge when working only with tabular features.

Within the image-based approach, Random Forest delivered the highest accuracies for both ResNet152 (79.63%) and ResNet152V2 (77.31%), respectively. For DenseNet169, the Multi-Layer Perceptron (MLP) model excelled,

also reaching 79.63% accuracy, outperforming other models such as Decision Tree (57.87%), which exhibited lower performance across all feature extractors. The range of accuracies for DenseNet169 (57.87% to 79.63%) highlights the importance of selecting an appropriate prediction model. Consequently, Random Forest with ResNet152 and ResNet152V2, and MLP with DenseNet169, were chosen for the final training and the aggregation of results in the image-based approach.

The integration of tabular features with images (multimodality) significantly enhanced prediction performance. Notably, the accuracy of the MLP model with DenseNet169 improved from 79.63% in the image-based approach to 81.48% in the multimodal setting, a 1.85% increase. Similarly, XGBoost with ResNet152 saw around 1% improvement, reaching 80.56% accuracy in multimodal approach from 79.63% in the image-based approach. However, the image feature set from ResNet152V2 combined with tabular features underperformed compared to the Support Vector Classifier (SVC) achieving 75.93% in contrast to 77.31% in the image-based approach. Compared to the image-based approach's performance, Random Forest model struggled to properly generalize on multimodal input, potentially due to its sensitivity to the heterogeneous patterns between image-extracted and tabular features.

Experimental results reveal that the multimodal approach consistently outperformed both unimodal variants. While the image-based models demonstrated strong performance individually, the inclusion of tabular features further enhanced predictive accuracy by providing complementary information. This finding underscores the effectiveness of fusing heterogeneous data sources to improve classification performance in complex real-world tasks.

5.4. Egg Grading with Ensemble Approach

Finalizing the result for image-only and multimodal approaches, we applied majority voting to combine the top three prediction results, as detailed in Table 10. For the image-based approach, the ensemble of Random Forest with ResNet152 and ResNet152V2, and MLP with DenseNet169, achieved an accuracy of 85.19%. For the multimodal approach, the ensemble of XGBoost with ResNet152, MLP with DenseNet169, and SVC with ResNet152V2 yielded a superior accuracy of 86.57%, a 5.09% improvement over the best individual multimodal model (MLP with DenseNet169 at 81.48%). These results highlight the effectiveness of the multimodal ensemble approach in leveraging

Table 10: Comparative Egg Grade Performance of Ensemble Approach: Image-based vs. Multimodal

Approach	Data Modality	Feature Extractor	Best Classifier	Accuracy (%)
Ensemble	Image only	ResNet152	Random Forest	85.19
		DenseNet169	Multi-Layer Perceptron	
		ResNet152V2	Random Forest	
	Multimodal (image + weight and shape)	ResNet152	XGBoost	86.57
		DenseNet169	Multi-Layer Perceptron	
		ResNet152V2	SVC	

complementary strengths of image and tabular features to achieve superior performance.

The Figure 5 demonstrates the confusion matrices of the ensemble result of the models described in Table 10 in image-based and multimodal setting. For image input data modality in Figure 5a, the True Negative rate has increased significantly from 72.22% in tabular approach to 97.22%, a 25% increase. Besides, the False Positive rate has decreased by 25% and came down to 2.78%. Overall, the ensemble approach in image-based model has improve the prediction result notable but become slightly biased towards “Low” class.

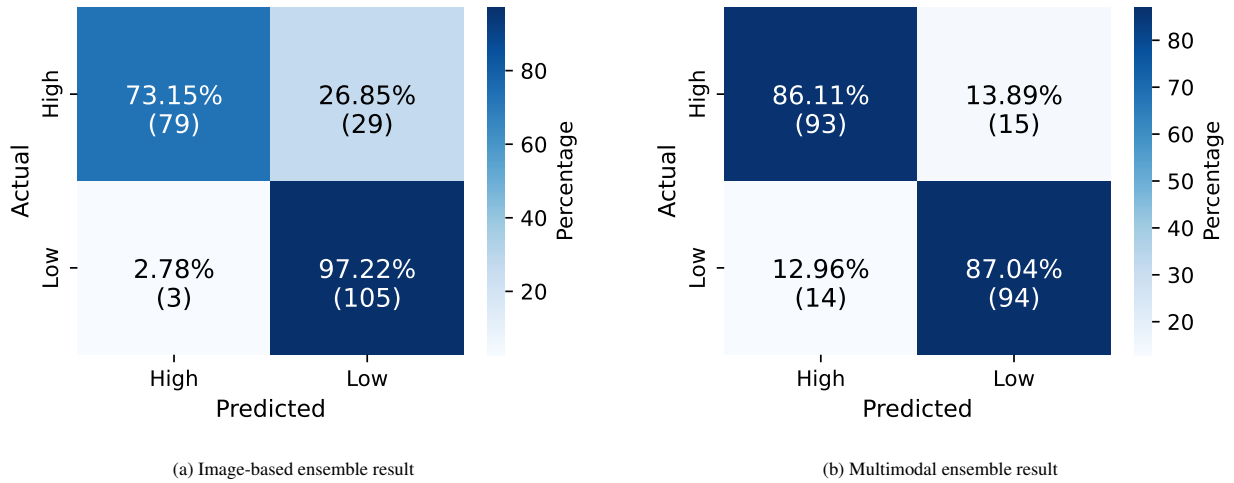


Figure 5: Confusion matrix of egg grade classification ensemble result for different data modality

The confusion matrix for the multimodal input-based ensemble approach, described in Figure 5b, demonstrates a well-balanced performance with a True Positive rate of 86.11% for the “High” class and a True Negative rate of 87.04% for the “Low” class.

Furthermore, the low False Positive rate (12.96%) and False Negative rate (13.89%) reflect minimal misclassification errors, indicating robust generalization across both target classes. The ROC-AUC curve in Figure 6, created using the average predicted probabilities, further confirms the superiority of the multimodal ensemble model compared to other approaches, highlighting its effectiveness in egg grading.

The figure compares the classification performance of two ensemble models (Image-only ensemble and Multimodal ensemble) for egg grading. Both ROC curves lie well above the diagonal reference line, indicating that the models perform significantly better than random guessing. The Image-only ensemble achieves an AUC of 0.88, reflecting strong discriminative ability based on visual features alone. Importantly, the Multimodal ensemble further improves the AUC to 0.91, demonstrating the added value of integrating tabular features (weight and shape) alongside image-derived features. This consistent upward shift in the ROC curve highlights better sensitivity–specificity trade-offs, particularly at lower false-positive rates, thereby providing strong evidence of robust generalization and suggesting that the multimodal approach achieves a more effective balance between sensitivity and specificity.

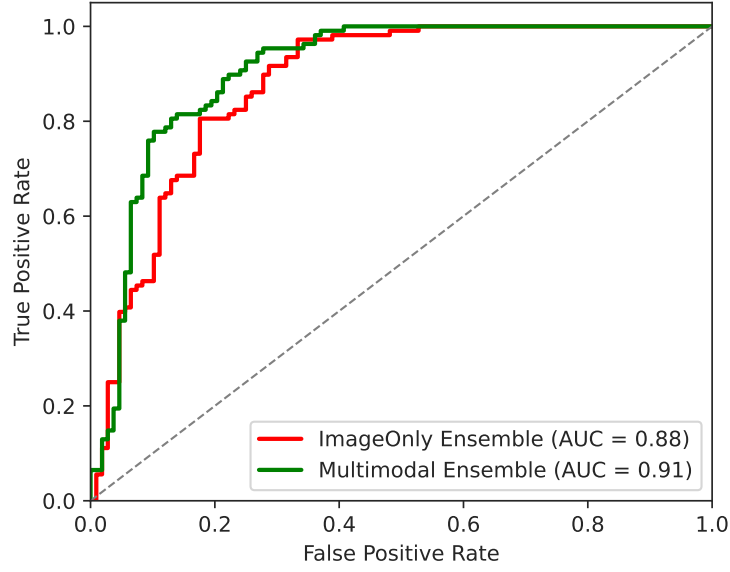


Figure 6: Combined ROC-AUC curve for Egg grade classification

5.5. Egg Freshness Classification - Tabular vs Image vs Multimodal

This subsection presents the results of the egg freshness classification based on the Yolk Index (YI), a critical indicator of internal egg quality, calculated as the ratio of yolk height to diameter. The dataset for this task comprised of 192 samples, with balanced representation across egg freshness categories, as detailed in Section 4.2. Consistent with the Haugh Unit (HU)-based classification task, multiple machine learning models, including Random Forest, Multi-Layer Perceptron (MLP), XGBoost, and others, were evaluated using 10-fold cross-validation. The analysis compares the effectiveness of individual input modalities (image-based and tabular data-based) and multimodal inputs in predicting egg freshness. These findings are significant for quality control in the poultry industry, where accurate freshness assessment ensures consumer safety and product reliability.

To evaluate the impact of different input modalities on predicting egg freshness, we conducted experiments using tabular features, image-based features, and multimodal fusion, with results presented in Table 11. The image-based and multimodal approaches utilized features extracted from ResNet152, DenseNet169, and ResNet152V2, which had demonstrated strong representation capability in earlier experiments (Table 9).

In the tabular-only setting, classification accuracy remained modest, with LightGBM achieving the highest score of 64.06%, followed closely by XGBoost (61.98%) and Logistic Regression/MLP (60.94%). These results highlight the limitations of relying solely on structural traits such as shape and weight, which, while informative, fail to capture the nuanced external characteristics that distinguish freshness levels.

The image-based models provided notable performance improvements. Using features extracted from DenseNet169, Logistic Regression reached the highest overall accuracy of 67.71%. Similarly, Random Forest performed best with ResNet152V2 features (67.71%), and MLP delivered competitive performance across architectures, peaking

Table 11: Classification accuracy (%) comparison across different input modalities (Tabular, Image only, and Multimodal) and classifiers for egg freshness. Image-based and multimodal approaches used features extracted from ResNet152, DenseNet169, and ResNet152V2. The highest accuracy for each modality is highlighted in bold.

Classifier	Tabular	Image only			Multimodal		
		ResNet152	DenseNet169	ResNet152V2	ResNet152	DenseNet169	ResNet152V2
Logistic Regression	60.94	60.94	67.71	66.67	67.71	68.75	59.38
Decision Tree	59.38	55.73	57.29	57.89	59.90	56.25	55.73
Random Forest	56.77	60.94	62.50	67.71	65.62	61.46	57.29
SVC	60.94	61.46	61.98	64.06	66.67	70.31	63.54
Gradient Boosting	60.42	56.77	57.81	58.33	65.62	55.73	55.21
Multi-Layer Perceptron	60.94	66.15	63.54	67.19	65.62	69.27	63.54
XGBoost	61.98	64.58	64.58	66.15	67.19	61.98	60.42
LightGBM	64.06	59.38	61.98	60.94	62.50	58.33	53.65
Adaptive Boosting	57.81	60.94	60.94	63.02	65.10	58.85	57.81

at 67.19% with ResNet152V2. These outcomes emphasize the discriminative power of visual features for freshness detection, though performance varied substantially depending on the choice of classifier and backbone CNN.

The multimodal fusion of image and tabular features consistently improved results compared to either unimodal approach. For instance, SVC with DenseNet169 features achieved 70.31%, the highest accuracy across all modalities, marking a clear improvement over its image-only counterpart (61.98%). Likewise, MLP with DenseNet169 improved from 63.54% (image-only) to 69.27%, while XGBoost with ResNet152 increased from 64.58% to 67.19%. Notably, some classifiers such as Random Forest and Gradient Boosting showed less benefit in the multimodal setting, suggesting potential difficulties in handling heterogeneous feature types. However, the overall accuracies were modest, with MLP and ResNet152V2 in the multimodal approach achieving the lowest accuracy of 63.54%, possibly due to challenges in integrating heterogeneous image and tabular features. Notably, DenseNet169 demonstrated robust performance across both image-only and multimodal modalities, underscoring its effectiveness as a feature extractor for egg freshness prediction.

These results underscore the superior effectiveness of a multimodal approach, likely due to their ability to capture visual representations by combining with physical features relevant to freshness, which has practical implications for automated quality assessment in the poultry industry.

5.5.1. Egg Freshness Assessment with Ensemble Approach

To enhance prediction performance, majority voting was applied to combine the top three prediction results from image-only and multimodal approaches, as shown in Table 12. The image-based ensemble, comprising MLP with ResNet152, Logistic Regression with DenseNet169, and Random Forest with ResNet152V2, achieved an accuracy of 67.71%, showing no improvement over the best individual model. The multimodal ensemble, including XGBoost with

ResNet152, SVC with DenseNet169, and MLP with ResNet152V2, reached a slightly higher accuracy of 70.83%, a marginal 0.52% increase over the best individual multimodal model.

Table 12: Comparative Egg Freshness Performance of Ensemble Approach: Image-Only vs. Multimodal

Approach	Data Modality	Feature Extractor	Best Prediction Model	Accuracy (%)
Ensemble	Image only	ResNet152	Multi-Layer Perceptron	67.71
		DenseNet169	Logistic Regression	
		ResNet152V2	Random Forest	
	Multimodal (image + weight and shape index)	ResNet152	XGBoost	70.83
		DenseNet169	SVC	
		ResNet152V2	Multi-Layer Perceptron	

The confusion matrices in Figure 7 and the ROC curve in Figure 8 reveal distinct performance trends across the three egg freshness classification approaches. The AUC score has been calculated by taking the average of the three prediction probabilities. The image-based ensemble achieved the highest discriminative ability with an AUC of 0.75, outperforming the multimodal ensemble (AUC = 0.73), suggesting that visual features can also capture freshness cues robustly without any extra information. However, the multimodal ensemble demonstrated improved classification for “Old” eggs, correctly identifying 70.83% (68 instances) compared to the image-based model’s 68.75% (66 instances). In both approaches, the number of fresh eggs that were also identified as fresh was the same, which is 70.83%. The multimodal approach exhibited marginally better balance in precision-recall trade-offs. High accuracy of multimodal ensemble indicates that the model’s final prediction is often correct. However, the low AUC reveals that the model’s probability scores are not well-separated among the classes. This means the model frequently assigns high probabili-

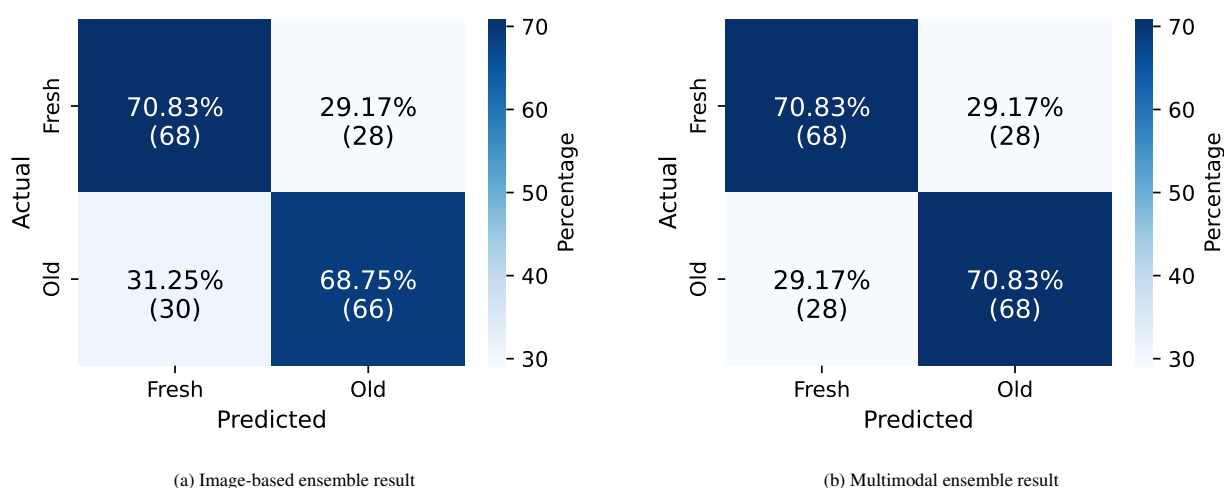


Figure 7: Confusion matrix of egg freshness classification ensemble result for different data modality

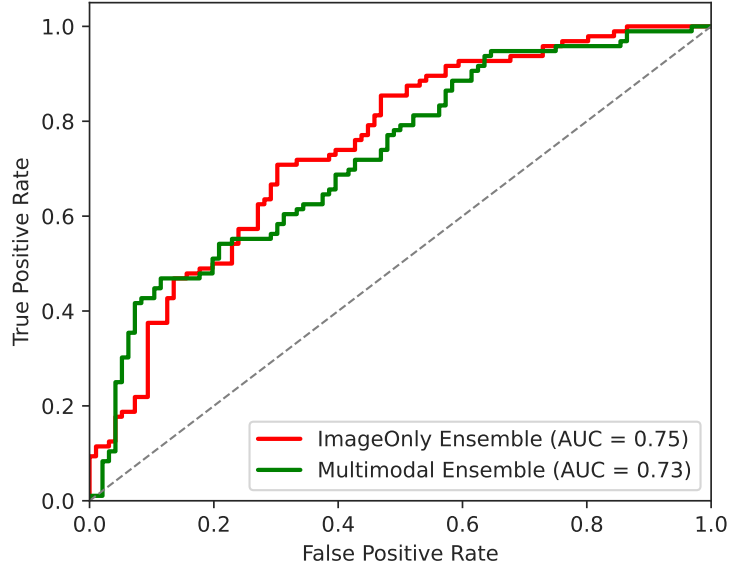


Figure 8: Combined ROC-AUC curve for Egg freshness classification

ties to one class for instances that actually belong to other classes, hurting its ranking ability. Hence, the multimodal ensemble model is effective at making the final “hard” decision but poor at quantifying its certainty.

The limited improvement in ensemble performance suggests potential overlap in the predictive information provided by the models and insufficient diversity in their outputs. These modest results may reflect challenges in capturing subtle freshness variations based on the Yolk Index, highlighting the need for further exploration of alternative feature extractors, additional tabular features, or more sophisticated ensemble techniques to improve classification performance in future work.

6. Discussion

6.1. Findings and Observations

This study presents several novel and significant findings that advance the understanding of egg freshness and grading using computer vision and multimodal learning. Most notably, our study provides the first systematic evidence that the external RGB appearance of an intact egg encodes reliable information about both its grade (interior quality) and freshness (temporal deterioration). We demonstrate that computer-vision pipelines can extract subtle spectral and textural signatures that correlate with Haugh Unit and Yolk Index values, achieving 86.5% and 70.83%. This finding opens an entirely non-invasive avenue for quality assessment that bypasses the need for candling or destructive testing as presented in previous researches [20, 40, 71, 74, 44, 1, 69].

A key observation is that traditional tabular features, weight and shape index, long-standing tabular benchmarks in egg-grading standards [1, 2, 43], are insufficient for reliable classification, as demonstrated by their poor predictive

performance in this study. In contrast, image-based features significantly improve model generalizability, suggesting that visual cues play a dominant role in determining egg quality. This aligns with recent trends in food quality assessment, where deep learning has uncovered non-intuitive visual biomarkers for perishable goods [80, 62, 61, 60],

Interestingly, our results challenge the conventional approach of directly using pre-trained model embeddings for prediction [51, 47]. Directly feeding ImageNet embeddings into a shallow classifier yielded mediocre results, whereas dimensionality reduction through PCA compression improved performance and appears to suppress nuisance variance such as illumination shifts or camera white-balance artifacts, while preserving the low-rank manifold on which egg-quality variations reside. This implies that while deep learning models capture rich representations, strategic feature transformation is critical to eliminate redundancy and amplify task-relevant signals. This observation also aligns with recent findings in few-shot agricultural vision tasks, where compact latent spaces often outperform high-dimensional raw features [45, 85, 83].

Furthermore, not all pre-trained architectures perform equally as feature extractors [23]. Our experiments reveal that ResNet-based models (ResNet152 and ResNet152V2) consistently outperform alternatives (e.g., InceptionV3, VGG19, EfficientNet) in this task. Residual skip connections evidently facilitate the capture of global contextual features and fine-grained object-specific details [79, 58, 21], a dual requirement that depthwise-separable or heavily down-sampled networks struggle to satisfy in this task.

A central objective of this study is to evaluate multimodal learning for egg classification. Our results demonstrate that combining image features with physical measurements (weight and shape index) yields better performance than unimodal inputs, underscoring the complementary nature of visual and tabular data. Thus, multimodal integration functions as a reliability-weighted ensemble, where the image data dominates when visual patterns are salient and the tabular data rescues edge cases that violate chromatic priors [19].

Surprisingly, we observed that egg grade classification (haugh unit-based) exhibits a stronger correlation with external imagery than freshness classification (yolk index-based), suggesting that yolk-index-related degradation is more challenging to identify than haugh-unit-related changes. This discrepancy may arise because grade-related traits manifest more visibly in shell appearance, whereas freshness indicators require deeper sensory analysis [16]. This finding has practical implications for prioritizing computer vision in grading over freshness assessment.

Finally, our analysis of egg source characteristics revealed a significant correlation between procurement channels and quality metrics. Specifically, eggs sourced from supermarkets and grocery stores exhibited consistently lower grades and freshness levels compared to those obtained from wholesale markets. This trend suggests that prolonged storage even under controlled environmental conditions leads to measurable degradation in perishable quality over time [5]. These findings highlight a critical gap in current retail inventory management practices, where extended shelf-life appears to compromise product quality [73]. From a practical standpoint, this discovery underscores the need for supermarkets and grocery retailers to implement more frequent restocking cycles. Such operational adjustments would not only improve the freshness of eggs available to consumers but also align with growing market demands for higher-quality perishable goods. This recommendation carries particular weight, given increasing con-

sumer awareness about food quality and the demonstrated ability of our models to detect subtle quality differences through visual inspection.

6.2. Limitations and Future Directions

Although this work introduces a new path of research for the poultry industry with egg grade and freshness classification with visual and multimodal features, there are some limitations of this work that should be acknowledged. Additionally, a few suggestions on future research directions in this domain could help fellow researchers to contribute meaningfully to similar tasks.

First, the dataset used in this research work is relatively smaller compared to other computer vision-based research works. Although the dataset encompasses egg collection from a wide variety of sources, the amount of data that could be ultimately used was limited. This occurred due to a lack of manpower and resources for storing and handling the eggs. Additionally, some eggs got broken in transportation and during the experiment. Therefore, we had to opt for an augmentation technique to increase the amount of data. A large dataset could help the models to better generalize for the task and improve prediction performance, opening the door for industry-grade applications.

Second, our team collected eggs from different sources without knowing the exact time difference between egg collection and the time the chicken laid those eggs. Therefore, it was found that the dataset contained more eggs of “Low” grade and “Old” freshness class than its counterpart.

Third, from our experimental results, it becomes evident that egg external RGB images carry more relational visual features with grade than with freshness. In this work, we had assumed that the visual feature significance could be of equal importance. Consequently, we only evaluated the usability of pre-trained CNNs for the grading task and used their best-performing feature extractors directly for freshness tasks. An in-depth experiment for the freshness assessment could reveal more interesting facts about egg freshness classification using external RGB images.

Fourth, in this study, we used pre-trained CNN models exclusively for visual feature extraction rather than conducting end-to-end training on RGB images. This design choice was necessitated by the limited size of our dataset, as fine-tuning deep CNN architectures typically requires substantially larger training sets to achieve robust generalization. While transfer learning through pre-trained models provided effective feature representations, we acknowledge that more advanced approaches such as vision transformers (ViTs) [13], few-shot learning [46], or zero-shot learning paradigms [78], could potentially yield superior performance given adequate training data. These emerging techniques may offer enhanced capability in capturing subtle discriminative features for egg quality classification, particularly in scenarios where dataset constraints can be addressed through expanded collection or synthetic data augmentation. Future work should explore these directions to further advance the accuracy and generalizability of visual-based egg quality assessment systems.

Fifth, we implemented feature fusion through the simple concatenation of image-derived and tabular features. While this approach provided a baseline for multimodal integration, we recognize that more sophisticated fusion techniques, such as element-wise multiplication, weighted addition, or learned fusion via convolutional layers, could

potentially enhance the model's predictive performance. These alternative fusion strategies may better capture non-linear interactions between heterogeneous feature types, thereby improving the model's ability to learn complementary relationships between visual and physical egg characteristics. Future investigations should systematically evaluate these fusion methods to optimize information integration while maintaining computational efficiency.

7. Conclusion

This study proposed ELMF4EggQ, an ensemble-based multimodal machine learning framework for non-destructive egg quality assessment, focusing on both egg grade and freshness classification. The framework integrates visual features from external egg images with structural attributes (shape and weight), leveraging their complementary strengths through feature-level fusion and ensemble learning. A novel dataset of 186 brown-shelled eggs was developed for this purpose, with expert-labeled grade and freshness annotations derived from internal quality measurements, including yolk index and Haugh unit. This dataset, along with the full codebase, has been made publicly available to promote transparency, reproducibility, and further research in this domain.

Through extensive experiments, we evaluated and compared unimodal (image-only and tabular-only) and multimodal inputs, with and without ensemble learning. Among unimodal approaches, the image-only model without ensemble achieved 79.63% accuracy for grade and 67.71% for freshness classification. Incorporating both image and tabular features improved grade classification to 81.48%, though freshness classification slightly decreased to 66.15%. The most significant performance gains were observed with ensemble learning. The image-based ensemble achieved 85.19% (grade) and 67.71% (freshness), while the multimodal ensemble further boosted results to 86.57% and 70.83%, respectively. These findings confirm the effectiveness of multimodal fusion and ensemble strategies in enhancing classification robustness and generalization. Importantly, the results also suggest that image-based features contribute more strongly to predictive performance than tabular features, particularly in grade classification. This may be attributed to the visual correlation between Haugh unit-based grading and external egg appearance—an association that is less pronounced in freshness classification.

The proposed framework demonstrates strong potential for real-world deployment in commercial egg processing facilities, retail quality assurance systems, and supply-chain optimization, enabling scalable, data-driven decision-making. Future work will explore larger and more diverse datasets, real-time inference on edge devices, and broader cross-market applicability.

References

- [1] Ab Nasir, A.F., Sabarudin, S.S., Majeed, A.P.A., Ghani, A.S.A., 2018. Automated egg grading system using computer vision: Investigation on weight measure versus shape parameters, in: IOP Conference Series: Materials Science and Engineering, IOP Publishing, p. 012003.
- [2] Abacı, S.H., Tunç, T., Önder, H., Erensoy, K., Sarıca, M., 2023. Examination of the relationships between internal and external egg quality traits: A structural equation model. *Turkish Journal of Agriculture-Food Science and Technology* 11, 2446–2451.

- [3] Abdanan Mehdizadeh, S., Minaei, S., Hancock, N.H., Karimi Torshizi, M.A., 2014. An intelligent system for egg quality classification based on visible-infrared transmittance spectroscopy. *Information Processing in Agriculture* 1, 105–114. doi:[10.1016/j.inpa.2014.10.002](https://doi.org/10.1016/j.inpa.2014.10.002).
- [4] Aragua, A., Mabayo, V.I., 2018. A cost-effective approach for chicken egg weight estimation through computer vision. *International Journal of Agriculture Environment and Food Sciences* 2, 82–87.
- [5] Aschemann-Witzel, J., 2018. Consumer perception and preference for suboptimal food under the emerging practice of expiration date based pricing in supermarkets. *Food quality and preference* 63, 119–128.
- [6] Attard, M.R.G., Bowen, J., Portugal, S.J., 2023. Surface texture heterogeneity in maculated bird eggshells. *Journal of the Royal Society Interface* 20, 20230293. doi:[10.1098/rsif.2023.0293](https://doi.org/10.1098/rsif.2023.0293).
- [7] Atwa, E.M., Xu, S., Rashwan, A.K., Abdelshafy, A.M., ElMasry, G., Al-Rejaie, S., Xu, H., Lin, H., Pan, J., 2024. Advances in emerging non-destructive technologies for detecting raw egg freshness: A comprehensive review. *Foods* 13, 3563.
- [8] Brasil, Y.L., Cruz-Tirado, J., Barbin, D.F., 2022. Fast online estimation of quail eggs freshness using portable nir spectrometer and machine learning. *Food Control* 131, 108418. doi:[10.1016/j.foodcont.2021.108418](https://doi.org/10.1016/j.foodcont.2021.108418).
- [9] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16, 321–357.
- [10] Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.
- [11] Chowdhury, A., Jiang, M., Chaudhuri, S., Jermaine, C., 2021. Few-shot image classification: Just use a library of pre-trained feature extractors and a simple classifier, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9445–9454.
- [12] Dai, D., Jiang, T., Lu, W., Shen, X., Xiu, R., Zhang, J., 2020. Nondestructive detection for egg freshness based on hyperspectral scattering image combined with ensemble learning. *Sensors* 20, 5484.
- [13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [14] Dulal, R., Zheng, L., Kabir, M.A., 2025. Mhaff: Multi-head attention feature fusion of cnn and transformer for cattle identification. *arXiv preprint arXiv:2501.05209*.
- [15] Feddern, V., Prá, M., Mores, R., Nicoloso, R., Coldebella, A., Abreu, P., 2017. Egg quality assessment at different storage conditions, seasons and laying hen strains. *Ciencia e Agrotecnologia* 41, 322–333. doi:[10.1590/1413-70542017413034816](https://doi.org/10.1590/1413-70542017413034816).
- [16] Gao, Z., Zheng, J., Xu, G., 2025. Research progress and technological application prospects of comprehensive evaluation methods for egg freshness. *Foods* 14, 1507.
- [17] Guanjun, B., Mimi, J., Yi, X., Shibo, C., Qinghua, Y., 2019. Cracked egg recognition based on machine vision. *Computers and electronics in agriculture* 158, 159–166.
- [18] Guérin, J., Thiery, S., Nyiri, E., Gibaru, O., Boots, B., 2021. Combining pretrained cnn feature extractors to enhance clustering of complex natural images. *Neurocomputing* 423, 551–571.
- [19] Hager, P., Menten, M.J., Rueckert, D., 2023. Best of both worlds: Multimodal contrastive learning with tabular and imaging data, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23924–23935.
- [20] Harnsoongnoen, S., Jaroensuk, N., 2021. The grades and freshness assessment of eggs based on density detection using machine vision and weighing sensor. *sci rep* 11, 16640 (2021).
- [21] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- [22] Heckler, L., König, R., Bergmann, P., 2023. Exploring the importance of pretrained feature extractors for unsupervised anomaly detection and localization, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2917–2926.
- [23] Holliday, A., Dudek, G., 2020. Pre-trained cnns as visual feature extractors: A broad evaluation., in: *CRV*, pp. 78–84.
- [24] Hsiao, W.T., Lin, H.H., Lai, L.H., 2021. Application of visual radiographic analysis of quality grade of table eggs. *Applied Sciences* 11, 2815. doi:[10.3390/app11062815](https://doi.org/10.3390/app11062815).

- [25] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
- [26] Huang, Q., Qiu, N., Ma, M., Jin, Y., Yang, H., Geng, F., Sun, S., 2012. Estimation of egg freshness using s-ovalbumin as an indicator. *Poultry Science* 91, 739–743. doi:[10.3382/ps.2011-01639](https://doi.org/10.3382/ps.2011-01639).
- [27] Huang, Y., Luo, Y., Cao, Y., Lin, X., Wei, H., Wu, M., Yang, X., Zhao, Z., 2023. Damage detection of unwashed eggs through video and deep learning. *Foods* 12, 2179.
- [28] Javadikia, P., Dehrouyeh, M.H., Naderloo, L., Rabbani, H., Lorestani, A.N., 2011. Measuring the weight of egg with image processing and anfis model, in: Swarm, Evolutionary, and Memetic Computing: Second International Conference, SEMCCO 2011, Visakhapatnam, Andhra Pradesh, India, December 19-21, 2011, Proceedings, Part I 2, Springer. pp. 407–416.
- [29] Jiang, M.L., Wu, P.L., Li, F., 2021. Detecting dark spot eggs based on cnn googlenet model, in: Simulation Tools and Techniques: 12th EAI International Conference, SIMUtools 2020, Guiyang, China, August 28-29, 2020, Proceedings, Part II 12, Springer. pp. 116–126.
- [30] Jiang, Y., Fu, D., Ma, M., 2022. Egg freshness indexes correlations with ovomucin concentration during storage. *Journal of Food Quality* 2022, 1–8. doi:[10.1155/2022/9562886](https://doi.org/10.1155/2022/9562886).
- [31] Jones, D., . Haugh unit: gold standard of egg quality. *Natl. Egg Qual. Sch. Proc* 7, 47–51.
- [32] Kim, T.H., Kim, J.H., Kim, J.Y., Oh, S.E., 2022. Egg freshness prediction model using real-time cold chain storage condition based on transfer learning. *Foods* 11, 3082.
- [33] Liu, Y., Ren, X., Yu, H., Cheng, Y., Guo, Y., Yao, W., Xie, Y., 2020. Non-destructive and online egg freshness assessment from the egg shell based on raman spectroscopy. *Food Control* 118, 107426.
- [34] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976–11986.
- [35] Ma, L., Sun, K., Tu, K., Pan, L., Zhang, W., 2017. Identification of double-yolked duck egg using computer vision. *PloS one* 12, e0190054.
- [36] Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (pca). *Computers & Geosciences* 19, 303–342.
- [37] Mota-Grajales, R., Torres-Peña, J., Camas-Anzueto, J., Pérez-Patricio, M., Coutiño, R.G., López-Estrada, F., Escobar-Gómez, E., Guerra-Crespo, H., 2019. Defect detection in eggshell using a vision system to ensure the incubation in poultry production. *Measurement* 135, 39–46.
- [38] Mutlag, W.K., Ali, S.K., Aydam, Z.M., Taher, B.H., 2020. Feature extraction methods: a review, in: Journal of Physics: Conference Series, IOP Publishing. p. 012028.
- [39] Nasiri, A., Omid, M., Taheri-Garavand, A., 2020. An automatic sorting system for unwashed eggs using deep learning. *Journal of Food Engineering* 283, 110036.
- [40] Nematinia, E., Abdanan Mehdizadeh, S., 2018. Assessment of egg freshness by prediction of haugh unit and albumen ph using an artificial neural network. *Journal of Food Measurement and Characterization* 12, 1449–1459.
- [41] Okinda, C., Sun, Y., Nyalala, I., Korohou, T., Opiyo, S., Wang, J., Shen, M., 2020. Egg volume estimation based on image processing and computer vision. *Journal of Food Engineering* 283, 110041.
- [42] Olakanmi, S., Karunakaran, C., Jayas, D., 2023. Applications of x-ray micro-computed tomography and small-angle x-ray scattering techniques in food systems: A concise review. *Journal of Food Engineering* 342, 111355. doi:<https://doi.org/10.1016/j.jfoodeng.2022.111355>.
- [43] de Oliveira-Boreli, F.P., Pereira, D.F., Gonçalves, J.A., da Silva, V.Z., de Alencar Nääs, I., 2023. Non-destructive assessment of hens' eggs quality using image analysis and machine learning. *Smart Agricultural Technology* 4, 100161.
- [44] Omid, M., Soltani, M., Dehrouyeh, M.H., Mohtasebi, S.S., Ahmadi, H., 2013. An expert egg grading system based on machine vision and artificial intelligence techniques. *Journal of food engineering* 118, 70–77.
- [45] Ouyang, X., Zhuang, J., Gu, J., Ye, S., 2025. Few-shot data augmentation by morphology-constrained latent diffusion for enhanced nematode recognition. *Computers* 14, 198.
- [46] Parnami, A., Lee, M., 2022. Learning from few examples: A summary of approaches to few-shot learning. [arXiv:2203.04291](https://arxiv.org/abs/2203.04291).

- [47] Poola, R.G., Pl, L., et al., 2023. Covid-19 diagnosis: A comprehensive review of pre-trained deep learning models based on feature extraction algorithm. *Results in Engineering* 18, 101020.
- [48] Priyadumkol, J., Kittichaikarn, C., Thainimit, S., 2017. Crack detection on unwashed eggs using image processing. *Journal of food engineering* 209, 76–82.
- [49] Qin, H., Wang, W., Chu, X., Jiang, H., Zhao, X., Jia, B., Yang, Y., Kimuli, D., Dong, A., Wang, B., et al., 2018. Research on the nondestructive detection of egg freshness based on image processing, in: 2018 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers. p. 1.
- [50] Rafed, R., Abedi, M.H., Mushfiq, S., 2024. Nutritional value of eggs in human diet. *J. Res. Appl. Sci. Biotechnol* 3, 172–176.
- [51] Rahman, M., Cao, Y., Sun, X., Li, B., Hao, Y., 2021. Deep pre-trained networks as a feature extractor with xgboost to detect tuberculosis from chest x-ray. *Computers & Electrical Engineering* 93, 107252.
- [52] Reddy, P., Reddy, V., Reddy, C., Rao, P., 1981. Egg weight, shape index and hatchability in khaki campbell duck eggs. .
- [53] Ren, H.L., Zhao, X.Y., Di, K.Q., Li, L.H., Hao, E.Y., Chen, H., Zhou, R.Y., Nie, C.S., Wang, D.H., 2023. Eggshell translucency in late-phase laying hens and its effect on egg quality and physiological indicators. *Frontiers in Veterinary Science* Volume 10 - 2023. doi:[10.3389/fvets.2023.1133752](https://doi.org/10.3389/fvets.2023.1133752).
- [54] Rho, T.G., Cho, B.K., 2024. Non-destructive evaluation of physicochemical properties for egg freshness: A review. *Agriculture* 14, 2049.
- [55] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520.
- [56] Sapkota, S., Kolakshyapati, M.R., Devkota, N.R., Gorkhali, N.A., Bhattarai, N., 2020. Evaluation of external and internal egg quality traits of indigenous sakini chicken in different generations of selection. *International Journal of Agriculture and Forestry* 10, 41–48. doi:[10.5923/j.ijaf.20201002.01](https://doi.org/10.5923/j.ijaf.20201002.01).
- [57] Sehirli, E., Arslan, K., 2022. An application for the classification of egg quality and haugh unit based on characteristic egg features using machine learning models. *Expert Systems with Applications* 205, 117692.
- [58] Shafiq, M., Gu, Z., 2022. Deep residual learning for image recognition: A survey. *Applied sciences* 12, 8972.
- [59] Sharaf Eddin, A., Ibrahim, S.A., Tahergorabi, R., 2019. Egg quality and safety with an overview of edible coating application for egg preservation. *Food Chemistry* 296, 29–39. doi:[10.1016/j.foodchem.2019.05.182](https://doi.org/10.1016/j.foodchem.2019.05.182).
- [60] Shehzad, K., Ali, U., Munir, A., 2025. Computer vision for food quality assessment: Advances and challenges. Available at SSRN 5196776 .
- [61] Shen, C., Jin, Q., Zhou, G., Wang, R., Wang, Z., Liu, D., Cai, K., Xu, B., 2025. Advanced deep learning algorithms in food quality and authenticity. *TrAC Trends in Analytical Chemistry* , 118374.
- [62] Shu, Y., Zhang, J., Wang, Y., Wei, Y., 2025. Fruit freshness classification and detection based on the resnet-101 network and non-local attention mechanism. *Foods* 14, 1987.
- [63] Siswanto, J., Hilman, M., Widiastri, M., 2017. Computer vision system for egg volume prediction using backpropagation neural network, in: *IOP Conference Series: Materials Science and Engineering*, IOP Publishing. p. 012002.
- [64] Soltani, M., Omid, M., Alimardani, R., 2015. Egg quality prediction using dielectric and visual properties based on artificial neural network. *Food analytical methods* 8, 710–717.
- [65] Sun, L., Yuan, L.m., Cai, J.r., Lin, H., Zhao, J.w., 2015. Egg freshness on-line estimation using machine vision and dynamic weighing. *Food Analytical Methods* 8, 922–928.
- [66] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: *Proceedings of the AAAI conference on artificial intelligence*.
- [67] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.
- [68] Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International conference on machine learning*, PMLR. pp. 6105–6114.
- [69] Tang, W., Zhang, H., Chen, H., Fan, W., Wang, Q., 2024. A non-destructive detection and grading method of the internal quality of preserved

eggs based on an improved convnext. *Foods* 13, 925.

- [70] Thipakorn, J., Waranusast, R., Riyamongkol, P., 2017. Egg weight prediction and egg size classification using image processing and machine learning, in: 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE. pp. 477–480.
- [71] Tolentino, L.K.S., Enrico, E.J.G., Listanco, R.L.M., Ramirez, M.A.M., Renon, T.L.U., Samson, M.R.B., 2018. Development of fertile egg detection and incubation system using image processing and automatic candling, in: TENCON 2018-2018 IEEE Region 10 Conference, IEEE. pp. 0701–0706.
- [72] Turkoglu, M., 2021. Defective egg detection based on deep features and bidirectional long-short-term-memory. *Computers and Electronics in Agriculture* 185, 106152.
- [73] Vandevijvere, S., Van Dam, I., Inač, Y., Smets, V., 2023. Unhealthy food availability, prominence and promotion in a representative sample of supermarkets in flanders (belgium): a detailed assessment. *Archives of Public Health* 81, 154.
- [74] Vasileva, A., Gorbunova, E., Vasilev, A., Peretyagin, V., Chertov, A., Korotaev, V., 2018. Assessing exterior egg quality indicators using machine vision. *British poultry science* 59, 636–645.
- [75] Wang, Q., Yang, Z., Liu, C., Sun, R., Yue, S., 2025. Research progress on non-destructive testing technology and equipment for poultry eggshell quality. *Foods* 14, 2223. doi:[10.3390/foods14132223](https://doi.org/10.3390/foods14132223).
- [76] Widiarsi, M., Santoso, L.P., Siswanto, J., 2019. Computer vision system in measurement of the volume and mass of egg using the disc method, in: IOP Conference Series: Materials Science and Engineering, IOP Publishing. p. 012050.
- [77] World Egg Organization, 2025. World Egg Day. <https://www.internationalegg.com/our-work/world-egg-day/>. Accessed on October 2, 2025.
- [78] Xian, Y., Schiele, B., Akata, Z., 2017. Zero-shot learning-the good, the bad and the ugly, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4582–4591.
- [79] Xu, G., Wang, X., Wu, X., Leng, X., Xu, Y., 2024. Development of skip connection in deep neural networks for computer vision and medical image analysis: A survey. *arXiv preprint arXiv:2405.01725*.
- [80] Yang, G., Wang, J., Nie, Z., Yang, H., Yu, S., 2023a. A lightweight yolov8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* 13, 1824.
- [81] Yang, X., Bist, R.B., Subedi, S., Chai, L., 2023b. A computer vision-based automatic system for egg grading and defect detection. *Animals* 13, 2354.
- [82] Yuan, L., Fu, X., Yang, X., Chen, X., Huang, G., Chen, X., Shi, W., Li, L., 2023. Non-destructive measurement of egg's haugh unit by vis-nir with ipls-lasso selection. *Foods* 12, 184. doi:[10.3390/foods12010184](https://doi.org/10.3390/foods12010184).
- [83] Zeng, S., Xia, Y., Gu, S., Liu, F., Zhou, J., 2025. Few-shot classification for soil images: Prototype correction and feature distance enhancement. *Computers and Electronics in Agriculture* 233, 110162.
- [84] Zhang, J., Lu, W., Jian, X., Hu, Q., Dai, D., 2023. Nondestructive detection of egg freshness based on infrared thermal imaging. *Sensors* 23, 5530.
- [85] Zhou, H., Li, W., Li, P., Xu, Y., Zhang, L., Zhou, X., Zhao, Z., Li, E., Lv, C., 2025. A novel few-shot learning framework based on diffusion models for high-accuracy sunflower disease detection and classification. *Plants* 14, 339.