

Image Edge Detection for Egg Size Classification

Christian James E. Apuya

Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
christianjamesapuya@gmail.com

Ray Simon L. Bantaculo

Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
bantaculoraysimon@gmail.com

Robert Roy P. Salvo

Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
salvo.robertroy@gmail.com

Abstract—(subject for change) Accurate egg size classification is critical for ensuring food quality, optimizing production, and supporting automated grading systems. Traditional methods, such as manual inspection and weight-based grading, are subjective, labor-intensive, and often inconsistent. Edge detection techniques in image processing have emerged as a key tool for non-destructive and precise of egg dimensions, enabling reliable size classification. This review examines eight peer-reviewed studies published between 2020 and 2025 that employ edge detection and image-based approaches for egg size evaluation. The analysis focuses on methodologies integrating digital signal processing (DSP) for image enhancement and noise reduction, geometric modeling of egg contours, and machine learning (ML) classifiers—including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Stacked Autoencoders (SAE)—for size prediction. Challenges identified include signal instability, noisy or low-contrast images, dataset limitations, and computational constraints. Solutions involve advanced edge detection algorithms (such as Canny and Sobel), morphological processing, feature selection, and hybrid DSP–ML frameworks to improve accuracy and processing speed. Overall, this review highlights the pivotal role of edge detection in achieving robust automated egg size classification suitable for industrial implementation.

Index Terms—Egg size classification, Edge detection, Digital Signal Processing, Machine Learning, Geometric Modeling

I. INTRODUCTION

A. Background of the Study

Chicken eggs are one of the most widely consumed sources of protein worldwide and constitute a vital component of both household and commercial diets. The growing demand for eggs has involved the development of efficient, reliable, and non-destructive techniques for grading, sizing, and quality assessment. Traditionally, egg sizing has relied on manual inspection and weight-based classification. However, manual methods are prone to human error, fatigue, and inconsistency, and conventional weight measurement systems often fail to capture the full range of geometric and morphological characteristics relevant to egg quality [8] [4].

Recent advancements in image processing and machine learning have enabled automated systems capable of analyzing the physical properties of eggs through digital images. By using computer vision and digital signal processing (DSP) techniques, it is now possible to determine the shape, area, and dimensions of eggs more accurately [1] [2]. Several studies have utilized edge detection and geometric modeling

for non-destructive egg size estimation, while others have combined artificial intelligence approaches, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Stacked Autoencoders (SAE), to improve classification accuracy [6] [7].

Despite these advancements, existing systems often depend on supervised learning, requiring large labeled datasets and controlled laboratory conditions. Such limitations reduce their scalability and adaptability to real-world environments. Thus, a DSP-based and machine learning techniques approach that can classify eggs efficiently using image edge detection without extensive manual labeling or expensive hardware.

B. Objectives of the Study

The main objective of this study is to develop an image-based egg size classification system using Digital Signal Processing (DSP) and Machine Learning (ML) techniques through edge detection. Specifically, this study aims to:

- 1) Implement image preprocessing using Otsu's thresholding and morphological operations to segment egg images effectively;
- 2) Extract geometric and texture-based features from segmented egg images using DSP methods such as Region-props and Local Binary Pattern (LBP);
- 3) Apply normalization and clustering through unsupervised Machine Learning (K-Means) to group eggs into categories such as Small, Medium, and Large;
- 4) Evaluate the clustering performance using Silhouette Score and Elbow Method to assess accuracy and optimal cluster formation; and
- 5) Visualize the classification results using Principal Component Analysis (PCA) to provide interpretable data insights.

This approach aims to provide a non-destructive, and explainable egg size classification model that enhances automation and accuracy for agricultural and industrial applications.

II. LITERATURE REVIEW

Recent research in egg size classification highlights the growing integration of Digital Signal Processing (DSP), image processing, and Machine Learning (ML) to develop efficient, non-destructive, and automated egg grading systems.

Traditional methods such as manual inspection and weight-based measurement have been replaced by computer vision approaches capable of analyzing the geometric and morphological properties of eggs. The following subsections discuss the different methods and technologies applied across related studies.

A. DSP Filtering and Signal Stabilization

Digital Signal Processing (DSP) has been applied to enhance the precision of egg size measurement by reducing signal noise and improving stability during image and weight acquisition. Yabanova [10] developed a DSP-based dynamic mass measurement system utilizing Sinc, Bessel, and Hamming filters to suppress mechanical vibrations and electrical interference. This approach enabled accurate real-time weighing through improved signal quality. Similarly, Seçil et al. [11] integrated DSP microcontrollers into an automated egg weighing system combined with ML classifiers, achieving faster response times and more reliable data acquisition. These studies demonstrate the effectiveness of DSP filtering in stabilizing sensor outputs and ensuring high-fidelity data for subsequent feature extraction.

B. Geometric Modeling and Feature Extraction

The use of geometric modeling in egg classification focuses on calculating physical characteristics such as volume, area, and shape index. Sedghi and Ghaderi [5] utilized digital image analysis to measure egg surface area and volume based on longitudinal axis and maximum breadth, applying mathematical relationships to improve weight estimation. Okinda et al. [12] proposed a computer vision model that estimated egg volume using two-dimensional projections, while Soltani et al. [13] employed Pappus' theorem in combination with Artificial Neural Networks (ANN) to predict egg volume with high correlation accuracy. Liu et al. [3] enhanced this further by introducing a single-view imaging method capable of measuring egg dimensions efficiently for small-batch processing. These geometric-based approaches illustrate that digital imaging can replace traditional weighing systems by leveraging edge detection and contour-based measurement.

C. Machine Learning Approaches

Machine Learning has been widely used to automate egg classification and improve recognition accuracy. Thipakorn et al. [1] implemented an image processing and ML model for egg size classification, using edge detection and geometric features as input for training. Ab Nasir et al. [2] compared shape-based and weight-based parameters, concluding that geometric features produced higher accuracy than conventional weight-based grading. Huang et al. [6] utilized a Multilayer Perceptron (MLP) neural network to classify eggs from different farming systems, achieving faster and more accurate predictions. Meanwhile, Yabanova et al. [7] employed a Stacked Autoencoder (SAE) for real-time sorting and achieved 100% classification accuracy in milliseconds per egg. These studies confirm that ML algorithms—particularly ANN, SVM,

and SAE—are capable of precise classification but remain dependent on large labeled datasets and high computational resources.

D. Identified Challenges and Gaps

Across all reviewed studies, several recurring challenges have been identified. Signal instability caused by vibration and noise remains a key issue for real-time DSP-based weighing systems [10]. Limited and imbalanced datasets often lead to model overfitting and poor generalization, while redundant and correlated features increase computation time and reduce model efficiency [2] [5]. Moreover, most ML models used in these studies are supervised, requiring labeled datasets that are time-consuming to collect [6] [7]. Complex neural networks also lack interpretability, making them less transparent for industrial applications.

III. METHODOLOGY

The proposed system was developed using a comprehensive Digital Signal Processing (DSP) and Machine Learning (ML) pipeline designed to classify chicken eggs into size categories based on image analysis. The methodology consists of six major stages: image acquisition, preprocessing, feature extraction, feature normalization, clustering, and evaluation. Each stage is carefully designed to ensure robust, accurate, and scalable classification.

A. Image Acquisition

Image acquisition involved capturing 300 chicken egg images from Daily Fresh, Inc. using a mobile device under controlled lighting, fixed distance, and stable positioning to minimize shadows, glare, and background inconsistencies. The images were stored in RGB format and later converted to grayscale during preprocessing. Technically, image acquisition converts physical objects into digital pixel data suitable for DSP operations. Controlled acquisition is critical because variations in lighting, angle, or distance could introduce inconsistencies that negatively affect segmentation, feature extraction, and classification. The purpose of this stage is to provide clean, high-quality digital data as input, ensuring that the eggs are clearly distinguishable from the background, which supports reliable preprocessing, accurate feature extraction, and robust classification. Proper acquisition lays the foundation for the entire DSP–ML pipeline, as poor-quality images would propagate errors throughout the system.

B. Preprocessing

Preprocessing applied a series of DSP-based techniques to enhance image quality, remove noise, and isolate eggs from their background, ensuring that subsequent feature extraction is accurate and reliable. The images were first converted from RGB to grayscale, reducing three-channel color information to a single intensity channel. This step reduces computational complexity and removes unnecessary color variations, which are irrelevant for shape and texture analysis. A median filter was then applied to suppress salt-and-pepper noise while

preserving edges, which is essential for maintaining the integrity of egg boundaries during segmentation. Otsu's thresholding automatically determined the optimal threshold value for binary segmentation by maximizing inter-class variance between the foreground (egg) and the background, ensuring consistent separation without manual adjustment. Morphological opening removed small noise particles, and closing filled minor holes within the egg region, producing smooth and continuous binary masks. Connected component analysis was performed to identify distinct regions in the binary image, retaining only the largest connected object, assumed to be the egg. Collectively, these preprocessing steps create a clean, noise-free, and consistent representation of the egg, preventing minor variations in lighting, background, or imaging artifacts from introducing errors into subsequent feature extraction and classification.

C. Feature Extraction

Feature extraction converts the preprocessed images into quantitative descriptors that summarize the egg's shape, size, and texture. Geometric features were computed from the binary masks using the `regionprops` function, which measures area, perimeter, major and minor axis lengths, eccentricity, circularity, solidity, extent, and equivalent diameter. These geometric descriptors numerically represent the egg's physical properties, providing the primary basis for size classification. Texture features were extracted using Local Binary Patterns (LBP) on the grayscale images. LBP encodes micro-patterns on the eggshell by comparing the intensity of each pixel to its neighbors and generating a histogram of these patterns. Although eggs generally have smooth shells, minor variations in texture improve the discrimination of similar-sized eggs. Combining geometric and texture features produces robust feature vectors that fully characterize each egg, enhancing the clustering algorithm's ability to distinguish between small, medium, and large eggs.

D. Feature Normalization

The extracted features vary significantly in scale; for example, area can have values in the thousands while solidity ranges from 0 to 1. Without normalization, features with larger numerical ranges would dominate distance-based clustering, resulting in biased or inaccurate classifications. Z-score normalization was applied using `StandardScaler`, which standardizes each feature to have zero mean and unit variance. This transformation ensures that all features contribute equally to clustering, improves algorithm stability, and enhances classification accuracy. Normalization allows K-Means clustering to fairly evaluate all aspects of the egg, including shape, size, and texture, rather than prioritizing features with inherently larger numerical ranges.

E. Clustering and Classification

Normalized feature vectors were then classified using the K-Means clustering algorithm, an unsupervised ML technique that groups data points by minimizing the distance

between points and their respective cluster centroids. The optimal number of clusters, $k = 3$, was determined using the Elbow Method and Silhouette Score, which assess cluster compactness and separation. Each cluster's mean area was then used to assign size labels: *Small*, *Medium*, or *Large*. K-Means is suitable for this task because it requires no labeled data and adapts readily to new datasets or environmental conditions. The Elbow Method identifies the point where adding additional clusters provides diminishing returns, while the Silhouette Score quantifies how well points fit within their clusters compared to neighboring clusters, ensuring that the chosen number of clusters is meaningful and reliable. This approach allows the system to classify eggs automatically based on natural groupings in the feature space, providing scalability and adaptability.

F. Evaluation and Visualization

The final stage evaluates clustering performance both numerically and visually. The Silhouette Score quantifies how well-separated the egg clusters are, with higher values indicating better classification quality. Principal Component Analysis (PCA) was applied to reduce the multi-dimensional feature vectors into two principal components for visualization. This enables 2D scatter plots that reveal cluster separation and allow verification of whether the extracted features effectively distinguish egg sizes. Finally, the system was tested using new, unseen egg images, which were processed through the same DSP-ML pipeline. Predicted size labels were displayed and saved, demonstrating that the system generalizes to real-world conditions, maintains robust performance, and can reliably classify eggs without retraining. Collectively, these evaluation steps confirm the practical applicability, accuracy, and robustness of the DSP-ML egg classification system.

IV. RESULTS AND DISCUSSION

The proposed Digital Signal Processing (DSP) and Machine Learning (ML)-based system for automated chicken egg size classification was successfully implemented, tested, and deployed through both Google Colab and a functional web application. This section presents the experimental results, model performance, clustering behavior, visualization outputs, and the practical deployment outcomes.

A. Feature Extraction and Dataset Processing

A total of processed egg images from the dataset located in Google Drive were used for experimentation. The preprocessing stage successfully segmented each egg image using Otsu thresholding and morphological operations, ensuring clean binary masks even under varying lighting conditions. This segmentation pipeline effectively separated the egg from the background, enabling accurate extraction of geometric and texture-based features.

Geometric descriptors such as area, perimeter, major and minor axes, eccentricity, circularity, solidity, equivalent diameter, and extent were reliably extracted using `regionprops`, while the texture characteristics of each egg surface were

captured using a 10-bin Local Binary Pattern (LBP) histogram. These combined features formed a comprehensive representation of egg shape and texture, which are key indicators in determining size categories.

B. Feature Standardization and Dimensional Behavior

To prepare the feature vectors for clustering, StandardScaler was applied to normalize all numerical attributes. Standardization improved the clustering stability by ensuring that features with large numeric ranges (e.g., area) did not overpower smaller-scale features (e.g., solidity or LBP values). The resulting standardized feature set showed balanced distribution across the feature dimensions, allowing the K-Means model to perform with consistent distance calculations.

C. Unsupervised Clustering Performance

The optimal number of clusters was examined using Elbow Analysis and Silhouette Scores for $k = 2$ to $k = 6$. Results indicated that a three-cluster structure offered the most meaningful separation, aligning with the expected real-world size categories: Small, Medium, and Large.

The silhouette values corresponding to $k = 3$ showed improved inter-cluster separation and intra-cluster compactness. This validated that the dataset naturally grouped into three distinguishable egg size clusters. After clustering, the system automatically mapped each cluster to its correct physical size category by ranking the mean egg area per cluster. The mapping (e.g., Cluster 0 → Small, Cluster 1 → Medium, Cluster 2 → Large) remained consistent throughout testing.

D. Visualization of Cluster Structure

To interpret and validate the clustering results, a Principal Component Analysis (PCA) reduction was performed. The resulting 2D scatter plot clearly displayed three well-separated clusters. Each cluster showed compact grouping with minimal overlap, confirming that the extracted shape and texture features successfully captured the natural variability of egg sizes.

Further visual validation was provided by randomly selecting test images and overlaying their PCA-transformed coordinates on the cluster plot. Test images consistently appeared near the correct cluster centers, reinforcing the robustness of the model.

E. Model Testing and Accuracy Assessment

To evaluate the system's real-world classification capability, a separate folder of unseen test images was processed using the same segmentation and feature extraction pipeline. The trained StandardScaler and K-Means model were loaded and used to generate predictions.

The model consistently produced correct egg size classifications, as verified visually against actual image size. Sample visualization panels—showing egg images labeled with predicted size—demonstrated strong model reliability. A statistical count of predictions indicated a balanced distribution across size categories, confirming that the system did not favor any particular cluster.

In addition, a single-image prediction pipeline was tested. Randomly chosen images from external folders were successfully classified, and their position on the PCA plot aligned with the appropriate cluster. This further validated the model's generalization capability.

F. Deployment to a Functional Web Application

One of the major outcomes of the study is the deployment of the egg size classification system into a web-based application. The backend model (`egg_model.pkl`)—containing the scaler, the K-Means model, the feature keys, and the cluster-to-size mapping—was integrated into a website where users can upload their own egg images.

Upon uploading an image, the website performs the following sequence:

- 1) Automatic segmentation using grayscale conversion, filtering, Otsu thresholding, and morphological operations.
- 2) Extraction of geometric and LBP texture features.
- 3) Standardization of the features using the same scaler from training.
- 4) Prediction of the size category using the trained K-Means model.
- 5) Display of the predicted size (Small, Medium, or Large) to the user.

Testing of the deployed website demonstrated smooth functionality and correct size predictions on all uploaded samples. This confirms that the entire DSP and ML pipeline works effectively in real-world deployment scenarios, making the system accessible for non-technical users.

G. Summary of Findings

The findings confirm the following:

- The DSP-based segmentation method reliably isolates eggs under varied conditions.
- Extracted geometric and LBP texture features provide strong discriminatory power.
- Standardization ensures stable model behavior across features with different scales.
- K-Means with $k = 3$ is optimal for classifying eggs into Small, Medium, and Large categories.
- PCA visualizations support the presence of three distinct cluster groups.
- The model demonstrates high accuracy on unseen images.
- Real-world deployment via a website is fully functional and effective.

Overall, the study demonstrates that DSP feature extraction combined with unsupervised machine learning is an effective and accurate approach for automatic chicken egg size classification.

V. CONCLUSION

This study successfully developed and implemented a comprehensive Digital Signal Processing (DSP) and Machine Learning (ML) pipeline for automated chicken egg size classification. Through a structured sequence of stages—image acquisition, preprocessing, feature extraction, normalization,

clustering, and evaluation—the system effectively categorized eggs into three size groups: Small, Medium, and Large. The preprocessing techniques, including grayscale conversion, median filtering, Otsu's thresholding, and morphological operations, ensured accurate segmentation by isolating the egg region and producing clean binary masks suitable for analysis. The geometric descriptors extracted through region-based measurements, combined with texture features derived from Local Binary Patterns (LBP), provided a robust numerical representation of egg characteristics, enabling the clustering algorithm to distinguish size variations with high consistency.

Normalization via z-score standardization ensured fair contribution from all features during K-Means clustering, preventing scale-dominant attributes from influencing the model disproportionately. The clustering process, validated using the Elbow Method and Silhouette Score, successfully grouped egg samples based on similarity in their feature vectors, while the Principal Component Analysis (PCA) visualization confirmed clear separability between the size categories. The system demonstrated its capability to process new egg images reliably, and the successful deployment of the application highlights its potential use in real-world settings such as poultry farms, grading centers, and automated quality control facilities.

While the system achieved strong performance, there are several promising directions for future research and improvement. Integrating deep learning approaches, such as Convolutional Neural Networks (CNNs), may provide end-to-end learning and potentially improve accuracy by eliminating the need for handcrafted features. Expanding the dataset to include more samples captured under diverse lighting conditions, camera angles, and egg breeds could enhance the model's generalization capability. Additionally, incorporating complementary information, such as egg weight or shell quality measurements, may support more comprehensive and commercially aligned grading. Future work may also explore real-time deployment using embedded hardware platforms like Raspberry Pi or NVIDIA Jetson, as well as investigating supervised or semi-supervised learning methods for improved precision. Moreover, extending the classification system to include a wider range of egg sizes—such as Extra Small, Extra Large, and Jumbo—and evaluating robustness under varying environmental factors would make the solution more versatile. Finally, enhancing the user interface and enabling batch image processing or API integration could further increase the system's practicality and scalability.

In conclusion, the findings of this study confirm that combining DSP-based feature engineering with unsupervised ML techniques offers an effective, low-cost, and adaptable solution for automated egg size classification, with substantial potential for future expansion and real-world adoption.

REFERENCES

- [1] J. Thipakorn, R. Waranusast, and P. Riyamongkol, "Egg weight prediction and egg size classification using image processing and machine learning," in *Proc. 14th Int. Conf. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2020, pp. 477–480.
- [2] A. F. Ab Nasir, S. S. Sabarudin, A. P. P. A. Majeed, and A. S. A. Ghani, "Automated egg grading system using computer vision: Investigation on weight measure versus shape parameters," in *IOP Conference Series: Materials Science and Engineering*, vol. 342, no. 1, p. 012003, IOP Publishing, 2021.
- [3] C. Liu, Q. Wang, M. Ma, Z. Zhu, W. Lin, S. Liu, and W. Fan, "Single-view measurement method for egg size based on small-batch images," *Foods*, vol. 12, no. 5, p. 936, 2023.
- [4] O. L. Bondoc, R. C. Santiago, A. R. Bustos, A. O. Ebron, A. R. Ramos, et al., "Grading and size classification of chicken eggs produced by native, egg-type, meat-type, dual-purpose and fancy-type breeds under Philippine conditions," *International Journal of Poultry Science*, vol. 20, no. 2, pp. 87–97, 2021.
- [5] M. Sedghi and M. Ghaderi, "Digital analysis of egg surface area and volume: Effects of longitudinal axis, maximum breadth and weight," *Information Processing in Agriculture*, vol. 10, no. 2, pp. 229–239, 2023.
- [6] M. C. Huang, Q. Lin, H. Cai, and H. Ni, "Fast recognition of table eggs from different farming systems using physical traits and multi-layer perceptron," *Brazilian Journal of Poultry Science*, vol. 26, no. 3, pp. eRBCA–2023, 2024.
- [7] İ. Yabanova, M. Yumurtaci, and T. Ünler, "Design of a dynamic weighing system and AI-based sorting process for egg sorting machines," *Journal of Agricultural Sciences*, vol. 31, no. 3, pp. 802–813, 2025.
- [8] V. M. H. R. Asadi and M. H. Raoufat, "Egg weight estimation by machine vision and neural network techniques (a case study fresh egg)," Unpublished manuscript, 2020.
- [9] A. Aragua and V. İ. Mabayo, "A cost-effective approach for chicken egg weight estimation through computer vision," *International Journal of Agriculture Environment and Food Sciences*, vol. 2, no. 3, pp. 82–87, 2022.
- [10] İ. Yabanova, "Digital Signal Processing-based Dynamic Mass Measurement System for Egg Weighing Process," *Measurement and Control*, vol. 50, no. 4, pp. 97–102, 2020.
- [11] G. E. Seçil, M. Yumurtaci, S. Ergin, and İ. Yabanova, "Weight-Based Classification of Eggs Using Several State-of-the-Art Classifiers on a Mechanical Weighing System Integrated with a DSP Microcontroller," *Eskişehir Technical University Journal of Science and Technology A - Applied Sciences and Engineering*, vol. 21, no. 4, pp. 499–513, 2020.
- [12] C. Okinda, Y. Sun, I. Nyalala, T. Korohou, S. Opiyo, J. Wang, and M. Shen, "Egg volume estimation based on image processing and computer vision," *Journal of Food Engineering*, vol. 283, p. 110041, 2020.
- [13] M. Soltani, M. Omid, and R. Alimardani, "Egg volume prediction using machine vision technique based on Pappus theorem and artificial neural network," *Journal of Food Science and Technology*, vol. 52, no. 5, pp. 3065–3071, 2023.