Image-Based Egg Size Classification Using DSP Feature Extraction and Unsupervised Machine Learning

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*—This study presents a complete Digital Signal Processing (DSP) and Machine Learning (ML) based process for automated chicken egg size classification using image analysis. The system addresses the limitations of traditional manual and weight-based grading by integrating image preprocessing, geometric and texture feature extraction, unsupervised learning, and dimensionality reduction. A dataset of egg images was processed through grayscale conversion, median filtering, Otsu's thresholding, and morphological operations to produce clean binary masks suitable for quantitative analysis. Geometric descriptors such as area, perimeter, axis lengths, eccentricity, solidity, extent, and equivalent diameter, combined with Local Binary Pattern (LBP) texture features, formed a comprehensive numerical representation for each egg. Z-score normalization ensured balanced feature contribution during clustering.**

K-Means clustering was evaluated using the Elbow Method, Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index, and it successfully grouped samples into Small, Medium, and Large categories. Principal Component Analysis (PCA) confirmed clear cluster separability. Testing on unseen images demonstrated strong generalization performance, and the model was deployed as a functional web application for real-time egg size prediction. Although the system performed reliably, its limitations include a moderate dataset size and reliance on handcrafted features. Future research may explore deep learning-based feature extraction, integration of complementary data such as weight or shell quality, and deployment on embedded hardware platforms to improve robustness and scalability.

***Index Terms*—Egg size classification, Edge detection, Digital Signal Processing, Machine Learning, Geometric Modeling, Feature Extraction, K-Means Clustering**

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 Regionprops 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 demonstrates a strong trend toward integrating Digital Signal Processing (DSP), image processing, and Machine Learning (ML) techniques to develop automated, non-destructive, and efficient grading systems for eggs. Traditional methods, such as manual inspection and weight-based classification, are now being replaced by advanced computer vision and processing techniques that analyze the geometric, morphological, and textural properties of eggs. This shift aims to achieve more precise, scalable, and automated grading systems that are less prone to human error.

A. DSP Filtering and Signal Stabilization

Digital Signal Processing (DSP) plays a crucial role in improving the precision of egg size measurement by reducing noise and enhancing stability during the acquisition of images and weight data. Effective DSP is particularly important in real-time industrial applications, where sensor noise and mechanical vibrations can significantly affect measurement accuracy. Yabanova et al. [7] developed a DSP-based dynamic mass measurement system using Sinc, Bessel, and Hamming filters to suppress high-frequency noise, minimize phase distortion, and smooth the frequency spectrum [10]. Similarly, Seçil et al. [11] implemented a DSP-integrated weighing system combined with machine learning classifiers to stabilize signals and improve classification speed and accuracy [11]. These studies highlight the necessity of robust signal processing to ensure reliable measurements under real-world conditions, and they provide a foundation for integrating DSP with modern ML algorithms for automated egg grading.

B. Geometric Modeling and Feature Extraction

Geometric modeling involves calculating physical characteristics such as volume, area, perimeter, and shape index to categorize eggs. Sedghi and Ghaderi [5] measured egg surface area and volume using digital image analysis of the longitudinal axis and maximum breadth, demonstrating strong correlations with weight [5]. Okinda et al. [12] estimated egg volume from two-dimensional projections of egg contours, showing that even simpler 2D imaging methods can provide accurate volume approximations [12]. Soltani et al. [13] applied Pappus' Theorem in combination with Artificial Neural Networks (ANNs) to predict egg volume, achieving high accuracy and demonstrating the benefit of integrating classical geometric modeling with ML techniques [13]. More recently, Liu et al. [3] introduced a single-view imaging method using SegFormer-based deep learning segmentation and single-view metrology to extract major and minor axes, achieving R^2 values of 0.969 and 0.926 respectively [3]. These approaches illustrate a progression from basic geometric feature extraction to advanced, ML-assisted modeling that allows accurate, non-destructive, and scalable egg classification.

C. Machine Learning Approaches

ML algorithms are increasingly applied to automate egg classification and improve recognition accuracy. Thipakorn et al. [1] employed edge detection and 13 geometric features as inputs for an SVM classifier, achieving a classification accuracy of 87.58% and strong weight prediction correlations (0.9915) [1]. Ab Nasir et al. [2] compared shape-based versus weight-based parameters, finding geometric features such as area, perimeter, and axis length yielded higher classification accuracy [2]. Huang et al. [6] implemented a Multilayer Perceptron (MLP) to classify eggs from different farming systems, demonstrating faster and more accurate predictions than traditional methods [6]. Yabanova et al. [7] utilized a Stacked Autoencoder (SAE) for real-time sorting, achieving 100% classification accuracy per egg in milliseconds [7]. These studies highlight the evolution of ML applications from classical SVM-based models to deep learning approaches capable of real-time industrial performance, while also emphasizing the challenges of data availability, computational requirements, and interpretability.

D. Identified Challenges and Gaps

Despite these advancements, several challenges remain. Signal instability from mechanical vibrations and electrical interference continues to affect real-time processing [10] [11]. Limited and imbalanced datasets restrict model generalization, particularly when eggs vary in size, type, or shell color [2] [5]. Redundant or highly correlated features can increase computational complexity and risk overfitting [2]. Most ML models rely on supervised learning and require large labeled datasets, which are time-consuming and costly to collect, especially in industrial environments [6] [7]. Additionally, advanced neural networks often act as “black boxes,” lacking interpretability for industrial validation and quality control. Addressing these gaps is critical for developing robust, scalable, and transparent egg classification systems suitable for diverse real-world conditions.

TABLE I: Summary of Reviewed Studies and Methods Used

Author, Year	Methods						
	DSP Filtering	**Geometric Modeling**	**SVM**	**k-NN**	**ANN**	**SAE**	**Feature Selection**
Thipakorn et al., 2020		✓	✓				
Ab Nasir et al., 2021		✓	✓	✓		✓	
Liu et al., 2023		✓					
Bondoc et al., 2021							✓
Sedghi & Ghaderi, 2023		✓					
Huang et al., 2024					✓		
Yabanova et al., 2025	✓					✓	
Asadi & Raoufat, 2020		✓			✓		✓
Aragua & Mabayo, 2022					✓		
Yabanova, 2020	✓						
Seçil et al., 2020	✓		✓	✓			
Okinda et al., 2020		✓					
Soltani et al., 2023		✓			✓		

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 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. Additionally, images with an alpha channel (transparency) were converted into standard RGB format to ensure compatibility with the processing pipeline. 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 correct classification. Proper acquisition lays the foundation for the entire DSP–ML process, 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 loaded and converted from RGB to grayscale, reducing the 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 blur was then applied to suppress salt-and-pepper noise while preserving edges, which is essential for maintaining the integrity of egg boundaries during segmentation. Following this, Otsu's thresholding method was applied to automatically determine the optimal threshold value for binary segmentation. The goal of this step was to separate the egg (foreground) from the background by maximizing the interclass variance between the two, ensuring consistent separation without manual adjustment.

Morphological operations were applied next to clean up the binary mask. A 9x9 elliptical kernel was used for both morphological closing and opening. Closing filled minor holes within the egg region, and opening removed small noise particles. These morphological operations ensured a smooth and continuus binary mask for the egg region.

Connected component analysis was performed on the mask to identify distinct regions in the binary image, retaining only the largest connected component, assumed to be the egg. This process effectively isolated the egg from other potential noise or background objects. 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. The geometric features include values such as area, perimeter, and eccentricity, while LBP histograms capture texture variations that help improve classification accuracy.

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.

Sufficient statistical power for this clustering task was ensured by employing a sample size of 30 samples per cluster, which is adequate for reliably detecting subgroups when the separation between clusters is large ($\Delta \geq 4$). This sample size is consistent with findings from recent cluster analysis studies, which recommend 20–30 observations per subgroup to achieve sufficient statistical power for detecting clustering, provided that the effect size (separation between clusters) is large enough [14].

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. Additionally, the Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI) were used to further evaluate clustering performance. The DBI measures the average similarity between clusters, with lower values indicating better separation, while the CHI score evaluates the ratio of between-cluster dispersion to within-cluster dispersion, where higher values indicate better-defined clusters. Both indices suggested that $k = 3$ provides the optimal clustering solution.

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. Special handling was implemented for images with alpha channels (PNG images), which were converted into standard RGB format to ensure consistency across the dataset.

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 clustering stability by ensuring that features with larger numerical ranges, such as area, did not overpower smaller-scale features, such as solidity or LBP values. This transformation allowed K-Means clustering to perform with consistent distance calculations across all features, enhancing model accuracy and ensuring that all features contribute equally to the clustering process.

C. Unsupervised Clustering Performance

The optimal number of clusters was examined using Elbow Analysis and Silhouette Scores for $k = 2$ to $k = 6$. The Elbow Method 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 Score 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 $\rightarrow$ Small, Cluster 1 $\rightarrow$ Medium, Cluster 2 $\rightarrow$ Large) remained consistent throughout testing.

D. Elbow and Silhouette Analysis

The Elbow Method and Silhouette Score were used to determine the optimal number of clusters.

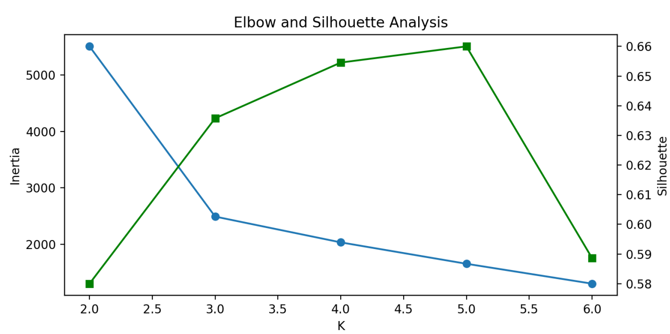

Fig. 1: Elbow and Silhouette Analysis for determining the optimal number of clusters. Inertia (blue) and Silhouette Score (green) are plotted against the number of clusters k. The elbow point and peak silhouette score at $k = 3$ suggest that three clusters provide the best fit for the dataset.

Figure 1 displays the Elbow (inertia) and Silhouette Score plotted against different values of k. As seen in the figure, the Inertia (blue curve) decreases as the number of clusters increases, with a significant inflection point at $k = 3$. This suggests that adding more clusters beyond 3 does not provide substantial improvements. The Silhouette Score (green curve) peaks at $k = 3$, indicating that the clusters are best defined with three groups. This analysis supports the choice of $k = 3$ as the optimal number of clusters for classifying eggs into size categories.

The Silhouette Score for $k = 3$ was calculated to be `0.6358`, indicating a good separation between the clusters. The confusion matrix further validated the clustering results. The matrix, shown below, compares the true sizes (Small, Medium, Large) with the predicted sizes from the K-Means model:

Confusion Matrix (True Size vs Predicted Size:)

$$\begin{bmatrix} 252 & 0 & 0 \\ 0 & 148 & 0 \\ 0 & 0 & 54 \end{bmatrix}$$

The confusion matrix shows perfect classification with no misclassifications for Small, Medium, or Large eggs, confirming the reliability of the model.

To further validate the choice of $k = 3$, the Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI) were calculated for different values of k ranging from 2 to 6. The DBI and CHI scores are important clustering evaluation metrics: DBI should be minimized, while CHI should be maximized.

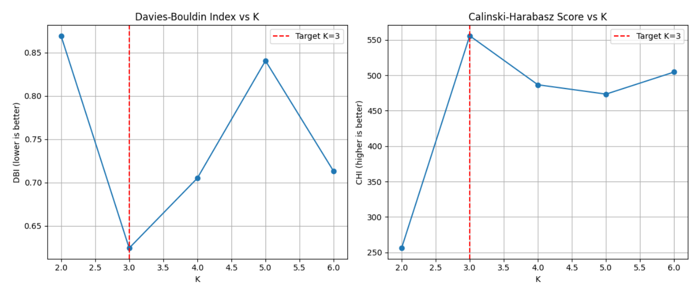

Fig. 2: Davies-Bouldin Index (left) and Calinski-Harabasz Score (right) vs k. Both metrics suggest that $k = 3$ provides the optimal clustering solution. The red dashed line highlights the optimal value of $k = 3$ for both indices.

As shown in Figure 2, both metrics strongly support $k = 3$ as the optimal number of clusters. The DBI (left plot) reaches its minimum at $k = 3$, indicating the best separation between clusters, while the CHI score (right plot) peaks at $k = 3$, confirming the maximization of cluster dispersion.

To further interpret the clustering results, the following size label mapping was used to assign each cluster to an egg size category: Cluster 0 corresponds to Small, Cluster 1 corresponds to Medium, and Cluster 2 corresponds to Large. The system's classification performance was evaluated by comparing predicted sizes to expected sizes for specific test weights. For instance, a test egg weighing 55g was predicted to be Medium, which matched the expected size. Similarly, a 60g egg was predicted as Medium, and the prediction aligned with the expected size.

Additionally, the DBI value for the optimal number of clusters $k = 3$ was found to be 0.6243, indicating good cluster separation (lower is better). The CHI score for $k = 3$ was calculated as 555.7526, which is high and further confirms that the clusters are well-defined with minimal overlap (higher is better). These metrics, combined with the visual evidence from the Elbow and Silhouette analysis, confirm that $k = 3$ is the optimal choice for classifying eggs into Small, Medium, and Large categories.

E. 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. The clusters were color-coded for easy visual distinction, with each cluster representing one of the size categories (Small, Medium, Large).

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.

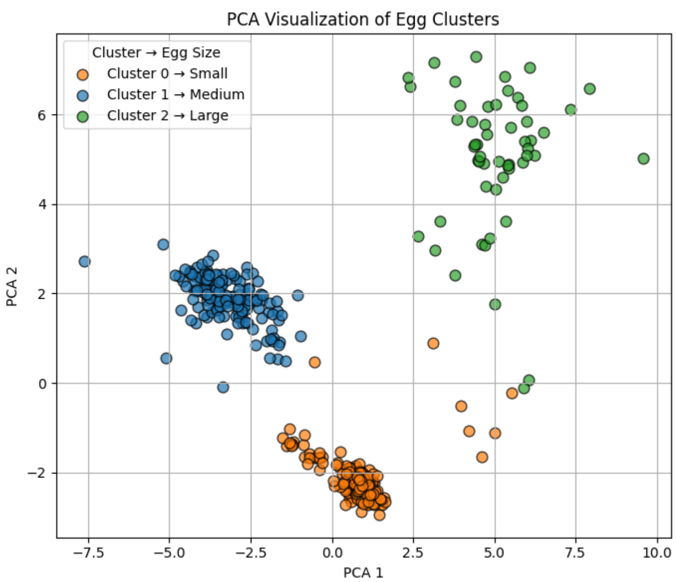

Fig. 3: PCA Visualization of Egg Clusters. The plot shows three well-separated clusters, corresponding to the size categories of Small (orange), Medium (blue), and Large (green).

The PCA visualization also confirms the optimal clustering solution at $k = 3$. As shown in Figure 3, the three clusters (small, medium, and large) are well-separated in the feature space. This separation aligns with the results from the Elbow Method, Silhouette Score, Davies-Bouldin Index (DBI), and Calinski-Harabasz Index (CHI), all of which suggested that $k = 3$ is the optimal choice for clustering. The PCA axes (PC1 and PC2) represent the most significant features in the dataset, with the horizontal axis (PC1) primarily separating small and large eggs, and the vertical axis (PC2) distinguishing the medium-sized eggs.

F. Model Testing and Accuracy Assessment

To evaluate the system's 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. A statistical count of predictions indicated a balanced distribution across size categories, confirming that the system did not favor any particular cluster. Additionally, a confusion matrix was used to compare predicted size categories with actual sizes, providing insight into the model's classification accuracy.

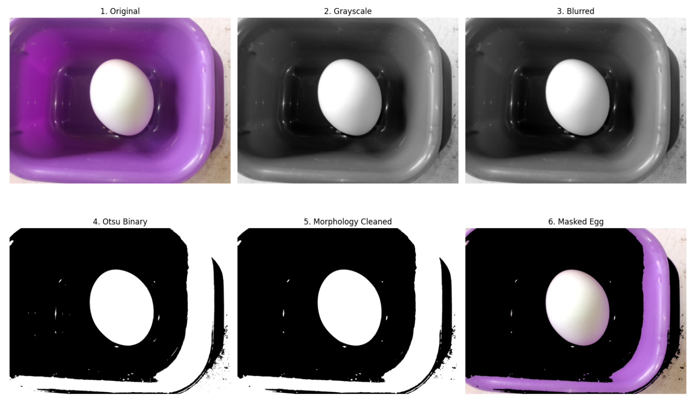

Fig. 4: Preprocessing pipeline for egg image segmentation. The steps include: 1) Original, 2) Grayscale, 3) Blurred, 4) Otsu Binary, 5) Morphology Cleaned, 6) Masked Egg.

Before performing classification on the test images, each image underwent several preprocessing steps, which are illustrated in Figure 4. Initially, the raw images were captured and converted to grayscale, simplifying the data by removing the color dimension. A median blur was applied to reduce noise while preserving important structural details, such as the boundaries of the egg. The Otsu thresholding method was then applied to create a binary mask that separates the egg from the background. After thresholding, morphological operations were applied to clean the binary mask by removing small noise particles and filling minor holes within the egg region. Finally, the isolated egg was obtained by masking out the background, resulting in a clean image where only the egg is visible.

These preprocessing steps were applied consistently to both the training and test images. The images were then processed through feature extraction, utilizing both geometric features (such as area, perimeter, etc.) and texture features (extracted using Local Binary Patterns). After feature extraction, the data was normalized using the StandardScaler and passed into the K-Means clustering model to predict the egg size. This pipeline ensures that the system operates reliably and consistently during both training and testing phases.

Figure 5 shows the PCA visualization of the clusters with the test image. The red star indicates the test image's position, which was classified as Small and placed near the Small cluster.

Figure 6 shows the predicted size of a test image, which was classified as `Small`. The egg is visually consistent with

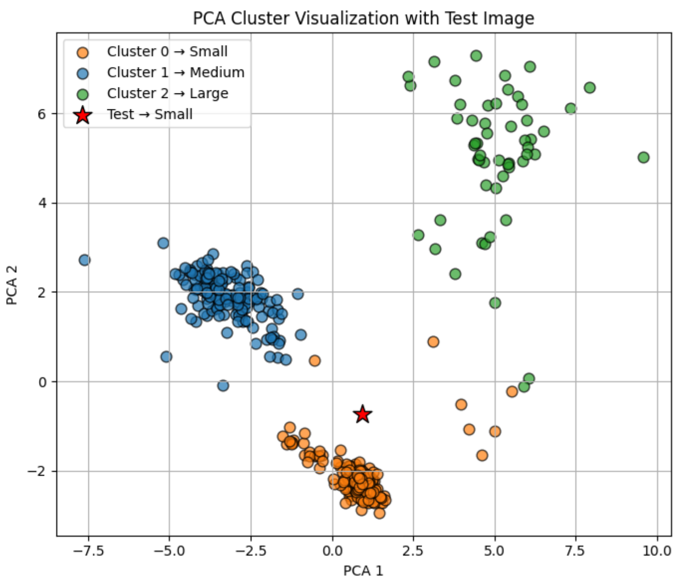

Fig. 5: PCA Cluster Visualization with Test Image. The red star indicates the test image, which was classified as Small, and positioned near the Small cluster.

Fig. 6: Test image with predicted size label.

its predicted size category.

The following features were extracted from the test image of the 6: area (16360.0000), perimeter (655.4396), major axis length (220.4199), minor axis length (119.1972), eccentricity (0.8412), equivalent diameter (144.3267), circularity (0.4786), solidity (0.9124), extent (0.4760), and 10 LBP texture features. These features were used by the model to classify the egg size, which is Small.

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.

G. 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 easily upload their own egg images. Upon uploading an image, the website automatically segments the egg, extracts features, and classifies it into size categories: Small, Medium, or Large.

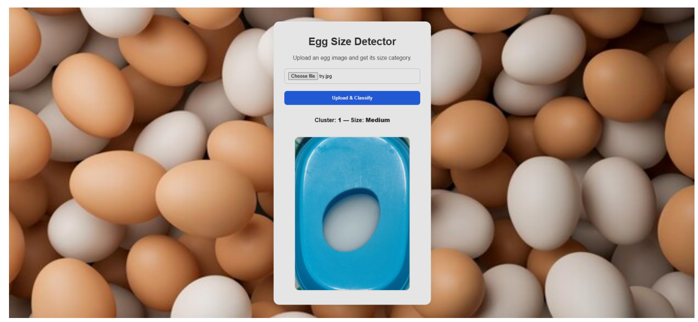

Fig. 7: Test image with predicted size label.

Figure 7 shows the website application which predicted the size category and is then displayed 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 processes works effectively in classifying egg sizes.

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, and the system's deployment as a web application highlights its practical usability.

V. Conclusion

This study successfully developed and implemented a complete Digital Signal Processing (DSP) and Machine Learning (ML) pipeline for automated chicken egg size classification using image-based analysis. By integrating image acquisition, preprocessing, feature extraction, feature normalization, unsupervised clustering, and performance evaluation, the system accurately categorized eggs into three major size groups: Small, Medium, and Large. The preprocessing operations—such as grayscale conversion, median filtering, Otsu's thresholding, and morphological cleaning—effectively isolated the egg region and produced consistent binary masks suitable for further analysis. Through the extraction of geometric descriptors (area, perimeter, major and minor axis lengths, eccentricity, solidity, extent, and equivalent diameter) and texture characteristics using Local Binary Patterns (LBP), the system generated a comprehensive feature set that reliably captured the shape and surface characteristics of each egg.

Normalization using z-score standardization ensured equal contribution among features, preventing scale-dominant attributes from biasing the clustering results. The K-Means algorithm, validated through the Elbow Method, Silhouette Score, Davies-Bouldin Index (DBI), and Calinski-Harabasz Index (CHI), demonstrated strong clustering performance with clear separation among the three egg size categories. Principal Component Analysis (PCA) visualization further confirmed the natural groupings within the extracted features. Testing on unseen images verified the model's generalization capability, and the deployment of the classification system as a functional web application illustrated its practicality and potential for real-world integration in poultry farms, grading facilities, and automated quality control systems.

Despite its demonstrated effectiveness, the system has certain limitations. The dataset consisted of approximately 300 images captured under controlled conditions, which may limit the robustness of the model when applied to more diverse environments. The reliance on handcrafted geometric and texture features, while effective, may not fully capture complex visual variations among eggs. Additionally, the use of a standard mobile device for image capture introduces variability in lighting, angle, and distance that could affect consistency. The absence of weight measurements, shell quality indicators, or multi-angle imaging also restricts the model's ability to perform comprehensive grading beyond size classification.

Future researchers may address these limitations by expanding the dataset to include a wider variety of egg types, lighting conditions, and imaging scenarios. Incorporating deep learning-based methods, such as Convolutional Neural Networks (CNNs), could enable automatic feature learning and potentially improve accuracy under more complex conditions. Combining image-based features with complementary data, such as weight, shell thickness, or defect detection may lead to a holistic egg grading framework. Future studies may also extend the classification range to include additional commercial categories, such as Extra Small, Extra Large, and Jumbo, and evaluate system performance under varying environmental factors. Improving the web application, adding batch processing, and supporting API-based integration would further enhance usability and scalability.

In conclusion, the study demonstrates that combining DSP-based feature engineering with unsupervised ML techniques provides an accurate, cost-effective, and adaptable solution for automated egg size classification. With further refinements and expanded scope, the system holds significant potential for industrial adoption and advancement in precision agriculture.

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. Yumurtacı, 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. Yumurtacı, 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.

[14] Dalmaijer, E.S., Nord, C.L. & Astle, D.E. Statistical power for cluster analysis. *BMC Bioinformatics 23, 205 (2022). https://doi.org/10.1186/s12859-022-04675-1*