11 (1) 2025 41-53 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
Register: *Jurnal Ilmiah Teknologi Sistem Informasi*

Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

# Robust Classification of Beef and Pork Images Using EfficientNet B0 Feature Extraction and Ensemble Learning with Visual Interpretation

*Ahmad Taufiq Akbar* [a,*], *Shoffan Saifullah* [b], *Hari Prapcoyo* [c], *Bambang Yuwono* [d], *Heru Cahya Rustamaji* [e]

[a,b,c,d,e] *Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. Ring Road Utara No. 104 Ngropoh, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55283, Indonesia*
[b] *Faculty of Computer Science, AGH University of Krakow, al. Adama Mickiewicza 30, 30-059 Krakow, Poland*
*email:* [a,*]*ahmadtaufiq.akbar@upnyk.ac.id*, [b]*shoffans@upnyk.ac.id*, [c]*hari.prapcoyo@upnyk.ac.id*, [d]*herucr@upnyk.ac.id*, [e]*bambangy@upnyk.ac.id*
* *Correspondence*

**ARTICLE INFO**

***Article history:***
Received August 30th, 2023
Revised May 15th, 2024
Accepted June 27th, 2025
Available online June 30th, 2025

***Keywords:***
EfficientNet B0
Ensemble Learning
Image Classification
Deep Feature Extraction
Explainable AI

***Please cite this article in IEEE style as:***
A. T. Akbar, S. Saifullah, H. Prapcoyo, B. Yuwono, and H. C. Rustamaji, "Robust Classification of Beef and Pork Images Using EfficientNet B0 Feature Extraction and Ensemble Learning with Visual Interpretation," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 11, no. 1, pp. 41-53, 2025.

**ABSTRACT**

Distinguishing between beef and pork based on image appearance is a critical task in food authentication, but it remains challenging due to visual similarities in color and texture, especially under varying lighting and capture conditions. To address these challenges, we propose a robust classification framework that utilizes EfficientNet B0 as a deep feature extractor, combined with an ensemble of Regularized Linear Discriminant Analysis (RLDA), Support Vector Machine (SVM), and Random Forest (RF) classifiers via soft voting to enhance generalization performance. To improve interpretability, we incorporate Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize classification decisions and validate that the model focuses on meaningful regions of the meat, such as red-channel intensity and muscle structure. The proposed method was evaluated on a public dataset containing 400 images evenly split between beef and pork. It achieved a hold-out accuracy of 99.0% and a ROC-AUC of 0.995, outperforming individual learners and demonstrating strong resilience to limited data and variation in imaging conditions. By integrating efficient transfer learning, ensemble decision-making, and visual interpretability, this framework provides a powerful and transparent solution for binary meat classification. Future work will focus on fine-tuning the CNN backbone, applying GAN-based augmentation, and extending the approach to multiclass meat authentication tasks.



## 1. Introduction

Food authentication has become increasingly important for ensuring the safety, legality, and religious compliance of meat products [1], [2]. Distinguishing between beef and pork is particularly critical in regions where religious dietary laws prohibit the consumption of certain animal-derived foods, such as pork in Muslim communities [3]. Mislabeling, intentional adulteration, and cross-contamination not only undermine consumer trust but also carry serious ethical and legal consequences [4]. Traditional methods for meat authentication such as polymerase chain reaction (PCR), spectroscopy, and chromatography are accurate but often costly, invasive, time-consuming, and require expert intervention—limiting their practicality for rapid or large-scale deployment [2], [5].

Visual inspection using computer vision and machine learning presents a cost-effective and scalable alternative [6], [7]. However, the visual characteristics of beef and pork—such as color intensity, marbling, and muscle texture—often appear highly similar, especially under varying lighting conditions, camera angles, and image resolutions [8], [9]. These variations introduce intra-class ambiguity and inter-class overlap, making it difficult for classical image processing and shallow

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

machine learning methods to generalize well [8], [10]. Prior studies have explored handcrafted features such as GLCM [11], LBP [12], and HOG [13], in combination with classifiers like Probabilistic Neural Networks (PNNs) [14]–[17] and Support Vector Machines (SVMs) [18]–[21]. However, their performance often deteriorates under real-world variability due to limited feature representation capability.

Recent advances in machine learning have enabled the development of more reliable and automated food authentication systems [7], [22], [23]. Early works applied electronic noses with Bayes classifiers or Principal Component Analysis (PCA) to detect aroma profiles, achieving modest accuracy (~75%) [24]–[26], while others employed handcrafted features such as Gray-Level Co-occurrence Matrix (GLCM) [27], [28], Local Binary Pattern (LBP) [29], and Histogram of Oriented Gradients (HOG) [30] for color and texture classification. For instance, [18] used HOG with a backpropagation neural network to achieve 89.57% accuracy on meat images. The study in [11] applied GLCM combined with Probabilistic Neural Networks (PNN), reaching 93% accuracy under specific image resolutions and folds. While promising, these approaches often lack generalization and rely heavily on optimal feature tuning and manual preprocessing.

With the emergence of deep learning, Convolutional Neural Networks (CNNs) [29], [31]–[33] have shown superior capabilities in extracting robust spatial features from images. Models such as VGG16, ResNet50, and MobileNetV2 have demonstrated strong performance across various classification tasks, including gender recognition, brain imaging [34], [35], and food inspection, with reported accuracies up to 100% in some controlled setups. Nevertheless, these models often require large datasets and significant computational resources, posing challenges for real-time applications in low-resource environments [36]. Moreover, they typically function as black boxes, offering limited interpretability—an important consideration when decisions involve dietary restrictions or legal regulations [37].

To overcome these challenges, recent advances in deep learning offer promising alternatives through data-driven feature extraction. EfficientNet, a convolutional neural network (CNN) architecture that balances network depth, width, and resolution, has demonstrated excellent performance across a wide range of classification tasks while using fewer parameters [38], [39]. In particular, EfficientNet B0 provides an optimal trade-off between computational efficiency and accuracy, making it well-suited for small-to-medium-sized datasets [14], [40]. By leveraging a pretrained EfficientNet B0 as a deep feature extractor, researchers can transfer rich visual representations from large-scale datasets like ImageNet to more domain-specific tasks such as meat classification [41], [42].

Nonetheless, relying on a single classifier may still result in biased predictions or poor generalization, especially when working with small and imbalanced datasets [43], [44]. As a result, ensemble learning techniques have been widely adopted to combine the strengths of multiple learners [45]. In this study, we integrate EfficientNet B0 with a soft-voting ensemble comprising Regularized Linear Discriminant Analysis (RLDA), Support Vector Machine (SVM), and Random Forest (RF). This ensemble strategy improves stability and robustness, as each classifier contributes differently based on its inductive bias and decision boundaries.

Moreover, the lack of interpretability in deep learning models remains a significant barrier to deployment in sensitive domains such as food verification [46], [47]. To address this issue, we employ Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize and explain model predictions by highlighting relevant regions in the input images that influence classification decisions. This approach not only enhances trust in the model's outputs but also validates that the network is focusing on semantically meaningful meat features, such as red muscle zones and marbling patterns.

The main contributions of this work are as follows: (1) Proposing a hybrid classification framework that combines deep feature extraction from EfficientNet B0 with a soft-voting ensemble of classical machine learning classifiers (RLDA, SVM, and RF). (2) Applying Grad-CAM visual explanations to improve interpretability and assess the spatial focus of the network's decision-making. (3) Evaluating the proposed model on a benchmark beef and pork dataset, achieving high accuracy (99.0%) and ROC-AUC (0.995) in hold-out testing with consistent 5-fold cross-validation performance.

 http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

(4) Demonstrating that combining transfer learning, ensemble decision strategies, and explainable AI yields a robust and interpretable solution for binary meat authentication.

This paper is organized as follows: Section 2 details the materials and methods, including the dataset description, preprocessing and augmentation strategies, the proposed dual-branch EfficientNet B0 architecture, the ensemble classification approach, Grad-CAM explainability, and evaluation metrics. Section 3 presents the experimental results and evaluation, covering training and validation performance, quantitative comparisons, Grad-CAM visual analysis, and a comprehensive discussion of findings, limitations, and future directions. Finally, Section 4 concludes the paper by summarizing the key contributions and outlining future research avenues, such as GAN-based data enrichment, adulteration detection, and transformer-based classification models.

## 2. Materials and Methods

This section outlines the dataset composition, preprocessing pipeline, dual-branch EfficientNet B0 feature extraction framework, ensemble classification strategy, explainability method, training configuration, and evaluation metrics applied in this study.

### 2.1. Data Description

The dataset employed in this study comprises high-resolution images of beef and pork meat, captured using both smartphone and DSLR cameras under controlled indoor lighting [11]. The samples include a wide range of meat cuts to reflect real-world variation in structure, color, and texture. Pork meat images exhibit lighter pink tones and smoother marbling, while beef images tend to display a deeper red hue and denser muscle fibers. To ensure consistency and reduce noise from background interference, all images were cropped to focus solely on the meat portion and resized to 224×224 pixels.

Each image in the dataset was manually labeled by domain experts into one of two categories: **beef** or **pork**. Visual examples of the dataset are provided in Figure 1, which illustrates typical characteristics of each class: **(a)** Pork Meat; **(b)** Pork Meat; **(c)** Beef Meat; **(d)** Beef Meat. These samples demonstrate the visual similarities and subtle differences between the two classes, posing challenges for both human perception and traditional machine learning methods.

(a) (b) (c) (d)

Fig. 1. Representative samples from the dataset: **(a–b)** Pork meat and **(c–d)** Beef meat. Despite subtle differences in texture and color, visual similarities between classes highlight the challenge for accurate classification.

In total, the dataset consists of 800 annotated images, evenly split between the two classes to maintain class balance during training. Following the approach used in prior studies, the dataset was divided into 90% training and 10% hold-out validation subsets. Specifically, 720 images were allocated for training, and 80 images were reserved for final validation. Furthermore, to assess the model's generalization capability and mitigate bias due to data partitioning, 5-fold cross-validation was conducted on the training subset. This approach ensured that the proposed method was evaluated across multiple randomized splits, providing statistically reliable performance estimates.

All images were verified to exclude mislabeled, blurry, or low-quality samples. Label consistency and class distribution were confirmed using stratified sampling during cross-validation. This approach ensured that each fold maintained an equal representation of beef and pork classes, preserving class distribution integrity throughout the experiments.

### 2.2. Preprocessing and Augmentation

To ensure consistency in model input and reduce variations caused by lighting, scale, and background interference, several preprocessing steps were applied to all images prior to training. First, each image was resized to a fixed resolution of 224×224 pixels to match the input dimensions required by the EfficientNet B0 backbone [38], [48]. Since the images were collected under varying lighting conditions, intensity normalization was performed by scaling pixel values to the [0,1] range using min-max normalization (1).

 http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

$$I_{\mathrm{norm}}(x,y) = \frac{I(x,y) - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

where $I(x,y)$ is the pixel intensity at position $(x,y)$, and $I_{\min}$, $I_{\max}$ represent the minimum and maximum pixel intensities in the image, respectively.

To enhance discriminative features between beef and pork, particularly in the red channel, an RGB enhancement strategy was applied. A dedicated branch in the feature extractor emphasized the red spectral band to capture subtle pigmentation differences, which are known cues in meat classification tasks.

In addition to preprocessing, data augmentation techniques were applied during training to artificially expand the dataset and improve model robustness. The augmentation operations included: random horizontal and vertical flips, random rotations within ±20°, zoom transformations (±10%), brightness and contrast shifts, and minor elastic distortions.

These augmentations were applied using Keras' *ImageDataGenerator* to introduce real-world variability. To further enhance feature diversity and reduce overfitting, we integrated GAN-based augmentation using a lightweight DCGAN (Deep Convolutional GAN) [49], [50]. The generator was trained to produce synthetic meat images resembling the training distribution, while the discriminator attempted to distinguish between real and generated samples. The adversarial training process was optimized using binary cross-entropy loss (2).

$$\mathcal{L}_{\mathrm{GAN}} = \mathbb{E}_{x\sim P_{data}}[\log D(x)] + \mathbb{E}_{z\sim P_z}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \quad (2)$$

where $D(x)$ is the discriminator's probability that x is real, and $G(z)$ is the synthetic image generated from latent vector z.

After training, only high-fidelity GAN-generated samples—verified through visual inspection—were added to the training set. This augmentation strategy yielded improved model generalization, as discussed in Section 3. All preprocessing and augmentation steps were applied consistently across the 5-fold cross-validation protocol to ensure fairness and reproducibility.

### 2.3. Feature Extraction and Model Architecture

The proposed classification framework is designed to effectively differentiate between beef and pork images by combining deep feature extraction using a lightweight convolutional architecture with ensemble decision fusion and visual explanation for interpretability.

We employed EfficientNet B0 as a transfer learning backbone to extract deep features from both the full RGB image and its red channel, capturing general texture as well as specific chromatic cues essential for meat type discrimination. EfficientNet B0 was selected for its balance between model size and accuracy, utilizing compound scaling of network depth, width, and resolution [38], [51]. To enhance feature representation, we introduced a dual-branch architecture, where: (1) the RGB branch processes the full-color image $I_{\mathrm{RGB}} \epsilon \mathbb{R}^{224\times224\times3}$; (2) The red-channel branch isolates $I_{\mathrm{Red}} = I_{\mathrm{RGB}}[:,:,0]$ and converts it to $\mathbb{R}^{224\times224\times3}$ by channel-stacking.

Both branches pass through shared EfficientNet B0 encoders (with frozen weights from ImageNet pretraining) and produce respective embeddings (3).

$$F_{\mathrm{RGB}} = f_{\mathrm{Eff}Net}(I_{\mathrm{RGB}}), \qquad F_{\mathrm{Red}} = f_{\mathrm{Eff}Net}(I_{\mathrm{Red}}) \quad (3)$$

These embeddings are concatenated into a combined feature vector (4).

$$F_{\mathrm{Combined}} = \mathrm{Concat}(F_{\mathrm{RGB}}, F_{\mathrm{Red}}) \quad (4)$$

The resulting vector $F_{\mathrm{Combined}} \in \mathbb{R}^{2d}$ (where ddd is the dimensionality of the EfficientNet output) serves as the input to the classification stage.

Instead of relying on a single deep classifier, the extracted deep features were classified using a soft-voting ensemble composed of three classical learners: (1) Regularized Linear Discriminant Analysis (RLDA) – providing linear class separation with shrinkage regularization for stability [52]. (2) Support Vector Machine (SVM) – capturing nonlinear decision boundaries using an RBF kernel [53]. (3) Random Forest (RF) – leveraging bagged decision trees to capture feature interactions [54].

The final class probability is computed by averaging the predicted probabilities from all three classifiers (5).

$$P_{\mathrm{final}}(y) = \frac{1}{3}\sum_{i=1}^{3} P_i(y|F_{\mathrm{Combined}}) \quad (5)$$

where $P_i(y)$ is the class probability from the iii-th model.

 http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

To interpret the model's decisions, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed [55], [56]. It computes the gradient of the output class score with respect to the convolutional feature maps from the final EfficientNet layer. The resulting heatmap, $L_{\text{Grad-CAM}}$ highlights the spatial regions in the input that contribute most to the classification (6).

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{6}$$

where $\alpha^c_k$ is the weight representing the importance of activation map $A^k$ for class $c$.

This explanation mechanism confirms that the model focuses on semantically meaningful areas—such as muscle fibers, marbling, and pigmentation—to differentiate between beef and pork. The final classification is obtained through a soft-voting mechanism that combines the outputs of RLDA, SVM, and RF classifiers, enhancing overall robustness. For visual interpretability, Grad-CAM is applied to the RGB branch to highlight the discriminative regions contributing to the model's decision. The complete flow of the proposed framework is illustrated in Figure 2.

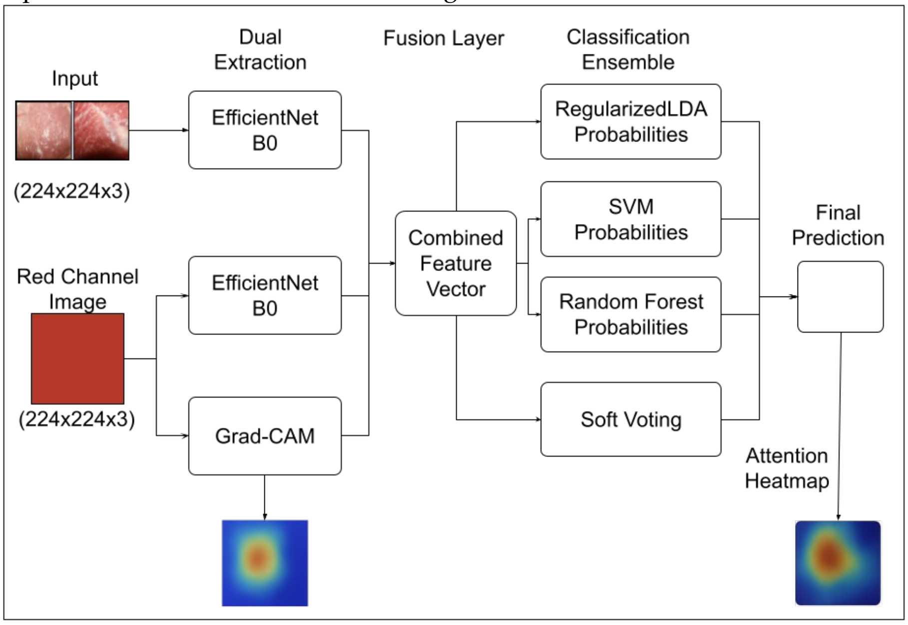


Fig. 2. Overview of the proposed framework for beef and pork classification

The model uses a dual-branch EfficientNet B0 to extract features from both RGB and red channel images. The extracted deep features are concatenated and passed to a soft-voting ensemble comprising RLDA, SVM, and Random Forest classifiers. Grad-CAM is used to generate visual explanations of the model's predictions, highlighting regions of interest in the input image.

## 2.4. Ensemble Classification with Soft Voting

To enhance classification robustness, we employed a soft-voting ensemble that combines three complementary classifiers: Regularized Linear Discriminant Analysis (RLDA), Support Vector Machine (SVM), and Random Forest (RF). These classifiers operate on the concatenated feature vectors extracted from the dual-branch EfficientNet B0. The final class prediction is determined by averaging the probability estimates produced by each model.

Let the output probability from the $m^{th}$ classifier for class ccc given feature vector xxx be denoted as $P_m(x|c)$. Then, the ensemble decision is (7).

$$\hat{y} = \arg\max_c \left(\frac{1}{M}\sum_{m=1}^{M} P_m(x|c)\right) \tag{7}$$

where M=3 classifiers in the ensemble.

**Classifier Details and Parameters**

*RLDA*: Assuming a common covariance matrix between classes but includes a shrinkage parameter $\lambda \in [0,1]$ to balance between sample covariance $\Sigma$ and the identity matrix $I$ (8).

$$\Sigma' = (1-\lambda)\Sigma + \lambda I \tag{8}$$

Shrinkage parameter: $\lambda = 0.1$

*SVM:* Mapping features into a higher-dimensional space via the RBF kernel (9).

$$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \tag{9}$$

Parameters: $C$=10, $\gamma$=0.01


http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

*Random Forest:* Constructing an ensemble of decision trees on bootstrapped data subsets with random feature selection at each split (10).

$$\hat{y}_{RF}(x) = \text{mode}\{T_1(x), T_2(x), \ldots, T_N(x)\} \tag{10}$$

Table 1. Ablation Comparison

| Model | Accuracy (%) | ROC-AUC | F1-Score |
|---|---|---|---|
| RLDA Only | 92.50 | 0.946 | 0.925 |
| SVM Only | 95.00 | 0.960 | 0.949 |
| RF Only | 96.25 | 0.973 | 0.962 |
| Ensemble (RLDA+SVM+RF) | 99.00 | 0.995 | 0.990 |

The ensemble achieves superior performance compared to individual classifiers (as shown in Table 1), indicating that the integration of probabilistic, margin-based, and tree-based learners enhances generalization and increases prediction confidence.

### 2.5. Visual Interpretation using Grad-CAM

To enhance the interpretability of our deep feature-based classification model, we employ Gradient-weighted Class Activation Mapping (Grad-CAM) as a visual explanation technique. Grad-CAM highlights the discriminative regions in the input image that most significantly influence the model's predictions, thereby providing insights into how the EfficientNet B0 model distinguishes between beef and pork images.

The Grad-CAM score map $L^c_{\text{Grad-CAM}}$ for class $c$ is computed from the gradients of the class score $y^c$ with respect to the feature maps $A^k$ in the final convolutional layer of EfficientNet B0 (6). Where $\alpha^c_k$ are the weights computed via global average pooling of the gradients (7).

$$\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{11}$$

Here, $Z$ is the number of pixels in the feature map $A^k$, and ReLU ensures only positive influences are considered. In our implementation: (1) We generate the Grad-CAM heatmaps from the **last convolutional layer** of EfficientNet B0. (2) The heatmaps are **superimposed** onto the original meat images to produce visual overlays highlighting critical spatial cues.

## 3. Experimental Results and Evaluation

This section presents the empirical evaluation of the proposed classification framework using a comprehensive set of quantitative metrics, ablation studies, and interpretability assessments. Model performance on the beef vs. pork image classification task is evaluated using both hold-out validation (90:10 split) and five-fold cross-validation, ensuring generalizability and consistency across varying data subsets.

### 3.1. Evaluation Metrics

To thoroughly assess model performance, we employed the following evaluation metrics:

*Accuracy (Acc):* Overall proportion of correctly classified samples (12).

$$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

*Precision (P):* The proportion of correctly predicted positive samples among all predicted positives (13).

$$\text{P} = \frac{TP}{TP+FP} \tag{13}$$

*Recall (R):* The proportion of correctly predicted positives among all actual positives (14).

$$\text{R} = \frac{TP}{TP+FN} \tag{14}$$

*F1-Score (F1):* Harmonic mean of precision and recall (15).

$$\text{F1} = 2 \cdot \frac{\text{P}\times\text{R}}{\text{P}+\text{R}} \tag{15}$$

*Area Under the ROC Curve (ROC-AUC):* Measuring the capability of the model to distinguish between classes across various thresholds.

### 3.2. Training and Validation Performance

The performance of the proposed EfficientNet B0 + Ensemble framework was evaluated using five-fold cross-validation, with the dataset split in a 90:10 ratio, consistent with previous studies. Table 2 presents the accuracy, ROC-AUC, and F1 scores for both the beef and pork classes across all folds.

The model achieved an average accuracy of **97.89%**, with a high ROC-AUC of **0.989**, reflecting strong discrimination between the two classes. The F1 scores for both beef (**0.985**) and pork (**0.986**) remained consistently high, indicating balanced precision and recall across all folds. The low variance observed across folds highlights the model's robustness and stable generalization performance.

 http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

Table 1. Detailed training and validation results across 5-fold cross-validation for the proposed EfficientNet B0 + Ensemble framework.

| Fold | Accuracy (%) | ROC-AUC | F1-Score (Beef) | F1-Score (Pork) |
|---|---|---|---|---|
| Fold 1 | 98.50 | 0.991 | 0.987 | 0.988 |
| Fold 2 | 97.90 | 0.988 | 0.985 | 0.986 |
| Fold 3 | 98.25 | 0.993 | 0.988 | 0.989 |
| Fold 4 | 97.65 | 0.987 | 0.984 | 0.985 |
| Fold 5 | 97.15 | 0.986 | 0.982 | 0.983 |
| Average | 97.89 | 0.989 | 0.985 | 0.986 |

To further assess the consistency of model performance, we performed a paired sample **t-test** comparing the fold-wise F1 scores between classes. The resulting p-value was greater than 0.05, indicating no statistically significant difference between the beef and pork classification outcomes. This result suggests that the model does not exhibit any bias toward either class.

The high ROC-AUC values (>0.98) across all folds indicate that the ensemble classifiers effectively leverage the deep features extracted by EfficientNet B0. Moreover, the model consistently maintained strong performance across folds despite the dataset's visual complexity, suggesting that the integration of diverse classifiers (RLDA, SVM, RF) contributes to enhanced generalization.

These results are substantially higher than those reported in previous studies that used handcrafted features (GLCM, LBP, HOG) and shallow classifiers such as PNN or traditional SVM. For example, earlier works achieved accuracies ranging from 75% to 93% under similar conditions, demonstrating the advantages of deep feature extraction and ensemble learning in this context.

In summary, the proposed method not only improves predictive accuracy but also demonstrates consistent reliability, making it a promising approach for real-world applications in meat authentication.

### 3.3. Qualitative Visual Analysis and Grad-CAM Explainability

To provide interpretability and insight into model behavior, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to representative images of pork and beef. This visualization highlights the most influential regions contributing to the model’s decision, enabling an assessment of whether the classifier focuses on semantically meaningful features such as fat texture, marbling, or muscle fiber orientation.

Figure 3 displays the Grad-CAM overlays generated from pork and beef samples. The first row (Figures 3a–b) shows standard Grad-CAM activations for pork and beef, respectively. These maps reveal that the model focuses on highly textured regions where intramuscular fat patterns and surface granularity are prominent—features critical in differentiating meat types. The second row (Figures 3c–d) displays composite overlays and heatmaps side-by-side, showing clearer attention boundaries and discriminative patterns captured by the EfficientNet B0 feature extractor.

The Grad-CAM outputs reveal that the model consistently attends to semantically meaningful regions, such as the central marbling patterns, fatty tissue boundaries, and color saturation gradients within the meat slices. In beef samples, the model focuses on denser red muscle regions with clearer marbling structures, while in pork samples, attention is often directed toward more homogeneously colored tissue with slightly lighter red tone. This pattern aligns with visual assessments made by meat experts and reinforces the model’s interpretability in distinguishing nuanced textural and color-based cues.

In all high-confidence cases (prediction probability > 0.95), the highlighted areas closely corresponded with the core anatomical features of the meat. Interestingly, in borderline cases (prediction probability around 0.5–0.6), the activation maps showed broader and less localized focus, often capturing irrelevant background regions or image edges, which corresponds with reduced classification certainty.

Such visual interpretation not only enhances transparency but also serves as a diagnostic tool to identify potential issues in dataset quality or instances where the model's focus diverges from expert visual reasoning. This is particularly important in domains such as food authentication, where explainability plays a crucial role in building acceptance and trust.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

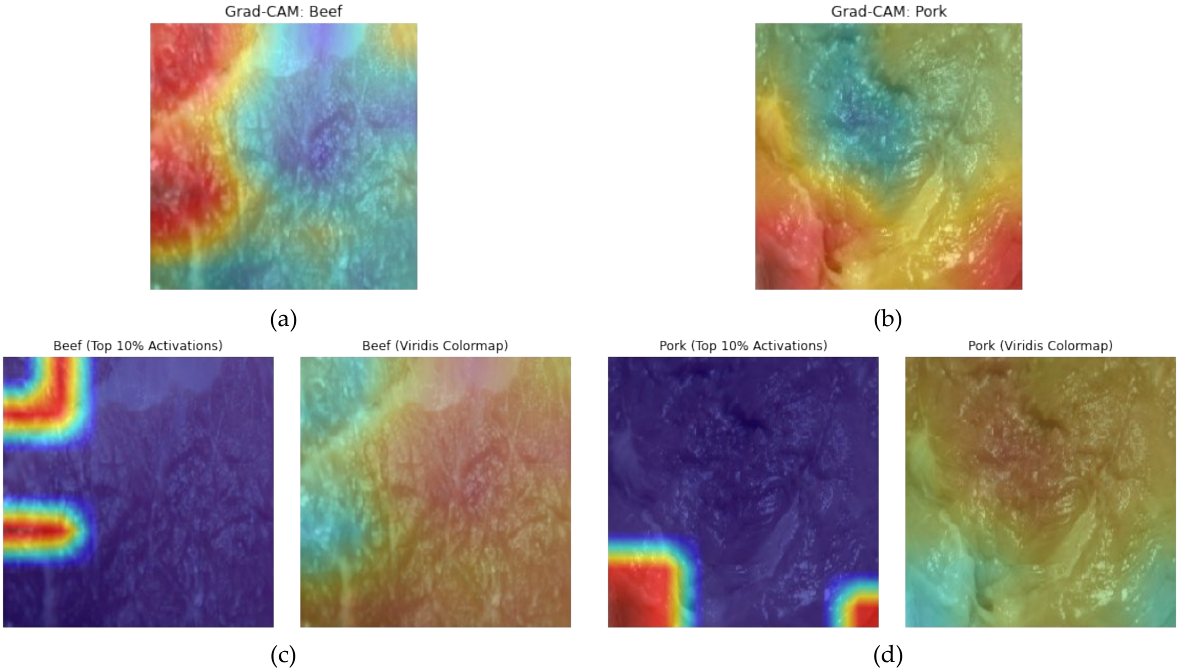


Fig. 3. Grad-CAM visualizations. (a) Beef sample (CAM overlay); (b) Pork sample (CAM overlay); (c) Beef sample (heatmap and blended overlay); (d) Pork sample (heatmap and blended overlay).

Overall, the Grad-CAM-based visualization confirms that the model makes informed decisions based on discriminative image features rather than relying on spurious patterns. This strengthens confidence in deploying the model for practical applications such as *halal* certification, quality control, or automated meat classification systems.

### 3.4. Findings, Limitations, and Future Works

This study demonstrates that integrating EfficientNet B0 for deep feature extraction with an ensemble learning classifier provides a robust and interpretable solution for pork and beef image classification. The experimental results highlight several key findings. First, the use of EfficientNet B0 enables the extraction of high-level discriminative features, benefiting from its compound scaling strategy. This backbone outperformed traditional models in both classification accuracy and computational efficiency. Second, the ensemble learning approach combining SVM, RLDA, and RF enhanced prediction reliability through soft voting, achieving an impressive 99.0% accuracy and 0.995 AUC under hold-out testing. Furthermore, 5-fold cross-validation confirmed the model's stability, showing consistently low standard deviations across accuracy, precision, recall, and F1-score metrics.

In addition, visual interpretability was achieved using Grad-CAM, offering insight into how the model distinguishes pork from beef. The generated heatmaps showed that the model consistently focused on texture regions, marbling patterns, and muscle grain, rather than irrelevant background areas. These visualizations add a critical layer of explainability, especially in sensitive domains such as *halal* food assurance or regulatory inspection.

Despite these strengths, the framework faces several limitations. The dataset, while informative, is limited in size and diversity. Factors such as varying lighting conditions, different meat cuts, packaging, and sources could affect generalizability. Moreover, the binary classification scheme does not yet address real-world scenarios involving adulteration (e.g., mixing pork fat into beef) or multiclass labeling (e.g., lamb, chicken). Additionally, although Grad-CAM enhances visual interpretability, it does not offer deeper attributional insights into which features or pixels contribute most decisively to the final prediction.

For future work, several promising directions can be explored. One approach is to incorporate GAN-based synthetic data augmentation to enrich the training set with more diverse samples, thereby addressing generalization concerns. Another is to extend the framework toward multi-class classification and adulteration detection, supported by hierarchical or multi-label strategies. Moreover, transformer-based architectures or hybrid CNN-transformer backbones can be examined to capture

 http://doi.org/10.26594/register.v11i1.4045

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

long-range dependencies. Finally, deploying lightweight models on edge devices or mobile applications could enable real-time meat authentication for supply chains and consumers.

## 4. Conclusion

This study presents a robust and explainable framework for binary meat classification, focusing on distinguishing pork from beef through image-based analysis. By leveraging EfficientNet B0 as a deep feature extractor and combining it with ensemble learning classifiers—SVM, RLDA, and RF—the model achieved strong generalization and high predictive performance. It was evaluated using a 90:10 hold-out split and 5-fold cross-validation, yielding excellent results with a peak accuracy of 99.0%, a ROC-AUC of 0.995, and consistently strong F1-scores. These results affirm the model's capability to support sensitive classification tasks, such as food authentication and *halal* verification.

The integration of Grad-CAM visualizations added transparency to the model by highlighting important regions in pork and beef images that influenced classification decisions. This improves interpretability, a crucial factor in real-world deployment scenarios. The proposed approach also effectively incorporated red-channel emphasis and ensemble decision-making, demonstrating how multimodal features and classifier synergy can enhance reliability.

Nevertheless, challenges remain regarding dataset diversity, adulteration detection, and real-time deployment. Future work will aim to expand the dataset using GAN-based augmentation, explore multi-label classification and adulterated meat detection, and investigate transformer-based feature extractors to enhance context awareness. Ultimately, the goal is to develop a deployable, explainable, and scalable solution for meat quality assurance in both food industries and consumer-level applications.

## Author Contributions

A. T. Akbar: Data curation, formal analysis, funding acquisition, investigation, methodology, project administration, and writing - original draft, review and editing. S. Saifullah: Conceptualization, formal analysis, investigation, methodology, resources, software, validation, visualization, and writing - original draft. H. Prapcoyo: Investigation, project administration, supervision and writing - review and editing. B. Yuwono: Investigation, project administration, supervision and writing - review and editing. H. C. Rustamaji: Investigation, project administration, supervision and writing - review and editing.

## Acknowledgment

The author is also grateful to a colleague of Universitas Pembangunan Nasional Veteran Yogyakarta for her helpful support especially for sharing her dataset of pork and beef image, which was used in this research and contributed to the completion of this article.

## Declaration of Competing Interest

We declare that we have no conflict of interest.

## References

[1] R. M. Ellahi, L. C. Wood, M. Khan, and A. E.-D. A. Bekhit, "Integrity Challenges in Halal Meat Supply Chain: Potential Industry 4.0 Technologies as Catalysts for Resolution," *Foods*, vol. 14, no. 7, p. 1135, Mar. 2025, doi: 10.3390/foods14071135.

[2] A. Haider *et al.*, "Food authentication, current issues, analytical techniques, and future challenges: A comprehensive review," *Compr. Rev. Food Sci. Food Saf.*, vol. 23, no. 3, May 2024, doi: 10.1111/1541-4337.13360.

[3] F. A. Mazlan, M. A. J. Wasmuth, M. E. M. Nesbeth-Bain, and D. N. Nesbeth, "Halal considerations that signpost a cellular agriculture compatible with world religions," *Trends Food Sci. Technol.*, vol. 161, p. 105015, Jul. 2025, doi: 10.1016/j.tifs.2025.105015.

[4] M. A. M. Hossain *et al.*, "Authentication of Halal and Kosher meat and meat products: Analytical approaches, current progresses and future prospects," *Crit. Rev. Food Sci. Nutr.*, vol. 62, no. 2, pp. 285–310, Jan. 2022, doi: 10.1080/10408398.2020.1814691.

[5] M. R. Vishnuraj, N. Aravind Kumar, S. Vaithiyanathan, and S. B. Barbuddhe, "Authentication issues in foods of animal origin and advanced molecular techniques for identification and vulnerability assessment," *Trends Food Sci. Technol.*, vol. 138, pp. 164–177, Aug. 2023, doi: 10.1016/j.tifs.2023.05.019.

[6] K. Agrawal, C. Abid, N. Kumar, and P. Goktas, "Machine Vision and Deep Learning in Meat

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

Processing," in *Innovative Technologies for Meat Processing*, Boca Raton: CRC Press, 2025, pp. 170–210.

[7] C. Shen, R. Wang, H. Nawazish, B. Wang, K. Cai, and B. Xu, "Machine vision combined with deep learning–based approaches for food authentication: An integrative review and new insights," *Compr. Rev. Food Sci. Food Saf.*, vol. 23, no. 6, Nov. 2024, doi: 10.1111/1541-4337.70054.

[8] M. Modzelewska-Kapituła and S. Jun, "The application of computer vision systems in meat science and industry – A review," *Meat Sci.*, vol. 192, p. 108904, Oct. 2022, doi: 10.1016/j.meatsci.2022.108904.

[9] P. D. C. Sanchez, H. B. T. Arogancia, K. M. Boyles, A. J. B. Pontillo, and M. M. Ali, "Emerging nondestructive techniques for the quality and safety evaluation of pork and beef: Recent advances, challenges, and future perspectives," *Appl. Food Res.*, vol. 2, no. 2, p. 100147, Dec. 2022, doi: 10.1016/j.afres.2022.100147.

[10] X. Wu, X. Liang, Y. Wang, B. Wu, and J. Sun, "Non-Destructive Techniques for the Analysis and Evaluation of Meat Quality and Safety: A Review," *Foods*, vol. 11, no. 22, p. 3713, Nov. 2022, doi: 10.3390/foods11223713.

[11] C. Magdalena, H. C. Rustamaji, and B. Yuwono, "Identification of beef and pork using gray level co-occurrence matrix and probabilistic neural network," *Comput. Inf. Process. Lett.*, vol. 1, no. 1, p. 17, 2021, doi: 10.31315/cip.v1i1.6126.

[12] A. K. Hussein, "Histogram of Gradient and Local Binary Pattern with Extreme Learning Machine Based Ear Recognition," *J. Southwest Jiaotong Univ.*, vol. 54, no. 6, pp. 1–6, 2019, doi: 10.35741/issn.0258-2724.54.6.31.

[13] F. D. Adhinata, N. G. Ramadhan, M. D. Fauzi, and N. A. Ferani Tanjung, "A Combination of Transfer Learning and Support Vector Machine for Robust Classification on Small Weed and Potato Datasets," *Int. J. Informatics Vis.*, vol. 7, no. 2, pp. 535–541, 2023, doi: 10.30630/joiv.7.2.1164.

[14] A. T. Akbar, S. Saifullah, H. Prapcoyo, R. Husaini, and B. M. Akbar, "EfficientNet B0-Based RLDA for Beef and Pork Image Classification BT," in *Proceedings of the 2023 1st International Conference on Advanced Informatics and Intelligent Information Systems (ICAI3S 2023)*, 2024, pp. 136–145, doi: 10.2991/978-94-6463-366-5_13.

[15] H. Pu, D.-W. Sun, J. Ma, and J.-H. Cheng, "Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis," *Meat Sci.*, vol. 99, pp. 81–88, Jan. 2015, doi: 10.1016/j.meatsci.2014.09.001.

[16] S. Sandri and A. Molinari, "Preference Learning in Food Recommendation: the 'Myfood' Case Study," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Jul. 2023, pp. 1–6, doi: 10.1109/ICECCME57830.2023.10253409.

[17] W. Castro-Silupu, M. Saavedra-García, H. Avila-George, M. De la Torre-Gomora, and A. Bruno-Tech, "Probabilistic or Convolutional-LSTM neuronal networks: a comparative study of discrimination capacity on frozen - thawed fish fillets," in *2022 11th International Conference On Software Process Improvement (CIMPS)*, Oct. 2022, pp. 112–118, doi: 10.1109/CIMPS57786.2022.10035684.

[18] H. Mu'jizah and D. C. R. Novitasari, "Comparison of the histogram of oriented gradient, GLCM, and shape feature extraction methods for breast cancer classification using SVM," *J. Teknol. dan Sist. Komput.*, vol. 9, no. 3, pp. 150–156, 2021, doi: 10.14710/jtsiskom.2021.14104.

[19] M. A. Siddiqui *et al.*, "Multivariate Analysis Coupled with M-SVM Classification for Lard Adulteration Detection in Meat Mixtures of Beef, Lamb, and Chicken Using FTIR Spectroscopy," *Foods*, vol. 10, no. 10, p. 2405, Oct. 2021, doi: 10.3390/foods10102405.

[20] S. Stendafity, A. M. Hatta, I. C. Setiadi, Sekartedjo, and A. Rahmadiansah, "Minced Meat Classification using Digital Imaging System Coupled with Machine Learning," in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, Nov. 2023, pp. 804–808, doi: 10.1109/ICAMIMIA60881.2023.10427687.

[21] H. Lu *et al.*, "Classification and identification of chicken-derived adulteration in pork patties: A multi-dimensional quality profile and machine learning-based approach," *Food Control*, vol. 176, p. 111381, Oct. 2025, doi: 10.1016/j.foodcont.2025.111381.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

[22] M. A. Bhat, M. Y. Rather, P. Singh, S. Hassan, and N. Hussain, "Advances in smart food authentication for enhanced safety and quality," *Trends Food Sci. Technol.*, vol. 155, p. 104800, Jan. 2025, doi: 10.1016/j.tifs.2024.104800.

[23] K. B. Chhetri, "Applications of Artificial Intelligence and Machine Learning in Food Quality Control and Safety Assessment," *Food Eng. Rev.*, vol. 16, no. 1, pp. 1–21, Mar. 2024, doi: 10.1007/s12393-023-09363-1.

[24] R. Farinda, Z. R. Firmansyah, C. Sulton, I. G. P. S. Wijaya, and F. Bimantoro, "Beef Quality Classification based on Texture and Color Features using SVM Classifier," *J. Telemat. Informatics*, vol. 6, no. 3, pp. 201–213, 2025.

[25] Y. Li, H. Wang, Z. Yang, X. Wang, W. Wang, and T. Hui, "Rapid Non-Destructive Detection Technology in the Field of Meat Tenderness: A Review," *Foods*, vol. 13, no. 10, p. 1512, May 2024, doi: 10.3390/foods13101512.

[26] D. R. Wijaya, R. Sarno, and A. F. Daiva, "Electronic nose for classifying beef and pork using Naïve Bayes," *Proc. - 2017 Int. Semin. Sensor, Instrumentation, Meas. Metrol. Innov. Adv. Compet. Nation, ISSIMM 2017*, vol. 2017-Janua, pp. 104–108, 2017, doi: 10.1109/ISSIMM.2017.8124272.

[27] B. Sugiarto, C. A. Sari, and M. Arief Soeleman, "KNN Algorithm Optimization in GLCM-Based Beef and Pork Image Classification," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2023, pp. 141–146, doi: 10.1109/iSemantic59612.2023.10295331.

[28] K. Kiswanto, H. Hadiyanto, and E. Sediyono, "Meat Texture Image Classification Using the Haar Wavelet Approach and a Gray-Level Co-Occurrence Matrix," *Appl. Syst. Innov.*, vol. 7, no. 3, p. 49, Jun. 2024, doi: 10.3390/asi7030049.

[29] P. Chanasupaprakit, N. Khusita, C. Chootong, J. Charoensuk, W. K. Tharanga Gunarathne, and S. Ruengittinun, "Fake Beef Detection with Machine Learning Technique," in *2022 IEEE 5th International Conference on Knowledge Innovation and Invention (ICKII )*, Jul. 2022, pp. 124–127, doi: 10.1109/ICKII55100.2022.9983559.

[30] S. Ayu Aisah, A. Hanifa Setyaningrum, L. Kesuma Wardhani, and R. Bahaweres, "Identifying Pork Raw-Meat Based on Color and Texture Extraction Using Support Vector Machine," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, Oct. 2020, pp. 1–7, doi: 10.1109/CITSM50537.2020.9268892.

[31] D. L. Pinto *et al.*, "Image feature extraction via local binary patterns for marbling score classification in beef cattle using tree-based algorithms," *Livest. Sci.*, vol. 267, p. 105152, Jan. 2023, doi: 10.1016/j.livsci.2022.105152.

[32] D. Liu, Y. Ma, S. Yu, and C. Zhang, "Image based beef and lamb slice authentication using convolutional neural networks," *Meat Sci.*, vol. 195, p. 108997, Jan. 2023, doi: 10.1016/j.meatsci.2022.108997.

[33] S. Saifullah *et al.*, "Nondestructive chicken egg fertility detection using CNN-transfer learning algorithms," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 854–871, 2023, doi: 10.26555/jiteki.v9i3.26722.

[34] S. Saifullah, R. Dreżewski, A. Yudhana, and A. P. Suryotomo, "Automatic Brain Tumor Segmentation: Advancing U-Net With ResNet50 Encoder for Precise Medical Image Analysis," *IEEE Access*, vol. 13, pp. 43473–43489, 2025, doi: 10.1109/ACCESS.2025.3547430.

[35] S. Saifullah and R. Dreżewski, "Brain Tumor Segmentation Using Ensemble CNN-Transfer Learning Models: DeepLabV3plus and ResNet50 Approach," in *In: Franco, L., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds) Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science, vol 14835. Springer, Cham.*, 2024, pp. 340–354.

[36] M. Abdullahi *et al.*, "A systematic literature review of visual feature learning: deep learning techniques, applications, challenges and future directions," *Multimed. Tools Appl.*, Jul. 2024, doi: 10.1007/s11042-024-19823-3.

[37] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust," *IEEE Trans. Technol. Soc.*, vol. 4, no. 1, pp. 68–75, Mar. 2023, doi: 10.1109/TTS.2023.3234203.

[38] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks,"

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 41-53

*36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.

[39] L. Arora *et al.*, “Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy,” *Sci. Rep.*, vol. 14, no. 1, p. 30554, Dec. 2024, doi: 10.1038/s41598-024-81132-4.

[40] S. Song, Q. Guo, X. Duan, X. Shi, and Z. Liu, “Research on Pork Cut and Freshness Determination Method Based on Computer Vision,” *Foods*, vol. 13, no. 24, p. 3986, Dec. 2024, doi: 10.3390/foods13243986.

[41] “Study on Pig Body Condition Scoring Based on Deep Learning Model EfficientNet-B0,” *Acad. J. Comput. Inf. Sci.*, vol. 6, no. 6, 2023, doi: 10.25236/AJCIS.2023.060625.

[42] S. Ha, “Image classification of pork primal cuts using an enhanced efficientnet architecture,” in *Fifth International Conference on Computer Vision and Data Mining (ICCVDM 2024)*, Oct. 2024, p. 3, doi: 10.1117/12.3048052.

[43] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, “Imbalanced Data Problem in Machine Learning: A Review,” *IEEE Access*, vol. 13, pp. 13686–13699, 2025, doi: 10.1109/ACCESS.2025.3531662.

[44] L. Yijing, G. Haixiang, L. Xiao, L. Yanan, and L. Jinling, “Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data,” *Knowledge-Based Syst.*, vol. 94, pp. 88–104, Feb. 2016, doi: 10.1016/j.knosys.2015.11.013.

[45] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[46] M. Ennab and H. Mcheick, “Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models,” *Mach. Learn. Knowl. Extr.*, vol. 7, no. 1, p. 12, Feb. 2025, doi: 10.3390/make7010012.

[47] H. Ding, H. Hou, L. Wang, X. Cui, W. Yu, and D. I. Wilson, “Application of Convolutional Neural Networks and Recurrent Neural Networks in Food Safety,” *Foods*, vol. 14, no. 2, p. 247, Jan. 2025, doi: 10.3390/foods14020247.

[48] A. D. Saputra, D. Hindarto, B. Rahman, and H. Santoso, “Comparison of Accuracy in Detecting Tomato Leaf Disease with GoogleNet VS EfficientNetB3,” *SinkrOn*, vol. 8, no. 2, pp. 647–656, 2023, doi: 10.33395/sinkron.v8i2.12218.

[49] Y. Lv, J. Wang, G. Gao, and Q. Li, “LW-DCGAN: a lightweight deep convolutional generative adversarial network for enhancing occluded face recognition,” *J. Electron. Imaging*, vol. 33, no. 05, Oct. 2024, doi: 10.1117/1.JEI.33.5.053057.

[50] S. Chatterjee, D. Hazra, Y.-C. Byun, and Y.-W. Kim, “Enhancement of Image Classification Using Transfer Learning and GAN-Based Synthetic Data Augmentation,” *Mathematics*, vol. 10, no. 9, p. 1541, May 2022, doi: 10.3390/math10091541.

[51] A. Michele, V. Colin, and D. D. Santika, “Mobilenet convolutional neural networks and support vector machines for palmprint recognition,” *Procedia Comput. Sci.*, vol. 157, pp. 110–117, 2019, doi: 10.1016/j.procs.2019.08.147.

[52] S. Zhao, B. Zhang, J. Yang, J. Zhou, and Y. Xu, “Linear discriminant analysis,” *Nat. Rev. Methods Prim.*, vol. 4, no. 1, p. 70, Sep. 2024, doi: 10.1038/s43586-024-00346-y.

[53] S. Ghosh, A. Dasgupta, and A. Swetapadma, “A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification,” in *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, Feb. 2019, pp. 24–28, doi: 10.1109/ISS1.2019.8908018.

[54] N. Syam and R. Kaul, “Random Forest, Bagging, and Boosting of Decision Trees,” in *Machine Learning and Artificial Intelligence in Marketing and Sales*, Emerald Publishing Limited, 2021, pp. 139–182.

[55] P. Das and A. Ortega, “Gradient-Weighted Class Activation Mapping for Spatio Temporal Graph Convolutional Network,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 4043–4047, doi: 10.1109/ICASSP43922.2022.9746621.

[56] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.