International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 4, December 2026, pp. 1648~1663
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i4.pp1648-1663

From data to decisions: ensemble deep learning framework for maize seed variety classification with explainable AI

Emmanuel Asante[1], Obed Appiah[2], Ezekiel Mensah Martey[1], Emmanuel Effah[3], Eric Opoku[4], Ernest Kwame Ampomah[1]
[1]Department of Computing and Data Analytics, University of Mines and Technology, Tarkwa, Ghana
[2]Department of Computer Science, University of Energy and Natural Resources, Sunyani, Ghana
[3]Department of Computer Science and Engineering, University of Mines and Technology, Tarkwa, Ghana
[4]Bredesen Center, University of Tennessee, Knoxville, USA

Article Info

Article history:

Received Jan 10, 2026
Revised Aug 4, 2026
Accepted Aug 28, 2026

Keywords:

Ensemble deep learning
Explainable agricultural AI
Low-resource rgb imaging
Maize seed variety classification
Precision farming
Seed variety identification

ABSTRACT

Accurate and non-destructive identification of maize seed varieties is essential for maintaining seed purity, improving crop productivity, and supporting agricultural quality assurance. Although recent studies report high accuracies using advanced imaging and deep learning techniques, many rely on hyperspectral sensors or controlled acquisition environments, or offer limited interpretability, thereby restricting real-world adoption. This study investigates maize seed variety classification using low-cost RGB imaging and an ensemble deep learning framework, focusing on three widely cultivated Ghanaian varieties: Bhihilifa, Sanzal Sima, and Wang Dataa. In collaboration with Heritage Seeds Ghana, an expert-verified dataset of 17,724 maize seed images was curated under realistic acquisition conditions with an Apple iPhone 12. A baseline convolutional neural network (CNN) was developed for benchmarking, followed by transfer learning with DenseNet169, MobileNetV2, and Xception architectures. An ensemble framework integrating DenseNet169 and Xception was proposed to improve robustness and generalization. Across five-fold cross-validation, the ensemble achieved a mean accuracy of 98.89±0.17%, precision of 98.86±0.18%, recall of 98.80±0.17%, and F1-score of 98.83±0.17%. To enhance transparency, Grad-CAM and saliency maps were employed to visualize discriminative features influencing model decisions. Results demonstrate that accurate and interpretable classification of maize seed varieties is achievable using standard RGB imagery without specialized sensing equipment.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Emmanuel Asante
Department of Computing and Data Analytics, School of Railway and Infrastructural Development
University of Mines and Technology
Tarkwa, Ghana
Email: easante@umat.edu.gh

1. INTRODUCTION

Maize (*Zea mays* L.) is one of the most important cereal crops globally, serving as a staple food, livestock feed, and industrial raw material [1], [2]. Global maize production now accounts for approximately 39% of total cereal output, a growth driven by advances in hybrid breeding and seed improvement technologies [3], [4]. However, these advances have also led to a rapid proliferation of varieties with highly similar morphological characteristics, making reliable identification of seed varieties increasingly challenging.

Accurate identification of maize seed varieties is critical for maintaining varietal purity, optimizing yield, and ensuring economic returns for farmers. Seed variety mixing during growth, harvesting, transportation, and storage poses serious challenges to seed purity and quality control [4], [5]. Since different maize varieties are adapted to specific soil types, climates, and cultivation practices, ensuring varietal purity before planting is essential for both agricultural productivity and economic sustainability [5], [6].

Traditional identification methods, including DNA molecular marker analysis, protein electrophoresis, gas chromatography, and expert visual inspection, are highly accurate but labor-intensive, costly, and dependent on specialized laboratory equipment [3], [7]. These constraints make them impractical for routine, large-scale seed inspection, creating a pressing demand for rapid, non-destructive, and cost-effective alternatives.

In response, researchers have increasingly explored image-based non-destructive techniques. Machine vision, near-infrared spectroscopy (NIRS), and hyperspectral imaging (HSI) have demonstrated strong classification performance [5], [8]–[11]. Machine vision extracts morphological features such as size, shape, color, and texture from RGB images [12], while NIRS captures chemical-bond absorption in the 780–2500 nm range [4]. HSI integrates both spatial and spectral information, providing rich representations for distinguishing visually similar varieties [4], [5]. Despite their effectiveness, hyperspectral and spectroscopic systems are expensive, computationally demanding, and complex to deploy [13], [14], limiting adoption in low-resource agricultural environments. Standard RGB imaging, readily available via smartphones and low-cost cameras, offers a far more accessible alternative.

Deep learning has emerged as a powerful tool for image-based seed classification, with convolutional neural networks (CNNs) and advanced architectures reporting high accuracies across multiple studies [5]-[7], [15], [16]. However, many existing maize seed classification studies suffer from one or more of the following limitations: (i) reliance on expensive imaging modalities that restrict deployment on low-resource or mobile devices; (ii) lack of evaluation on region-specific varieties, particularly those cultivated in sub-Saharan Africa; and (iii) insufficient model interpretability, which is increasingly important for trust and adoption in agricultural decision-making systems.

This study addresses these gaps by investigating accurate, interpretable, and resource-efficient maize seed variety classification using low-cost RGB imaging. We focus on three widely cultivated Ghanaian varieties, Bhihilifa, Sanzal Sima, and Wang Dataa, captured with a standard smartphone camera. In collaboration with Heritage Seeds Ghana, we curated a large, expert-verified dataset of 17,724 maize seed images under realistic acquisition conditions. To our knowledge, this is the first publicly available, expert-labeled RGB dataset of Ghanaian maize seed varieties, addressing a critical gap in region-specific agricultural AI research.

We propose an ensemble deep learning framework integrating complementary feature representations from fine-tuned DenseNet169 [17] and Xception [18] architectures. To promote transparency and trust in AI-driven agricultural systems, we incorporate explainable AI techniques, specifically Grad-CAM [19] and saliency maps [20], to visualize the discriminative seed features influencing model predictions.

The main contributions of this work are as follows:

- A large, region-specific maize seed image dataset for three Ghanaian varieties, collected using low-resource RGB imaging and expert labeling, and made publicly available.
- A computationally efficient ensemble deep learning model that achieves 98.89±0.17% cross-validated accuracy without specialized sensing hardware.
- An interpretable classification framework that provides visual explanations of model decisions via Grad-CAM and saliency maps, supporting trust and adoption in agricultural practice.

By explicitly addressing accessibility, interpretability, and real-world applicability, this study advances maize seed variety classification beyond laboratory-scale solutions with direct relevance for seed quality assurance in Ghana and neighboring West African countries. The remainder of this paper is organized as follows: Section 2 describes the research method, dataset, and experimental setup. Section 3 presents results and discussion. Section 4 concludes the paper.

2. METHOD

This section describes the complete methodological pipeline for maize seed variety classification, covering data acquisition, preprocessing, model development, ensemble construction, and evaluation as shown in Figure 1. The study employed a multi-phase design: (i) dataset curation under realistic conditions using a 12MP phone camera; (ii) development of a baseline CNN for benchmarking; (iii) transfer learning with three pre-trained architectures; (iv) ensemble construction via stacking; and (v) evaluation of the proposed ensemble using five-fold cross-validation and explainability visualization.

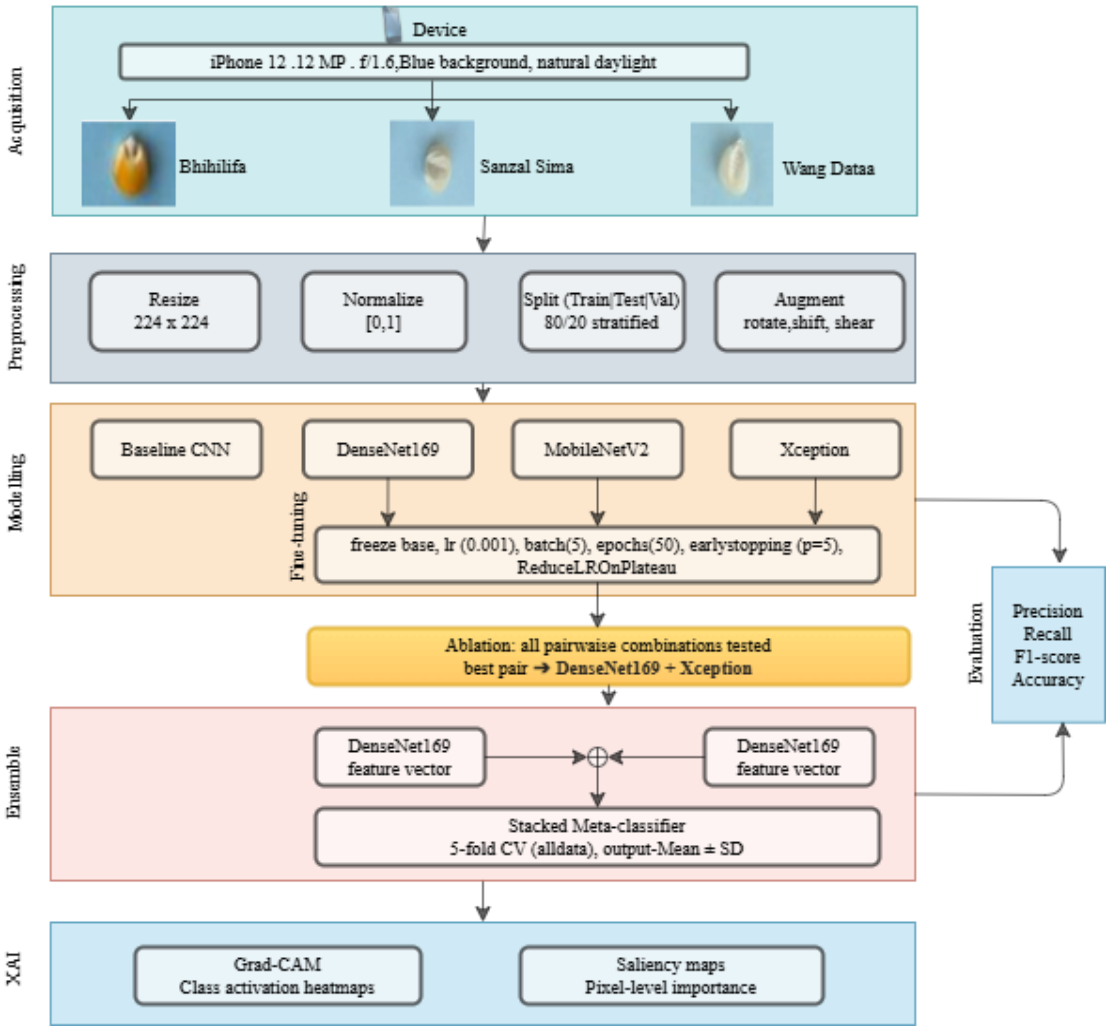


Figure 1. End-to-end pipeline for Ghanaian maize seed variety classification, from expert-verified RGB image acquisition through preprocessing, transfer learning, ensemble construction, five-fold cross-validation, and explainability analysis via Grad-CAM and saliency maps

2.1. Experimental setup and environment

All experiments were conducted on a Dell XPS 13 (2024) laptop equipped with an Intel Core Ultra 7 155H processor (up to 4.8 GHz, 16 cores), 16 GB LPDDR5x RAM (7467 MT/s), and Intel Arc integrated graphics. The software environment consisted of Python 3.10.12, TensorFlow 2.13.0 with the integrated Keras API (tf.keras), NumPy 1.24.3, scikit-learn 1.3.0, Matplotlib 3.7.2, and OpenCV 4.8.0. Model training was performed on a CPU due to the absence of a discrete GPU in the XPS 13 configuration, thereby validating the feasibility of the proposed approach on accessible hardware.

2.2. Dataset

2.2.1. Image acquisition

The dataset comprises images of three maize seed varieties, Wang Dataa, Sanzal Sima, and Bhihilifa, sourced from Heritage Seeds Ghana. These cultivars are frequently grown in northern Ghana. Seeds were manually categorized and labeled by domain experts at Heritage Seeds Ghana prior to image capture. Images were acquired using an Apple iPhone 12 smartphone (12 MP wide-angle sensor, f/1.6 aperture, 26 mm equivalent focal length, optical image stabilization, Apple A14 Bionic processor), producing original JPEG images at 4032×3024 pixels. A uniform blue background was used during capture to enhance seed visibility and reduce background noise under natural daylight conditions. All images were subsequently resized to 224×224 pixels to ensure uniformity across the dataset. The expert-verified labeling process was carried out over two weeks to ensure the validity and accuracy of all annotations. The image acquisition setup is illustrated in Figure 2, and the curation pipeline is shown in Figure 3.

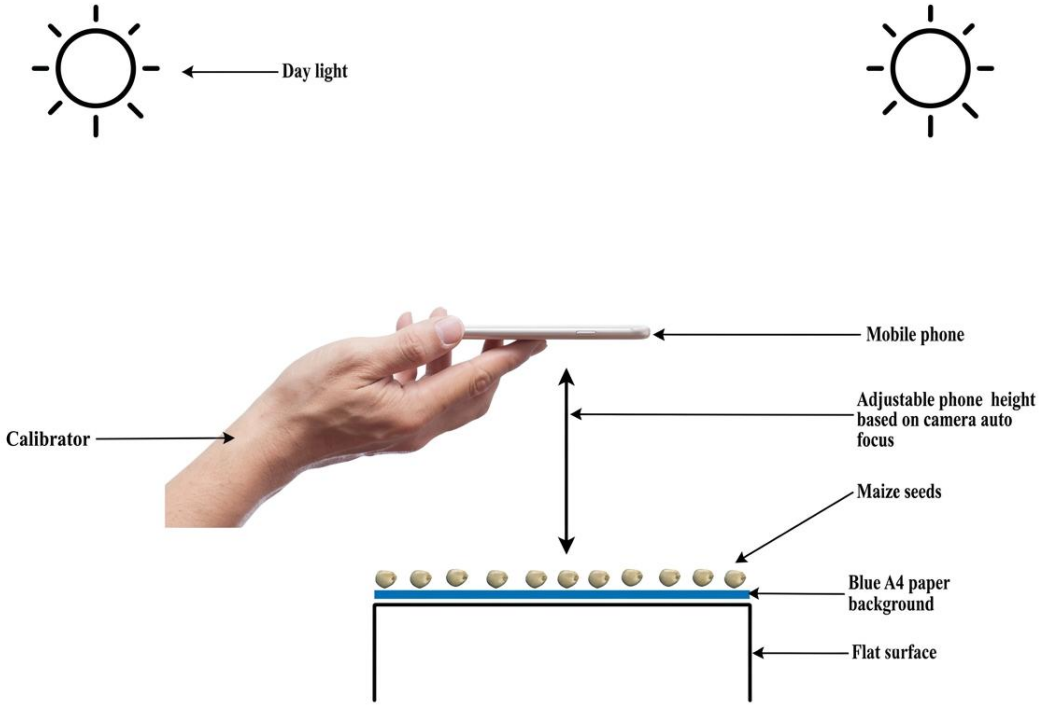


Figure 2. Image acquisition setup

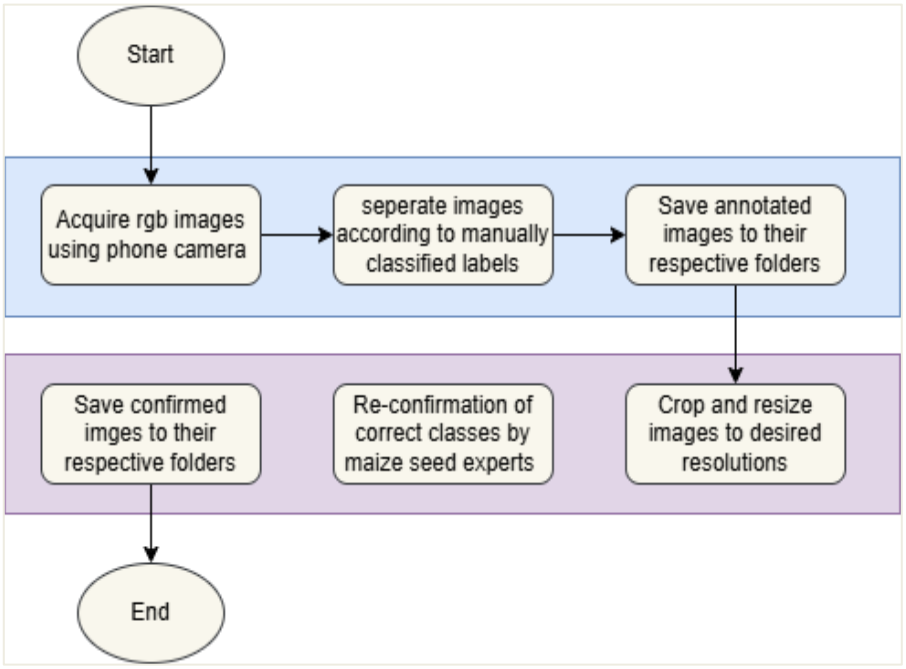


Figure 3. Dataset curation pipeline

2.2.2. Dataset description and statistics

The initial dataset comprised 3,202 images distributed across three varieties: 1,328 Bhihilifa, 850 Sanzal Sima, and 1,024 Wang Dataa. Table 1 summarizes the distribution before and after augmentation. For the baseline CNN and individual transfer learning models, an 80/20 stratified split was used to separate training and test data, with an additional 20% of the training set reserved for validation. The class distribution across these splits is shown in Figure 4.

For final ensemble evaluation, five-fold cross-validation was conducted using StratifiedKFold (n_splits=5, shuffle=True, random_state=42) to ensure consistent class proportions across all folds. In each fold, the ensemble was trained from scratch on the training portion of that fold and evaluated on the held-out test portion. This procedure ensures that no test data was seen during training in any fold, providing an unbiased estimate of generalization performance. Macro-averaged precision, recall, F1-score, and accuracy were computed per fold, with final performance reported as mean ± standard deviation across all five folds.

Table 1. Dataset description before and after augmentation

Maize Variety	Initial Dataset Size	Augmented Dataset Size
Bhihilifa	1,328	6,480
Sanzal Sima	850	5,100
Wang Dataa	1,024	6,144



Figure 4. Class distribution of Bhihilifa, Sanzal Sima, and Wang Dataa images across training, validation, and test splits, confirming consistent class proportions following stratified sampling

2.2.3. Data augmentation

To enhance dataset diversity and reduce overfitting, standard augmentation transformations were applied using Keras's ImageDataGenerator: rotation_range=20°, width_shift_range=0.2, height_shift_range=0.2, shear_range=0.2, and channel_shift_range=20. These transformations simulate realistic variations in seed orientation, scale, and illumination encountered during practical seed handling. After augmentation, the dataset expanded as indicated in Table 1, providing sufficient variability for robust model training.

2.2.4. Data preprocessing

All image file paths and labels were converted to NumPy arrays to streamline manipulation. An ImageDataGenerator pipeline normalized pixel values by scaling 1/255 to the [0, 1] range, stabilizing gradient updates during backpropagation. Class labels were one-hot encoded to be compatible with the categorical cross-entropy loss function. Images were resized to 224×224×3 (RGB) pixels as required by the pre-trained architectures.

2.3. Evaluation metrics

The models in this work are evaluated using four widely adopted performance indicators: Accuracy (1), Precision (2), Recall (3), and F1-score (4). These metrics collectively measure the correctness, sensitivity, and balance of the classification performance. The evaluation was conducted using 5-fold cross-validation, and the reported results represent the mean ± standard deviation across folds to ensure statistical reliability and robustness of the ensemble model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 * \frac{P*R}{P+R} \quad (4)$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

2.4. Model development

2.4.1. Baseline model

A baseline CNN, as shown in Figure 5, was developed to provide a performance benchmark for comparison with the transfer learning and ensemble approaches. The architecture consists of two convolutional blocks, each containing a Conv2D layer (3×3 kernels, ReLU activation), a MaxPooling2D layer for spatial down-sampling, and a Dropout layer to mitigate overfitting. The first block uses 32 filters and the second uses 64. After the convolutional blocks, a global average pooling (GAP) layer reduces the spatial dimensions, followed by a Dense layer (128 units, ReLU), a Dropout layer, and a softmax output layer with 3 units corresponding to the 3 maize varieties. The model was compiled with the Adam optimizer (learning rate=0.001) and categorical cross-entropy loss. Training used batch size 5, 50 epochs, and early stopping with patience=5. The baseline model served three purposes: (i) establishing a performance lower bound; (ii) validating dataset quality and learnability; and (iii) guiding subsequent architectural and hyperparameter decisions.

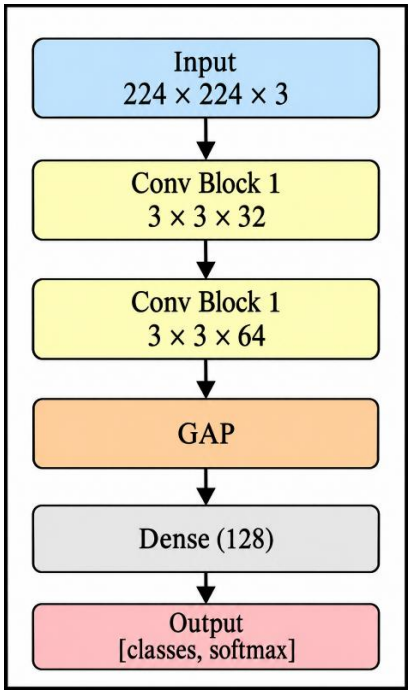


Figure 5. Baseline CNN architecture for Ghanaian maize seed variety classification

2.4.2. Pre-trained models

Access to large, well-annotated agricultural datasets is limited in many African contexts, making transfer learning a highly suitable strategy. Transfer learning enables models pre-trained on ImageNet [21] to adapt their learned feature representations to smaller, domain-specific datasets such as maize seed images, substantially reducing training time, preventing overfitting, and improving generalization [22].

Three pre-trained architectures were selected for their complementary strengths in visual feature extraction: (i) MobileNetV2 [23], chosen for its suitability for mobile and embedded systems via inverted residuals and linear bottlenecks; (ii) DenseNet169 [17], selected for its dense connectivity that encourages feature reuse and efficient gradient flow; and (iii) Xception [18], chosen for its depthwise separable convolutions that improve representational capacity while reducing computational cost. All three have been extensively validated on seed, crop, and plant classification tasks [1], [8], [22].

Fine-tuning was performed by freezing the convolutional base of each model to preserve learned low-level features, and appending task-specific layers: a GAP layer, a Dense layer (128 units, ReLU), a Dropout layer (rate=0.2), and a three-class softmax output layer. Training hyperparameters were consistent across all models: batch size=5, initial learning rate=0.001, maximum 50 epochs, early stopping (patience=5), and ReduceLROnPlateau scheduling (factor=0.2, minimum learning rate=1×10^{-4}) to adapt the learning rate when validation loss plateaued. The fine-tuning process is illustrated in Figure 6.

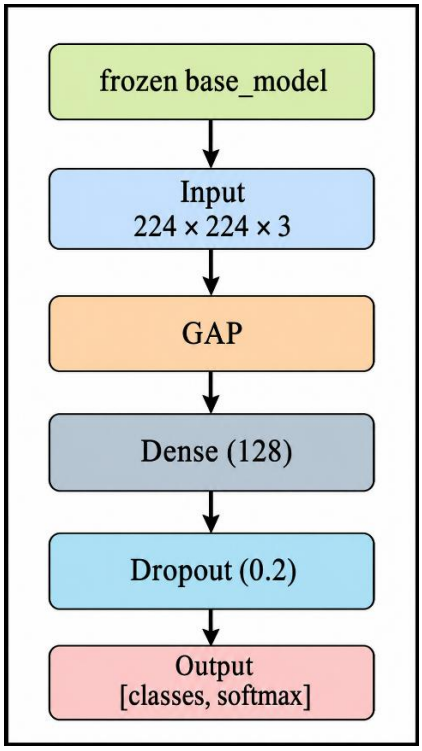


Figure 6. Fine-tuning pipeline for the pre-trained transfer learning architectures (DenseNet169, MobileNetV2, and Xception)

2.5. Proposed ensemble approach

An ensemble learning strategy was implemented to combine the predictive strengths of multiple fine-tuned models. The stacking technique aggregated the probabilistic outputs from the two best-performing individual models, DenseNet169 and Xception, into a meta-model that generates the final prediction as shown in Figure 7. This approach enhances accuracy and robustness by leveraging the complementary learning patterns of the two architectures.

Each fine-tuned model produced a probability distribution over the three classes. These outputs were concatenated to form a single feature vector, which was then passed to a meta-classifier consisting of a Dense layer (ReLU activation), a Dropout layer (rate=0.2), and a softmax output layer.

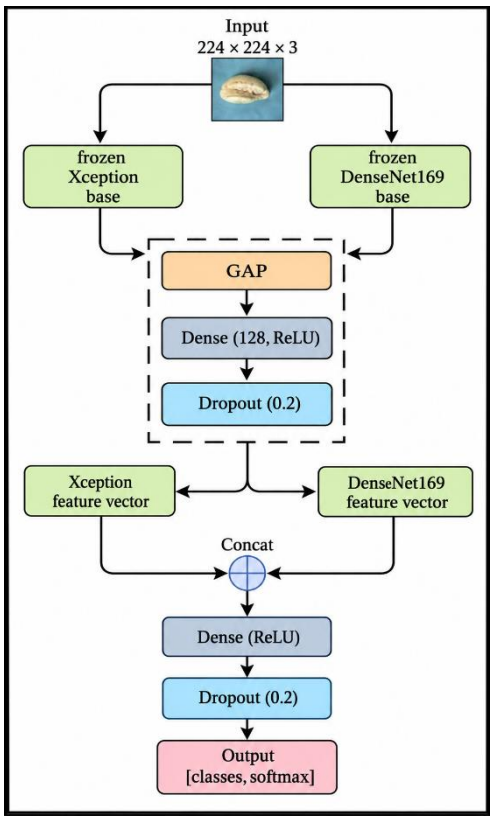


Figure 7. Proposed ensemble architecture combining DenseNet169 and Xception via feature concatenation and a stacked meta-classifier

The ensemble process is formalized as (5)-(10):

$$F_{Xception} = \mathcal{G}(B_{Xception}(\mathrm{I})) \quad (5)$$

$$F_{DenseNet169} = \mathcal{G}(B_{DenseNet169}(\mathrm{I})) \quad (6)$$

$$F_{concat} = \mathrm{concat}(F_{Xception}, F_{DenseNet169}) \quad (7)$$

$$D = \mathrm{ReLU}(D_d F_{concat}, b_d) \quad (8)$$

$$F_{concat} = Dropout\ (D, rate = 0.2) \quad (9)$$

$$P = \mathrm{softmax}(W_o D_{drop}, b_o) \quad (10)$$

Where,
I: Input tensor with shape (224, 224, 3).
$B_{Xception}$: The base model Xception without the top layers.
$B_{DenseNet169}$: The base model DenseNet169 without the top layers.
$\mathcal{G}$: The GAP operation.
$F_{Xception}$: Output feature map from the Xception base model after GAP.
$F_{DenseNet169}$: Output feature map from the DenseNet169 base model after GAP.
F_{concat}: Concatenated feature vector from all three base models.
W_d: Weight matrix of the dense layer with ReLU activation., b_d: Bias vector of the dense layer with ReLU activation.
D: Output of the dense layer with ReLU activation., D_{drop}: Output after applying dropout to D.
W_o: Weight matrix of the final dense layer (output layer) with softmax activation.
b_o: Bias vector of the final dense layer (output layer) with softmax activation.
P: Final class probabilities after applying the softmax function.

2.5.1. Ensemble candidate selection (Ablation)

An ablation study was conducted to identify the optimal ensemble configuration. All pairwise and triple combinations of the three fine-tuned models were evaluated on the test set. As summarized in Table 2, the Xception + DenseNet169 (X+D) combination achieved the best performance across all metrics (accuracy, precision, recall, F1-score = 0.97) with a moderate trainable parameter footprint of 3.18 MB, justifying its selection for the proposed ensemble.

Table 2. Ablation study: performance comparison of ensemble configurations on test data

Models	Accuracy	Precision	Recall	F1-score	Trainable Params (MB)
D+M	0.95	0.95	0.95	0.95	2.52
X+M	0.95	0.95	0.95	0.95	2.85
X+D	0.97	0.97	0.97	0.97	3.18
X+D+M	0.96	0.96	0.96	0.96	4.27

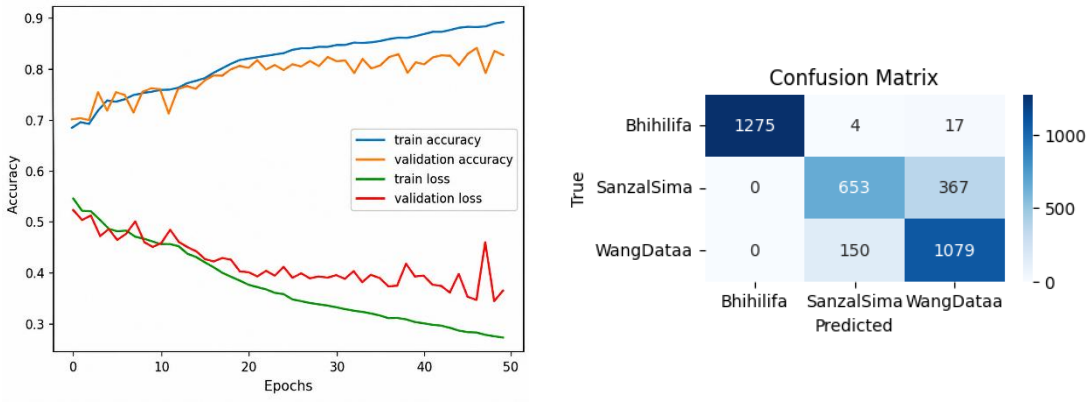
3. RESULTS AND DISCUSSION

This section presents and interprets the experimental results across four sub-studies: (3.1) baseline CNN performance, (3.2) comparative evaluation of individual pre-trained models and the ensemble, (3.3) class-wise prediction analysis and statistical validation, and (3.4) explainability visualization. Each sub-section contextualizes findings relative to the study's objectives and prior literature.

3.1. Baseline results

Figure 8 presents the baseline CNN's training curves Figure 8(a) and confusion matrix on the test set Figure 8(b). The model was trained for 50 epochs and achieved a training accuracy of approximately 0.90 and a validation accuracy of approximately 0.85. Both training and validation losses decreased consistently, with training loss converging to approximately 0.25. The close alignment of the accuracy and loss curves indicates adequate generalization without signs of severe overfitting.

The baseline CNN achieved an accuracy of 85% on the test set. Table 3 shows that Bhihilifa was classified near-perfectly (precision=1.00, recall=0.98, F1=0.99), whereas Sanzal Sima and Wang Dataa achieved substantially lower F1-scores of 0.72 and 0.77, respectively. The confusion matrix shows that 653 Sanzal Sima images were correctly classified, 367 were misclassified as Wang Dataa, 1,079 Wang Dataa images were correctly identified, and 150 were misclassified as Sanzal Sima. This pattern indicates that the two varieties share overlapping morphological characteristics that challenge shallow feature extractors. While the baseline demonstrated that meaningful patterns exist in the dataset, its limited discriminative capacity motivated the use of deeper architectures.



(a) (b)

Figure 8. Baseline training curve (a) and confusion matrix and (b) on test set

3.2. Comparative study of deep learning approaches

Table 3 and Figure 9 summarize the performance of the four deep learning models, DenseNet169, MobileNetV2, Xception, and the proposed ensemble, on the test set. Figure 9 shows the training curves for each model, and Figure 10 shows their corresponding confusion matrices.

Among the individual pre-trained architectures, DenseNet169 achieved the highest performance at 94% across all metrics, followed by Xception at 93% and MobileNetV2 at 92%. All three significantly outperformed the baseline CNN (85%), confirming the benefit of ImageNet-pretrained feature representations for domain-specific classification tasks.

DenseNet169 achieved the most stable and consistent training behavior, with training and validation accuracies converging near 0.94 and steadily decreasing losses (Figure 9(a)). Its confusion matrix (Figure 10(a)) shows 1,290 correctly classified Bhihilifa, 918 Sanzal Sima, and 1,128 Wang Dataa samples, reflecting strong discriminative ability due to its dense connectivity and feature reuse mechanism. MobileNetV2 reached a final accuracy of 0.92 after 25 epochs (Figure 9(b)), with more misclassifications between Sanzal Sima and Wang Dataa (Figure 10(b)), consistent with its lightweight design, which prioritizes computational efficiency over representational depth. Xception attained 0.93 after 20 epochs (Figure 9(c)), achieving improved class separation over MobileNetV2 through its depthwise separable convolutions (Figure 10(c)).

The proposed ensemble further improved performance to 0.97 across all metrics (Figure 9(d), Figure 10(d)). This improvement reflects the principle of ensemble diversity: DenseNet169's dense feature propagation and Xception's efficient spatial separation capture complementary discriminative patterns, and their combination reduces individual model biases, particularly for the Sanzal Sima–Wang Dataa confusion that persisted in single models. The ensemble achieved 1,291 correct Bhihilifa, 953 Sanzal Sima, and 1,195 Wang Dataa classifications, with very few errors across classes.

Table 3. Comparison of DL models using weighted-average Accuracy, Precision, Recall, and F1-Score on test data

No	Model	Accuracy	Precision	Recall	F1-score
1	MobileNetv2	92	92	92	92
2	Xception	93	93	93	93
3	DenseNet169	94	94	94	94
4	Ensemble	97	97	97	97

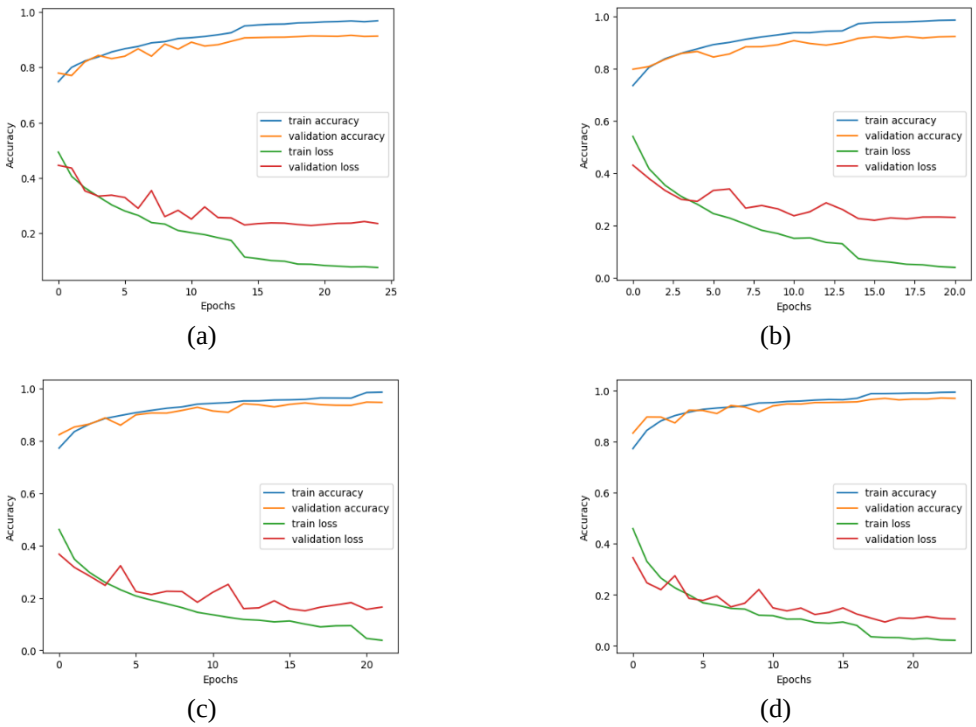


Figure 9. Training insights of (a) MobileNetv2, (b)Xception, (c) DenseNet169, and (d) ensemble model

3.3. Class-wise predictions and statistical validation

Per-class classification report for all models, as shown in Table 4, indicates that the progressive improvement across architectures is consistent and interpretable. The baseline CNN achieved near-perfect

Bhihilifa performance (F1=0.99) but struggled substantially with Sanzal Sima (F1=0.72) and Wang Dataa (F1=0.77), confirming that shallow feature extraction is insufficient for distinguishing visually similar varieties.

The ensemble's per-class results confirm balanced and robust performance: perfect scores for Bhihilifa (precision, recall, F1=1.00), and high F1-scores of 0.95 and 0.96 for Sanzal Sima and Wang Dataa, respectively. This represents the greatest improvement over the baseline for the two harder classes, confirming that the ensemble effectively exploits complementary feature representations to resolve inter-class visual ambiguity.

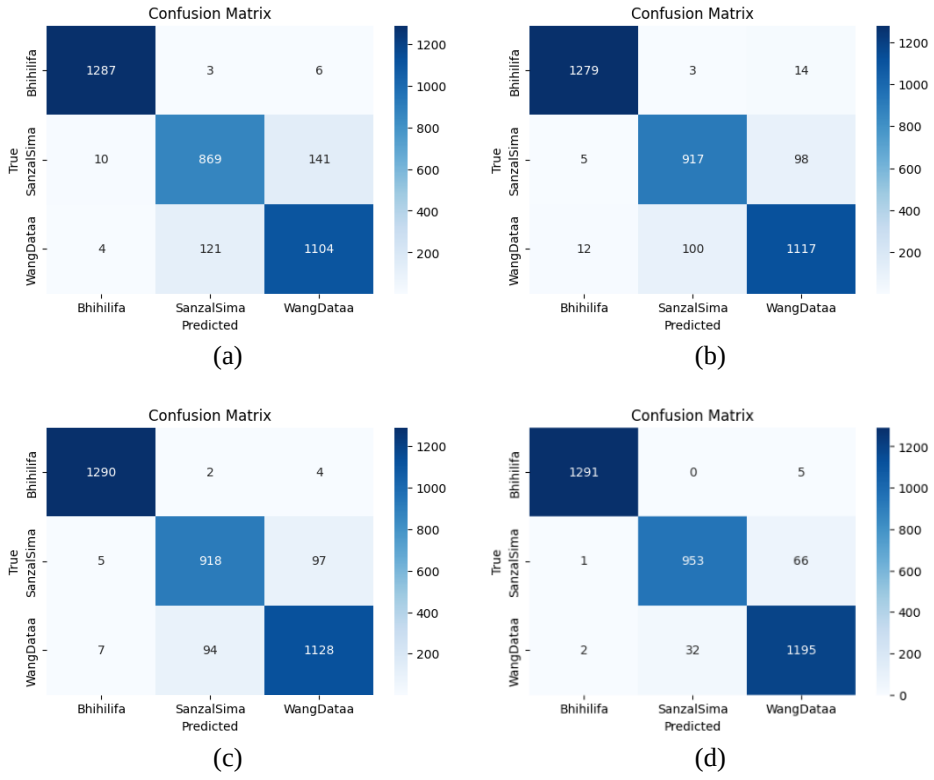


Figure 10. Confusion matrix of, (a) MobileNetv2, (b)Xception, (c) DenseNet169, and (d) ensemble model on test data

Table 4. Per-class precision, recall, and F1-score for all models on test data

Model	Classes	Precision	Recall	F1-score
Baseline	Bhihilifa	0.99	0.98	0.98
	SanzalSima	0.73	0.72	0.72
	WangDataa	0.76	0.78	0.77
MobileNetV2	Bhihilifa	0.99	0.99	0.99
	SanzalSima	0.88	0.85	0.86
	WangDataa	0.88	0.90	0.89
Xception	Bhihilifa	0.99	0.99	0.99
	SanzalSima	0.90	0.90	0.90
	WangDataa	0.91	0.91	0.91
DenseNet-169	Bhihilifa	0.99	1.00	0.99
	SanzalSima	0.91	0.90	0.90
	WangDataa	0.92	0.92	0.92
Ensemble	Bhihilifa	1.00	1.00	1.00
	SanzalSima	0.97	0.93	0.95
	WangDataa	0.94	0.97	0.96

Table 5 reports the five-fold cross-validation results for the proposed ensemble. The ensemble achieved a mean accuracy of 98.89±0.17%, precision of 98.86±0.18%, recall of 98.80±0.17%, and F1-score of 98.83±0.17%. The consistently low standard deviations across all folds confirm that the model generalizes well to unseen data without overfitting to specific training subsets. Figure 11 illustrates the confusion

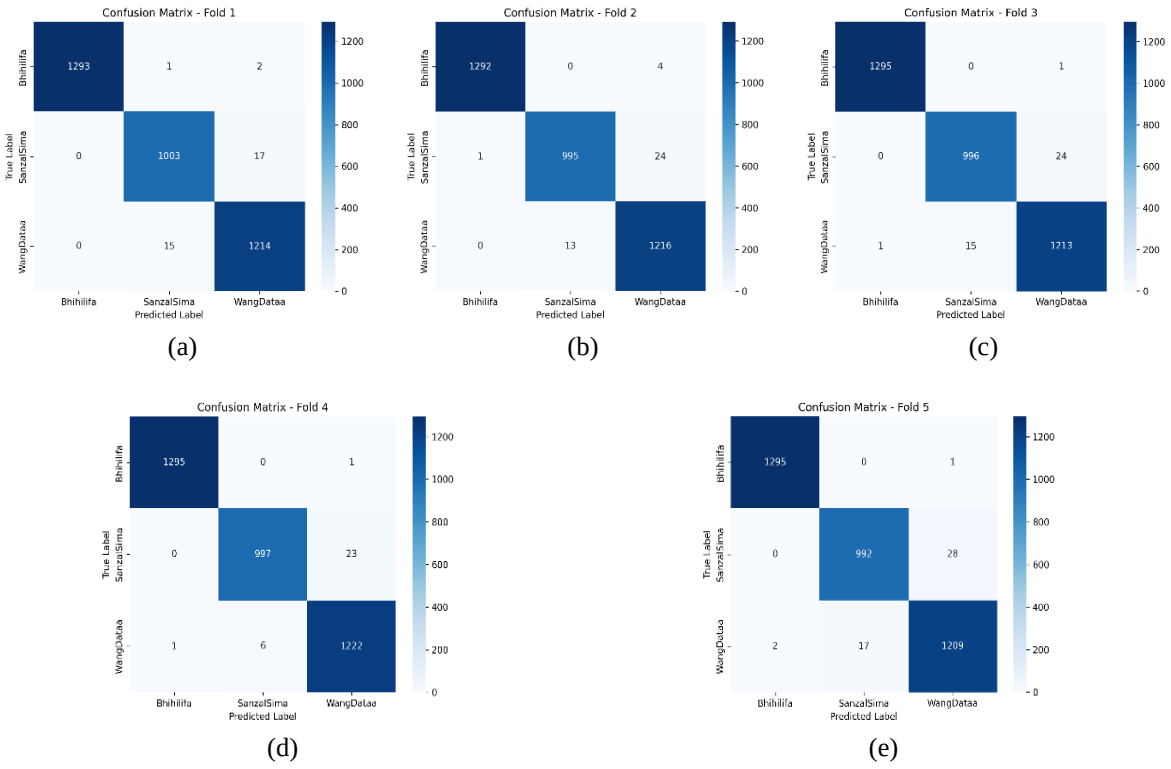
matrices for each fold, Figures 11(a) Fold 1, Figure 11(b) Fold 2, Figure 11(c) Fold 3, and Figure 11(d) Fold 4 demonstrating consistent classification accuracy across Bhihilifa, Sanzal Sima, and Wang Dataa.

Table 5. Five-fold cross-validation results for the proposed ensemble

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	99.01	98.96	98.96	98.96
2	98.82	98.79	98.73	98.76
3	98.84	98.81	98.76	98.78
4	99.13	99.13	99.03	99.08
5	98.65	98.61	98.54	98.57
Mean ± SD	98.89 ± 0.17	98.86 ± 0.18	98.80 ± 0.17	98.83 ± 0.17

3.4. Comparison with state-of-the-art methods

Table 6 compares the proposed ensemble with state-of-the-art maize seed classification methods. The comparison is organized by imaging modality and methodology to enable systematic interpretation. The proposed ensemble achieves 99.13% accuracy on the best single-fold run, matching or exceeding all competing methods. Several observations are noteworthy. First, while methods such as P-ResNet [15] and VGG-16 [14] report comparable accuracy, they were evaluated on non-Ghanaian, non-field-condition datasets and do not address model interpretability [14]. The practical barrier of accessibility is also overlooked in these studies: both rely on controlled laboratory imaging setups. Second, the CNN-LSTM approach [13] employs HSI, a far more expensive and operationally complex modality, yet achieves lower accuracy (95%) than the proposed ensemble. This highlights the key advantage of the proposed approach: achieving superior accuracy using only a commodity smartphone camera. Third, the proposed framework combines competitive accuracy with Grad-CAM- and saliency-map explainability, enabling end users such as farmers and quality-assurance personnel to understand and verify model decisions. While [15] and [24] also employs Grad-CAM-based explainability but does not combine it with smartphone-acquired RGB imaging. This combination of performance, accessibility, and transparency positions the proposed system as particularly well-suited for deployment in low-resource agricultural settings.



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

Figure 11. Confusion matrix of the ensemble model derived from 5-fold cross-validation (a) Fold 1, (b) Fold 2, (c) Fold 3, (d) Fold 4, and (e) Fold 5, illustrating consistent correct predictions across the three maize seed varieties (Bhihilifa, Sanzal Sima, and Wang Dataa)

Table 6. Performance comparison of the proposed ensemble with state-of-the-art maize seed classification methods. P = Precision, R = Recall, F = F1-Score, A = Accuracy; "-" indicates metric not reported in the original study, mAP = mean average precision

Methods	Modality	P (%)	R (%)	F (%)	A (%)
CNN-LSTM [13]	Hyperspectral	-	-	-	>95.00
VGG-16 [14]	RGB	-	-	-	>99.00
GoogleNet [25]	RGB	-	-	-	95
P-ResNet [15]	RGB	99.78	99.55	99.71	99.70
CNN [24]	RGB	96.74	96.67	-	96.67
E-YOLOv8 [26]	RGB	-	-	-	mAP=96.2
Ensemble (Proposed)	RGB (smartphone)	99.13	99.03	99.08	99.13

3.5. Model interpretability via explainable AI

To assess the model's decision-making process, gradient-weighted class activation mapping (Grad-CAM) and saliency maps were used to visualize the image regions that most strongly influenced the classification decisions. A combined attention map was generated by averaging the activations from the final convolutional layers of DenseNet169 and Xception. Two randomly selected test images per class (Bhihilifa, Sanzal Sima, Wang Dataa) were analyzed to ensure representative visualization across classes rather than isolated examples, as shown in Figure 12.

The Grad-CAM visualizations in Figure 12 indicate that for Bhihilifa, the model attended primarily to the seed's apex and surface grooves (highlighted in yellow and red). For Sanzal Sima and Wang Dataa, the model focused on kernel edges and surface texture patterns, subtle morphological features that are difficult to distinguish visually but correspond to known botanical differences between these varieties. The saliency maps confirmed these findings, showing strong pixel-level activations around morphologically distinctive regions. These results validate that the model learned biologically meaningful features rather than spurious correlations, supporting its trustworthiness for agricultural deployment.

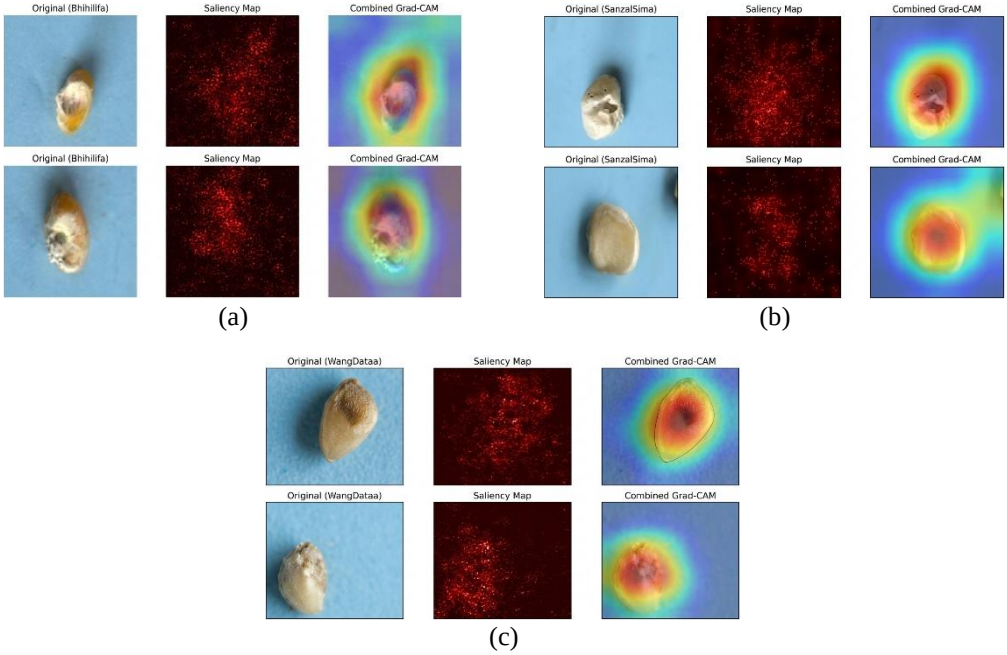


Figure 12. Visualization of model explainability using saliency maps and combined grad-CAM for the ensemble model across, (a) Bhihilifa, (b) SanzalSima, and (c) WangDataa varieties

4. CONCLUSION

This study presented a comprehensive, interpretable, and resource-efficient system for classifying Ghanaian maize seed varieties using low-cost RGB imaging and an ensemble deep learning framework combining DenseNet169 and Xception. The proposed system achieved a mean accuracy of 98.89±0.17%, precision of 98.86±0.18%, recall of 98.80±0.17%, and F1-score of 98.83±0.17% across five-fold cross-validation, outperforming all individual models and matching or exceeding competing methods from the literature, including those using HSI. Explainable AI techniques (Grad-CAM and saliency maps)

confirmed that the model's decisions are grounded in biologically meaningful morphological features, supporting its reliability for practical deployment.

Several limitations of this study merit acknowledgment. First, the dataset is limited to three maize varieties cultivated in northern Ghana; performance on other regional varieties or across a broader set of Ghanaian cultivars has not been evaluated. Second, images were captured under controlled conditions (daylight, uniform background); variability in illumination, background clutter, or camera distance in fully uncontrolled field environments may degrade performance. Third, all experiments were conducted on CPU-only hardware, which limits training speed; GPU-accelerated environments would considerably reduce training time. Fourth, the ensemble's 3.18 MB trainable parameter footprint, while moderate, should be evaluated further for deployment on resource-constrained edge devices.

Several promising directions for future work emerge from these findings. Expanding the dataset to include additional Ghanaian and West African maize varieties, as well as images captured under diverse field conditions, would improve generalization and real-world robustness. Investigating model compression techniques, such as quantization and knowledge distillation, would support deployment on mobile and edge devices for real-time field use. Extending the framework to other staple crops cultivated in sub-Saharan Africa (e.g., sorghum, millet, cowpea) would broaden its agricultural impact. Finally, integrating the system into a farmer-facing mobile application with an intuitive, explainability-driven interface would translate the technical contributions of this work into directly actionable tools for smallholder farmers and seed quality-assurance personnel.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Emmanuel Asante	✓	✓		✓	✓	✓		✓	✓	✓	✓			
Obed Appiah		✓				✓		✓	✓	✓		✓		
Ezekiel Mensah Martey				✓			✓			✓	✓	✓	✓	
Emmanuel Effah		✓		✓		✓				✓		✓		
Eric Opoku	✓	✓								✓	✓		✓	
Ernest Kwame Ampomah				✓			✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Mendeley Data at http://doi.org/10.17632/nfhbbjx6b7.1.

REFERENCES

[1] R. Ma *et al.*, "Identification of maize seed varieties using MobileNetV2 with improved attention mechanism CBAM," *Agriculture (Switzerland)*, vol. 13, no. 1, 2023, doi: 10.3390/agriculture13010011.

[2] M. Aqil *et al.*, "Integration of smartphone technology for maize recognition," *IOP Conference Series: Earth and Environmental Science*, vol. 911, no. 1, 2021, doi: 10.1088/1755-1315/911/1/012037.

[3] S. Gebeyehu and Z. S. Shibeshi, "Maize seed variety identification model using image processing and deep learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, pp. 990–998, 2024, doi: 10.11591/ijeecs.v33.i2.pp990-998.

[4] L. Zhang, S. Zhang, J. Liu, Y. Wei, D. An, and J. Wu, "Maize seed variety identification using hyperspectral imaging and self-supervised learning: a two-stage training approach without spectral preprocessing," *Expert Systems with Applications*, vol. 238, 2024, doi: 10.1016/j.eswa.2023.122113.
[5] C. Xia, S. Yang, M. Huang, Q. Zhu, Y. Guo, and J. Qin, "Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis," *Infrared Physics and Technology*, vol. 103, 2019, doi: 10.1016/j.infrared.2019.103077.
[6] Q. Liu, Z. Wang, Y. Long, C. Zhang, S. Fan, and W. Huang, "Variety classification of coated maize seeds based on Raman hyperspectral imaging," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 270, 2022, doi: 10.1016/j.saa.2021.120772.
[7] Q. Zhou, W. Huang, X. Tian, Y. Yang, and D. Liang, "Identification of the variety of maize seeds based on hyperspectral images coupled with convolutional neural networks and subregional voting," *Journal of the Science of Food and Agriculture*, vol. 101, no. 11, pp. 4532–4542, 2021, doi: 10.1002/jsfa.11095.
[8] C. Bi, N. Hu, Y. Zou, S. Zhang, S. Xu, and H. Yu, "Development of deep learning methodology for maize seed variety recognition based on improved swin transformer," *Agronomy*, vol. 12, no. 8, 2022, doi: 10.3390/agronomy12081843.
[9] B. Veeramani, J. W. Raymond, and P. Chanda, "DeepSort: deep convolutional networks for sorting haploid maize seeds," *BMC Bioinformatics*, vol. 19, 2018, doi: 10.1186/s12859-018-2267-2.
[10] P. Xu *et al.*, "Identification of defective maize seeds using hyperspectral imaging combined with deep learning," *Foods*, vol. 12, no. 1, 2023, doi: 10.3390/foods12010144.
[11] H. Yang, C. Wang, H. Zhang, Y. Zhou, and B. Luo, "Recognition of maize seed varieties based on hyperspectral imaging technology and integrated learning algorithms," *PeerJ Computer Science*, vol. 9, 2023, doi: 10.7717/peerj-cs.1354.
[12] Y. Altuntaş, Z. Cömert, and A. F. Kocamaz, "Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach," *Computers and Electronics in Agriculture*, vol. 163, 2019, doi: 10.1016/j.compag.2019.104874.
[13] Y. Wang and S. Song, "Variety identification of sweet maize seeds based on hyperspectral imaging combined with deep learning," *Infrared Physics and Technology*, vol. 130, 2023, doi: 10.1016/j.infrared.2023.104611.
[14] K. Tu *et al.*, "A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968′ using machine vision combined with deep learning," *Computers and Electronics in Agriculture*, vol. 182, 2021, doi: 10.1016/j.compag.2021.106002.
[15] P. Xu, Q. Tan, Y. Zhang, X. Zha, S. Yang, and R. Yang, "Research on maize seed classification and recognition based on machine vision and deep learning," *Agriculture (Switzerland)*, vol. 12, no. 2, 2022, doi: 10.3390/agriculture12020232.
[16] A. Ali *et al.*, "Machine learning approach for the classification of corn seed using hybrid features," *International Journal of Food Properties*, vol. 23, no. 1, pp. 1097–1111, 2020, doi: 10.1080/10942912.2020.1778724.
[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.
[18] F. Chollet, "Xception: deep learning with depthwise separable convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800–1807, 2017, doi: 10.1109/CVPR.2017.195.
[19] 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," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 618–626, 2017, doi: 10.1109/ICCV.2017.74.
[20] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, 2014.
[21] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 248–255, 2009, doi: 10.1109/CVPR.2009.5206848.
[22] Y. Gulzar, Z. Ünal, S. Ayoub, and F. A. Reegu, "Exploring transfer learning for enhanced seed classification: pre-trained xception model," *Lecture Notes in Civil Engineering*, vol. 458 LNCE, pp. 137–147, 2024, doi: 10.1007/978-3-031-51579-8_14.
[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
[24] O. Akbarpour, S. Akbari, D. Fanourakis, and A. Taheri-Garavand, "Deep learning-based seed variety classification: a case study in maize," *BMC Plant Biology*, vol. 26, no. 1, 2026, doi: 10.1186/s12870-025-07919-3.
[25] S. Huang, X. Fan, L. Sun, Y. Shen, and X. Suo, "Research on classification method of maize seed defect based on machine vision," *Journal of Sensors*, vol. 2019, 2019, doi: 10.1155/2019/2716975.
[26] S. Niu *et al.*, "High accuracy identification of maize seed varieties based on a lightweight improved YOLOv8," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-00499-0.

BIOGRAPHIES OF AUTHORS



Emmanuel Asante received his Bachelor of Science degree and a Master of Philosophy degree in Computer Science from the University of Energy and Natural Resources (UENR), Ghana, and a Postgraduate Diploma in Education (PGDE) from the University of Education, Winneba, Kumasi Campus. He is currently an Assistant Lecturer in the Department of Computing and Data Analytics at the University of Mines and Technology (UMaT), Tarkwa, Ghana, on the SRID Campus, and is pursuing a Ph.D. in Computer Science and Engineering at UMaT. His research interests include artificial intelligence, computer vision, deep learning, and data science, with particular emphasis on applications in education and agriculture. He is actively involved in projects aimed at developing smart systems to enhance efficiency and sustainability in Ghana's agricultural sector. He can be contacted at email: easante@umat.edu.gh.



Obed Appiah received his Ph.D. in Computer Science from Kwame Nkrumah University of Science and Technology (KNUST), Ghana. He is currently a Senior Lecturer at the University of Energy and Natural Resources (UENR), Sunyani, and a Research Scientist with the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL). His research interests include computer vision, machine learning, and digitalisation, and he has published extensively in reputable international journals and conference proceedings. He can be contacted at email: obed.appiah@uenr.edu.gh.



Ezekiel Mensah Martey received his Doctor of Engineering degree in Software Engineering from the University of Electronic Science and Technology of China, PR China. He is currently a Lecturer in the Department of Computing and Data Analytics at the University of Mines and Technology (UMaT), Tarkwa, Ghana, on the SRID Campus. His research interests include information retrieval, pattern recognition, computer graphics, image processing, and computer vision. He can be contacted at email: emmartey@umat.edu.gh.



Emmanuel Effah received his Ph.D. in Computer Science from Gaston Berger University in Senegal, an MSc in Information Technology from the Open University of Malaysia, and a BSc in Electrical and Electronic Engineering from the University of Mines and Technology (UMaT), Ghana. He is currently a Lecturer in the Department of Computer Science and Engineering at UMaT, Ghana. His research interests include wireless sensor networks (WSNs), Agricultural Internet of Things (Agri-IoT), IoT, smart system design technology, and machine learning. He has authored many research papers published in reputable journals and conference proceedings and has participated in several international research projects. Currently, Emmanuel is the principal investigator (PI) of the World Bank-RSIF/JIRA-funded Agri-IoT Tech Project (Grant Ref No.: RSIF/JIRA/004) and the Senior Personnel of the NSF-funded DISES Project (Award Number (FAIN): 2307870). He can be contacted at email: eeffah@umat.edu.gh.



Eric Opoku received an MSc in Data Science and Analytics from Grand Valley State University, USA, an MPhil in Computer Science from Kwame Nkrumah University of Science and Technology (KNUST), Ghana, and a BSc in Computer Science from the University of Energy and Natural Resources (UENR), Ghana. He is currently a PhD student in Data Science and Engineering at the University of Tennessee, Knoxville (Bredesen Center). His research focuses on computer vision and multimodal artificial intelligence for precision livestock monitoring, with particular emphasis on early detection of bovine respiratory disease and abnormal calf behaviors through the integration of video, acoustic, sensor, and contextual data, using resource-aware and interpretable models for real-world deployment. His work on automated cocoa bean grading received an AI Ghana Award and support to present at NeurIPS. He can be contacted at email: eopoku2@vols.utk.edu.



Ernest Kwame Ampomah received his Doctor of Philosophy degree in Software Engineering from the University of Electronic Science and Technology of China (UESTC), PR China, a Master of Philosophy degree in Computational Science and Engineering from the University of Ghana, a Bachelor of Science degree in Geomatics Engineering from Kwame Nkrumah University of Science and Technology (KNUST), and a Bachelor of Technology degree in Computer Technology from Kumasi Technical University. He also holds a Postgraduate Diploma in Education (PGDE) from Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development (AAMUSTED), Kumasi. He is currently a Lecturer in the Department of Computing and Data Analytics at the University of Mines and Technology (UMaT), Tarkwa, Ghana, on the SRID Campus. His research interests include data science, computational intelligence, artificial intelligence and deep learning, smart systems, and cybersecurity. He can be contacted at email: ekampomah@umat.edu.gh.