IJAMEC

INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 13
Issue 03

September, 2025

Research Article

A Hybrid Model Approach Based on Swin Transformer and EfficientNetV2 for Maize Variety Classification

Huseyin BULDUK [a,*], ***Kadir SABANCI*** [a]

[a] *Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman, Türkiye*

ARTICLE INFO

Article history:
Received 07 August 2025
Accepted 29 September 2025
Keywords:
Classification,
Hybrid Deep Learning Model,
Swin Transformer,
EfficientNetV2,
Maize Classification

ABSTRACT

In this study, two different deep learning-based models were proposed for the classification of the maize varieties Chulpi Cancha, Indurata, and Rugosa. In the first stage, a single model was developed using the Swin Transformer architecture with an attention mechanism. This model was then integrated with EfficientNetV2 to create a hybrid structure. The developed models were tested on a dataset consisting of 1050 images with a fixed background and high resolution. The Swin Transformer model produced successful results with 99.37% accuracy, while the hybrid model achieved 100% test accuracy, accurately classifying all samples. The findings demonstrate that the Swin Transformer and EfficientNetV2-based hybrid architectures offer high discrimination power and generalization capacity in image-based classification of maize varieties. Future studies are recommended to conduct additional tests using images taken under different environmental conditions and larger datasets encompassing a wider range of varieties.

1. Introduction

In recent years, AI in agriculture has made significant strides, particularly in early plant disease detection, yield prediction, and smart farming applications. For example, the review "Revolutionizing Agriculture" notes that image processing, machine learning (ML), and deep learning (DL) models are being used to detect leaf diseases in vegetables such as tomatoes, eggplants, and cucumbers, while realistic field applications using IoT sensors are becoming increasingly widespread [1]. Corn (Zea mays) is an important cereal grain species widely cultivated and utilized in various fields worldwide [2]. Besides being a staple food source in human nutrition, it is also widely used as animal feed and a raw material for many industrial products [3]. As the third most produced agricultural product worldwide after wheat and rice, corn is of strategic importance both in agricultural production processes and in the economy [3], [4], [5]. Accurately classifying corn and performing quality analyzes are crucial for the product's market value and commercialization. Depending on different environmental conditions, corn varieties with different characteristics are grown in various geographical regions [4], [6]. Therefore, classification processes are essential for critical agricultural applications such as crop monitoring, yield estimation, and determination of seed purity. Furthermore, such classification systems support the production of high-quality products and contribute to the more effective management of processes such as storage, processing, and pricing in the agricultural supply chain [3], [5], [7]. In a study conducted by Yang et al. [8], it was aimed to classify corn varieties using morphological and texture-based features obtained from visible and near-infrared hyperspectral images. In this context, support vector machines (SVM) and partial least squares-discriminant analysis (PLS-DA) models were applied, and over 96.3% classification accuracy was achieved with the SVM method. In another study conducted by Zhao et al. [9], classification was proposed by extracting morphological features such as colour, texture, and shape from images of corn seeds. In this direction, genetic algorithms and support vector machines were used

* Corresponding author. E-mail address: *huseyinbulduk@kmu.edu.tr*
DOI: 10.58190/ijamec.2025.132

together, and an accuracy rate of 94.4% was achieved. In the study by Zhang et al. [10], a hyperspectral imaging method was utilized for the fast and non-destructive classification of four different corn seed varieties. Images obtained in the wavelength range of 450–979 nm were analyzed with different models such as deep convolutional neural network (DCNN), k nearest neighbors (k-NN), and SVM. The results obtained reveal that the DCNN model outperforms other methods and achieves a classification accuracy of 94.4%.

In a study conducted by Velesaca et al. [11], corn kernels were classified into three categories: "good," "dirty," and "spoiled" using the Mask R-CNN algorithm. In this context, the Mask R-CNN method was compared with the VGG16 and ResNet50 models, and the average accuracy of the proposed deep learning-based approach was reported as 95.6%. Ni et al. [12] applied SVM, artificial neural network (ANN), principal component analysis (PCA), and ResNet architectures to analyze corn images obtained using a dual-camera system. According to the comparison results, the ResNet model showed the best performance with an accuracy rate of 98.2%. In another study developed by Javanmardi et al. [13], a deep learning-based CNN approach was adopted to classify corn varieties. The features obtained with this model are as follows: The samples were classified using various machine learning algorithms such as cubic SVM, quadratic SVM, weighted k-nearest neighbors (kNN), boosted tree, bagged tree, and linear discriminant analysis (LDA). The findings revealed that the features extracted with deep learning provided higher accuracy compared to traditional methods. In this context, the CNN–ANN classifier stood out as the most successful method with an accuracy rate of 98.1%. In the study conducted by Isik et al. [14], it was emphasized that seed purity is one of the main elements that increases agricultural yield, and in this direction, it was stated that the accurate classification of corn (maize) varieties is an important problem. Six different classification models were developed to solve this problem. A special dataset consisting of a total of 14,469 images belonging to four classes was created for training the models. The images cover four different corn varieties: BT6470, CALIPOS, ES_ARMANDI, and HIVA, provided by BIOTEK. In the classification process, AlexNet and ResNet50 architectures were used with the transfer learning method. To improve model performance, these architectures were hybridized with Directional Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) algorithms. According to the results, the highest classification accuracy of 98.10% was obtained from the ResNet50 + BiLSTM hybrid model.

In a study by Kiratiratanapruk et al. [15], a system for the detection of defects in corn seeds was developed by utilizing machine vision techniques. For this purpose, seeds were classified using a specially designed imaging device. According to the obtained results, defective seeds were identified with an accuracy rate of 80.6%, while healthy seeds were identified with an accuracy rate of 95.6%. Zhao et al. [16] proposed an approach in which morphological features such as color, texture, and shape were extracted for the classification of corn seed types. In this context, the classification process was carried out by combining genetic algorithms (GA) and support vector machines (SVM), and an accuracy rate of 94.4% was achieved. In another study by Kai et al. [17], image processing techniques were used to detect corn leaf diseases. The color and texture features obtained from the images were classified using a backpropagation artificial neural network (BPNN), and an accuracy rate of over 98% was achieved. Yang et al. [18] analyzed hyperspectral image data to assess the quality of waxy maize seeds. Five morphological and eight texture features were extracted from the images, and these features were classified using support vector machines (SVM) and partial least squares discriminant analysis (PLS-DA). High classification success rates were achieved in the study, ranging from 96.3% to 98.2%. Huang et al. [19] conducted a study aimed at determining seed purity and improving yield using hyperspectral images of four maize varieties produced in different years. The classification processes were carried out with the Least Squares Support Vector Machine (LSSVM) model, and an accuracy rate of 94.4% was achieved. Williams et al. [20] classified kernels of the same maize variety into three groups based on their hardness characteristics: hard, medium, and soft. The PLS-DA method was used in the analysis of hyperspectral images, and the classification success rates ranged from 85% to 96%. Wu et al. [21] compared five different classification algorithms with image processing techniques in their study on the determination of corn grain quality. These algorithms were determined as SVM, SVM-grid search, SVM-genetic algorithm (GA), SVM-particle swarm optimization (PSO) and backpropagation artificial neural network (BPNN). The accuracy rates obtained were reported as 92.31%, 94.87%, 97.44%, 97.44% and 92.31%, respectively. Daskalov et al. [22] designed an automatic inspection system for the detection of corn seeds damaged by Fusarium Moniliforme. In the study, which was carried out using images of corn varieties produced in Bulgaria, preprocessing, feature selection and classification steps were performed, and accuracy rates ranging from 91.6% to 92.8% were achieved with SVM and k-nearest neighbors (KNN) algorithms. Li et al. [23] developed a classification approach based on computer vision and machine learning techniques to identify various types of damaged corn kernels. A total of 17 features, 12 colors and 5 shapes were extracted from the images, and classification accuracy was between 74.76% and 96.67% using these features.

In a study conducted by Lopes et al. [24], a deep learning-based computer vision system was developed for the classification of cocoa beans. This system used the ResNet18 architecture, and the model demonstrated high classification performance with an accuracy rate of 96.82%. In another study conducted by Oliveira et al. [25], computer vision techniques were used in the classification of fermented cocoa beans. In this study, the effect of the number of samples per class on the classification performance was analyzed, and an accuracy rate of 92% was achieved. Lopes et al. [26] developed a computer vision-based approach for the classification of barley flour samples. In the study, the Spatial Pyramid Partition Ensemble method was used, considering 55 image features; the classification processes achieved 75% accuracy with the k-nearest neighbors (k-NN) algorithm and 100% accuracy with the J48 algorithm. Avuçlu et al. [2] proposed a hybrid model for the classification of three different maize cultivars of the Zea mays species. In this study, 12 different morphological features of corn kernels were extracted, and these data were classified using machine learning (ML) algorithms. In the standard classification process, test accuracy rates were 96.66% for Decision Tree (DT), 97.32% for Random Forest (RF), and 96.66% for Naive Bayes (NB). The proposed hybrid model demonstrated that these rates reached 100% for all three algorithms. Statistical analyzes revealed that the overall accuracy was 97.67% for the standard classification and 100% for the hybrid model. These results demonstrate the high efficiency of the developed hybrid corn classification system. Avuçlu et al. [27] used deep learning-based ResCNN, DAG-Net, and ResNet-18 models to classify three different corn varieties: Chulpi Cancha, Indurata, and Rugosa. Classifications were performed using 1050 corn images, each with a fixed background. The images were divided into three separate datasets: normal color images (Colour Images - CI), images generated with the Canny edge detection algorithm (CEDA), and images generated with the Sobel edge detection algorithm (SEDA). In classifications performed using normal color images (CI), the DAG-Net model achieved 100% accuracy for the Indurata variety. Accuracy rates for other corn varieties and models ranged from 99.33% to 99.52%. When images generated with Canny edge detection (CEDA) were used, the DAG-Net model achieved approximately 99.90% accuracy for the Indurata variety. Other models performed above 99% for the Chulpi Cancha and Rugosa varieties. In experiments conducted with Sobel edge detection (SEDA), the DAG-Net and ResNet-18 models achieved 100% accuracy for the Indurata variety. Accuracy rates for all other models and corn variety combinations were generally recorded between 99.33% and 99.52%. One of the key findings of the study is that models trained on datasets generated with edge detection algorithms (CEDA and SEDA) took less time to train than models trained on regular color images. This significantly increases the efficiency of the classification process.

Besides corn, various agricultural products have been classified using deep learning algorithms. For example, Sönmez et al. [28] classified wheat varieties and hybrids such as Ahmetbuğdayı, Cesare, BC1F6, and BC2F5 using color features. Approximately 99% accuracy has been achieved with machine learning algorithms such as artificial neural network (ANN), support vector machines (SVM), decision tree (DT), k-nearest neighbors (k-NN), and random forest (RF). It has also been shown that different wheat hybrids can be modelled based on color. Das and Rupa [29] performed disease classification on strawberry leaf images using ResNet architectures. The ResNet-50 model achieved 88% accuracy, the ResNet-101 model 82% accuracy, and the ResNet-152 model 80% accuracy. Altan [30] classified diseases on pepper plant leaves using a Capsule Network (CapsNet) based artificial neural network model. The developed model achieved high success in detecting diseases affecting pepper production, achieving accuracies between 95% and 97%. Sun et al. [31] classified tomato leaf diseases with high accuracy by combining EfficientNetV2 and Swin Transformer architectures. The proposed model outperformed both individual models and previously proposed combined architectures with an accuracy rate of 99.70%. Li et al. [32] classified potato leaf diseases (early and late mildew) with the Swin Transformer architecture and achieved 97.7% accuracy. Zhao et al. [33] classified fungal images belonging to 114 classes with a Swin Transformer-based model and achieved an accuracy rate of 87.66%.

The most significant innovation of this work is the first application of the Swin Transformer and EfficientNetV2 hybrid architectures to the classification of corn varieties. Studies on the classification of corn seeds in the literature have largely been limited to single CNNs or classical machine learning methods. The proposed hybrid model combines a Transformer-based global attention mechanism with CNN-based optimized convolutional filters to provide a more robust feature representation. In this respect, the study surpasses existing approaches in agricultural image classification and adds value to both academic literature and agricultural applications by achieving high accuracy, particularly in determining varietal purity.

In recent years, artificial intelligence-supported methods have been increasingly applied in agriculture. Deep learning and machine learning-based approaches are effectively used in early diagnosis of plant diseases [31], [32], crop yield estimation [6], smart irrigation systems, and agricultural robotics applications. These methods allow the development of highly accurate classification and decision support systems not only in laboratory conditions but also at the field level. The application of next-generation deep learning architectures, particularly

Swin Transformer and EfficientNet, in agriculture offers significant advantages in accurate classification of crop varieties and quality control processes [27], [31]. In this context, the hybrid approach proposed in our study aligns with current trends in the literature and contributes to agricultural AI applications.

2. MATERIAL AND METHODS

2.1. Hardware and Software Environment

The model's training, testing, and evaluation processes were implemented using a two-stage system architecture. Preliminary processing, dataset management, and model design were conducted on a Windows 10 PC with an Intel Core i7-13700 processor and 32 GB of RAM. To meet the high computational power requirements, the model's training and evaluation processes were performed on the GPU-accelerated Google Colab platform using an NVIDIA A100 graphics processing unit. The Python programming language was used in the software development process, and the PyTorch library was used to create deep learning models. Visualization of the obtained results was achieved using data visualization tools such as Matplotlib and Seaborn.

2.2. Dataset

The dataset used in this study consists of color images of three different corn varieties: Chulpi Cancha, Indurata, and Rugosa. These three varieties have morphologically distinguishable characteristics and are important in industrial and commercial applications. The dataset contains a total of 1050 images, with an equal number of examples for each class (350 images/class). All images were included in the system at a fixed resolution, high quality, and in RGB format.

The dataset was divided into two groups: training (80%) and testing (20%). Basic data augmentation techniques were applied to the training data to ensure class balance and reduce the risk of overfitting. In this context, operations such as random rotation and horizontal and vertical flips were performed. Figure 1 presents sample images of Chulpi Cancha, Indurata, and Rugosa.

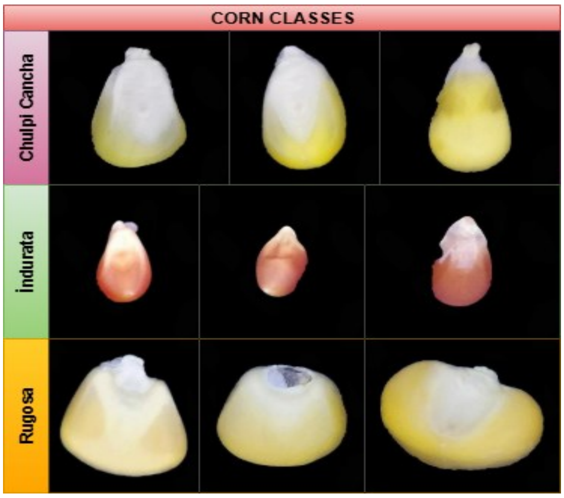

Figure 1. Corn Varieties

2.3. Deep Learning Models Used

In this study, three different deep learning-based architectures were used to classify corn varieties: Swin Transformer and Swin Transformer + EfficientNetV2 (hybrid model). The structural features of these models are summarized below.

2.3.1. Swin Transformer

Swin Transformer (Shifted Window Transformer) is a model designed to address some of the limitations of the Vision Transformer (ViT) architecture and to perform more efficiently, particularly on low-resolution images. Swin Transformer divides the input image into fixed-sized windows and applies the Multi-head Self-Attention (MSA) mechanism locally within these windows. This approach both reduces computational cost and enables more efficient learning of local visual details. Thanks to the Window-based Multi-head Self-Attention (W-MSA) and Shifted Window Multi-head Self-Attention (SW-MSA) layers within Swin Transformer, the model can effectively capture both local and global context. These features significantly increase classification accuracy, especially in datasets with limited sample sizes.

2.3.2. EfficientNetV2

EfficientNetV2 is an improved version of the EfficientNet family of architectures and stands out as a faster-training, more parameter-efficient deep convolutional neural network (CNN) architecture. This model balances depth, width, and resolution parameters based on scalability.

In this study, EfficientNetV2 is integrated with the Swin Transformer in a hybrid architecture based on the feature fusion method. This integration aims to increase classification accuracy by supporting the Swin Transformer's attention mechanism with the powerful feature extraction capabilities of EfficientNetV2.

2.3.3. Swin Transformer – EfficentNetV2 Hybrid Model

In this study, a hybrid deep learning architecture was developed to leverage the complementary feature extraction capabilities of the Swin Transformer and EfficientNetV2 models. Both models were used with pretrained versions, and feature extraction was performed using only the backbone layers, removing the final classification layers. The "swin_tiny_patch4_window7_224" architecture was chosen as the Swin Transformer component, and the head layer at the end of the model was replaced with nn.Identity(). This resulted in a 768-dimensional feature vector extracted solely from the transformer-based backbone layers. On the EfficientNetV2 side, the "efficientnetv2_rw_s" variant was used, removing the classifier layer and preserving only the backbone layers. A 1280-dimensional feature vector was extracted from this model. These two vectors obtained from both models were concatenated in a feedforward step, creating a combined feature representation with a total dimension of 2048. This combined vector was fed to a single-layer fully connected classifier (linear layer), performing a three-class classification (Chulpi Cancha, Indurata, Rugosa). Thus, the structural information captured by the Swin Transformer's global attention mechanism was complemented by low- and mid-level visual information extracted by EfficientNetV2's optimized convolutional filters, resulting in a richer feature representation. This technical structure increased the model's generalization capacity and positively contributed to classification accuracy.

2.4. Model Training and Evaluation

All models were trained using the same training and test datasets, and comparative analyzes were conducted accordingly. Hyperparameters were kept constant throughout the training process; training was conducted for 5 epochs, the mini-batch size was set to 16, and the learning rate was set to 1:4. The Adam algorithm was used in the optimization process, and a cross-entropy loss function was used appropriate to the classification task. Additionally, various data augmentation techniques were applied to the training data to improve the generalization ability of the model. In this context, the training set was diversified using Random Horizontal Flip, Random Rotation, and ColorJitter, which simulates color variations. Table 1 shows the training parameters of the models. Figure 2 shows the flowchart of the study.

Table 1. Training Hyperparameters of the Model

Hyperparameter	Swin Transformer	EfficentNetV2	Hybrid Model (Swin + EfficientNetV2
Input image size	224 × 224	224 × 224	224 × 224
Batch size	16	16	16
Epochs	5	5	5
Optimizer	Adam	Adam	Adam
Initial learning rate	1e-4	1e-4	1e-4
Learning rate schedule	StepLR (gamma=0.1, step=3)	StepLR (same)	StepLR (same)
Loss function	Cross-Entropy	Cross-Entropy	Cross-Entropy
Weight decay	1e-5	1e-5	1e-5
Data augmentation	Random Rotation, Flip, ColorJitter	Same	Same
Pretrained weights	ImageNet-1K	ImageNet-1K	Both pretrained backbones

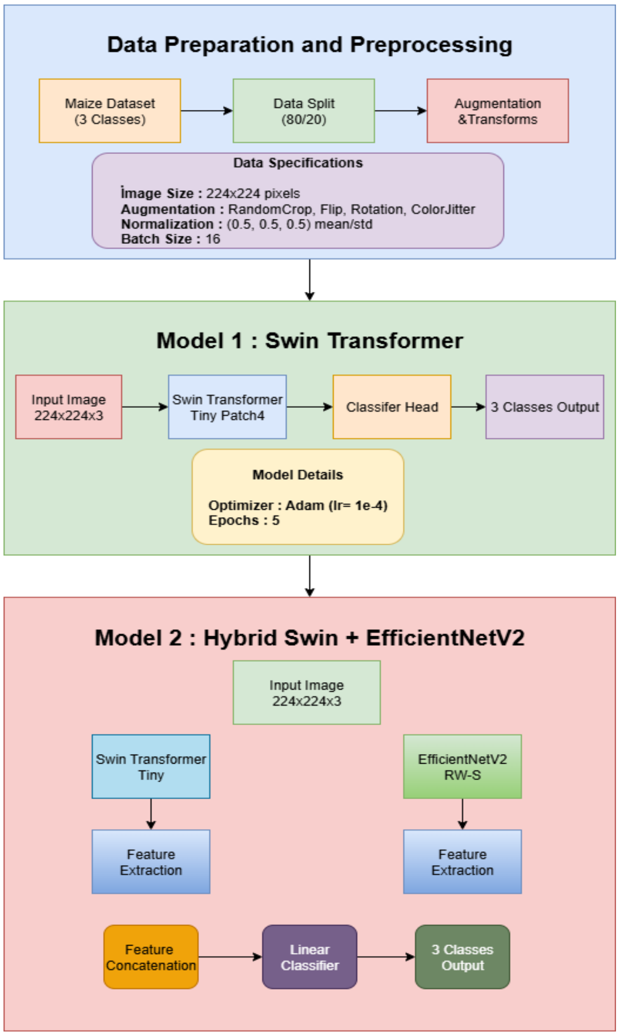

Figure 2. Flowchart of the Study

3. RESULTS AND DISCUSSION

3.1. Findings of the Swin Transformer Model

The Swin Transformer model demonstrated very successful performance in classifying the varieties Chulpi Cancha, Indurata, and Rugosa of the Zea mays species.

The 99.37% accuracy rate achieved on the test dataset demonstrates the model's high generalization capacity. Figure 3 shows the confusion matrix of the Swin Transformer model.

The confusion matrix, where the model misclassified only one of the 158 test samples, demonstrates that the Swin Transformer is capable of learning to distinguish between classes at a very high level. The presence of near-zero values in the out-of-class cells in each row demonstrates that the separation between classes is significant.

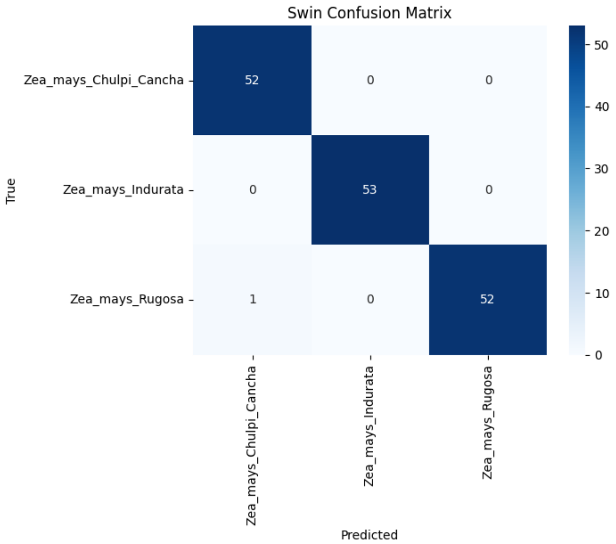

Figure 3. Swin Transformer Model Confusion Matrix

The model's lowest performance occurred when a single sample from the Rugosa class was confused with Chulpi Cancha. However, this error did not significantly affect the overall success rate of 99.37%. Furthermore, the fact that the error was only one-way (Rugosa → Chulpi Cancha) demonstrates that the model does not exhibit any systematic bias.

The confusion matrix is consistent with the findings in the classification report and visually supports the Swin Transformer model's high recall and precision values. In this context, it clearly demonstrates that the Swin Transformer model is an effective method that provides high accuracy, stability, and reliability in the classification of maize species.

To address the concern that the near-perfect accuracy might result from overfitting, additional performance metrics were examined. As shown in the classification report, precision, recall, and F1-scores for all three classes range between 0.98 and 1.00, indicating highly consistent results. Furthermore, both macro and weighted averages reach 0.99, confirming that the model performs uniformly across all classes rather than overfitting to a particular subset. Therefore, the high accuracy observed is attributable to the model’s effective learning capability rather than overfitting.

3.2. Findings of the Swin Transformer + EfficientNetV2 Hybrid Model

The model, which utilizes a hybrid of Swin Transformer and EfficientNetV2, demonstrated high success in classifying Chulpi Cancha, Indurata, and Rugosa varieties of the Zea mays species. The 100% accuracy rate achieved on the test data demonstrates the model's high level of generalization capacity.

The results confirm that the Swin Transformer model provides high accuracy in classifying corn varieties; however, they also reveal that the hybrid version with EfficientNetV2 distinguishes between classes more effectively, increasing the accuracy rate to 100%. In this context, it can be said that the proposed hybrid model stands out with its superior learning capacity and strong generalization ability in agricultural image classification problems.

The confusion matrix of the Swin Transformer + EfficientNetV2 hybrid model demonstrates the strong synergy achieved by combining these two architectures. Swin Transformer's attention-based global and local information extraction capability, combined with EfficientNetV2's optimized convolutional feature extraction, enabled complete and consistent learning of distinguishing features between classes.

While the previous confusion matrix obtained with the Swin Transformer model exhibited a classification error in a single example of the Rugosa class, this error is eliminated in this hybrid model. This demonstrates that the hybrid structure significantly enhances generalization ability.

Figure 4 presents the confusion matrix of the hybrid model.

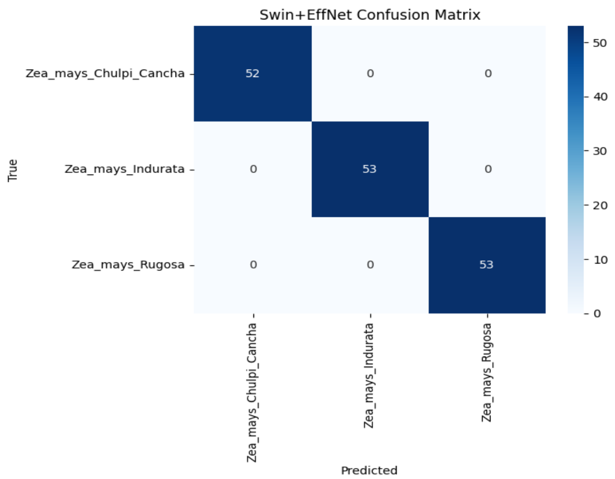

Figure 4. Swin Transformer - EfficientNetV2 Hybrid Model Mixed Matrix

The ROC curve and AUC values presented in Figure 5 are one of the strongest indicators supporting the classification performance of the Swin Transformer + EfficientNetV2 hybrid model.

The results clearly demonstrate that the Swin

Transformer and EfficientNetV2 architectures, when working together, provide a powerful complementary learning mechanism. Swin Transformer effectively learns global and local relationships thanks to its self-attention-based structure, while EfficientNetV2 provides efficient and detailed feature extraction with its depth, width, and resolution scaling advantages. Thanks to this hybrid structure, the developed model not only demonstrated high discrimination accuracy between classes but also demonstrated reliable and stable classifier performance, with no class interfering with the others.

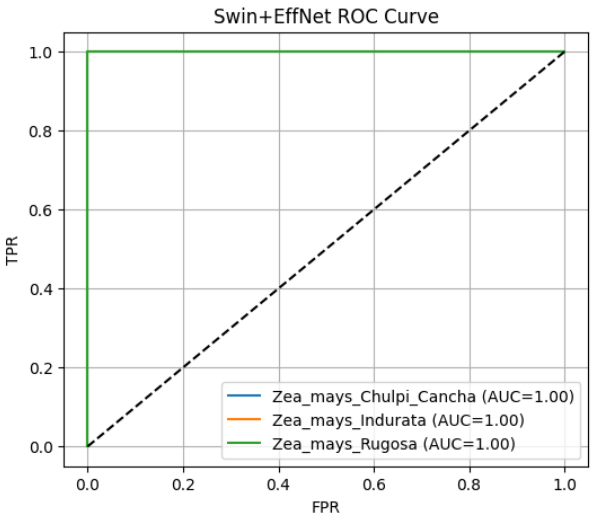

Figure 5. Swin Transformer - EfficientNetV2 Hybrid Model ROC Curve

Table 2 presents a comparison of the results of our study with the methods and accuracy rates reported in similar studies in the literature.

Table 2. Comparison of the proposed hybrid method with previous studies.

Study	Method	Accuracy (%)
Avuclu et al. [3]	Morphological Features + ML Hybrid	100
Lopes et al. [26]	J48	100
Avuclu et al. [27]	ResCNN, DAG-Net, ResNet18	100
Isik et al. [14]	ResNet50+BiLSTM	98.1
Ni et al. [12]	ResNet	98.2
Javanmardi et al. [13]	CNN-ANN	98.1
Yang et al. [8]	SVM + PLS-DA	98.2
Wu et al. [21]	SVM-GA	97.44
Yang et al. [18]	SVM	96.3
Lopes et al. [24]	ResNet18	96.82
Our Study	Swin Transformer	99.37
Our Study (Hybrid)	Swin + EfficientNetV2	100

3.3. Discussion

In this study, deep learning-based models were comparatively evaluated for the classification of Chulpi Cancha, Indurata, and Rugosa varieties of the Zea mays species. Among the implemented models, Swin Transformer, thanks to its attention-based architecture, achieved high accuracy in class discrimination. This model, which achieved 99.37% test accuracy, made an error in only one instance; specifically, while achieving 100% accuracy for the Indurata class, it classified a single instance of the Rugosa class as Chulpi Cancha. This result demonstrates that the Swin Transformer model is largely capable of learning subtle morphological differences between classes, but that class confusion can occur in some limited cases.

A model obtained by hybridizing Swin Transformer with the EfficientNetV2 architecture took classification performance one step further, achieving 100% test accuracy. The hybrid model accurately classified all test instances from all three classes, demonstrating high levels of both generalization and discriminative feature extraction. The combination of Swin Transformer's attention mechanism and EfficientNetV2's optimized convolutional structure increased the model's representational power at both local and global levels, enabling it to more clearly distinguish differences between morphologically similar classes. This study was the first to test Swin Transformer and EfficientNetV2 in the classification of hybrid maize varieties, achieving 100% accuracy.

The ROC curves for the Swin + EfficientNetV2 hybrid model showed an AUC value of 1.00 for each class. This demonstrates that the model not only provides high accuracy but also clearly distinguishes all classes with high confidence. This result supports the claim that deep learning-based hybrid model approaches provide effective and reliable solutions to agricultural classification problems.

When the findings obtained in this study are compared with similar approaches in the literature, it is seen that the proposed hybrid model provides a significant superiority. For example, Zhang et al. [10] achieved 94.4% accuracy with a CNN-based model on hyperspectral images, Ni et al. [12] reported 98.2% accuracy with a ResNet-based model, and Avuçlu and Köklü [27] reached up to 99.52% accuracy using DAG-Net and ResNet-18. The Swin Transformer + EfficientNetV2 hybrid model proposed in our study achieved 100% classification accuracy on all test samples. This result demonstrates that the hybrid approach not only increases classification accuracy but also maximizes inter-class discrimination by reaching a value of 1.00 for each class in the ROC-AUC analysis. Thus, it is confirmed that our model offers stronger generalization capacity beyond the methods reported in the existing literature.

Another important aspect to consider is the computational cost of the proposed models. While the experiments in this study were conducted using high-performance hardware (NVIDIA A100 GPU), real-world agricultural applications often rely on resource-constrained platforms such as mobile devices, embedded systems, or field robots. Therefore, future research should focus on optimizing the proposed hybrid architecture for lightweight deployment. Approaches such as model pruning, quantization, knowledge distillation, and the use of efficient backbone networks (e.g., MobileNetV3 or ShuffleNet) can significantly reduce memory and power consumption. This would enable the integration of the developed system into portable or edge-computing platforms, making it more practical for real-time field applications in precision agriculture.

The findings are not limited to the classification of corn varieties but also have a broader impact on agricultural AI applications. It appears that deep learning-based hybrid models can be successfully used in critical tasks such as disease detection, quality control, and yield estimation in various crops. In this regard, the hybrid Swin Transformer + EfficientNetV2 model proposed in our study has an adaptable structure for different agricultural data types (e.g., leaf images, hyperspectral data, or drone-based field imagery). Thus, the high accuracy achieved not only contributes to variety classification but also to the development of decision support systems in agricultural production and the dissemination of smart agricultural technologies.

4. CONCLUSIONS

This study demonstrated that deep learning-based hybrid models can achieve highly accurate classification of maize varieties. The Swin Transformer model alone reached 99.37% accuracy, while the proposed Swin Transformer + EfficientNetV2 hybrid achieved 100% accuracy by correctly classifying all test samples.

Beyond academic contributions, these findings have clear real-world implications. The proposed hybrid model can support critical agricultural processes such as seed purity assessment, quality control, and automated decision-making in precision farming. With optimization techniques (e.g., model compression, quantization, knowledge distillation), the model can be deployed on mobile devices, embedded systems, or agricultural robots, enabling real-time field applications.

Overall, the results highlight the potential of integrating next-generation attention-based architectures with optimized CNNs to develop lightweight, robust, and practical solutions for agricultural data analysis. Future work should expand the model to additional maize varieties and different data types (e.g., field images, hyperspectral data) to further validate scalability and adaptability.

Declaration of Ethical Standards

The authors confirm that this study adheres to all ethical standards, including proper authorship attribution, accurate citation, appropriate data reporting and the publication of original research.

Credit Authorship Contribution Statement

The conceptualization of the research and the data collection process were carried out by Huseyin Bulduk. The evaluation and analysis of the results were performed by Kadir Sabanci. The original draft of the manuscript was written by Huseyin Bulduk, while the review and editing were undertaken by Kadir Sabanci.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Funding / Acknowledgements

No funding or research grants were received during the preparation of this study.

Availability of Data and Material

The authors confirm that the data supporting the findings of this study are available within the manuscript. The data that support the findings of this study are available from this link: https://citedata.com/Corn_3_Classes_Image_Dataset.zip.

References

[1] A. Kumar, R. Singh, and S. S. Chauhan, "Revolutionizing Agriculture: Applications of Artificial Intelligence in Crop Management and Disease Detection," Frontiers in Plant Science, vol. 15, no. 1356260, 2024. https://doi: 10.3389/fpls.2024.1356260

[2] Ambrose, A., Kandpal, L. M., Kim, M. S., Lee, W.-H., & Cho, B.-K. (2016). High speed measurement of corn seed viability using hyperspectral imaging. Infrared Physics & Technology, 75, 173–179. https://doi.org/10.1016/j.infrared.2015.12.008

[3] Avuçlu, E., Taşdemir, Ş., & Koklu, M. (2023). A new hybrid model for classification of corn using morphological properties. European Food Research and Technology, 249(3), 835–847. https://doi.org/10.1007/s00217-022-04181-x

[4] Chivasa, W., O. Mutanga, and C. Biradar. 2019. "Phenology-Based Discrimination of Maize (Zea mays L.) Varieties Using Multitemporal Hyperspectral Data." Journal of Applied Remote Sensing 13, no. 1: 017504– 017504. https://doi.org/10.1117/1.JRS.13.017504.

[5] Kishore B, Yasar A, Taspinar YS, Kursun R, Cinar I, Shankar VG, Koklu M, Ofori I. Computer-Aided Multiclass Classification of Corn from Corn Images Integrating Deep Feature Extraction. Comput Intell Neurosci. 2022 Aug 10;2022:2062944. https://doi.org/10.1155/2022/2062944 .

[6] Isik, H., S. Tasdemir, Y. Selim Taspinar, et al. 2024. "Maize Seeds Forecasting With Hybrid Directional and Bi-Directional Long Short-Term Memory Models." Food Science & Nutrition 12, no. 2: 786–803. https://doi.org/10.1002/fsn3.3783.

[7] Wang, F., H. Zhao, C. Yu, J. Tang, W. Wu, and Q. Yang. 2020. "Determination of the Geographical Origin of Maize (Zea mays L.) Using Mineral Element Fingerprints." Journal of the Science of Food and Agriculture 100, no. 3: 1294–1300. https://doi.org/10.1002/jsfa.10144.

[8] Yang, X., Hong, H., You, Z., & Cheng, F. (2015). Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. Sensors, 15(7), 15578–15594. https://doi.org/10.3390/s150715578

[9] Zhao, M., Wu, W., Zhang, Y. Q., & Li, X. (2011). Combining genetic algorithm and SVM for corn variety identification. International Conference on Mechatronic Science, Electric Engineering and Computer, 990–993. https://doi.org/10.1109/MEC.2011.6025631

[10] Zhang, J., Dai, L. & Cheng, F. Corn seed variety classification

based on hyperspectral reflectance imaging and deep convolutional neural network. Food Measure 15, 484–494 (2021). https://doi.org/10.1007/s11694-020-00646-3.

[11] H. O. Velesaca, R. Mira, P. L. Suárez, C. X. Larrea and A. D. Sappa, "Deep Learning based Corn Kernel Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 294-302, https://doi.org/10.1109/CVPRW50498.2020.00041

[12] Ni, C., Wang, D., Vinson, R., Holmes, M., & Tao, Y. (2019). Automatic inspection machine for maize kernels based on deep convolutional neural networks. Biosystems Engineering, 178, 131–144. https://doi.org/10.1016/j.biosystemseng.2018.11.010

[13] Javanmardi, S., Ashtiani, S.-H. M., Verbeek, F. J., & Martynenko, A. (2021). Computer-vision classification of corn seed varieties using deep convolutional neural network. Journal of Stored Products Research, 92, 101800. https://doi.org/10.1016/j.jspr.2021.101800

[14] Isik H, Tasdemir S, Taspinar YS, Kursun R, Cinar I, Yasar A, Yasin ET, Koklu M. Maize seeds forecasting with hybrid directional and bi-directional long short-term memory models. Food Sci Nutr. 2023 Nov 9;12(2):786-803. doi: 10.1002/fsn3.3783. PMID: 38370035; PMCID: PMC10867492. https://doi.org/10.1002/fsn3.3783

[15] K. Kiratiratanapruk and W. Sinthupinyo, "Color and texture for corn seed classification by machine vision," 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), Chiang Mai, Thailand, 2011, pp. 1-5, https://doi.org/:10.1109/ISPACS.2011.6146100.

[16] Zhao, Y., S. Zhu, C. Zhang, X. Feng, L. Feng, and Y. He. 2018. Application of Hyperspectral Imaging and Chemometrics for Variety Classification of Maize Seeds. RSC Advances 8, no. 3: 1337–1345. https://doi.org/10.1039/C7RA05954J

[17] S. Kai, L. Zhikun, S. Hang and G. Chunhong, "A Research of Maize Disease Image Recognition of Corn Based on BP Networks," 2011 Third International Conference on Measuring Technology and Mechatronics Automation, Shanghai, China, 2011, pp. 246-249, https://doi.org/10.1109/ICMTMA.2011.66.

[18] Yang, X., Hong, H., You, Z., & Cheng, F. (2015). Spectral and Image Integrated Analysis of Hyperspectral Data for Waxy Corn Seed Variety Classification. Sensors, 15(7), 15578-15594. https://doi.org/10.3390/s150715578

[19] Huang M, Tang J, Yang B, Zhu Q (2016) Classifcation of corn seeds of diferent years based on hyperspectral imaging and model updating. Comput Electron Agric 122:139–145. https://doi.org/10.1016/j.compag.2016.01.029

[20] Williams PJ, Kucheryavskiy S (2016). Classifcation of corn kernels using NIR hyperspectral imaging. Food Chem 209:131–138. https://doi.org/10.1016/j.foodchem.2016.04.044

[21] Wu A, Zhu J, Yang Y, Liu X, Wang X, Wang L, Chen J (2018). Classifcation of corn kernels grades using image analysis and support vector machine. Adv Mech Eng 10(12):1687814018817642. https://doi.org/10.1177/1687814018817642

[22] Daskalov P, Kirilova E, Georgieva T (2018). Performance of an automatic inspection system for classifcation of Fusarium Moniliforme damaged corn seeds by image analysis. In: MATEC Web of Conferences, Vol. 210, EDP Sciences, p. 02014. https://doi.org/10.1051/matecconf/201821002014

[23] Li, X., Dai, B., Sun, H., & Li, W. (2019). Corn Classification System based on Computer Vision. Symmetry, 11(4), 591. https://doi.org/10.3390/sym11040591

[24] Lopes, J.F., da Costa, V.G.T., Barbin, D.F. et al. Deep computer vision system for cocoa classification. Multimed Tools Appl 81, 41059–41077 (2022). https://doi.org/10.1007/s11042-022-13097-3

[25] Marciano M. Oliveira, Breno V. Cerqueira, Sylvio Barbon, Douglas F. Barbin, Classification of fermented cocoa beans (cut test) using computer vision, Journal of Food Composition and Analysis, Volume 97, 2021, 103771, ISSN 0889-1575, https://doi.org/10.1016/j.jfca.2020.103771.

[26] Lopes, J. F., Ludwig, L., Barbin, D. F., Grossmann, M. V. E., & Barbon, S., Jr. (2019). Computer Vision Classification of Barley Flour Based on Spatial Pyramid Partition Ensemble. Sensors, 19(13), 2953. https://doi.org/10.3390/s19132953

[27] Avuçlu E, Köklü M. Fast and Accurate Classification of Corn Varieties Using Deep Learning With Edge Detection Techniques. J Food Sci. 2025 Jul;90(7):e70439. https://doi.org/10.1111/1750-3841.70439 . PMID: 40702919.

[28] M. E. Sönmez, K. Sabancı, and N. Aydın, "Classification of Wheat Rootstock and Their Hybrids According to Color Features by Machine Learning Algorithms", International Journal of Applied Mathematics Electronics and Computers, vol. 10, no. 2, pp. 39–48, 2022, https://doi.org/10.18100/ijamec.1098276.

[29] P. K. Das and S. S. Rupa. ResNet for Leaf-based Disease Classification in Strawberry Plant. J. Appl. Methods Electron. Comput., vol. 11, no. 3, pp. 151–157, Sep. 2023, https://doi.org/10.58190/ijamec.2023.42.

[30] G. Altan, "Performance Evaluation of Capsule Networks for Classification of Plant Leaf Diseases", International Journal of Applied Mathematics Electronics and Computers, vol. 8, no. 3, pp. 57–63, 2020, https://doi.org/10.18100/ijamec.797392

[31] Sun, Y., Ning, L., Zhao, B., & Yan, J. (2024). Tomato Leaf Disease Classification by Combining EfficientNetv2 and a Swin Transformer. Applied Sciences. https://doi.org/10.3390/app14177472.

[32] Li, L., & Tanone, R. (2023). Disease Identification in Potato Leaves using Swin Transformer. 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM), 1-5. https://doi.org/10.1109/IMCOM56909.2023.10035609.

[33] Zhao, K., Huo, Y., Xue, L., Yao, M., Tian, Q., & Wang, H. (2023). Mushroom Image Classification and Recognition Based on Improved Swin Transformer. 2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE), 225-231. https://doi.org/10.1109/ICISCAE59047.2023.10392483.