Orclever

Vol. 6 No. 1 (2026)
DOI: 10.56038/ejrnd2026890477
www.orclever.com

The European Journal of Research and Development
e-ISSN: 2822-2296

Challenges in Maize Root Phenotyping: Preprocessing Limits and Class Imbalance in Deep Learning

Hüdanur Engin[1*], Ali Murat Tiryaki[1]

[1] *Çanakkale Onsekiz Mart Üniversitesi*

iD 0009-0007-5900-1963

iD 0000-0001-8224-6319

* Corresponding author: hudanur.engin@gmail.com

DOI: 10.56038/ejrnd2026890477

Received: 2026-03-12 · Revised: 2026-05-21 · Accepted: 2026-06-20 · Published: 2026-06-22

Abstract

Doubled Haploid (DH) technology significantly accelerates the development of homozygous lines in maize breeding; however, its scalability is constrained by the reliable discrimination of haploid and diploid individuals. The widely used R1-nj anthocyanin marker at the seed stage is susceptible to genetic suppression and environmental variability, leading to high misclassification rates. This limitation has driven a shift toward seedling root morphology as a more robust phenotypic marker, yet it introduces major challenges, including complex image noise and severe class imbalance. In this study, we systematically evaluate the limitations of standard computer vision pipelines and baseline deep learning models for root-based classification. Automated background removal methods (HSV, Rembg) are shown to misinterpret fine root hairs as noise, resulting in significant morphological data loss. Additionally, experiments conducted under a realistic class imbalance (1:5.4) demonstrate that widely used CNN architectures (ResNet50, VGG16, EfficientNetB0, DenseNet121) exhibit strong majority class bias, with haploid recall dropping to 0.00% and 27.7%. These findings reveal a critical limitation in existing approaches and highlight the need for domain-informed datasets and imbalance-aware learning strategies for robust and scalable AI-based maize breeding systems.

Keywords: Deep Learning, Class Imbalance, Image Preprocessing, Minority Class Detection, Convolutional Neural Networks (CNN), Root Phenotyping, Agricultural Image Analysis

1. Introduction

The global climate crisis and the rapidly increasing world population are placing unprecedented pressure on the sustainability of agricultural productivity. In maize (*Zea mays* L.) production, which is a cornerstone of global food security, the development of high-yielding and stress-resilient varieties has become a critical necessity. In this context, Doubled Haploid (DH) technology has emerged as a revolutionary advancement in modern maize breeding, reducing the time required to obtain fully homozygous (pure) lines from 7–8 generations to only two generations [15]. Beyond accelerating breeding cycles, DH technology also facilitates the rapid purification of genetic material and supports compliance with Distinctness, Uniformity, and Stability (DUS) criteria required for plant variety protection [7].

Despite these advantages, DH technology introduces a major bottleneck: the accurate, rapid, and cost-effective discrimination of haploid and diploid individuals among large breeding populations [4]. Traditionally, this process relies on the visual identification of the R1-nj (Navajo) anthocyanin color marker at the seed stage [15, 2]. However, extensive studies have demonstrated that this pigmentation is often inconsistently distributed, can be suppressed by genetic inhibitors, and is influenced by environmental factors, resulting in high rates of misclassification (false negatives) [1]. Consequently, manual sorting based on human observation remains labor-intensive and unreliable, motivating the search for alternative and more robust diagnostic approaches.

Early computational attempts to address this limitation employed traditional machine learning algorithms such as Support Vector Machines (SVM) and Random Forests (RF), which achieved only limited performance in maize seed classification tasks [11]. With the emergence of deep learning, convolutional neural networks (CNNs) have been widely adopted to automate seed-based classification, including systems designed to detect the R1-nj marker (e.g., DeepSort-based approaches) [1, 19]. In parallel, hyperspectral and multispectral imaging techniques have been explored to overcome the limitations of visible-spectrum data, with studies reporting classification accuracies exceeding 90% using hybrid LSTM-CNN models and transfer learning [12, 20]. Similarly, Vis-NIR technologies have been proposed as promising alternatives [21]. Nevertheless, the high cost of such spectral imaging systems limits their applicability in real world breeding environments, confining them largely to controlled laboratory settings.

More recently, the use of standard RGB imaging combined with deep learning has emerged as a cost-effective alternative, significantly advancing digital agriculture applications [14]. State of the art CNN architectures such as VGG [17], ResNet [8], DenseNet [10], and EfficientNet [18] have demonstrated outstanding performance in general image classification tasks. However, these architectures are typically optimized for clean and balanced datasets. In agricultural scenarios, particularly in root and seedling imaging, complex backgrounds, residual artifacts, and fine morphological structures introduce significant noise. Automated preprocessing techniques such as HSV-based segmentation and background removal tools (e.g., Rembg) often misinterpret biologically meaningful features especially fine root hairs as noise, leading to critical information loss [13, 14].

An alternative line of research has therefore shifted focus away from seed-based markers toward phenotypic traits observable during early plant development. Among these, seedling root morphology has emerged as the most promising surrogate marker. Root system architecture (RSA), including root vigor and morphological traits influenced by genetic factors such as the Pl1 gene and heterosis effects in diploids, provides a biologically robust basis for haploid/diploid discrimination [4]. However, this shift introduces new computational challenges that must be addressed.

One of the primary challenges lies in image noise and preprocessing limitations. Unlike seeds, which can be imaged under relatively controlled and homogeneous conditions, seedling roots exhibit highly complex structures with fine root hairs and environmental residues, making accurate segmentation and feature extraction significantly more difficult [13, 14]. Another critical issue is the class imbalance problem inherent in real breeding populations [3], where haploids are rare (approximately 10–15%) and diploids dominate (approximately 80–90%) [6]. Under such conditions, deep learning models trained with standard loss functions tend to favor the majority class, leading to misleadingly high accuracy while failing to detect minority instances [9]. Therefore, evaluation metrics such as recall become essential for assessing true model performance, as highlighted in previous studies [5, 16].

Despite extensive efforts in the literature, a significant gap remains. Existing approaches either depend on costly spectral imaging systems or fail to effectively handle the combined challenges of noisy root images and severe class imbalance. Consequently, there is a clear need for algorithmic strategies that can operate on standard RGB root images while maintaining robustness against preprocessing-induced data loss and accurately detecting minority haploid classes under extreme imbalance conditions (e.g., 1:5.4).

In this study, we systematically investigate the limitations of standard computer vision techniques and baseline CNN architectures including ResNet50[8], VGG16[17], EfficientNetB0[18], and DenseNet121[10] in the classification of maize seedling root images. Specifically, we analyze the morphological data loss caused by automated background removal methods and experimentally demonstrate the failure of conventional models under realistic class imbalance conditions, highlighting the emergence of the accuracy paradox. The findings emphasize the necessity of domain expert annotated datasets and provide insights into the development of cost-sensitive and imbalance aware deep learning approaches for future digital agriculture applications.

2. Methodology

This section delineates the experimental framework developed to investigate the algorithmic challenges encountered in the digital phenotyping of maize seedling roots. The study is organized into five principal subsections: dataset characteristics, root image acquisition and structural properties, limitations of automated image preprocessing algorithms, the mathematical formulation of the class imbalance problem, and the baseline deep learning architectures evaluated.

2.1.Dataset

The dataset utilized in this study was compiled in collaboration with the Faculty of Agriculture at Çanakkale Onsekiz Mart University. Following the preprocessing stages, the final dataset comprises a total of 2,308 high-resolution RGB images capturing the root morphology of maize seedlings. In terms of genetic diversity, the data reflects the natural variations found in actual breeding lines, encompassing a wide array of genotypes within the population. The seedling roots were photographed during the early germination stage when they were on average 3 to 7 days old, a critical period where the morphological differences driven by heterosis and root vigor become biologically prominent.

To manage the severe class imbalance inherent in real-world industrial crosses and to ensure an objective evaluation of the models, the dataset was partitioned using Stratified Sampling. This technique guarantees that the original class distribution is strictly preserved across all subsets. The dataset was divided into training, validation, and testing sets with an 80:10:10 ratio, respectively. Within the entire collection, the majority class consists of 1,948 Diploid images (84.5%), while the rare minority target class contains 360 Haploid images (15.5%), establishing a severe class imbalance ratio of approximately 1:5.4 throughout the experimental pipeline.

2.2. Acquisition of Root Images and Structural Characteristics

The maize root images used in the experiments were acquired in a laboratory environment at the Faculty of Agriculture, Çanakkale Onsekiz Mart University, using standard RGB cameras. Seedlings taken from breeding populations were photographed on a constant laboratory background. Biologically examined, maize roots have a much more irregular morphology compared to seeds. Light reflections present in the environmental setting and the extremely fine “root hairs” at the tips of the roots create a highly complex and “noisy” background for computer vision algorithms.

2.3. Limitation Analysis of Automated Image Preprocessing Algorithms

The success of deep learning models depends directly on the quality of the training data and how clearly the target object can be separated from the background. In this study, standard automated preprocessing methods frequently preferred in the literature for isolating maize roots from the background were tested. Figure 1 shows the data losses and structural deformations caused by these automated preprocessing algorithms. The specific limitations analyzed during this process are as follows:

- **HSV Color Space Transformation and Thresholding:** Images in RGB format were converted to the HSV (Hue, Saturation, Value) color space, which is more resistant to light variations. However, the semi-transparent structure of the fine maize root tips and the glare effects in the background caused thresholding algorithms to miscalculate root boundaries. In regions where the contrast between the root and the background was low, the algorithms fragmented the root structure, producing disjointed pixels.
- **Rembg (U-2-Net Based) Deep Learning Algorithm:** The Rembg library, a universal model optimized for background removal, was tested. However, because such pre-trained models are trained on macroscopic objects like humans, vehicles, or animals, they failed on microscopic details in agricultural data. During the tests, it was determined that the Rembg algorithm perceived the “fine root hairs,” which are of vital importance in haploid/diploid discrimination of maize roots, as environmental noise and completely deleted them, leading to irrecoverable morphological data loss.

This situation proved that standard automation algorithms cannot be used to create datasets for maize phenotyping and that special reference datasets requiring domain-expert intervention are needed. During the experimental process, a reference subset, in which the morphological integrity was manually preserved by the researchers, was used to test the deep learning models.

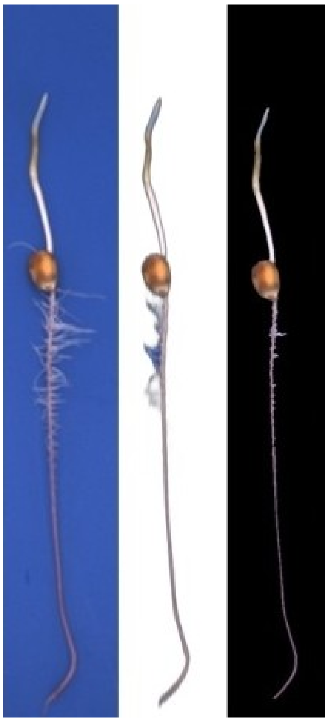

Figure 1: Illustration of Structural Deformation and Information Loss in Fine Root Hairs Induced by Automated Image Preprocessing Algorithms (Original–REMBG–HSV)

2.4. Mathematical Definition of the Class Imbalance Problem

The genetic distribution ratios in actual maize breeding greenhouses were perfectly reflected in the dataset of this study. The target class sought within the population, Haploid individuals, are rarely found (15.5%), while Diploid individuals form the dominant majority (84.5%). This situation creates a severe class imbalance in the dataset at a ratio of approximately 1:5.4.

In machine learning, class imbalance causes the majority class to dominate the loss function. The traditional Binary Cross-Entropy function is defined as follows [9]:

$$L = -[y \log(p) + (1 - y) \log(1 - p)] \tag{1}$$

In the absence of a mitigation strategy such as class weighting, the model tends to predominantly predict the majority Diploid (0) class in order to minimize the overall loss function (L), thereby giving rise to the so-called “lazy learning” phenomenon.

2.5. Evaluated Baseline Deep Learning Architectures

In order to observe the destructive effect of class imbalance, four of the most powerful convolutional neural network (CNN) architectures in the literature were tested without applying any imbalance strategy (in the Baseline scenario):

- **VGG16:** This architecture increases depth by using multiple 3x3 convolution filters. Although the standard structure of the model offers powerful hierarchical feature extraction to capture fine details in agricultural data, the fully connected networks in the final layer can rapidly overfit to the majority class in imbalanced datasets.
- **ResNet50:** This is an architecture that utilizes “skip connections” to solve the “vanishing gradient” problem in deep networks. Mathematically, the output of a block is expressed as $H(x) = F(x) + x$. Although this residual learning logic increases the stability of the model, it tends to ignore the minority class when trained with a standard loss function.
- **EfficientNetB0:** This is a modern architecture that simultaneously optimizes image resolution, network depth, and network width using a compound scaling coefficient. Despite being extremely powerful in terms of parameter efficiency, its standard configuration is assumed to be inadequate in detecting the rare class.
- **DenseNet121:** This is an architecture where each layer receives direct inputs from all preceding layers. The feed-forward structure is expressed by the formula $x_l = H_l([x_0, x_1, \ldots, x_{l-1}])$. Although this network, which is very strong

Vol. 6 No. 1 (2026)
DOI: 10.56038/ejrnd2026890477
www.orclever.com

The European Journal of Research and Development
e-ISSN: 2822-2296

in terms of feature reuse, appears resistant to overfitting, it stands out as one of the architectures that can exhibit the greatest bias towards the majority class under severe class imbalances.

3. Results

This section presents the test performances of baseline deep learning architectures trained on the maize seedling root dataset, which features a severe class imbalance ratio of 1:5.4, and analyzes the algorithmic biases exhibited by the models against the minority class (haploid).

3.1.Reconsidering the “Accuracy Paradox”: The Misleading Implications of Evaluation Metrics

Accuracy, the most widely used metric in classification problems, yields reliable results in scenarios where the dataset is balanced (1:1). However, in problems experiencing severe class imbalances, such as agricultural phenotyping, the Accuracy metric creates a misleading “illusion of success.” In this study, to measure the performance of the models realistically, the focus was placed not only on Accuracy values but also on the Sensitivity (Recall), Precision, and F1-Score metrics, which indicate the power to detect the rare class. Because missing a single haploid individual in maize breeding would disrupt the entire genetic process, the “Haploid Recall” value was accepted as the most critical evaluation criterion.

3.2. Performance Collapse of Baseline Architectures and Lazy Learning

The test set results of the models trained with standard loss functions, without any imbalance strategy applied, revealed a striking algorithmic vulnerability. This dramatic gap observed between the overall accuracy (Accuracy) values and the rare class detection (Haploid Recall) performances of the four different deep learning architectures is presented comparatively in Figure 2.

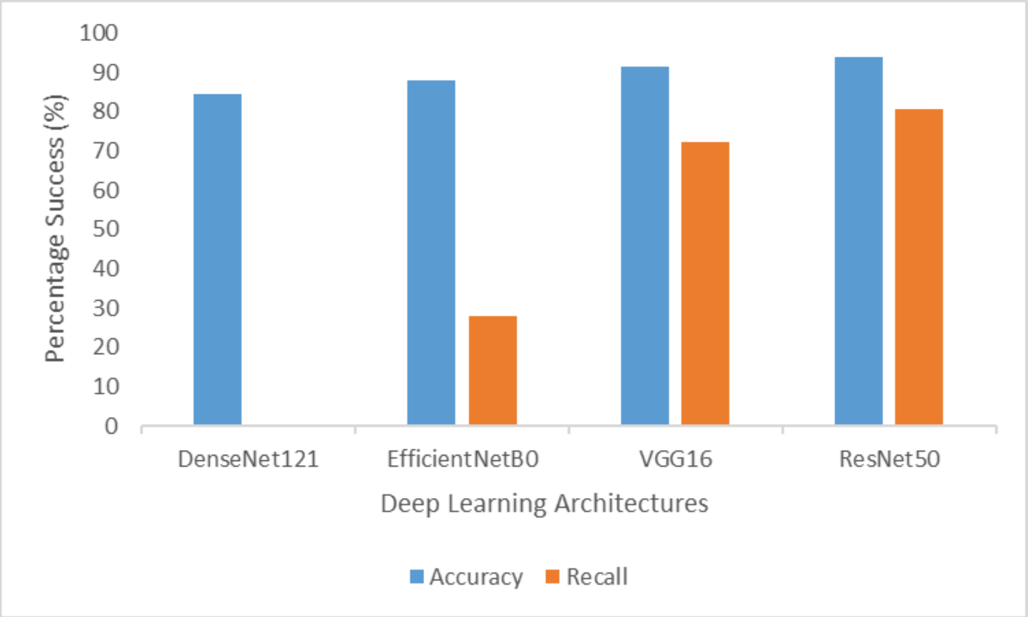

Figure 2: The “Accuracy Paradox” Observed in Baseline Architectures: Failure to Detect the Minority Class (Haploid) Despite High Accuracy Rates.

- **The Collapse of DenseNet121 (0.00% Recall):** Considered one of the most successful architectures for feature extraction in the literature, DenseNet121 literally collapsed in this agricultural problem. As clearly seen in the graphical distribution in Figure 2, although the overall Accuracy value of the model on the test set appeared to be at a high rate of 84.41%, the Haploid Recall value was realized at 0.00%. The model entered a “Lazy Learning” tendency under the intense class imbalance it encountered and assigned the “Diploid” (majority class) label to all samples in the test set to reduce the total loss value.
- **EfficientNetB0 and Insufficient Sensitivity:** The EfficientNetB0 architecture, prominent for its parameter efficiency, exhibited a similar “Majority Class Bias.” Although the overall accuracy of the model resulted in the 87.87% band, the Haploid Recall value remained at only 27.7%. In other words, the network missed approximately 73 out of every 100 actual haploid maize roots in the real world, mistaking them for diploids.

- **Instabilities of VGG16 and ResNet50:** Although VGG16 and ResNet50 architectures were able to show somewhat more resistance to the minority class compared to DenseNet, they could not reach the stability required to be used as an autonomous system in agricultural greenhouses and the industrial Recall reliability of over 90% in their standard forms.

3.3. Confusion Matrix Analysis

The Confusion Matrices presented in Figure 3 provide a clearer understanding of these dramatic failures of the models. When the matrices are examined, it is observed that deep learning models fall into a mathematical paradox during cross-entropy optimization. As clearly seen in Figure 3, the DenseNet121 model completely ignored the minority class and assigned the “diploid” label to all samples in the test set, while EfficientNetB0 was able to detect only 10 out of 36 rare haploid samples. Instead of expending effort to learn the complex root hairs and growth differences of the minority class haploids, the model discovered that statistically continuously predicting the majority class is a “cheaper” and “safer” path in the loss function.

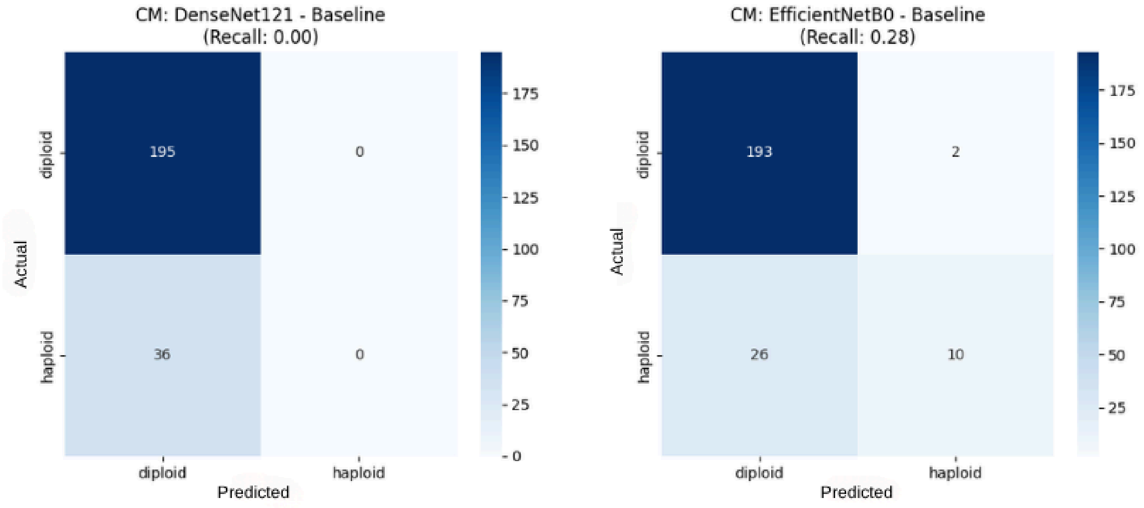

Figure 3: Test Set Confusion Matrices of Standard (Baseline) DenseNet121 and EfficientNetB0 Architectures Trained Under Class Imbalance and the Lazy Learning Anomaly.

3.4. Learning Dynamics and ROC Analysis

The classification capabilities of the baseline architectures and their ability to distinguish classes from each other were also validated through ROC (Receiver Operating Characteristic) curves and Area Under the Curve (AUC) scores. When the ROC curves presented in Figure 4 are examined, it is observed that the curves belonging to the DenseNet121 and EfficientNetB0 architectures, in particular, fail to sufficiently diverge from the random guess (chance) line and remain far below the ideal classification curve in the upper left corner. These low AUC scores (0.60 and 0.67, respectively) prove that the models lack the statistical feature extraction power (discrimination power) to distinguish the complex variations between diploid and haploid root morphologies.

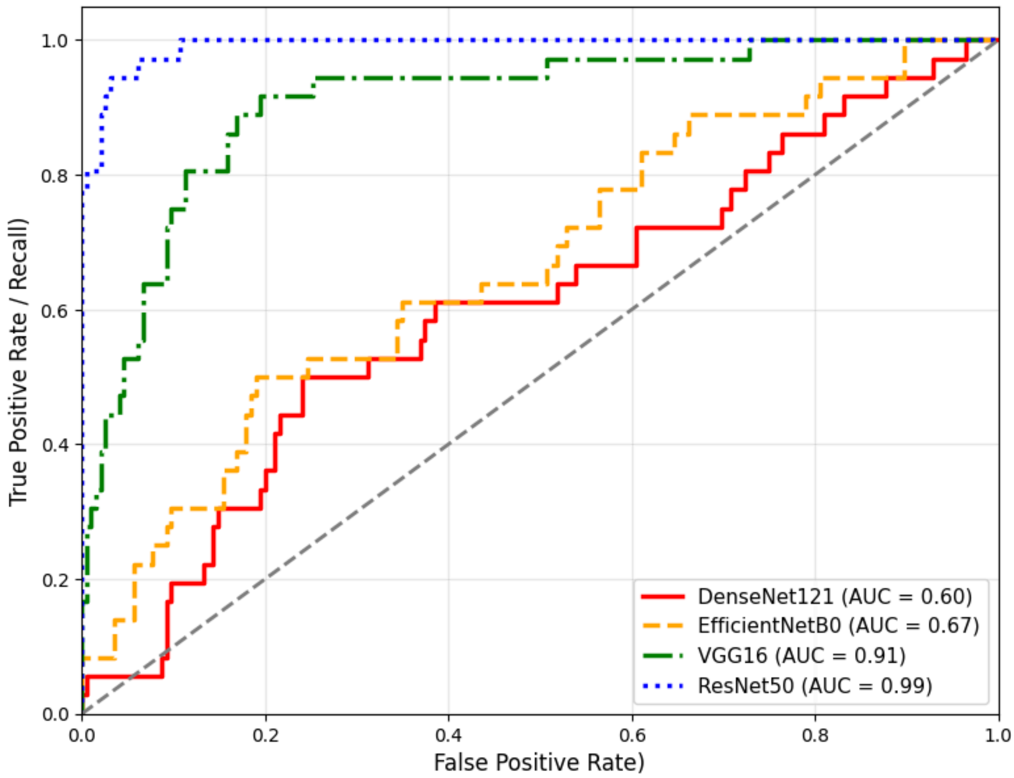

Figure 4: ROC Curves and AUC Performances of Baseline Architectures Trained Under Class Imbalance.

4. Discussion

These findings expose a fundamental flaw in current computer vision approaches in the literature. Although transitioning from seed phenotyping to root phenotyping is biologically the correct move, the environmental noise contained in root images (preprocessing failures) and the severe class imbalance inherent in maize populations render standard convolutional neural networks dysfunctional. Integrating a model into agricultural breeding processes merely by looking at the overall accuracy value is a massive engineering error that will result in the loss of rare and valuable haploid plants.

5. Conclusion

The development of Doubled Haploid (DH) lines in maize (*Zea mays* L.) breeding is of vital importance for agricultural genetics and global food security strategies. In this study, the analytical challenges brought about by shifting the autonomous haploid sorting process from the seed phenotype to root morphology were comprehensively examined along the axis of computer vision and standard deep learning architectures.

Experimental findings clearly demonstrated that the dense biological and environmental noise contained in agricultural root images (laboratory background, water droplets, etc.) could not be managed by pre-trained universal background removal algorithms such as HSV Color Space and Rembg. These algorithms caused data loss in the fine root hairs, which are critical in the haploid/diploid discrimination of maize, rendering autonomous preprocessing pipelines dysfunctional.

Furthermore, it was proven that VGG16, ResNet50, EfficientNetB0, and DenseNet121 architectures trained under a severe class imbalance of 1:5.4, which represents a real breeding population, experienced a systematic failure in detecting the minority class (haploid). In particular, the fact that the DenseNet121 architecture, one of the most powerful convolutional neural networks in the literature, yielded a 0.00% Haploid Recall despite reaching a high overall accuracy (Accuracy) of 84.41%, indisputably revealed the “Accuracy Paradox” in agricultural AI applications and the algorithmic bias of models towards the majority class (majority class bias).

Consequently, the industrial reliability of autonomous agricultural systems cannot be achieved by directly integrating universal deep learning models into agricultural data without applying any strategies. This study shows that the limits of standard methods in the field of digital phenotyping have been reached. In future studies, the creation of domain-expert-annotated reference datasets instead of automated algorithms, and the development of innovative strategies such as cost-sensitive learning that resolve class imbalance at the loss function level, are suggested as an urgent scientific necessity.

6. Acknowledgements

The authors would like to express their sincere gratitude to the Faculty of Agriculture at Çanakkale Onsekiz Mart University for providing the maize seedling dataset, laboratory facilities, and domain-expert support during the data acquisition phase of this study.

7. References

[1] Y. Altuntaş, Z. Cömert, and A. F. Kocamaz, “Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach,” Comput. Electron. Agric., vol. 163, p. 104874, 2019. doi: 10.1016/j.compag.2019.104874.

[2] B. W. Boote, D. J. Freppon, G. N. De La Fuente, T. Lübberstedt, B. J. Nikolau, and E. A. Smith, “Haploid differentiation in maize kernels based on fluorescence imaging,” Plant Breed., vol. 135, no. 4, pp. 439-445, 2016. doi: 10.1111/pbr.12382.

[3] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” Neural Networks, vol. 106, pp. 249-259, 2018. doi: 10.1016/j.neunet.2018.07.011.

[4] V. Chaikam, W. Molenaar, A. E. Melchinger, and P. M. Boddupalli, “Doubled haploid technology for line development in maize: technical advances and prospects,” Theor. Appl. Genet., vol. 132, no. 12, pp. 3227-3243, 2019. doi: 10.1007/s00122-019-03433-x.

[5] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genomics, vol. 21, no. 1, pp. 1-13, 2020. doi: 10.1186/s12864-019-6413-7.

[6] A. Dermail, M. Mitchell, T. Foster, M. Fakude, Y. R. Chen, K. Suriharn, U. K. Frei, and T. Lübberstedt, “Haploid identification in maize,” Front. Plant Sci., vol. 15, p. 1378421, 2024. doi: 10.3389/fpls.2024.1378421.

[7] H. H. Geiger and G. A. Gordillo, “Doubled haploids in hybrid maize breeding,” Maydica, vol. 54, no. 4, p. 485, 2009.

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[9] Y. Ho and S. Wookey, “The real-world-weight cross-entropy loss function: Modeling the costs of misclassification,” IEEE Access, vol. 8, pp. 4806-4813, 2019. doi: 10.1109/ACCESS.2019.2962617.

[10] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4700-4708).

[11] F. Kahrıman, A. M. Güz, and İ. Pehlivan, “Use of machine learning models-based image analysis to discriminate haploid and diploid seeds in maize,” Sci. Agric., vol. 80, 2023. doi: 10.1590/1984-70332023v23n4a44.

[12] W. Liao, X. Wang, D. An, and Y. Wei, “Hyperspectral imaging technology and transfer learning utilized in identification of haploid maize seeds,” Sensors, vol. 19, no. 5, p. 1167, 2019. doi: 10.48550/arXiv.1805.11784.

[13] J. Pace, N. Lee, H. S. Naik, B. Ganapathysubramanian, and T. Lübberstedt, “Analysis of maize (Zea mays L.) seedling roots with the high-throughput image analysis tool ARIA (Automatic Root Image Analysis),” PLoS ONE, vol. 9, no. 9, p. e108255, 2014. doi: 10.1371/journal.pone.0108255.

[14] M. P. Pound, J. A. Atkinson, A. J. Townsend, M. H. Wilson, M. Griffiths, A. S. Jackson, et al., “Deep machine learning provides state-of-the-art performance in image-based plant phenotyping,” Gigascience, vol. 6, no. 10, p. gix083, 2017. doi: 10.1093/gigascience/gix083.

[15] B. M. Prasanna, V. Chaikam, and G. Mahuku, **Doubled Haploid Technology in Maize Breeding: Theory and Practice**. CIMMYT, 2012.

[16] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” PLoS ONE, vol. 10, no. 3, p. e0118432, 2015. doi: 10.1371/journal.pone.0118432.

[17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[18] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[19] B. Veeramani, J. W. Raymond, and P. Chanda, "DeepSort: deep convolutional networks for sorting haploid maize seeds," BMC Bioinformatics, vol. 19, no. 9, pp. 1-10, 2018. doi: 10.1186/s12859-018-2267-2.

[20] X. Y. Wang, W. X. Liao, D. An, and Y. G. Wei, "Maize haploid identification via LSTM-CNN and hyperspectral imaging technology," arXiv preprint arXiv:1805.09105, 2018.

[21] L. Zhang, D. Wang, J. Liu, and D. An, "Vis-NIR hyperspectral imaging combined with incremental learning for open world maize seed varieties identification," Front. Plant Sci., vol. 13, p. 893630, 2022. doi: 10.3389/fpls.2022.893630.