www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(3), May-Aug 2025, 478-488. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i3.851

Evaluating Deep Learning Techniques for Sugarcane Disease Classification

Melesio Crespo-Sánchez
Universidad Autónoma de San Luis Potosí, San Luis Potosí, México
mel.crespo18@gmail.com

Abstract. Sugarcane is one of Mexico's principal crops (INEGI, 2025), playing a crucial role in the sugar industry and its derivatives. However, various diseases pose a threat to sugarcane cultivation, resulting in significant economic losses due to the large-scale eradication of crops. Early and accurate identification of diseases is essential for effective management, yet it remains challenging without specialised knowledge. Deep learning tools can facilitate the detection of such diseases. This study presents a comparative analysis of three state-of-the-art deep learning architectures—EfficientNetV2B0, DenseNet121, and ResNet101V2—for sugarcane disease detection. Using a dataset of 7,000 sugarcane leaf images categorised into five classes (healthy and four disease types), the evaluation of these models was based on multiple classification metrics. The findings highlight competitive performance among the models, showcasing their respective strengths and limitations in terms of accuracy and computational efficiency. This analysis offers valuable insights into deep learning-based approaches for sugarcane disease detection, supporting the development of practical solutions for the agricultural sector.

Keywords: Sugarcane diseases detection, Deep learning

Article Info
Received January 02, 2025
Accepted April 07, 2025

1 Introduction

A significant challenge Mexican farmers face is crop disease management, which reduces yields and results in expensive treatment measures. In sugarcane cultivation, diseases such as rust, yellow leaf virus, mosaic, and red rot have particularly severe economic consequences, threatening productivity and profitability (Viswanathan & Rao, 2011). Effective disease management relies on early detection, as timely intervention can minimize losses and prevent widespread outbreaks. However, traditional methods, such as expert visual inspection, are time-consuming, labor-intensive, and prone to human error, particularly in large-scale farms where manual monitoring is impractical. Recent advancements in deep learning offer promising solutions for the early detection and classification of plant diseases through automated image analysis. Convolutional neural networks (CNNs), in particular, have demonstrated high accuracy in identifying disease symptoms in plant leaves, enabling faster diagnoses than conventional techniques. By leveraging deep learning models trained on annotated datasets of sugarcane leaf images, farmers can implement cost-effective, scalable, and real-time disease monitoring systems. These innovations not only enhance agricultural productivity but also contribute to sustainable farming practices by reducing the excessive use of pesticides and other chemical treatments.

In the literature, various deep learning models designed to address the challenge of plant disease classification are found. In (Srivastava et al., 2020), the authors achieved an Area Under the Curve (AUC) of 90.2% using *Inception v3* and *VGG-19* models in combination with different classifiers, demonstrating the potential of CNN-based architectures for plant disease identification. More recently, (Li et al., 2023) introduced the *SLViT* model, a vision transformer-based approach, reaching an accuracy of 98.87%, highlighting the advancements in transformer architectures for image classification tasks. Similarly, (Hemalatha et al., 2022) achieved a 98% accuracy utilizing the *LeNet-5* model, reinforcing the efficacy of traditional CNNs in feature extraction for disease detection.

Beyond general plant disease classification, some studies have focused on sugarcane-specific applications. Authors in (Alencastre-Miranda et al., 2020) employed *AlexNet* for classifying sugarcane billets, achieving a Matthews Correlation

Coefficient (MCC) of 80%. Their work underscores the importance of robust models tailored for sugarcane disease detection, as accurate classification at different growth stages is crucial for disease management. Furthermore, research in related agricultural domains has provided valuable insights into transferable methodologies. For example, (Rao et al., 2021) achieved a remarkable 99.3% accuracy in mango and grape leaf disease classification using *AlexNet*, demonstrating that while crops may differ, the underlying deep learning techniques for disease detection remain broadly applicable. These studies collectively emphasize the potential of deep learning in precision agriculture, paving the way for more efficient, scalable, and automated disease diagnosis systems across various crops.

This paper presents a comparative study on the effectiveness of three state-of-the-art deep learning models—*EfficientNetV2B0* (Tan & Le, 2021), *DenseNet121* (Huang et al., 2017), and *ResNet101V2* (He et al., 2016)—in classifying sugarcane leaf diseases. Each of these models differs in architectural design, computational efficiency, and feature extraction capabilities, making their evaluation in this specific application essential. Since deep learning models for sugarcane disease detection are intended for field deployment, the selection of these three models for comparison was based on their performance characteristics. Specifically, for their ability to good problem generalization, memory usage, and their demonstrated strong performance in various image classification tasks (Deng et al., 2009). A brief description of each model is described below:

EfficientNetV2B0, known for its optimized compound scaling and improved training efficiency, offers a balanced trade-off between accuracy and computational cost. Its ability to achieve high performance with fewer parameters makes it particularly advantageous for deployment in real-world agricultural settings, such as mobile or edge devices. However, its reduced model size compared to deeper architectures may limit its feature extraction capacity in highly complex datasets, potentially impacting accuracy when distinguishing between visually similar disease symptoms.

DenseNet121 introduces dense connectivity, where each layer receives inputs from all preceding layers, improving gradient flow and feature reuse. This architecture enhances learning efficiency and often achieves superior performance with fewer parameters compared to traditional deep networks. A key advantage of *DenseNet121* is its ability to capture intricate leaf disease patterns while maintaining computational efficiency. However, the increased number of connections can lead to higher memory consumption during training, making it less ideal for resource-constrained environments.

ResNet101V2, an improved version of the original *ResNet*, leverages deep residual connections to address the vanishing gradient problem, allowing it to train very deep networks effectively. Its depth enables robust hierarchical feature extraction, which is beneficial for distinguishing complex disease patterns in sugarcane leaves. However, its larger model size and higher computational demands can slow down inference, making it less suitable for real-time or low-power applications.

To summarize these descriptions Table 1 presents a comparison of these models, providing a clearer understanding of their capabilities and limitations.

Based on the results obtained in this study, the key contributions of this work are outlined as follows:

1. **A comprehensive model evaluation**. By systematically comparing *EfficientNetV2B0*, *DenseNet121*, and *ResNet101V2* for sugarcane disease detection.
2. **Identification of the best model**. By demonstrating that *EfficientNetV2B0*, pre-trained on ImageNet, outperforms the other tested models in terms of classification accuracy and generalization.
3. **Performance insights**. Providing a detailed analysis of the tested models' strengths and limitations across various classification metrics offers valuable insights for researchers and practitioners in agricultural applications.
4. **A pre-processed dataset of sugarcane leaf images**. By making available a public ready-to-use dataset of 7,000 sugarcane leaf images categorized into five classes, contributing to the standardization of sugarcane disease classification tasks.
5. **Potential for real-world deployment**. Establishing a foundation for future implementation of the best-performing model in this work as a real-time sugarcane disease detection tool for practical agricultural use.

The remainder of this paper is organized as follows: Section 2 outlines the experimental methodology used to evaluate the tested models and determine the most effective approach for sugarcane disease detection. Section 3 details the implementation of this methodology and presents a statistical analysis of the obtained results. Finally, Section 4 shows the conclusions of this study.

Table 1. Comparison between the evaluated models

Feature	**EfficientNetV2B0**	**DenseNet121**	**ResNet101V2**
Architecture Type	Scaled CNN (EfficientNet family)	Densely Connected CNN	Residual CNN (ResNet family)
Model Depth	Shallow (smallest in EfficientNetV2)	Moderate (121 layers)	Deep (101 layers)
Number of Parameters	~7.1 millions	~8 millions	~44.6 millions
Memory Usage	Low (efficient use of parameters)	High (dense connections require more memory)	Very high (deep network, large footprint)
Training Speed	Fastest	Moderate	Slowest (due to depth)
Inference Speed	Fastest	Slow (due to feature reuse)	Slower than EfficientNet
Generalization Ability	Very high (optimized scaling)	High	High
Best for	Low-resource environments, and efficiency-focused applications	Tasks requiring efficient feature reuse.	Applications needing deep feature extraction.

2 Methodology

The primary objective of this study was to detect diseases in sugarcane leaves by evaluating a set of deep learning models from the literature and selecting the most effective one for accurate classification. A systematic comparison of multiple architectures was conducted to identify a model that balances high accuracy and robustness in real-world agricultural applications. The experimental methodology, illustrated in Figure 1, consisted of three key stages:

1. *Image Pre-processing*: This stage involves cleaning and preparing the input images to enhance their quality before feeding them into the deep learning models to increase dataset diversity, improve generalization, and mitigate potential overfitting during training.
2. *Model Training*: The selected deep learning architectures, as discussed in Section 1, were trained on the pre-processed dataset. Each model underwent hyperparameter tuning to maximize classification performance. Evaluation metrics such as *accuracy*, *precision*, *recall*, and *F1-score* were used to assess their effectiveness in distinguishing between healthy and diseased sugarcane leaves.
3. *Analysis of results*: After training, an analysis of the models' performance based on their classification metrics determined the best-performing model for this task.

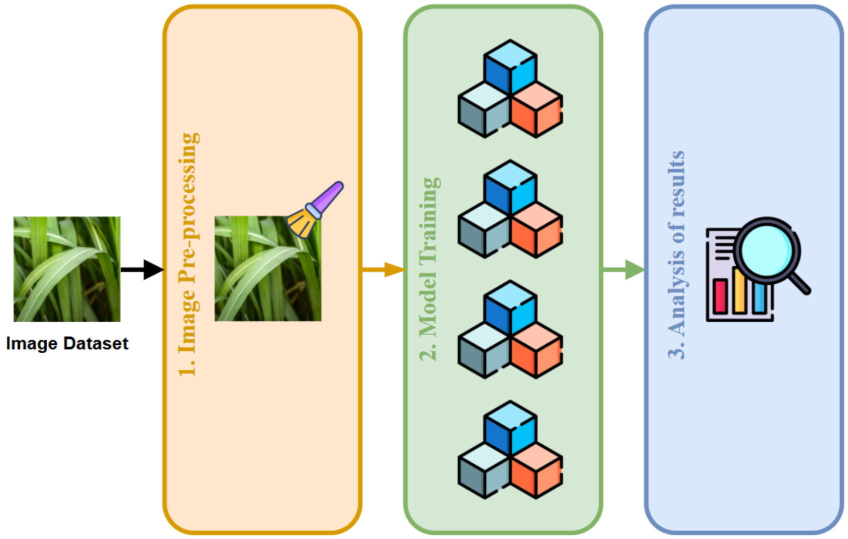

Figure 1. Experimental methodology

3 Experiments and Results

The experimentation followed the methodology described in Section 2, implemented with Python and TensorFlow. The experiments were run on a workstation with an Nvidia RTX 3090 Ti GPU, a 12th generation Intel Core i9 processor, 128GB of RAM, and Ubuntu 22.04 as the operating system. The detailed implementation of this setup is described below.

3.1 Dataset Description

This study used two datasets of sugarcane leaf images sourced from the Kaggle platform. The first dataset comprises images categorized into five classes: *Rust*, *Yellow Leaf Virus*, *Mosaic*, *Red Rot*, and *Healthy Leaves* (Daphal et al., 2022). The second dataset contains images, divided into two classes: *Healthy Leaves* and *Yellow Leaf Virus* (Ruhin, 2023). When combined, the total number of images reached 3,785, with the following class distribution: *Yellow Leaf Virus* (1,339), *Mosaic* (462), *Red Rot* (518), *Rust* (514), and *Healthy Leaves* (952). Figure 2 illustrates examples of the final dataset.

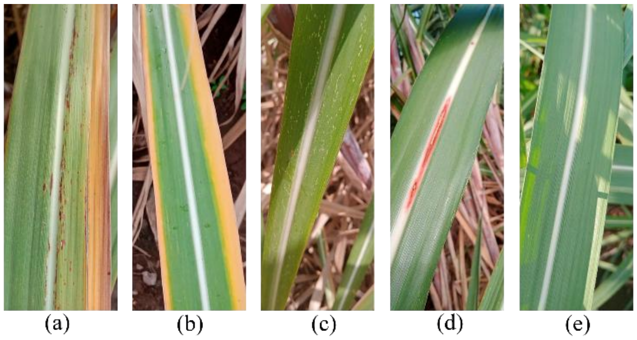
(a) (b) (c) (d) (e)

Figure 2. Image examples: a) Rust, b) Yellow Leaf Virus, c) Mosaic, d) Red Rot, and e) Healty Leaf

3.2 Image Pre-processing

The dataset pre-processing primarily involved data augmentation, a crucial step in improving the model's robustness and generalization capabilities. Given the natural variability in sugarcane leaf appearances due to differences in lighting conditions, angles, and environmental factors, the dataset's diversity was artificially increased through this task. This process helps mitigate overfitting by enabling the model to learn more invariant and discriminative features.

The techniques implemented included image rotation, which simulated different viewing angles of the leaves; zooming, which allowed the model to focus on varying levels of detail within the images; and image skewing, which introduced slight distortions to account for perspective changes and natural deformations. Additionally, transformations like horizontal and vertical flipping, brightness adjustments, and random cropping were used to further enhance variability in the dataset.

As a result of this augmentation process, the dataset was expanded, ensuring a more balanced and comprehensive training set. Each class, was increased to 1,500 images, providing sufficient samples for deep learning models to effectively learn disease-related patterns while reducing the risk of bias toward specific conditions. This enriched dataset played a pivotal role in improving the reliability and robustness of the trained models, making them more capable of accurately identifying sugarcane leaf diseases in real-world scenarios. The final version of this pre-processed dataset is available in an external public repository (Crespo-Sanchez, 2025).

3.3 Model Training

The implemented models consist of two main components: (1) a feature extraction stage that leverages the architectures of the three tested models—*EfficientNetV2B0*, *DenseNet121*, and *ResNet101V2*—implemented in *TensorFlow* with their default parameters as defined by the original authors, and (2) a classification stage using a dense neural network, whose parameters were determined empirically based on classification performance. The experiments include both pre-trained models initialized with *ImageNet* weights (Deng et al., 2009) and models with randomly initialized weights for feature extraction, resulting in six distinct configurations. Table 2 provides a detailed overview of these.

Table 2. Tested model configurations

Configuration	Feature Extraction Model	Classifier Architecture
1	EfficientNetV2B0 (Pre-trained)	A hidden layer of 860 dense units with a ReLU activation function, followed by an output layer with 5 units and a softmax activation function.
2	DenseNet121 (Pre-trained)	A hidden layer with 680 dense units, L2 regularization of 0.01, a dropout rate of 0.4, and ReLU activation, followed by an output layer with 5 units and softmax activation.
3	ResNet101V2 (Pre-trained)	A hidden layer of 800 dense units with a ReLU activation function, followed by an output layer with 5 units and a softmax activation function.
4	EfficientNetV2B0 (Random)	A hidden layer of 860 dense units with a ReLU activation function, followed by an output layer with 5 units and a softmax activation function.
5	DenseNet121 (Random)	A hidden layer with 680 dense units, L2 regularization of 0.01, a dropout rate of 0.4, and ReLU activation, followed by an output layer with 5 units and softmax activation.
6	ResNet101V2 (Random)	A hidden layer of 800 dense units with a ReLU activation function, followed by an output layer with 5 units and a softmax activation function.

To ensure a rigorous and unbiased evaluation of the deep learning models, the *stratified k-fold cross-validation* strategy with a value of $k = 5$ was employed during training and testing. With this value of k, the dataset was partitioned into 80% for training and 20% for testing in each fold, ensuring that the final evaluation was conducted on an independent subset of data not seen during training. Also, within each fold training set, a further split was applied, where 80% of the data was used for model training, and the remaining 20% was allocated for validation. Following this partitioning strategy, each fold consisted of 4,800 images for training, 1,200 for validation, and 1,500 for testing. Using a class-stratified approach ensures that the training, validation, and testing sets maintain the same proportion of images per class. This also allowed to assess the model's generalization ability while reducing the risk of overfitting specific data distributions.

Each configuration underwent 31 repetitions of the *stratified k-fold cross-validation* process, leading to a total of 155 ($k \times 31$) executions. By performing multiple runs, variations due to weight initialization, stochastic optimization, and random data splits were accounted for, ensuring a comprehensive evaluation of each model's performance.

For each execution, standard performance metrics, including *accuracy*, *precision*, *recall*, and *F1-score*, were used to assess the model's classification effectiveness. These metrics were computed in *micro-average* and *macro-average* formats to provide a more detailed understanding of model performance across all disease categories. The *micro-average* aggregates contributions from all classes to calculate a global metric, making it particularly useful when dealing with class imbalances. In contrast, the *macro-average* treats each class equally, providing insight into how well the model performs across different disease categories, regardless of class distribution.

This extensive validation approach ensured the robustness and reliability of the results, allowing to confidently identify the most effective tested model for sugarcane leaf disease classification.

3.4 Analysis of Results

The classification results were analyzed to evaluate each model's performance. A comparative assessment of model *accuracy* is presented below, while additional performance metrics, including *precision*, *recall*, and *F1-score*, as described in Section 3.3, can be found in Appendix B. This supplementary resource provides a more detailed breakdown of the models' behavior across various evaluation criteria, ensuring transparency and facilitating further research.

Figure 3 illustrates the *accuracy* distribution of each model across the training, validation, and test sets, with the corresponding numerical configurations detailed in Table 2. This analysis revealed that all models achieved stable learning behavior throughout the training process, with no signs of overfitting or underfitting. The absence of overfitting suggests that the models effectively generalized to unseen data. The lack of underfitting indicates models successfully captured the essential patterns required for accurate sugarcane leaf disease classification.

When comparing model performances, it is observed that configurations 1, 2, and 4 consistently outperformed 3, 5, and 6, achieving higher average accuracy across all dataset partitions. Notably, configurations 1, 2, and 4 also demonstrated lower variance in their accuracy distributions, signifying stable learning dynamics and reliable convergence toward optimal classification performance. The reduced variance suggests that these models are less sensitive to fluctuations in training conditions and can maintain strong predictive capabilities across different samples.

However, this analysis also identified a bias in the *accuracy* distribution across all configurations, suggesting potential underlying factors affecting performance. This bias is evident in the boxplots of Figure 3, where all accuracy distributions exhibit left skewness. This skewness may arise from variations in disease symptom visibility or inherent architectural biases within the models. Further investigation is needed to determine whether additional augmentation strategies or alternative feature extraction methods could mitigate this effect and enhance overall classification robustness.

By analyzing these findings, valuable insights into the strengths and limitations of each model are gained to refine this approach to sugarcane leaf disease classification and improve the reliability of automated disease detection systems.

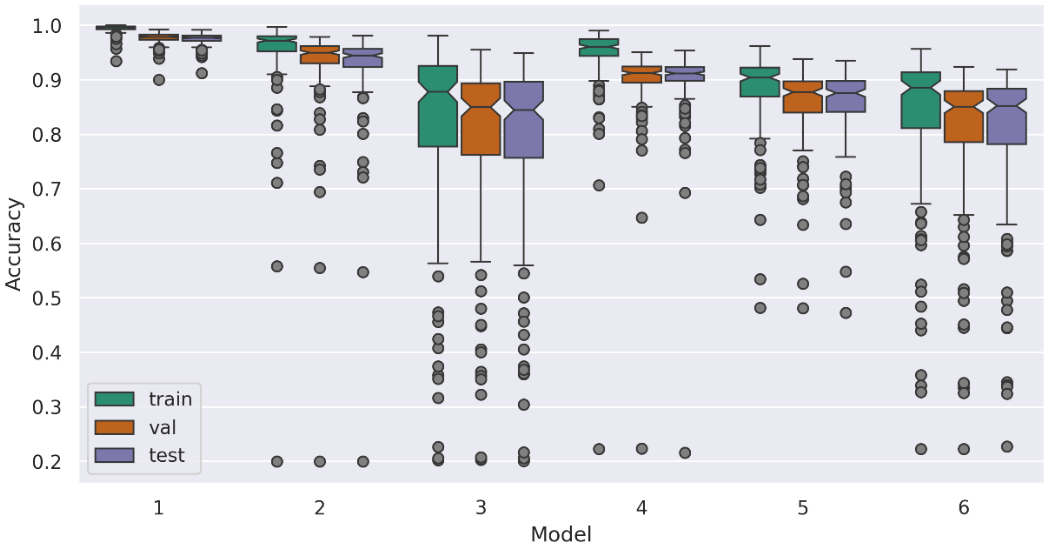

Figure 3. Tested configurations' accuracy distribution

The skewness in the accuracy distributions prompted a normality test to check for conformity to a normal distribution. The Shapiro-Wilk test was applied to each set of results —training, validation, and test accuracies for all configurations— which is commonly used to detect deviations from normality, particularly in small to moderate sample sizes.

The results of the Shapiro-Wilk test revealed that, in all cases, the computed *p-values* were less than 0.001, indicating strong evidence against the null hypothesis of normality. This confirms that none of the distributions satisfied the normality assumption, reinforcing the initial observation of skewed distributions. Given this finding, non-parametric statistical measures provided a more robust approach to summarize and compare models' performance.

Accordingly, Table 3 presents a statistical summary of the accuracy results, reporting the median ($\tilde{x}$) and interquartile range (IQR) for each model across the training, validation, and test image sets. The median, as a central tendency measure, provides a more reliable indicator of model performance in the presence of skewed data, as it is less sensitive to outliers compared to the mean. The IQR, which represents the range between the first and third quartiles, captures the spread and variability of the accuracy values, offering insights into each model's consistency. A smaller IQR suggests more stable performance, while a larger IQR indicates greater variability in accuracy across different training iterations.

These statistical measures help provide a more accurate and fair assessment of model effectiveness, ensuring that conclusions are not disproportionately influenced by extreme values or distributional biases.

Table 3. Accuracy results for each model architecture

Configuration	Accuracy					
	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	**0.996**	**0.005**	**0.979**	**0.009**	**0.977**	**0.009**
2	0.972	0.028	0.950	0.031	0.944	0.033
3	0.877	0.148	0.850	0.131	0.844	0.139
4	0.960	0.031	0.912	0.029	0.912	0.025
5	0.904	0.053	0.877	0.057	0.876	0.056
6	0.885	0.102	0.850	0.093	0.852	0.101

Among the six configurations tested, configuration 1 demonstrated the best overall performance, achieving the highest $\tilde{x}$ across all dataset partitions: the training, validation, and test sets. Additionally, this one exhibited the smallest IQR, suggesting not only strong overall performance but also consistent accuracy across different folds and data splits. This consistency is particularly valuable for ensuring that the model can generalize well to unseen data, reducing the risk of overfitting to specific subsets of the training data. While configuration 1 clearly outperformed the others, it is worth noting that all of them yielded reasonably substantial results, with the lowest $\tilde{x}$ in the test set being 0.844, achieved by configuration 3. This still represents a strong level of classification performance, indicating that all were capable of accurately detecting sugarcane leaf diseases, albeit with varying degrees of effectiveness.

Despite the relatively strong performance of the models, a more nuanced evaluation was required to determine if there were any statistically significant differences in their performance. To this end, the Kruskal-Wallis test helped assess whether significant differences existed among the performances on the test image set. The Kruskal-Wallis test is particularly appropriate when normality assumptions are violated, as was the case in these execution results.

Following the Kruskal-Wallis test, Dunn's post hoc test was applied to conduct pairwise comparisons between the distributions of the test set accuracies. This allowed for a more detailed exploration of which configuration specifically differed from each other in terms of performance. The results of these tests are illustrated in Figure 4, providing a visual representation of the pairwise differences and highlighting whether any specific configuration outperformed others with statistical significance. Conducting these additional analyses ensured that the observed differences in performance were not due to random chance and could be confidently attributed to the inherent characteristics of the models themselves.

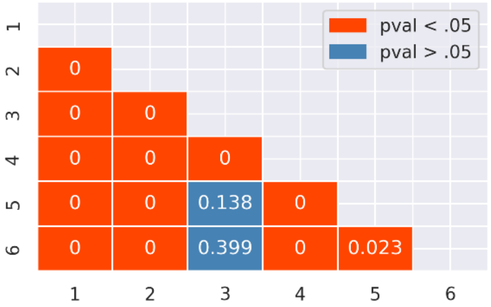

Figure 4. Dunn's post hoc test p-values

The p-values obtained from the post hoc tests (see Figure 4) provide critical insight into the statistical significance of the differences in performance between the six configurations. Most of these pairs showed significant differences, with p-values being less than 0.05, indicating that the observed differences in accuracy were unlikely to have occurred by chance. This suggests that certain models consistently outperformed others across multiple folds and data partitions. However, there were notable exceptions, particularly when comparing configurations 3 vs. 5 and 3 vs. 6. In these cases, the p-values were higher than the 0.05 threshold, indicating no statistically significant differences in performance between these pairs. This outcome suggests that either configuration 3, 5, or 6 could be chosen interchangeably without significantly impacting the overall performance concerning the metrics considered in this study.

Given that configuration 1 achieved the highest $\tilde{x}$ across all dataset partitions —training, validation, and test— and demonstrated statistically significant differences in performance compared to the others in the post hoc tests, this one was selected as the best of the tested models for the task of detecting diseases in sugarcane. The superior performance of configuration 1, coupled with the consistent results observed across different data splits, suggests that it is the most reliable choice for real-world deployment in disease detection applications. Its ability to distinguish sugarcane leaf diseases with high accuracy and consistency, integrated with its statistically significant advantage over the other configurations, makes it the best model, at least in this study, for improving disease management in sugarcane agriculture.

4 Conclusions

This study presented a comparative analysis of deep learning models from the literature to evaluate their effectiveness in detecting sugarcane diseases. Among the tested models, *EfficientNetV2B0* pre-trained on the *ImageNet* dataset achieved the highest performance across commonly used classification metrics. In addition to delivering the best results, this model demonstrated strong generalization capabilities for sugarcane disease detection.

The experiments demonstrated that these techniques are computationally feasible and can be successfully applied to agricultural disease detection tasks. By leveraging deep learning models for sugarcane leaf disease classification, the study highlighted the potential for developing robust tools for early disease detection. Such tools are critical for supporting the growth and sustainability of the agricultural sector, as early identification of diseases can lead to more timely and efficient interventions, ultimately improving crop yield and reducing the economic impact of crop diseases. The findings suggest that deep learning-based image classification models are not only effective but also scalable, offering a promising solution for large-scale deployment in real-world agricultural settings. These advancements could significantly contribute to more informed decision-making and precision agriculture, enhancing productivity and developing sustainable farming practices.

Future work includes deploying the top-performing model as a real-time sugarcane disease detection system. This could be implemented as a publicly accessible web or mobile application, allowing farmers and agricultural professionals to quickly diagnose plant health by capturing and analyzing leaf images. Additionally, integrating this model into an easy-to-use interface would facilitate widespread adoption, making advanced deep learning-based diagnostics more accessible to the agricultural sector.

References

INEGI, 2025, Economía y Sectores Productivos. Accessed Mar 05, 2025, from https://www.inegi.org.mx/temas/agricultura/.

Viswanathan, R., & Rao, G. P. (2011). Disease scenario and management of major sugarcane diseases in India. *Sugar Tech*, *13*, 336-353.

Srivastava, S., Kumar, P., Mohd, N., Singh, A., & Gill, F. S. (2020). A novel deep learning framework approach for sugarcane disease detection. *SN Computer Science*, *1*, 1-7.

Li, X., Li, X., Zhang, S., Zhang, G., Zhang, M., & Shang, H. (2023). SLViT: Shuffle-convolution-based lightweight Vision transformer for effective diagnosis of sugarcane leaf diseases. *Journal of King Saud University-Computer and Information Sciences*, *35*(6), 101401.

Hemalatha, N. K., Brunda, R. N., Prakruthi, G. S., Prabhu, B. B., Shukla, A., & Narasipura, O. S. J. (2022). Sugarcane leaf disease detection through deep learning. In *Deep Learning for Sustainable Agriculture* (pp. 297-323). Academic Press.

Alencastre-Miranda, M., Johnson, R. M., & Krebs, H. I. (2020). Convolutional neural networks and transfer learning for quality inspection of different sugarcane varieties. *IEEE Transactions on Industrial Informatics*, *17*(2), 787-794.

Rao, U. S., Swathi, R., Sanjana, V., Arpitha, L., Chandrasekhar, K., & Naik, P. K. (2021). Deep learning precision farming: grapes and mango leaf disease detection by transfer learning. *Global transitions proceedings*, *2*(2), 535-544.

Tan, M., & Le, Q. (2021, July). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.

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).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14* (pp. 630-645). Springer International Publishing.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

Daphal, Swapnil; Koli, Sanjay (2022), “Sugarcane Leaf Disease Dataset”, Mendeley Data, V1, doi: 10.17632/9424skmnrk.1. Accessed Jan 28, 2025, from https://data.mendeley.com/datasets/9424skmnrk/1.

Ruhin, W., (2023), Sugarcane leaf diseases. Accessed Jan 28, 2025, from https://www.kaggle.com/datasets/wardaruhin/sugarcane-leaf-diseases.

Crespo-Sanchez, M. (2025). Sugarcane Disease Classification Dataset (Versión 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14964141.

Appendix A Nomenclature definition

Table 4 shows the definitions of the acronyms used in this paper.

Table 4. Acronym definitions

Acronym	**Definition**
CNNs	Convolutional Neural Networks
AUC	Area Under the Curve
MCC	Matthews Correlation Coefficient
ReLU	Rectified Linear Unit
L2 regularization	regularization technique to avoid overfitting
IQR	Interquartile range
$\tilde{x}$	Median

Appendix B Classification metrics' results

This section shows the rest of the performance metrics used for the evaluation of the tested models.

Table 5. Micro average precision results

Micro Average Precision						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.979	0.010	0.977	0.010
2	0.972	0.028	0.950	0.032	0.945	0.033
3	0.878	0.148	0.850	0.131	0.845	0.140
4	0.961	0.031	0.913	0.030	0.912	0.025
5	0.904	0.053	0.878	0.058	0.876	0.057
6	0.886	0.102	0.851	0.094	0.853	0.101

Table 6. Micro average recall results

Micro Average Recall						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.979	0.010	0.977	0.010
2	0.972	0.028	0.950	0.032	0.945	0.033
3	0.878	0.148	0.850	0.131	0.845	0.140
4	0.961	0.031	0.913	0.030	0.912	0.025
5	0.904	0.053	0.878	0.058	0.876	0.057
6	0.886	0.102	0.851	0.094	0.853	0.101

Table 7. Micro average F1-Score

Micro Average F1 Score						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.979	0.010	0.977	0.010
2	0.972	0.028	0.950	0.032	0.945	0.033
3	0.878	0.148	0.850	0.131	0.845	0.140
4	0.961	0.031	0.913	0.030	0.912	0.025
5	0.904	0.053	0.878	0.058	0.876	0.057
6	0.886	0.102	0.851	0.094	0.853	0.101

Table 8. Macro average precision results

Macro Average Precision						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.980	0.009	0.978	0.009
2	0.973	0.025	0.953	0.029	0.948	0.028
3	0.896	0.116	0.868	0.103	0.862	0.108
4	0.962	0.028	0.916	0.028	0.916	0.024
5	0.913	0.040	0.892	0.041	0.889	0.039
6	0.896	0.072	0.867	0.057	0.866	0.063

Table 9. Macro average recall

Macro Average Recall						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.979	0.010	0.977	0.010
2	0.972	0.028	0.950	0.032	0.945	0.033
3	0.878	0.148	0.850	0.131	0.845	0.140
4	0.961	0.031	0.913	0.030	0.912	0.025
5	0.904	0.053	0.878	0.058	0.876	0.057
6	0.886	0.102	0.851	0.094	0.853	0.101

Table 10. Macro average F1-Score

Macro Average F1-Score						
Configuration	Training		Validation		Test	
	$\tilde{x}$	IQR	$\tilde{x}$	IQR	$\tilde{x}$	IQR
1	0.997	0.005	0.979	0.010	0.977	0.010
2	0.972	0.028	0.950	0.032	0.944	0.033
3	0.878	0.150	0.849	0.136	0.846	0.141
4	0.961	0.031	0.913	0.029	0.912	0.026
5	0.904	0.054	0.877	0.058	0.876	0.058
6	0.886	0.110	0.850	0.101	0.849	0.100