The Impact of Image Input Size on the Efficiency and Suitability of Deep Learning Models in Diabetic Retinopathy*

Thiago Araújo[1], Beatriz Schaan[12], Carla M. D. S. Freitas[1], and Philippe Navaux[1]

[1]Informatics Institute, Federal University of Rio Grande do Sul, Brazil,

[2]Clinical Research Center, Hospital de Clínicas de Porto Alegre, Brazil

{tsaraujo, carla, navaux}@inf.ufrgs.br, bschaan@hcpa.edu.br

Abstract. *The high computational cost and environmental impact of deep learning (DL) models hinder their adoption for diabetic retinopathy (DR) classification, underscoring the need for models that strike a balance between performance and efficiency. This study evaluates how image input size affects the effectiveness of convolutional neural networks (CNNs) in classifying DR. Using a dataset of 2,579 Brazilian fundus images from the Federal University of São Paulo (UNIFESP), we assessed the performance of the EfficientNetV2B0 and MobileNet models across five input image sizes. We measured the area under the ROC curve (AUC), energy consumption, and training time. For the MobileNet model, an input resolution of 300×300 achieved an AUC of 0.91, resulting in a 69% reduction in energy consumption and a 74% reduction in training time. EfficientNetV2B0 achieved an AUC of 0.93 at the same resolution, with reductions in energy consumption and training time of 66% and 70%, respectively. These results highlight the importance of optimizing input size as a simple yet effective way to improve the efficiency and sustainability of DL diagnostic models while maintaining high predictive accuracy.*

1. Introduction

DL has significantly advanced medical image analysis, enabling automated DR screening systems with performance comparable to that of ophthalmologists. However, the computational cost of training and deploying these models limits their large-scale adoption, especially in public healthcare systems where infrastructure limitations and operational expenses are significant constraints.

Focusing just on performance neglects other vital aspects of model suitability. Therefore, metrics such as energy consumption and training time are vital for developing efficient models. This focus aligns with the principles of Green AI, which emphasizes reducing the environmental and computational impacts of AI system development [Bolón-Canedo et al. 2024]. Although hyperparameter optimization is common to improve accuracy in medical imaging [Asiri et al. 2024, Julian and Devipriya 2024], its potential to improve efficiency is still largely unexplored [Viviani and Ranganathan 2024].

*This work was supported by CAPES - Funding Code 001, CIARS RITEs/FAPERGS and CI-IA FAPESP-MCTIC-CGI-BR projects. Some experiments in this work used the PCAD infrastructure, http://gppd-hpc.inf.ufrgs.br, at INF/UFRGS.

This work investigates the effect of input image size on the performance and efficiency of DR classification using *MobileNet* and *EfficientNetV2-B0*. We tested different image sizes for each model, measuring AUC, training time, and total energy consumption. The results indicate that image size has a significant impact on computational efficiency. Choosing the right image resolution is therefore important for balancing diagnostic performance with training time and energy consumption.

2. Methodology

2.1. Dataset and Preprocessing

We conducted an experiment using a UNIFESP dataset comprising 2,579 color fundus images collected from diabetic patients between 2010 and 2020. The dataset focused on images centered on the macula and included labels from clinical experts for the binary task of identifying referable DR. To prevent data leakage and ensure a rigorous evaluation, we employed a patient-stratified 70-30 train-test split. This approach ensured that all images from a single patient were assigned exclusively to either the training set or the test set, thereby avoiding overlap.

2.2. Model Architecture

All experiments employed a pipeline architecture that facilitates the interchange of backbone networks. This pipeline consists of the following components:

- **Input Layer**: Accepts retinal fundus images at a specified resolution.
- **Backbone Network**: A pre-trained CNN for feature extraction. This study utilizes the *MobileNet* and *EfficientNetV2B0* architectures, selected for their effective balance of performance and efficiency.
- **Global Average Pooling**: Reduces the spatial dimensions of the final feature maps.
- **Dense Output Layer**: A single neuron with a **sigmoid** activation function to produce the final prediction.

The hyperparameter under investigation was the input image size, a key factor for efficiency explored in related literature [Wojciuk et al. 2024, Bahrami et al. 2025]. We evaluated five input resolutions for each model, ranging from 100×100 to 500×500 pixels, to measure its impact on performance and computational efficiency.

2.3. Experimental Environment

Our experiments were conducted on an NVIDIA Grace Superchip, utilizing only its CPU cores. To ensure reproducibility, we containerized the training environment using Singularity, with the code available on GitHub[1]. Each configuration was trained for 200 epochs over ten independent runs, employing a batch size of 288 due to memory constraints. We report the mean and standard deviation of the metrics across all runs.

[1]https://github.com/tsilvaaraujo/eramia-rs2025

3. Results

The experiments show a balance between diagnostic accuracy and computational efficiency. Both the input image size and the model's architecture affect this balance. Larger images generally improve performance, but both models reach a point where higher resolutions offer little additional benefit. Performance improves noticeably as image size increases from 100×100 to 300×300 pixels, but further increases bring only minor gains in AUC while significantly increasing computational demands.

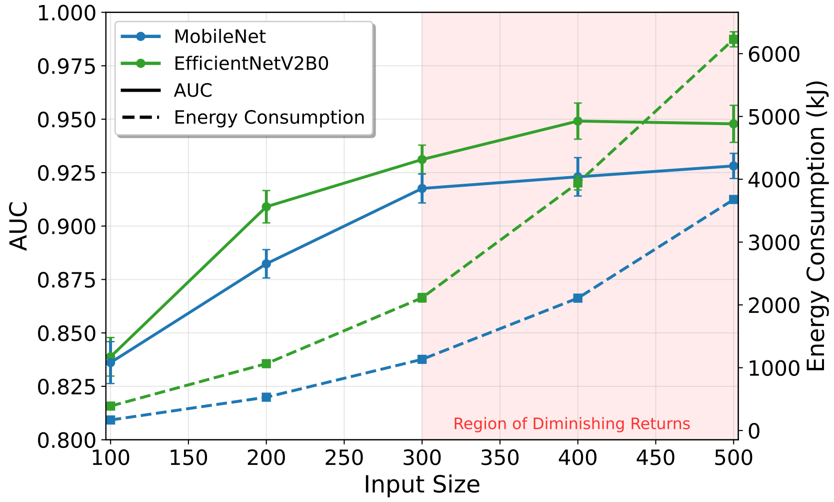

Figure 1. AUC and energy consumption of each input size resolution.

As depicted in Figure 1, where solid lines represent model AUC and dashed lines indicate energy consumption, along with the data presented in Table 1, it is clear that the input size significantly influences the suitability of models for various operational contexts. The findings reveal that a resolution of 300×300 pixels serves as a critical threshold. At this resolution, MobileNet provides an efficient solution, achieving a robust AUC of 0.918 while consuming only 1,133 kJ of energy and completing training in 3,744 seconds. Conversely, EfficientNetV2B0, though it achieves a higher accuracy (0.931 AUC) at the same resolution, incurs a substantial cost, consuming 2,114 kJ of energy and requiring 7,639 seconds of training time, almost double the resources for a marginal performance improvement. This trade-off illustrates that input size is not merely a hyperparameter; rather, it is a fundamental factor in aligning model selection with application priorities, whether they emphasize computational efficiency or optimal diagnostic performance.

In summary, the resolution of the input image plays an important role in determining the operational profile of a deep learning model for diabetic retinopathy screening. The 300×300 pixel threshold consistently stands out as the most significant configuration, marking the point at which computational costs begin to outweigh diagnostic advantages. This insight offers a clear guideline for implementation: practitioners can choose a model based on their specific resource limitations and performance needs, utilizing input size as a key lever to balance efficiency and accuracy.

4. Conclusion

This study demonstrates that the input image size plays a crucial role in striking a balance between diagnostic accuracy and computational efficiency in diabetic retinopathy

Table 1. Training time for each input size.

Model	Size	Training time (s)
EfficientNetV2B0	100x100	1,354.00 +- 32.88
	200x200	3,501.09 +- 63.37
	300x300	7,639.08 +- 66.01
	400x400	15,823.01 +- 170.77
	500x500	25,724.10 +- 265.28
MobileNet	100x100	578.21 +- 24.57
	200x200	1,698.74 +- 11.89
	300x300	3,743.78 +- 14.39
	400x400	7,322.23 +- 23.27
	500x500	14,560.78 +- 34.97

screening. Performance improves as image resolution increases, but the advantages beyond 300×300 pixels are minimal and come at a significant cost in terms of energy use and training time. At this resolution, MobileNet provides a strong balance of accuracy and efficiency, while EfficientNetV2B0 achieves slightly better accuracy at nearly twice the computational expense. Improving model suitability is relevant for public health systems, where limited resources and growing demand make sustainable solutions essential. Future research will extend this efficiency-oriented framework to other medical imaging domains and investigate automated optimization methods for dynamically balancing accuracy, cost, and energy efficiency.

References

Asiri, A. A., Shaf, A., Ali, T., Aamir, M., Irfan, M., and Alqahtani, S. (2024). Enhancing brain tumor diagnosis: an optimized cnn hyperparameter model for improved accuracy and reliability. *PeerJ Computer Science*, 10:e1878.

Bahrami, A., Rakhshaninejad, M., Ghousi, R., and Atashi, A. (2025). Enhancing machine learning performance in cardiac surgery icu: Hyperparameter optimization with metaheuristic algorithm. *PloS one*, 20(2):e0311250.

Bolón-Canedo, V., Morán-Fernández, L., Cancela, B., and Alonso-Betanzos, A. (2024). A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, 599:128096.

Julian, A. and Devipriya, R. (2024). Exploring hyperparameter tuning strategies for optimizing model performance. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies*, pages 1–4. IEEE.

Viviani, L. A. and Ranganathan, P. (2024). Evaluating the suitability of lstm models for edge computing. In *2024 Cyber Awareness and Research Symposium (CARS)*, pages 1–7. IEEE.

Wojciuk, M., Swiderska-Chadaj, Z., Siwek, K., and Gertych, A. (2024). Improving classification accuracy of fine-tuned cnn models: Impact of hyperparameter optimization. *Heliyon*, 10(5).