Transfer Learning with EfficientNet-B0 for Histopathological Patch Classification (PathMNIST): Proposal and Implementation

José Alejandro Torres Muñoz - Fiorella Valentina Trujillo Abello

1. Summary

The automatic classification of histopathological images is a key challenge in digital pathology and computer-assisted diagnosis. This paper presents a transfer learning-based method using EfficientNet-B0, a model pre-trained on ImageNet and adapted to classify the 9 classes of the PathMNIST dataset. The solution is based on modifying the output layer and applying data augmentation techniques to improve generalization.

Experimental results show that the model achieves 90.24% accuracy on the test set, along with a macro F1 score of 84.83% and a weighted F1 score of 89.78%, demonstrating solid and balanced performance on both frequent and minority classes. Furthermore, a detailed analysis of the confusion matrix reveals that classification errors are concentrated in tissues with morphological similarities, which is clinically significant.

It is concluded that EfficientNet-B0 represents an efficient and competitive alternative to heavier architectures such as ResNet or DenseNet, contributing to the literature with a lower computational cost solution while maintaining high performance.

2. Introduction

Histopathology is an inherent diagnostic tool for many disorders, particularly cancers. Manual examination of hematoxylin and eosin (H&E)-stained slides is tedious, dependent on the pathologist's experience, and prone to interobserver variability. Therefore, computer vision methods aim to improve the objectivity, speed, and reproducibility of the diagnostic process.

The PathMNIST dataset, derived from MedMNIST, is a commonly used benchmark in the field for testing automatic classification techniques on histological images. Since it consists of patches representative of various tissue types and conditions, it allows for the evaluation of the generalization ability of models.

This work suggests the use of EfficientNet-B0 as an efficient convolutional neural network that, through its composite scaling approach, achieves an optimal compromise between accuracy and computational cost. The proposal consists of adapting this model through transfer learning, adjusting the final layer for 9 classes and applying data augmentation to improve learning.

3. Related jobs

The most prominent approaches in the literature include:

- • ResNet — (He et al., 2015). Residual architectures that allow deep networks to be trained without gradient degradation.
- • DenseNet — (Huang et al., 2016). Dense connectivity that improves gradient flow and promotes feature reuse.
- • EfficientNet — (Tan & Le, 2019). Efficient model that introduces composite scaling, offering competitive performance with fewer parameters.
- • Vision Transformers — (ViT, Dosovitskiy et al., 2020). Self-attention-based architectures, powerful in scenarios with large volumes of data.

In general, studies agree that transfer learning from pre-trained models in ImageNet is essential in the biomedical domain, where data availability is limited. Our contribution lies in demonstrating that EfficientNet-B0, despite its lower complexity, achieves performance close to or superior to that of heavier models, validating its suitability for environments with limited computational resources.

4. Materials and methods

4.1. Data set.

- Dataset: PathMNIST (MedMNIST).
- Important features: H&E patches in RGB, 9 classes, official splits (train/val/test).
- This was used because PathMNIST is a standardized benchmark; using the official splits allows for comparability with previous work and ensures that the evaluation is consistent with the literature.

4.2. Preprocessing

a) Resize to 224×224 px
   - This is because EfficientNet-B0 was pre-trained on ImageNet with 224×224; using this resolution allows full advantage to be taken of the pre-trained weights and architecture (same input dimensions, filters, and adapted layers).
   - The expectation was that this would result in better convergence and transferability of low-level filters (edges, textures).
   - Normalization with ImageNet mean/standard deviation.

   Normalization with ImageNet mean/standard deviation

   Normalization equalizes the statistical distribution of inputs seen by the model during pretraining, stabilizing training and avoiding large jumps in activations.

b) Stain normalization

   This is because H&E slides show staining variations between laboratories; staining standardization reduces domain shift and improves generalization in clinical and multicenter settings.

c) Data increases

- RandomHorizontalFlip, RandomRotation, ColorJitter (brightness/contrast), etc. were applied here.
- With the aim of:
  - Rotation/flip simulate variability in orientation during acquisition.
  - ColorJitter simulates small differences in staining and exposure.
  - And the benefit increases robustness against variability and reduces overfitting.

4.3. Architecture and fine-tuning strategy

a) Backbone - Pre-trained EfficientNet-B0: Offers a good balance between accuracy and efficiency; lower computational cost than ResNet50 while maintaining high representation capacity.
b) Replacing the classification head: This maintains convolutional weights and retrains the head, allowing the model to leverage general features and learn domain-specific separation.
c) Training strategy:
   - Phase 1: Freeze the convolutional layers and train only the head for N epochs with moderate LR, which allowed us to adapt the head without disturbing overall weights.
   - Phase 2: Gradually thawing out final blocks and fine-tuning with lower LR helped us improve adaptation without "forgetting" pre-training.

   Progressive fine-tuning reduces the risk of catastrophic learning and leads to better results than complete retraining from scratch when data is limited.

4.4. Hyperparameters

- Batch size = 32
  Offers a good balance between gradient stability and memory usage on GPUs such as Tesla T4; reduces gradient variance with respect to very small batches..
- Optimize: Adam (lr = 0.0005)
  This optimizer is adaptive and robust for initial learning rates; in fine-tuning, it usually works well with LR values less than 1e-3.
  LR = 5e-4 It is small enough to avoid destabilizing pretrained weights, but large enough to converge in a few epochs.
- Epochs: 2
  It is a short time, but it already shows partial convergence..
- Loss function: CrossEntropyLoss.
  This is a standard for multi-class classification; if there is imbalance, weight can be used in the loss or FocalLoss.
  can be used.

4.5. Contribution to the literature

The contribution of our work with respect to what has been previously published in the literature focuses on helping to establish an efficient and lightweight model, such as EfficientNet-B0, to present metrics capable of tackling a very complex biomedical challenge, such as PathMNIST, without necessarily having to rely on deeper and more costly architectures, such as those of ResNet-50, DenseNet-121, or even Transformers. Our configuration demonstrates that a low-complexity structure is robust enough to deliver solid performance, while its advantage is that it requires fewer computational devices and allows for faster training overall. This is a matter of practical interest in biomedical practice, as it is true that the state-of-the-art hardware of many institutions in this field is limited and diagnostic accuracy must be combined with cost-effectiveness.

Another essential contribution is the clinical consistency of the results. The calculation of the confusion matrix shows that the model errors are not arbitrary, as they are concentrated in histologically similar classes, such as cancer-associated stroma confused with normal colonic mucosa or adenocarcinomatous epithelium. These error patterns reflect real limitations in clinical practice and add value by demonstrating that the neural network is capturing morphologically significant features. Thus, the contribution is not only an efficient technical solution, but also a model whose results make sense from a biological and clinical point of view, which favors its acceptance in hospital settings.

Finally, the proposal contributes at the methodological level by implementing a clear and reproducible process based on transfer learning, applicable data augmentations, and a structured validation scheme. This provides researchers or healthcare professionals with a practical roadmap for reproducing or adapting the model to different biomedical contexts. Unlike studies that present highly complex solutions that are difficult to replicate outside the academic sphere, this work emphasizes applicability and accessibility, making a direct contribution to the existing literature by opening up the possibility of applying computer vision models in institutions with limited resources without compromising diagnostic capability.

5. Validation methodology.

5.1. Validation protocol applied.

The official PathMNIST splits (train/val/test) were used to ensure comparability with the literature. The best checkpoint was selected based on the validation metric (val_acc/val_loss). Inference on the Test Set with the model saved as best_model and calculation of metrics: Accuracy, F1_macro, F1_weighted.

6. Results.

The final consolidated results are presented:

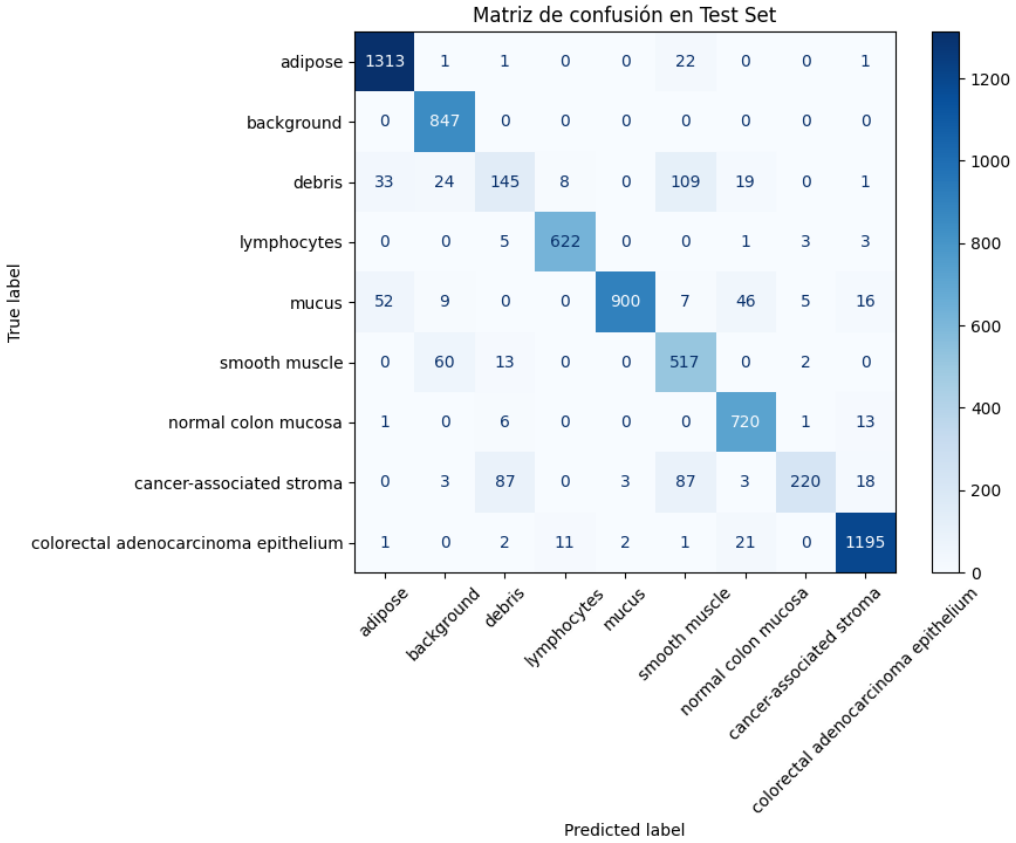
- Accuracy (Test): 90.24%

- F1_macro: 84.83%
- F1_weighted: 89.78%

An accuracy of 90.24% indicates that 9 out of 10 patches were correctly classified in the Test Set, which is competitive for PathMNIST and consistent with a well-tuned transfer learning pipeline.

The difference between F1_macro (0.8483) and F1_weighted (0.8978) shows that the model performs better in classes with greater representation (increases the weight). The lower F1_macro suggests that there are less represented classes where F1 is significantly lower, which is common in unbalanced multiclass datasets.

Confusion matrix:



Adipose, background, and colorectal adenocarcinoma epithelium show strong diagonal patterns with very few confusions; the model distinguishes these textures and structures well.

Cancer-associated stroma presents confusions with normal colon mucosa and colorectal adenocarcinoma epithelium. This is interpreted as being because the tumor stroma shares a similar extracellular composition and cellular arrangement with the affected mucosa and changes in the tumor epithelium; the model reflects this ambiguity.

Debris and mucus show errors with background or adipose in some cases, which may be due to artifacts in the preparation or similarity of color/tone.

The errors do not appear arbitrary but rather aligned with histological complexity: this makes it easier to prioritize which classifications should undergo human verification.

7. Conclusions.

The EfficientNet-B0 model demonstrates that it is possible to achieve solid results in PathMNIST without resorting to complex architectures. Its 90.24% test accuracy positions it as a competitive alternative to ResNet and DenseNet, but with lower computational cost.

The confusion matrix analysis adds value: errors occur in classes with real morphological similarities, indicating that the model's predictions make clinical sense. This not only supports the model's effectiveness but also opens the door to practical applications in hospital settings.

The analysis confirms that classification errors are concentrated in histologically similar classes, reinforcing the validity of the model. Compared to the literature, our approach provides computational efficiency and a viable solution for resource-limited environments.

8. Referencias

- He, K. et al. (2015). Deep Residual Learning for Image Recognition. CVPR.
- Huang, G. et al. (2016). Densely Connected Convolutional Networks. CVPR.
- Tan, M. & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for CNNs. ICML.
- Dosovitskiy, A. et al. (2020). An Image is Worth 16x16 Words. ICLR.
- MedMNIST: https://medmnist.github.io/