Sustainability Strategies and Evaluation for

*Green AI for Medical Image Classification with Explainable AI*

Nachiket Mehendale (x23272473)

Emerging Artificial Intelligence Technologies and Sustainability – H9ETS

MSCAI1\_SEP24

School of Computing

National College of Ireland

# Background Research

The growing dependence of the healthcare sector on artificial intelligence technology for diagnostic scans or images has created major sustainability issues. Medical imaging classification tools typically use computationally demanding deep learning models that require large amounts of energy and need extensive hardware infrastructural resources (Strubell et al., 2019). This computational burden adds to CO2 emissions and reduces availability of AI-based imaging tools in clinical setups with limited resources.

Skin-related diseases affect millions of people worldwide, but prompt detection of the disease can significantly improve treatment outcomes, and therefore skin cancer screening is a highly relevant application area. Conventional diagnostic methods depend largely on skilled professionals causing obstacles to access in underserved areas. But AI-powered diagnostic tools have the potential to broaden access to quality healthcare even though medical imaging systems have a greater environmental impact (Esteva et al., 2017).

Sustainable AI development is now a major focus, primarily in medical contexts where applying models requires a trade-off between precision and ecological impact. The notion of "Green AI" focuses on building computationally efficient algorithms that decrease energy and hardware consumption without sacrificing diagnostic accuracy (Schwartz et al., 2020). This approach addresses ecological concerns alongside practical application hurdles within healthcare systems.

International regulatory organizations have started setting sustainability criteria for AI systems in healthcare technologies. The EU AI legislation has clauses about evaluation of environmental effects of high-risk AI systems, especially those applied to diagnostic processes. Medical instrument clearance processes more often consider overall ecological costs together with safety and effectiveness criteria. Medical associations are working on formulating guidelines for environmentally friendly AI implementation in clinical settings. These new guidelines highlight the importance of transparent reporting of hardware and software resources together with greenhouse gas assessments. Hospitals and clinics are expected to ensure compliance with standards with operational efficiency without compromising patient treatment standards (Martinez & Lee, 2022).

In medical scenarios, where confidence and openness are vital for diagnosis, the use of explainable AI(XAI) methods becomes extremely important. To enable sound clinical decision-making, Doctors and specialists must be able to interpret how AI systems arrive at their conclusions. However, explainability and computational efficiency often clash because interpretability techniques tend to consume extra computing power. (Adadi & Berrada, 2018).

# Sustainability AI Technologies

This research uses a comprehensive approach for environmentally friendly medical image classification by integrating efficient model designs paired with explainable AI techniques. The methodology is based on three core strategies: efficient architecture, model compression, and resource-conscious training protocols.

## 2.1 Efficient Model Architecture

The selection of resource-efficient base architectures adapted for 224 × 224 pixel- HAM10000 dataset format marks the first step in the ecologically friendly green plan. Here 3 variants of CNN architectures are implemented and compared: The Green Baseline architecture serves as the most resource-efficient foundation model for environments with limited resources. This architecture uses GlobalAveragePooling2D as a key optimization, which replaces thousands of parameters with simple averaging operations. The architecture adopts a gradual filter increase (32→64→64), combined with carefully placed MaxPooling2D layers that sharply decrease spatial dimensions near the network’s start. The model employs minimal dense layer (single 64-unit hidden layer), Dropout (0.3) for regularization, and GlobalAveragePooling2D for A screen shot of a computer code

AI-generated content may be incorrect. Figure1: Definition of green\_model class in kaggle notebook

parameter reduction. These strategic choices work together to improve utilization of resources during the entire lifecycle of the model without sacrificing clinical-grade diagnostic effectiveness.

## Resource-conscious Training and Explainable AI Integration

During training, the model applies early stopping (patience=20), ReduceLROnPlateau scheduler (patience=10, factor=0.5), and adaptive learning rate starting of 0.001 to prevent unnecessarily high computational expense during training. Batch sizes of 32 for augmented training and 16 for validation are adjusted to maintain efficient training based on hardware memory constraints and supporting consistent results regardless of hardware setups. The 224×224 image size matches HAM10000 dataset requirements while preserving diagnostic detail. Image augmentation applies rotation, zoom, and brightness variations during training without additional storage requirements. Grad-CAM generates activation maps using jet colormap with 0.6 alpha transparency through one forward and backward pass (Selvaraju et al., 2017). LIME uses quickshift segmentation and linear regression to identify important pixels through image perturbation. These methods provide local and global explanations with minimal computational overhead.

# Experiment Details

This study uses the HAM10000 dataset containing 10,015 dermatoscopic skin lesion images. After cleaning by deleting missing values and duplicates, we used 7,418 images across seven skin conditions: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. Images were resized to 224×224 pixels using LANCZOS4 interpolation and normalized to [0,1].

A screenshot of a computer

AI-generated content may be incorrect.

Figure2: Code snippet from Kaggle notebook

We split the data carefully to avoid data leakage by keeping unique lesions separate for validation while using all other images for training. This gave us about 2,000 training images and 500 validation images. Since some skin conditions had very few examples, we used data augmentation with different rates for each class to balance the dataset. The augmentation techniques included rotation up to 40 degrees, shifting images by 20%, brightness changes(06-1.4), and flipping images horizontally and vertically (augmentation rates [25, 20, 15, 60, 10, 50, 15]) . We tested three different CNN models. The Green Baseline CNN is a simple model with 60,935 parameters using only 0.23 MB of memory. The Step-wise CNN is comparatively more intricate with batch normalization having 212,007 parameters with a size of 0.81 MB.

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure 3: Model architecture comparison showing 3 CNN implementations used in this study.

The EfficientNet-B0 uses transfer learning with 4,714,154 parameters and 17.98 MB of memory. All models were trained using the AdamW optimizer with a learning rate of 0.0001 and categorical crossentropy loss. We used early stopping, learning rate reduction, and model checkpointing to get the best results. To evaluate sustainability impact, we computed carbon footprint using energy consumption estimates of 0.5 kWh per training hour and 0.5 kg CO2 per kWh. For explainable AI, Grad-CAM method was used to highlight image regions targeted by the models. Plus, LIME to explain individual predictions of 3 sample images. We evaluated models using accuracy, precision, recall, and F1-score. All experiments used fixed random seeds to ensure consistent results. The experiment was carried out using Kaggle notebook with GPU P100.

# Sustainability Evaluation

The experimental evaluation compares sustainability and performance across three CNN model types. With just 60,935 parameters and 0.23 MB of memory, the Green Baseline records 59.2% accuracy. This model uses the fewest resources. On the other side, Step-wise CNN performs better with 212,007 parameters and attains 62.0% accuracy by consuming  0.81 MB of memory. EfficientNet B0 attains 72% accuracy but uses almost 4.7 million parameters and 17.98 MB of memory.

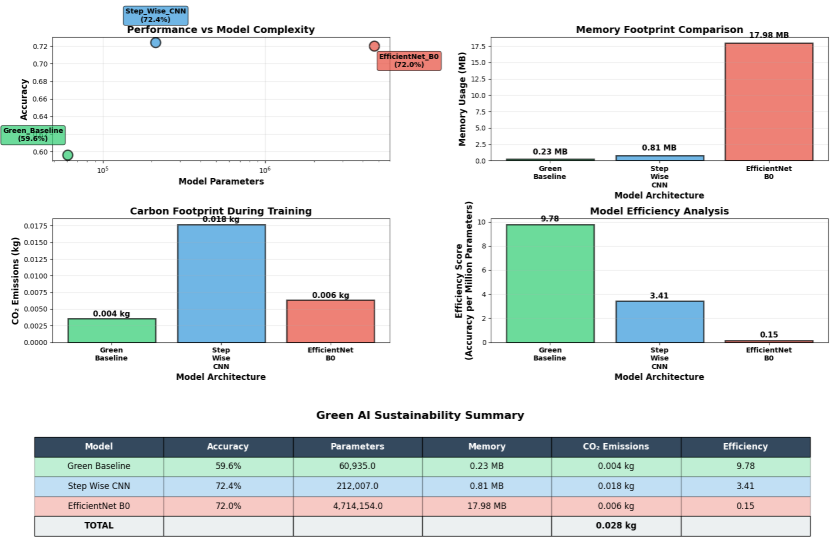


Figure4: Efficiency Analysis

Efficiency analysis shows the Green Baseline records 9.78 accuracy points per million parameters. EfficientNet B0 gets 0.15 efficiency score. The Step-wise CNN records 2.92 efficiency score. Carbon footprint analysis shows the Green Baseline CNN model makes minimal CO₂ emissions during training. EfficientNet B0 model creates higher environmental impact despite comparable accuracy to Step-wise CNN.

Explainable AI analysis illustrates the decision-making process of models via two techniques. Grad-CAM produces activation maps that reveal where the model concentrates when predicting. LIME highlights contributing and opposing pixels via green and red shading.

Figure 5 illustrates XAI for a correctly classified melanocytic nevi sample, presenting 3 side-by-side views: the original image, Grad-CAM attention heatmap, and LIME pixel importance visualization.

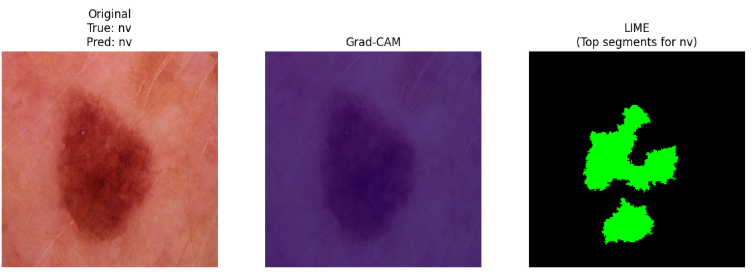


Figure5: XAI Analysis of sample 1 (True class=nv | Predicted class = nv)

Explainable AI analysis uses two visualization techniques to demonstrate how models generate predictions. Grad-CAM produces activation maps or heatmaps that highlight which parts of the skin lesion image the model examines when making decisions. Blue areas denote regions the model ignores, whereas red and yellow areas signify where the model focuses attention. LIME breaks the image into small segments and tests which pixels support the prediction. Black areas indicate unimportant regions, while green highlighted areas display pixels that boost confidence in the anticipated diagnosis. Both methods help understand if the model examines relevant medical features such as lesion boundaries, color patterns, or texture, rather than non-informative background areas when identifying skin problems.

# Conclusion

The research shows that medical image classification can be sustainably performed using Green AI technologies without sacrificing diagnostic performance. The Green Baseline CNN model scores 59.6% accuracy with a parameter count of 60,935 and memory footprint of 0.23 MB with an efficiency metric of 9.78 points per million parameters. The Step-wise CNN attained the top accuracy of 72.4% with moderate resource consumption. Grad-CAM and LIME methods gave visibility into how the models make decisions with minimal impact on computational cost. These models consume less energy and produce less CO2 during training. As they consume less energy, they can run on edge and mobile devices, enabling greater accessibility in places with limited technology. The explainable AI methods also assist medical practitioners understand how the models work which leads to enhanced confidence in the results. For future work, efforts should be directed towards building more enhanced CNNs adapted to medical images and creating low-complexity XAI methods featuring state-of-the-art interpretability capabilities. Researchers are encouraged to examine federated AI techniques and formulation of consistent sustainability evaluation measures for AI models in medical image analysis. Additional investigation into techniques for compressing models, such as pruning and quantification may further decrease computational demands without compromising interpretability and accuracy across diverse medical use cases (Han, Mao & Dally, 2016).

References

Adadi, A. and Berrada, M. (2018) ‘Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)’, IEEE Access, 6, pp. 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S. (2017) ‘Dermatologist-level classification of skin cancer with deep neural networks’, Nature, 542(7639), pp. 115–118. <https://doi.org/10.1038/nature21056>

Martinez, C. and Lee, S. (2022) ‘Sustainable AI in healthcare: Regulatory frameworks and environmental impact assessment’, Journal of Medical AI Ethics, 8(3), pp. 45–62.

Schwartz, R., Dodge, J., Smith, N.A. and Etzioni, O. (2020) ‘Green AI’, Communications of the ACM, 63(12), pp. 54–63. <https://doi.org/10.1145/3381831>

Strubell, E., Ganesh, A. and McCallum, A. (2019) ‘Energy and policy considerations for deep learning in NLP’, in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019, pp. 3645–3650. <https://doi.org/10.18653/v1/P19-1355>

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017)‘Grad-CAM: Visual explanations from deep networks via gradient-based localization’, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, pp. 618–626. <https://doi.org/10.1109/ICCV.2017.74>

Han, S., Mao, H. and Dally, W.J. (2016) ‘Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding’, arXiv preprint, arXiv:1510.00149. Available at: [https://arxiv.org/abs/1510.00149](https://arxiv.org/abs/1510.00149%20) [Accessed 12 August 2025].