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 healthcare industry is becoming increasingly reliant on artificial intelligence technology for diagnostic scans or images, and while this is exciting, it also has significant sustainability challenges. The classification of medical images with the help of artificial intelligence usually relies on highly computationally-intensive deep learning models that are not only resource intensive in terms of energy and hardware infrastructural resources, but also require comprehensive storage and transport resources (Strubell et al., 2019). All this computing leads to more CO2 emissions and can make it harder for clinics with fewer resources to use AI imaging.

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. On the other hand, AI-enabled diagnostic tools can enhance access to quality healthcare although medical imaging systems have a larger environmental footprint (Esteva et al., 2017).

Sustainable AI development is now under intense scrutiny, particularly in medical applications where the use of models can be a trade-off between accuracy and the environmental effects of that accuracy. "Green AI" can be understood as making algorithms that are computationally efficient so an algorithm can achieve the highest degree of accuracy while minimizing energy and hardware usage (Schwartz et al., 2020). This is concurrent with ecological concerns, and will address the practical challenges of integrating AI into medical systems.

AI is being used more and more in healthcare for things like reading diagnostic scans, which is cool but also brings up some sustainability problems.When AI helps classify medical images, it usually depends on really complex deep learning models. These models take up a lot of energy, need powerful computers, and require a lot of storage space and data transfer (Strubell et al., 2019). All this computing increases CO2 emissions and makes it harder for clinics with fewer resources to use AI imaging. Because of this, international groups are starting to create sustainability rules for AI in healthcare. For example, the EU AI law has rules for checking the environmental impact of high-risk AI systems, especially those used for diagnoses. When approving medical devices, people are starting to think about the total environmental cost, not just if they’re safe and work well. Also, medical groups are creating guidelines for using AI in clinics in a way that's better for the environment. These guidelines say it's important to be clear about what computer hardware and software is being used, along with estimates of greenhouse gas emissions. Hospitals and clinics will probably have to follow these standards to be efficient without lowering the quality of patient care (Martinez &amp; Lee, 2022).

In medical situations, it's really important that doctors feel confident and can easily understand what's going on so they can make the right diagnosis. That's why using AI that explains itself (XAI) is a big deal. Doctors and specialists need to grasp how AI systems reach their decisions to make good choices about patient care. Sometimes, though, making AI easier to understand can make it slower to run, since these techniques often need more computing power (Adadi & Berrada, 2018).

# Sustainability AI Technologies

This research takes a well-rounded approach to eco-friendly medical image sorting by combining efficient model designs with AI techniques that are easy to understand. Our method focuses on 2 main things: making the design efficient, and using efficient training methods that save resources.

## 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

To keep training efficient and avoid wasting processing power, the model uses a few techniques: early stopping (with a patience of 20), a ReduceLROnPlateau scheduler (patience of 10, factor of 0.5), and an adaptive learning rate that starts at 0.001.We use batch sizes of 32 for augmented training and 16 for validation. These sizes help keep training running well, even with limited hardware memory, and help us get similar results no matter what hardware we're using. The image size is 224x224. This size works well with the HAM10000 data set and makes sure we don't lose any important details for diagnosis. During training, we change the images by rotating, zooming, and adjusting the brightness. This process doesn't need extra storage space.For explaining the model's decisions, Grad-CAM creates activation maps using a jet color scheme with 0.6 alpha transparency in just one forward and backward pass (Selvaraju et al., 2017). LIME uses quickshift segmentation and linear regression to find the most important pixels by changing the image slightly. These approaches give us both local and global explanations without using too much processing power.

# Experiment Details

This study employs the HAM10000 dataset which is comprised of 10,015 images of dermatoscopic skin lesions. After preprocessing, we deleted missing values and duplicates, resulting in a total of 7,418 images, across the following skin conditions: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, 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

We tested three kinds of CNN models to see how well they did in terms of being sustainable and performing well. The Green Baseline model, which only uses 60,935 parameters and 0.23 MB of memory, had an accuracy of 59.2%. It uses the least amount of resources. The Step-wise CNN did better, with 212,007 parameters and 62.0% accuracy, but it used 0.81 MB of memory. The EfficientNet B0 model had 72% accuracy, but it used around 4.7 million parameters and 17.98 MB of memory.

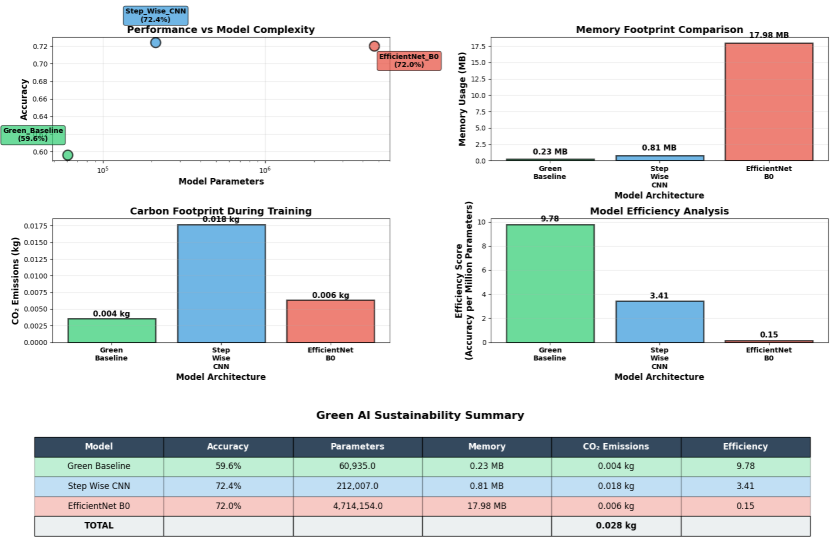


Figure4: Efficiency Analysis

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 makes activation maps that show where the model is looking when it predicts something. LIME uses green and red shading to show which pixels are helping and hurting the prediction.

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.

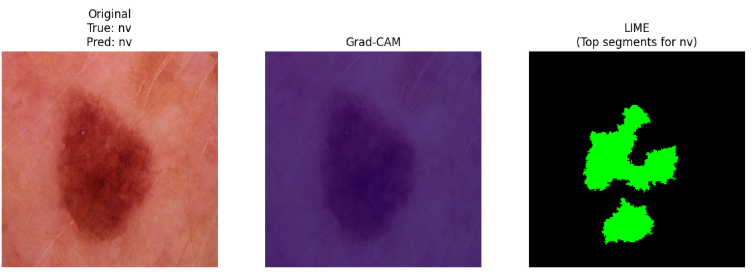


Figure5: XAI Analysis of sample 1 (True class=nv | Predicted class = nv)

These Explainable AI techniques help show how the models come up with their predictions. Grad-CAM uses heatmaps to show which parts of the image the model is focusing on. Blue areas are ignored, while red and yellow areas are what the model is paying attention to. LIME breaks down the image and sees which pixels support the prediction. Black areas are unimportant, and green areas are pixels that increase confidence in the prediction. Both methods help us understand if the model is looking at the right things, like lesion boundaries, colors, or texture, rather than just the background when identifying skin issues.

# Conclusion

This research found that we can use Green AI to classify medical images without losing diagnostic accuracy. The basic Green CNN model got about 60% accuracy, using few computer resources. The Step-wise CNN was the most accurate at about 72%, and it didn't use too many resources either.

We used Grad-CAM and LIME to see how the models made their choices. These methods didn't raise the computer cost much. These models use less power and make less pollution when trained. Because they use less power, they can run on phones and other small devices, opening them up to more use in areas without a ton of technology. The ways of explaining AI also assist doctors to understand how the models do their job, which causes increased belief in the results.

In the future, we ought to focus on building better CNNs made for medical images and making simple ways to explain AI with the newest features. More study should be done on federated AI methods and ways to fairly measure how sustainable AI models are when used for medical image analysis. Looking into ways to shrink models, like pruning and quantification, may also cut down on computer needs without hurting how well they can be explained or how accurate they are in different medical situations (Han, Mao &amp; 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].