

ISIC 2024 - Skin Cancer Detection with ML

Eduard Ferré Sánchez (100538716@alumnos.uc3m.es) - May 25, 2025

1 Introduction

Skin cancer is one of the most prevalent cancers worldwide, where early detection is vital. The ISIC 2024 challenge aims to automatically classify skin lesions as benign or malignant using dermoscopic images and patient metadata. This project applies *machine learning* and *deep learning* techniques to tackle challenges such as **severe class imbalance** and **multimodal data integration**, with the objective of building a **reliable** and **explainable** diagnostic support system aligned with medical screening practices.

2 Preprocessing & Data Imbalance

The ISIC dataset contains dermoscopic images and patient metadata, but is highly imbalanced, with approximately **400,616 benign** and only **343 malignant** cases. Addressing this **class imbalance** was a core focus of preprocessing, applying *SMOTE* to metadata for synthetic malignant samples, and extensive *data augmentation* (resizing, flipping, rotations, blurring, perspective distortion) to malignant images.

Metadata preprocessing also involved filtering out **irrelevant columns** (e.g., `patient_id`, `image_type`), *imputing missing values*, and encoding categorical variables to ensure high data quality and preserve sample size.

A **custom filter** enhanced lesion borders to improve diagnostic feature extraction. Furthermore, **weighted random samplers** were used in image data loaders, balancing class distribution in each batch to prevent bias toward the majority class during training.

3 Best Models & Ensemble Strategies

Initial benchmarking used *Multi-Layer Perceptrons* on metadata, images, and a multimodal combination. The multimodal MLP scored **79.80% accuracy**, highlighting the value of integrating data types.

3.1 Metadata Modeling

For metadata, models such as *Random Forests*, *Support Vector Machines*, and *gradient boosting* were tested. The best results arose from an ensemble of **LightGBM**, **CatBoost**, and **XGBoost**. These models were chosen for their **complementary strengths**: LightGBM for efficiency, CatBoost for native categorical handling and overfitting reduction, and XGBoost for robust regularization. The ensemble averaging improved generalization and maintained interpretability compared to image models.

3.2 Image Modeling

Image modeling demanded complex architectures due to unstructured data. After baseline CNNs and ResNet variants, **ConvNeXt** was selected for its advanced feature extraction, merging ResNet-like efficiency with Transformer-inspired design. Additionally, a hybrid **ConvNeXt + Vision Transformer (ViT)** model was developed, capturing both local and global features. These architectures excelled in recognizing subtle malignant signs, particularly with strong augmentation.

3.3 Ensemble Strategy for Final Prediction

Two ensemble methods were evaluated:

- **Hybrid Worst Case:** Outputs the max probability if any model exceeds a threshold (e.g., 0.6), else averages predictions—balancing sensitivity and false positives.
- **Worst Case:** Always selects the highest probability, prioritizing sensitivity.

The Hybrid Worst Case excelled on the public leaderboard; the Worst Case performed better on the private leaderboard.

4 Explainability Analysis for Medical Use

Given that this system targets doctors and medical professionals, **explainability** is paramount. The metadata models are inherently interpretable due to feature importance measures available in gradient boosting methods. For image models, techniques such as *Grad-CAM* and attention visualization on ConvNeXt+ViT help highlight lesion areas influencing predictions, offering intuitive explanations to clinicians. This transparency supports trust and aids clinical decision-making.

5 Methodology and Medical Justifications

This project combined **careful preprocessing**, **class imbalance mitigation**, and *advanced multimodal modeling* to improve skin cancer detection performance. By merging metadata and dermoscopic images, and employing robust **ensemble strategies**, the system enhanced accuracy and reliability.

In clinical practice, missing a malignant case has far graver consequences than false alarms. Thus, the ensembles were designed to **prioritize sensitivity**, aligning with medical screening standards to minimize false negatives.

Moreover, the chosen models and explainability methods ensure *interpretability and trustworthiness*, critical for AI acceptance in healthcare. This work underscores the potential of **AI-assisted diagnostic tools** to effectively support dermatological screening.

6 Final Remarks

The complete project, including code, preprocessing pipelines, and trained models, is publicly available on **GitHub** for transparency and reproducibility.

The hybrid ConvNeXt+ViT model was inspired by the approach presented in the paper: “**A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification**”.