An Ensemble Model for Robust Skin Cancer Detection with 3D Total-Body Photographs

Thomas He, Chunyu Li, Zongwei Huang, Xiaoqian Zhou, Mira Parikh
{mperform, chunyuli, zongweih, zhouxq, miparikh}@umich.edu

March 13 2025

Abstract

Skin cancer, particularly melanoma, remains among the deadliest cancers worldwide, and early detection is critical for improving patient outcomes. Traditional diagnostic methods relying solely on visual examination by dermatologists are subjective and often inaccessible in underserved regions. To address these challenges, we propose an ensemble machine learning approach utilizing the ISIC 2024 dataset, comprising dermoscopic lesion images and patient metadata. Our methodology integrates convolutional neural networks (CNNs) (EfficientNet-B3, ResNet50, and DenseNet121) for visual feature extraction, alongside gradient boosting models (XGBoost, LightGBM, CatBoost) for effectively modeling structured metadata. These two model types are combined through a weighted ensemble strategy, with performance evaluated primarily using partial Area Under the Curve (pAUC) above 80% True Positive Rate (TPR).

1 Introduction

Skin cancer remains one of the most prevalent and deadly forms of cancer worldwide, with melanoma being the most aggressive type. Early detection is critical for improving patient outcomes, as the survival rate significantly decreases when the disease progresses to advanced stages. Traditional diagnostic methods rely heavily on visual examination by dermatologists, followed by biopsies for confirmation. However, these approaches are time-consuming, subjective, and often inaccessible to individuals in underserved or remote areas.

The rapid advancement of machine learning, particularly deep learning, has opened new avenues for automating skin cancer detection. By leveraging large datasets of dermatological images, these technologies can assist in identifying malignant lesions with high accuracy, potentially reducing the burden on healthcare systems and improving diagnostic efficiency. However, in clinical practice, dermatologists also consider additional patient information—such as age, sex, and lesion location—which can enhance diagnostic accuracy. Integrating this metadata with CNN-based image analysis has been shown to further improve model performance in skin cancer classification tasks.

By developing a system that combines advanced deep learning techniques with comprehensive patient metadata, we aim to assist dermatologists in making more accurate and timely diagnoses. The successful implementation of this technology could revolutionize skin cancer screening, making it more accessible, efficient, and reliable, ultimately saving lives and reducing healthcare costs.

2 Proposed Method

In this project, we utilize the ISIC 2024 dataset, which comprises standardized clinical images derived from lesions captured via 3D Total Body Photography (TBP), along with structured patient metadata. The metadata includes patient demographics (e.g., age, sex) and lesion-specific details (e.g., anatomical location). Additionally, this metadata contains carefully engineered image-derived features. Notably, the dataset is highly imbalanced, with fewer than 0.1% of samples representing positive (malignant) cases, which requires careful handling during model training to ensure effective performance.

To address the skin cancer prediction task effectively, we propose a two-level ensemble modeling strategy that leverages both the strengths of convolutional neural networks (CNNs) for visual feature extraction and gradient boosting methods for metadata analysis. CNN-based models are proficient at capturing intricate visual patterns in lesion images, which may not be easily discernible to clinicians. Simultaneously, gradient boosting models (e.g., XGBoost, LightGBM) are particularly effective at modeling nonlinear relationships in structured metadata and are robust to issues such as missing values and severe class imbalance.

Our approach begins with rigorous data augmentation techniques applied to malignant lesion images, including flipping, cropping, rotation, and random erasing, to address the data imbalance issue in the dataset. Following augmentation, we implement our two-level ensemble approach:

Level 1 (Within-class ensembles):

- Image-based ensemble: CNN models with different backbone architectures, including EfficientNet-B3, ResNet50, and DenseNet121, will be individually fine-tuned on the lesion images. These models, pretrained on ImageNet, provide diverse visual embeddings.
- Metadata-based ensemble: Gradient boosting tree models, such as XGBoost, LightGBM, and CatBoost, will be independently trained on metadata and handcrafted image-derived features.

Level 2 (Between-class ensemble):

- Predictions from the CNN ensemble and the metadata ensemble are combined via weighted averaging. This final ensemble integrates complementary predictions from distinct modalities, providing a potentially more accurate and robust diagnostic output.

A schematic diagram for our overall model architecture is displayed in Fig 1.

Consistent with the ISIC 2024 challenge, our primary evaluation metric will be the partial Area Under the Curve (pAUC) above an 80% True Positive Rate (TPR). This metric takes values in [0, 0.2] and it specifically evaluates performance in high-sensitivity regions, which are crucial for clinical decision-making. This ensures our modeling efforts remain closely aligned with real-world diagnostic priorities.

3 Related Work

Skin cancer detection has been extensively studied in the field of medical image analysis, with deep learning playing a pivotal role in recent advancements. Traditional methods relied on handcrafted features extracted from dermoscopic images, while modern approaches leverage deep neural networks for automated feature extraction and classification.

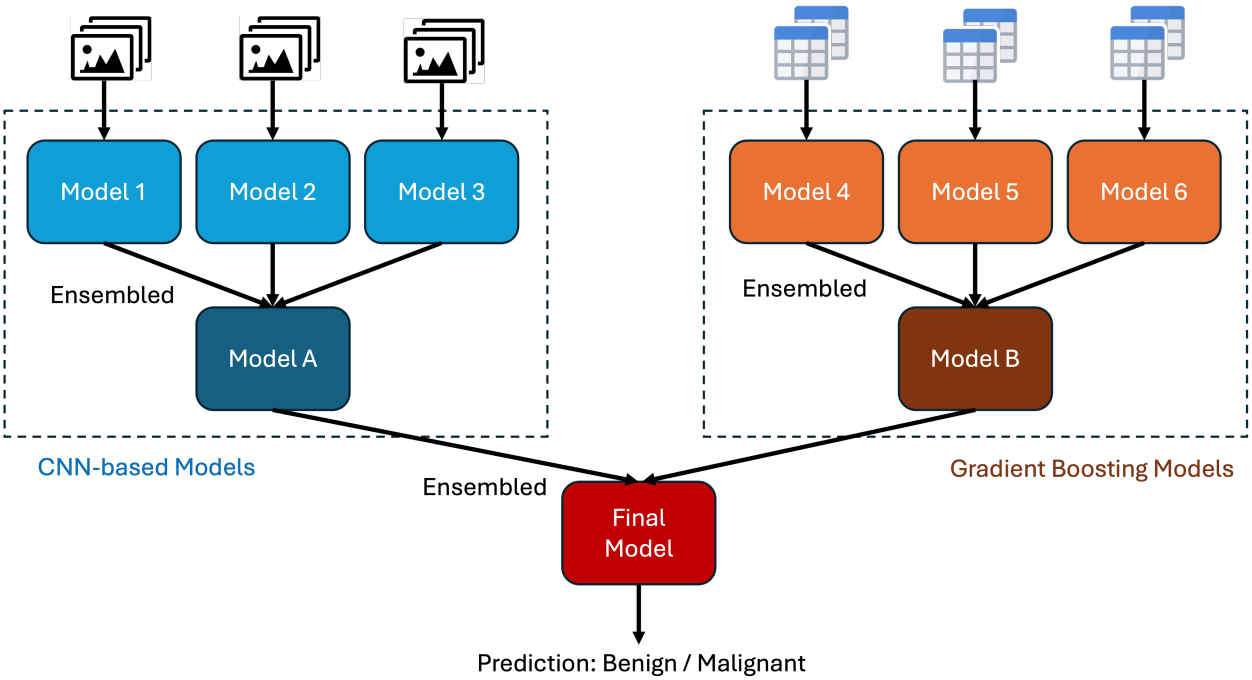

Figure 1: Model Architecture

Early methods for skin cancer detection primarily used image processing and machine learning techniques to extract features such as asymmetry, border irregularity, color variation, and texture patterns [1]. These features were then fed into classifiers like Support Vector Machines (SVMs) and Random Forests (RFs) for lesion classification. While these methods provided reasonable accuracy, their reliance on manually engineered features limited their generalizability.

The rise of deep learning significantly improved skin cancer detection by enabling automatic feature extraction. Convolutional Neural Networks (CNNs) have become the dominant approach, achieving state-of-the-art performance on datasets like ISIC, HAM10000, and PH2 [2]. Several CNN architectures, including ResNet, EfficientNet, and MobileNet, have demonstrated high accuracy in classifying malignant and benign skin lesions [4].

Recent studies have explored advanced deep learning techniques to further improve skin cancer classification: Pretrained CNN models like InceptionV3, VGG16, and EfficientNet have been fine-tuned on dermatological datasets to enhance accuracy while reducing the need for large labeled datasets [5]. Combining multiple CNN architectures in an ensemble model has been effective in boosting classification performance [3]. These approaches have been used to leverage unlabeled data and improve robustness, particularly in real-world mobile-based applications [4].

4 Preliminary Results

Up to this point, we have completed preliminary dataset exploration and established baseline models using only metadata. After performing a stratified train-validation-test split, missing values were imputed using K-Nearest Neighbors (KNN) imputation, followed by standardization of numeric features. Initially, three classical machine learning models—elastic net, random forest, and

	Elastic Net	Random Forest	XGBoost	Ensemble XGBoost Model
pAUC	0.1265	0.1347	**0.1466**	0.1370

Table 1: Preliminary Experiment Results

XGBoost—were trained without explicit resampling. To address the issue of severe class imbalance, the built-in class_weight='balanced' parameter was applied, assigning greater importance to the minority class (positive examples) during training.

Subsequently, explicit resampling strategies were adopted on the training dataset. Negative (majority) samples were down-sampled, reducing the imbalance to a negative-to-positive ratio of 10:1. The minority class (positive) samples were further augmented using the Synthetic Minority Over-sampling Technique (SMOTE), generating synthetic examples to improve the ratio to approximately 10:3. This resampling process was independently repeated five times, each resulting in a distinct training subset used to train separate XGBoost classifiers. Final predictions were obtained through a soft-voting ensemble of these five classifiers, where the predicted probability of a positive label was the equally weighted average of predictions. Hyperparameter tuning for all models was conducted using the validation set and the Optuna framework. Preliminary experimental results are summarized in Table 1.

These results reflect the predictive superiority of gradient boosting models. Given that the maximum possible partial AUC (pAUC) above 80% TPR is 0.20, these initial results clearly indicate substantial room for improvement. These baseline outcomes will serve as a valuable reference point for systematic comparisons and ablation studies as we continue refining our models, which we expect to support our assertion that integrating advanced ensemble approaches and careful handling of metadata can significantly improve skin cancer detection performance.

5 Future Milestones

March 14 - March 21: CNN model development and image augmentation

- Preprocess dataset to ensure correct data cleansing for training.
- Apply extensive data augmentation specifically tailored to address class imbalance (rotations, flipping, cropping, random erasing, and color jittering).
- Explore advanced data augmentation methods leveraging autoencoders and diffusion models.
- Fine-tune backbone CNN architectures (EfficientNet-B3, ResNet50, DenseNet121).

March 22 - March 28: Ensemble model implementation

- Develop ensemble frameworks combining CNN-based image models and previously trained gradient boosting models on metadata.
- Explore and experiment with multiple strategies for ensemble fusion, including weighted averaging and stacking methods.
- Perform hyperparameter tuning of CNN models (learning rates, batch sizes, dropout rates) and gradient boosting models (XGBoost, LightGBM, CatBoost).

- Finalize the data augmentation approach.
- Identify the best-performing ensemble model configurations based on validation partial AUC scores.

April 5 - April 11: Model evaluation and ablation studies

- Evaluate selected models rigorously on the held-out test dataset using the primary metric (partial AUC above 80% TPR).
- Conduct detailed comparative analyses and ablation studies to quantify each component's contribution (image vs. metadata, ensemble strategies).
- Create visualizations (ROC curves, confusion matrices, performance comparisons) to clearly demonstrate model performance.

April 12 - 18: Summarize results and findings into a well-structured report

- Document experimental outcomes, insights, and the incremental improvements achieved through the proposed ensemble methods.
- Finish a structured final report, emphasizing key findings, comparisons with baseline methods, and practical significance of results.
- Compile visualizations and clearly interpret results into a final presentation.

6 Conclusions

In summary, we have successfully established a baseline model utilizing patient metadata and classical machine learning techniques, specifically gradient boosting methods, demonstrating promising predictive capabilities on the highly imbalanced ISIC 2024 skin cancer detection dataset. Our preliminary experiments indicate room for improvement given that the maximum possible pAUC score is 0.20. Moving forward, we will extend our approach by integrating convolutional neural networks (CNNs) trained on lesion images with our metadata-based models, leveraging advanced ensemble strategies and data augmentation techniques, potentially including generative approaches such as autoencoders and diffusion models. By combining visual features with structured patient information, we anticipate our final model will substantially outperform current baselines, particularly in the critical high-sensitivity region measured by partial AUC above 80% True Positive Rate. Ultimately, we expect to demonstrate that our carefully designed ensemble framework can serve as a highly effective diagnostic support system for early and accurate skin cancer detection.

Author Contributions

X.Z. and Z.H. performed data preprocessing and conducted the literature review; C.L. and Y.H. developed the overall model pipeline and conducted preliminary experiments; M.P. assisted with data preprocessing and experimental implementation; all authors contributed to the writing of this report.

References

[1] Uzma Bano Ansari and Tanuja Sarode. Skin cancer detection using image processing. *International Research Journal of Engineering and Technology*, 4(4):2875–2881, 2017.

[2] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Abdul Hakeem M Saeed, Mohammed Olaythah Alraddadi, and Mater Hussen Mahnashi. Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, 18(10), 2021.

[3] Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar, and Sonali Patil. Skin cancer detection and classification. In *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, pages 1–6, 2017.

[4] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[5] M. Vidya and Maya V. Karki. Skin cancer detection using machine learning techniques. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–5, 2020.