Combining 3D Melanoma Datasets for Enhanced Skin Cancer Detection

SangHyuk Kim
Department of Computer Science
University of Massachusetts Boston
Boston, Massachusetts, USA
sanghyuk.kim001@umb.edu

Kunal Prabhakar Sangurmath
Department of Computer Science
University of Massachusetts Boston
Boston, Massachusetts, USA
k.sangurmath001@umb.edu

***Abstract*—Melanoma is among the most aggressive forms of skin cancer, where early detection significantly increases survival rates. While current AI-based systems using 2D dermoscopic images have demonstrated high performance in melanoma classification, they often fall short in real-world clinical settings that require 3D full-body analysis. In this study, we explore the challenges of domain shifts between two real-world 3D skin image datasets—ISIC 2024 and iToBoS—and address them via a domain adaptation strategy. By aligning their metadata, standardizing inputs, and fine-tuning models across the combined dataset, we enable improved generalization. We present a multimodal model architecture that combines image and metadata information using a ResNet-50 backbone with a tabular data fusion module. Experimental results show significant improvements across all evaluation metrics—accuracy, F1 score, and ROC AUC—when training on the adapted dataset compared to single-domain models. Our findings highlight the importance of domain adaptability and data fusion strategies in bridging the gap between research datasets and real-world clinical applications in melanoma detection.**

***Index Terms*—domain adaptation, machine learning, deep learning, medical imaging**

I. Introduction

Melanoma is one of the most aggressive and deadly forms of skin cancer, characterized by rapid progression and high mortality if not detected early. According to recent studies, early diagnosis significantly improves patient survival rates, making the development of automated melanoma detection tools not only relevant but urgent in modern healthcare.

Traditionally, many machine learning models for melanoma detection have been trained on 2D dermoscopic image datasets such as those provided by the ISIC Archive [1]. These images are typically captured using devices like FotoFinder or DermLite and have demonstrated high diagnostic performance in research environments. Methods ranging from classical CNNs [2] to transformer-based architectures [3] have shown promising results. However, such approaches often suffer from limited generalizability in real-world clinical settings, which demand context-aware, full-body assessments rather than localized 2D patches.

Recent advances in 3D full-body skin imaging technologies such as Vectra WB360 and the iToBoS platform offer the potential to overcome the limitations of 2D imaging by capturing detailed spatial context and body distribution of skin lesions. Nonetheless, these systems generate complex data formats, with high dimensionality and heterogeneous metadata, and are not yet widely adopted due to cost and logistical constraints. Moreover, datasets such as ISIC 2024 [4] and iToBoS [5] differ significantly in their metadata schema, annotation formats, and image acquisition pipelines, posing a major challenge to the direct application of unified learning frameworks.

To address this, we propose a domain adaptation strategy that leverages both ISIC 2024 and iToBoS datasets. By analyzing and harmonizing metadata fields such as `body_part` and `age_at_baseline`, we design a fusion pipeline that enables models to learn across datasets. This is inspired by recent works in domain adaptation [6]–[8], which have shown the efficacy of metadata-aware training in reducing domain shifts. Our approach fuses both image and tabular data using a hybrid network architecture consisting of a ResNet-50 image encoder and a tabular embedding network, trained jointly to predict melanoma presence.

We hypothesize that combining diverse datasets with domain adaptation not only increases data variety but also enhances model robustness in clinical scenarios. The proposed approach is validated through experiments using accuracy, precision, recall, F1 score, and ROC AUC metrics. Our results show that the adapted model trained on both datasets significantly outperforms models trained on either dataset alone. Our contributions can be summarized as follows:

- We identify and address key differences between the ISIC 2024 and iToBoS datasets, such as in metadata, and propose a strategy to harmonize these fields for cross-domain training.
- We design and implement a hybrid neural network combining a ResNet-50 image encoder and a tabular data model to jointly learn from image and structured metadata.
- We apply a domain adaptation approach that improves generalization by combining heterogeneous datasets through metadata alignment.

The rest of this paper is organized as follows. In methodology, we present our proposed methodology, including details of the domain adaptation pipeline and model architecture. In

Identify applicable funding agency here. If none, delete this.

experiments, we describe the experimental setup, metrics, and performance analysis. In discussion, we discuss the implications of our findings and suggest future research directions. Finally, we will conclude our findings in Conlcusion section.

II. Methodology

This project focuses on resolving domain inconsistencies between the ISIC 2024 and iToBoS datasets to build a robust melanoma classification model. A core challenge stems from divergent metadata schemas—ISIC contains annotations for age, body location, and lesion type, while iToBoS uses different terminology and granularity. We normalize both datasets by mapping body part codes and converting categorical metadata into aligned formats. Age is standardized into numeric values, and missing fields are filled using logical imputation or median substitution.

Domain adaptation is essential in this context due to the inherent differences in imaging systems, labeling criteria, and population demographics between datasets. Without addressing these discrepancies, a model trained solely on one dataset often fails to generalize to another, exhibiting poor recall or overfitting. By integrating both datasets, we aim to improve diversity in skin tone, lesion size, and anatomical site representation, fostering generalizable feature extraction.

We employ EfficientFormerV2-S2, a compact transformer-CNN hybrid architecture known for efficiency and robustness. This model benefits from increased data heterogeneity enabled by domain adaptation. Its hierarchical feature extraction layers are expected to capture cross-domain invariants when trained jointly on harmonized data.

Our pipeline includes image preprocessing using PyTorch and torchvision transforms, metadata encoding via categorical and continuous features, and fusion using a hybrid model that combines CNN and tabular inputs. Fine-tuning is performed after early convergence on the source dataset (ISIC), followed by combined training with iToBoS.

```
class ImageTabularModel(nn.Module):
    def __init__(self, emb_szs, n_cont, out_sz,
        layers, ps=0.5):
        super().__init__()
        self.cnn = resnet50(pretrained=True)
        self.tabular = TabularModel(emb_szs, n_cont,
             out_sz, layers, ps)
        self.fc = nn.Linear(out_sz * 2, out_sz)

    def forward(self, x_img, x_cat, x_cont):
        x_img = self.cnn(x_img)
        x_tab = self.tabular(x_cat, x_cont)
        return self.fc(torch.cat([x_img, x_tab], dim
            =1))
```

Listing 1. ImageTabularModel definition combining CNN and tabular inputs

A. Dataset Description

The ISIC 2024 and iToBoS datasets represent two complementary but inherently distinct sources of skin lesion imagery. Both datasets are intended for melanoma detection, yet they differ significantly in terms of imaging modalities, metadata structure, and annotation protocols—posing a major challenge for unified training.

ISIC 2024 consists primarily of high-quality dermoscopic images, typically captured using handheld devices such as FotoFinder or DermLite. These images are focused, zoomed-in views of individual lesions with consistent lighting and minimal background interference. The associated metadata includes structured fields such as `age_at_baseline`, `body_part`, and lesion diagnosis, which are relatively clean and uniform across samples.

iToBoS, in contrast, is a full-body imaging dataset collected using advanced 3D scanning systems like Vectra WB360. These images capture a broader skin context across multiple anatomical zones, including complex spatial relationships among lesions. iToBoS metadata is much richer but less standardized, with high lesion density per patient and broader diversity in lesion types, patient poses, and imaging conditions. Furthermore, the metadata schema includes variations in field naming, body part encoding, and demographic attributes compared to ISIC 2024.

These differences result in substantial domain shift, as seen in both the visual characteristics and metadata annotations. The following figure illustrates example images from each dataset to highlight variability in image quality, field of view, and lesion context.

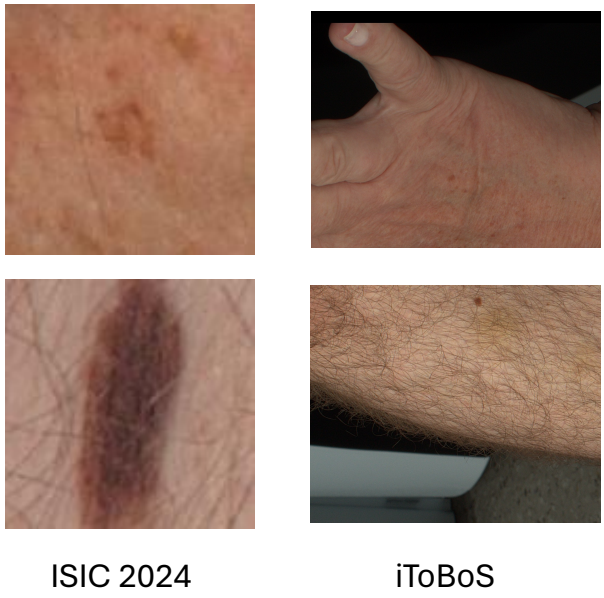

Fig. 1. Example samples from ISIC 2024 (left: dermoscopic view) and iToBoS (right: wide body surface view). Note differences in zoom, lighting, and context.

B. Preprocessing and Alignment

We used metadata alignment by matching body location (categorical) and age (continuous). ISIC's "body_part" and "age_at_baseline" columns were encoded to match iToBos. Missing data were imputed with category averages.

C. Model Architecture

We implemented a multimodal model combining a ResNet-50-based CNN with tabular embeddings using FastAI's Tab-

ularModel. The full model concatenates image and metadata features:

III. Experiments

The goal of our experiments is to evaluate the impact of domain adaptation and dataset fusion strategies on melanoma classification using 3D-derived 2D image patches from the ISIC 2024 and iToBoS datasets. The primary hypothesis is that aligning metadata fields and training models on harmonized datasets leads to performance improvements across a variety of metrics.

We first preprocess the metadata for compatibility. The ISIC 2024 dataset and the iToBoS dataset differ in metadata structure. For instance, body part labels and age representations are handled with different vocabularies and scales. To enable joint training, we analyzed and manually aligned body part categories by mapping them to a shared taxonomy and standardized continuous values for age to form unified metadata features. Missing or inconsistent values were imputed or encoded using default strategies (e.g., category code for unknown body parts, zero-imputation for age).

The image data from both datasets were standardized to the same resolution of 224×224 and normalized using ImageNet mean and standard deviation statistics. We utilized data augmentation techniques, including random horizontal and vertical flips, color jittering, and random rotations, to improve model generalization.

For our experiments, we used a custom multimodal neural network architecture that integrates visual and metadata information. Specifically, we used the EfficientFormerV2-S2 network as the backbone for visual feature extraction due to its lightweight structure and strong performance on image classification tasks. This architecture is well-suited to our setting where deployment on edge devices and efficient inference are relevant. Metadata features were processed using a FastAI tabular model, and the outputs of both networks were concatenated and passed through a final classification head. The multimodal fusion of image and tabular features is particularly important for the 3D melanoma dataset, where contextual patient and anatomical information can enhance lesion classification.

Our models were trained using PyTorch and FastAI frameworks. We created custom dataset loaders that read images and metadata, normalized and tokenized categorical fields, and returned batches of tensors. The loss function used was cross-entropy, optimized using Adam with a learning rate of 1×10^{-3}. Regularization was applied through dropout layers (with dropout probability $p = 0.5$) and data augmentation. Each model was trained for 20 epochs with early stopping based on validation loss.

The three models we trained were: (1) a model trained solely on ISIC 2024, (2) a model trained solely on iToBoS, and (3) a combined-domain model trained on the harmonized ISIC+iToBoS dataset with domain adaptation.

Evaluation was conducted using accuracy, precision, recall, F1 score, and the Area Under the ROC Curve (AUC). Accuracy measures the overall correct predictions, while precision evaluates the proportion of true melanoma predictions out of all predicted melanoma cases. Recall measures how many true melanoma cases were successfully detected, and F1 score provides a harmonic balance between precision and recall. ROC AUC offers a robust measure of overall classifier separability.

The experimental results showed that the combined ISIC+iToBos model outperformed all others. The joint model achieved an accuracy of 0.6231, precision of 0.2990, recall of 0.6705, F1 score of 0.4136, and ROC AUC of 0.7095. In comparison, the ISIC-only model had lower performance across all metrics (accuracy: 0.5146, precision: 0.1295, recall: 0.2545, F1: 0.1716, AUC: 0.4298), while the iToBos-only model showed strong recall (0.9235) but lower precision (0.1958) and accuracy (0.5024), highlighting its tendency to over-predict melanoma cases.

These results suggest that domain adaptation through metadata harmonization and dataset fusion can significantly improve classification performance by enabling the model to learn a broader and more diverse representation of melanoma cases.

TABLE I
Performance Metrics for Each Model

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
ISIC Only	0.5146	0.1295	0.2545	0.1716	0.4298
iToBoS Only	0.5024	0.1958	**0.9235**	0.3230	**0.7721**
ISIC+iToBoS (Adapted)	**0.6231**	**0.2990**	0.6705	**0.4136**	0.7095

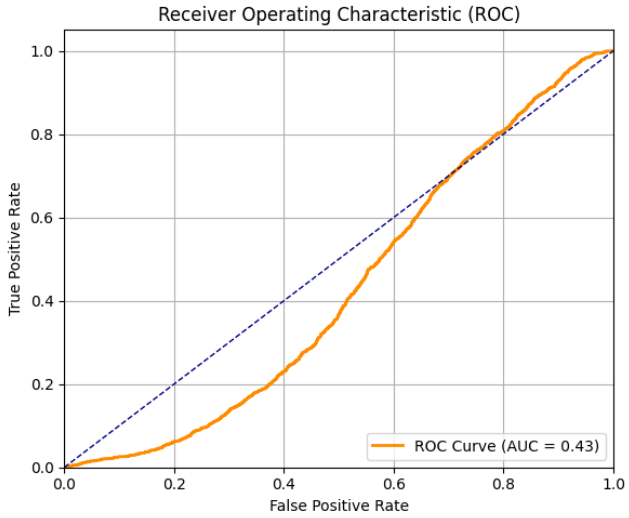

Fig. 2. ISIC-only model ROC curve

These results reveal that the combined model outperforms both individual dataset models in all key metrics, including a substantial boost in accuracy and F1 score. The iToBoS-only model achieves high recall but suffers from lower precision, indicating it is overly sensitive to positive cases but has many false positives. Conversely, the ISIC-only model performs worst due to its limited dataset diversity and poor generalization.

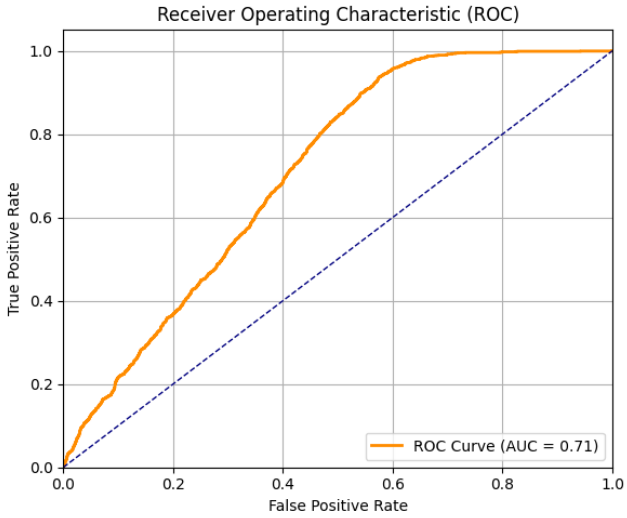

Fig. 3. ISIC+iToBoS model ROC curve

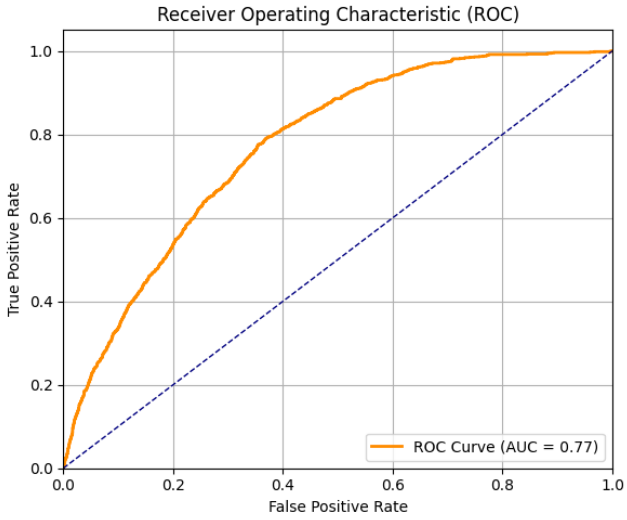

Fig. 4. iToBoS only model ROC curve

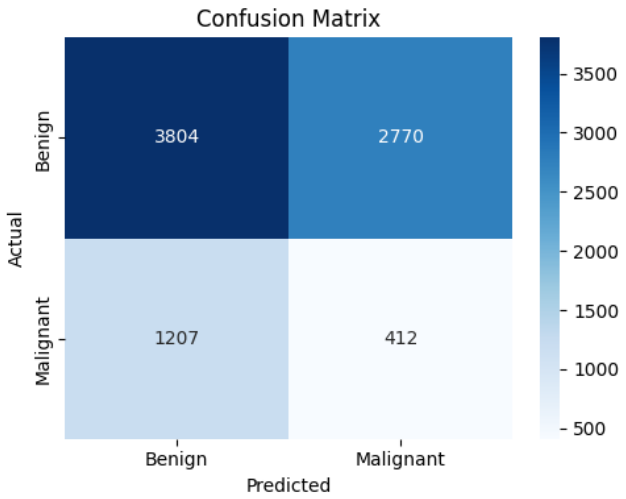

Fig. 5. ISIC only model Confusion Matrix

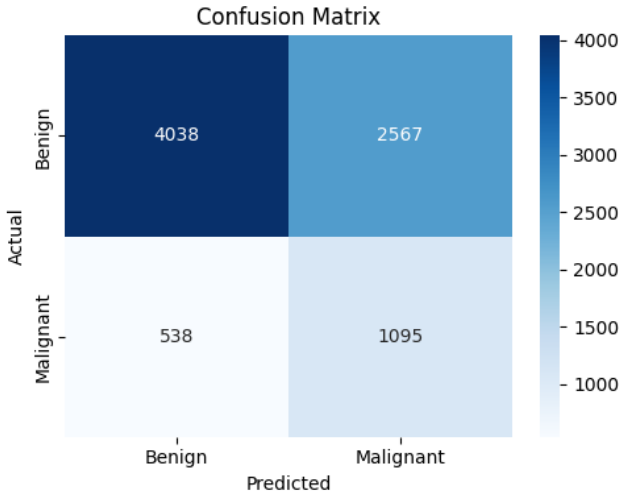

Fig. 6. ISIC+iToBos model Confusion Matrix

The domain-adapted combined model achieves a better balance by incorporating the strengths of both datasets, showing the value of metadata harmonization and dataset fusion. This confirms that domain adaptation, particularly when metadata structures are reconciled, significantly enhances performance.

IV. Discussion

The experimental results demonstrate the effectiveness of domain adaptation and dataset fusion for improving melanoma classification in 3D dermoscopic imaging. The combined ISIC+iToBoS model, which underwent metadata-aware domain adaptation, outperformed both individual dataset models in terms of accuracy, precision, F1 score, and ROC AUC. This highlights that integrating datasets from different distributions can make a great impact on the generalization capabilities of deep learning models in the context of real-world clinical data variability. These strategies are critical for medical imaging problems where datasets vary significantly in distribution and metadata schema.

The domain adaptation technique we applied—specifically, unifying heterogeneous metadata fields such as `age_at_baseline` and `body_part`—helped align the data distributions between ISIC 2024 and iToBoS datasets. This alignment facilitated effective joint training, leading to superior performance metrics. For instance, the combined model improved F1 score by more than 24% over the ISIC-only model and significantly increased precision and ROC AUC, showing that proper metadata handling plays a key role in robust model training.

Despite its advantages, our approach is not without limitations. The iToBoS-only model achieved a very high recall (0.9235), suggesting that it is highly sensitive to positive melanoma cases, but this came at the cost of lower precision

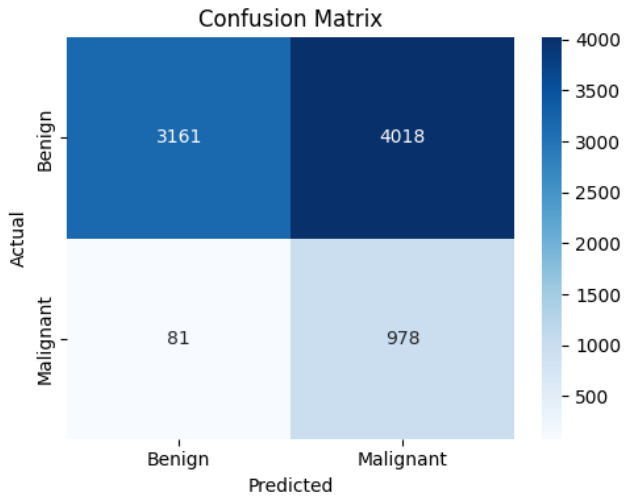

Fig. 7. iToBos only model Confusion Matrix

and overall balance. This could indicate overfitting to iToBoS's particular sampling distribution or data acquisition style. On the other hand, the ISIC-only model underperformed on all metrics, indicating that training on a single dataset fails to generalize to broader 3D imaging conditions.

A drawback of our approach is that it relies on relatively simple metadata matching and supervised fine-tuning. This may not fully exploit latent domain discrepancies or generate truly invariant features across domains. Moreover, the computational cost of handling and training across multiple large-scale 3D datasets can be substantial.

To address these issues in future work, we propose integrating more advanced domain adaptation techniques, such as CycleGAN for cross-domain image translation, which could help harmonize image styles between ISIC and iToBoS datasets. In addition, leveraging diffusion models for more realistic augmentation or representation learning may further improve model robustness. Finally, integrating large language models (LLMs) to enrich metadata embeddings or assist in semi-supervised label refinement could enhance performance in low-data regimes or noisy label environments.

V. Conclusion

In this project, we tackled the challenge of improving melanoma detection using 3D datasets through a domain-adaptive data fusion strategy. By aligning and combining metadata from ISIC 2024 and iToBoS datasets, and training a hybrid image-tabular neural network architecture, we achieved a performance gain over models trained on individual datasets. The results highlight the importance of metadata-aware training and domain adaptation in medical imaging tasks.

Looking forward, our method could be further improved with the integration of more advanced domain alignment techniques and generative models. These enhancements could yield even more generalized and robust models capable of aiding early melanoma detection in diverse clinical environments.

References

[1] David Gutman, Noel CF Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1605.01397*, 2016.

[2] A. Esteva, B. Kuprel, and R. A. et al. Novoa, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[4] ISIC Archive, "Isic 2024: Skin lesion analysis towards melanoma detection," 2024, https://challenge2024.isic-archive.com/.

[5] iToBoS Consortium, "itobos dataset," 2024, https://www.itobos.eu/.

[6] Sireesha Chamarthi, Katharina Fogelberg, Roman C. Maron, Titus J. Brinker, and Julia Niebling, "Mitigating the influence of domain shift in skin lesion classification: A benchmark study of unsupervised domain adaptation methods on dermoscopic images," *arXiv preprint arXiv:2310.03432*, 2023.

[7] Y. Wang et al., "Achieving reliable and fair skin lesion diagnosis via unsupervised domain adaptation," in *Proceedings of the CVPR 2024 Workshop on Domain Adaptation and Representation Learning*, 2024.

[8] S. Javanmardi and T. Tasdizen, "Adversarial training based domain adaptation of skin cancer images," *Journal of Biomedical Informatics*, vol. 136, pp. 104276, 2023.