

Skin Lesion Classification Using Dermoscopic Images and Clinical Metadata: Insights from Multimodal Models

Sakib Ahammed, Xia Cui, Wenqi Lu, and Moi Hoon Yap

Department of Computing and Mathematics, Manchester Metropolitan University
The Dalton Building, Chester Street, M1 5GD Manchester

SAKIB.AHAMMED@stu.mmu.ac.uk and {X.Cui, W.Lu, M.Yap}@mmu.ac.uk

Abstract

State-of-the-art methods primarily use dermoscopic skin images to classify lesions. Incorporating clinical metadata to improve melanoma diagnosis has increased popularity in the community. This study explores whether a single modality (dermoscopic images) or multimodality (dermoscopic images and clinical metadata) enhances performance for skin lesion classification. In particular, it tries to answer the question by comparing three multimodal architectures. This paper introduces the ISIC-DICM-17K dataset, a curated balanced dataset comprising 17,060 dermoscopic images and clinical metadata with an equal distribution of melanoma and non-melanoma classes. We evaluate the impact of including clinical metadata on multimodal model performance using both supervised and transfer learning models. Our results show that including metadata significantly enhances supervised learning models' performance, while all transfer learning models outperform supervised learning models with 3% to 12% performance improvement. Our statistical and visual analyses highlight the importance of clinical metadata in improving feature clustering and class separability, and the HeatMap_{index} analysis shows the effectiveness in identifying relevant features on lesion images. Our GitHub repository (accessible at <https://github.com/mmu-dermatology-research/isic-dicm-17k>) contains the ISIC-DICM-17K dataset and image IDs referencing the original ISIC dataset sources.

1. Introduction

Skin cancer is the most widespread cancer globally, with its continuously increasing and is the fifth most common form of cancer [4, 7, 19]. Melanoma, the deadliest form of skin cancer [14], is anticipated to rise dramatically, with almost half a million cases expected by 2040—an alarming 62% increase since 2018 [1]. This global epidemic

results in one skin cancer-related death every 4 minutes. It is a highly aggressive form of skin cancer that necessitates accurate and early detection to improve patient outcomes [21]. The International Skin Imaging Collaboration (ISIC) provides the most extensive collection of publicly accessible dermoscopic skin images, including clinical information¹. The increasing prevalence of dermoscopic imaging has promoted research, leading to the development of automated Computer-Aided Diagnosis (CAD) solutions for diagnosing melanoma and other cancers. These images have been instrumental in developing state-of-the-art models using Convolutional Neural Network (CNN) [43] and Vision Transformer (ViT) [11, 47] for skin lesion classification. Prior studies [13] have demonstrated that integrating clinical metadata with dermoscopic images can significantly improve the accuracy of melanoma detection. However, the challenge remains to effectively balance and preprocess multimodality datasets to ensure robust model training and evaluation.

Including clinical metadata with dermoscopic images for skin lesion analysis opens up possibilities, for instance, the ensemble models with metadata provide better model diversity. Dong et al. [13] underscore the advantages of integrating clinical metadata with dermoscopic images and propose a novel pipeline for segmentation and classification tasks, demonstrating improved performance in skin cancer diagnosis. However, Andre et al. [29] point out the limitations of integrating metadata into CNN models, emphasising that the efficacy depends on the concatenation methods. Cassidy et al. [6] highlight duplicated dermoscopic images in the ISIC dataset across testing and training sets, emphasising the need for a curated dataset for future research. These prior works highlight the importance of careful consideration and transparency in using clinical metadata alongside dermoscopic images in skin lesion classification.

The main contributions of this paper are as follows:

- We construct a curated and balanced dataset, the ISIC-

¹<https://www.isic-archive.com/>

DICM-17K (ISIC Dermoscopic Images and Clinical Metadata 17K) Dataset, derived from the ISIC Archive Gallery, comprising 17,060 dermoscopic images and clinical metadata (8,530 melanoma and 8,530 non-melanoma classes).

- We benchmark the proposed dataset using 9 state-of-the-art backbone models across 4 settings (image-only vs image+meta and supervised learning vs transfer learning models).
- We analyse the importance of clinical metadata in enhancing model performance and feature clustering with attention regions using statistical and visual measures.

2. Related Works

Deep neural networks have revolutionised image processing, excelling in segmentation, feature extraction, and classification [27]. Studies have explored various architectures, including ConvNeXt, GoogleNet, DenseNet201, and Swin Transformer, with ConvNeXt demonstrating notable performance [30]. The Intelligent Multilevel Thresholding with Deep Learning (IMLT-DL) model integrates preprocessing techniques with Inception v3-based feature extraction for enhanced accuracy [32]. Jojoa Acosta et al. [23] proposed a two-stage approach using Mask R-CNN and ResNet152, achieving high accuracy and balanced sensitivity and specificity. Afza et al. [2] developed a hierarchical framework combining superpixels, ResNet-50, and optimization algorithms, demonstrating improved accuracy across multiple datasets. Azeem et al. [5] introduced SkinLesNet, a novel multi-layer CNN outperforming benchmark models on various datasets. Salma et al. [34] presented a CAD system incorporating image preprocessing, segmentation, and feature extraction based on the ABCD rule, with ResNet50 and SVM achieving superior performance. Combined with preprocessing and feature extraction, ANU-Net demonstrated high accuracy in classification tasks [31]. A melanoma detection system leveraging k-means clustering, feature extraction, and deep learning effectively classified cases using the PH2 database [41]. Additionally, a two-stage approach integrating Mask R-CNN for region-of-interest cropping and ResNet152 for classification outperformed earlier models in accuracy and balance, particularly in the 2017 International Symposium on Biomedical Imaging challenge [23].

While deep learning has advanced skin lesion classification, traditional models often overlook a crucial element—clinical metadata. Sun et al. [39] recognised this gap and demonstrated that integrating patient information and augmentation metadata could significantly boost classification accuracy. Building on this, Dong et al. [13] introduced a novel approach, leveraging Multi-Scale Holistic Feature Exploration and Cross-Modality Collaborative Feature Exploration to refine segmentation and classification. However, clinical metadata remains challenging due

to handling missing values, varying encoding strategies, single-feature numerical encoding for age, image count, and image size, and the complexity of unifying heterogeneous data types into a consistent 15-dimensional representation. Meanwhile, Pacheco and Krohling [29] tackled the challenge head-on with the Metadata Processing Block, ensuring metadata actively enhances image-based feature extraction. Furthering this effort, Salma and Eltrass [34] developed an automated system that combined morphological filtering, Grab-cut segmentation, and the ABCD rule with pre-trained CNNs and SVM, underscoring the potential of metadata-driven classification. A graph-based intercategory and intermodality network that leverages relationships between clinical and dermoscopic images, including the 7-point checklist categories [15].

Wen et al. [45] highlight that integrating clinical metadata remains challenging due to data heterogeneity and standardisation issues. Clinical metadata (e.g., age, sex, lesion history, genetic factors) comes from diverse sources, making standardisation difficult. Variations in data formats, missing values, and inconsistent labelling across datasets hinder model training. Lack or incomplete clinical metadata can introduce bias and reduce model reliability [46]. Furthermore, adding metadata can introduce biases, such as overfitting to demographic factors and leading to unfair predictions. To ensure clinical trust, models must prioritise interpretability, allowing clinicians to understand metadata-driven decision-making [48]. Training deep learning models with image and metadata input requires additional computational resources, and metadata preprocessing (e.g., one-hot encoding, normalisation) must be optimised to avoid unnecessarily increasing model complexity [44]. Gessert et al. [16] showed a clear benefit to network performance from including patient metadata when training CNNs. However, effectively integrating clinical metadata remains challenging, as it requires careful encoding of categorical variables (e.g., anatomical site and sex) and thoughtful handling of missing values in numerical features (e.g., age). Improper encoding can introduce bias or reduce model robustness, underscoring the complexity and importance of metadata integration in deep learning pipelines.

Melanoma classification is often hindered by imbalanced datasets, where the scarcity of malignant cases skews model learning and reduces diagnostic reliability [3]. To mitigate these challenges, researchers have explored ensemble models tailored for heterogeneous datasets [38], along with data augmentation and oversampling techniques [35] to address data imbalance. Using residual neural networks, Yu et al. [50] tackled imbalance with a one-class classification approach. A deep clustering method with center-oriented margin-free triplet loss enhances class separation [28]. Jaisakthi et al. [37] also demonstrated that integrating metadata with EfficientNet and transfer learning sig-

Table 1. Clinical metadata features. #Records denotes the corresponding number of available records of the collected metadata.

Feature Name	Description	Type	#Records
isic_id	ISIC lesion image identification number	string	80,504
patient_id	Patient identification number	string	40,631
lesion_id	ISIC lesion identification number	string	62,919
sex	Patient sex	float	79,183
age_approx	Patient age approx	float	79,071
anatom_site_general	General anatomic location of the lesion	string	64,220
benign_malignant	Type of the lesion	string	72,418
dermoscopic_type	Type of dermoscopic image	string	21,274
diagnosis	Lesion diagnosis	string	53,035
diagnosis_confirm_type	A variable that describes how the diagnosis was classified	string	77,257
image_type	Type of image	string	80,504
mel_class	What was the subclassification of malignant melanoma?	string	1,010
mel_mitotic_index	What was the mitotic index of invasive malignant melanomas?	float	55
mel_type	Histologic subtype	string	198
mel_ulcer	Was there histopathologic ulceration present	float	217
melanocytic	Is an associated melanocytic nevus present with a melanocytic tumor?	boolean	51,619
nevus_type	What was the subclassification of the nevus	string	1,369
clin_size_long_diam_mm	The longest diameter of the lesion in mm	float	3,955
mel_thick_mm	Thickness is the depth of invasion in millimetres	float	678
personal_hx_mm	Personal history of melanoma	float	14,186
family_hx_mm	Family history of melanoma	float	13,650

Table 2. Summary of the ISIC Archive Gallery datasets distribution of images across different categories. #Images denotes the corresponding number of available images in the collected dataset.

Image Type	#Images	Total
dermoscopic	80,504	81,030
clinical: overview	432	
clinical: close-up	58	
TBP tile: overview	36	

nificantly improves classification performance. However, these methods only partially address the challenge, underscoring the necessity for a well-curated, balanced dataset incorporating clinical metadata to ensure robust and unbiased skin lesion classification. Despite these advancements, the full potential of incorporating metadata in deep learning models remains underexplored. This presents a significant opportunity to enhance melanoma prediction and skin lesion classification.

3. Datasets

The ISIC 2016 to 2020 Challenges [8–10, 17, 33, 42] have significantly advanced skin lesion analysis research by providing high-resolution, biopsy-confirmed digital skin lesion images with expert annotations and globally sourced metadata. These ISIC challenges have expanded available datasets and fostered a collaborative environment for re-

searchers and professionals in dermatology and medical imaging. The ISIC datasets are encompassed within the ISIC Archive Gallery² and ISIC 2016 to 2020 Challenge datasets. The number of images has increased substantially each year since introducing these challenges. This section augments the training dataset by incorporating the latest clinical metadata and dermoscopic images from the ISIC Archive Gallery and publicly available ISIC datasets to promote the field further.

The dermoscopic image types within the ISIC Archive Gallery dataset encompass various categories crucial for skin lesion research. Table 2 presents the dataset summary, a collection of 81,030 entries (last accessed on Nov 2023), encompasses four distinct image types: dermoscopic, clinical close-up, clinical overview, and TBP (total body photos) tile overview. The *Dermoscopic* category includes 80,504 images and high-resolution magnified views essential for detailed lesion analysis. This study focuses on the *Dermoscopic* images, one of the most commonly used image types in skin lesion diagnosis by clinicians.

The clinical metadata collection process sourced data from the ISIC Archive³. The metadata of 80,504 dermoscopic images includes 24 features directly retrieved from patient clinical information. To optimise the information for analysis, 21 features have been carefully selected (Table 1). Three features (attribution, copyright_license, and acquisi-

²<http://gallery.isic-archive.com/>

³<https://api.isic-archive.com/>

tion_day) were excluded due to their lack of relevance.

Based on the availability across datasets, this study considers three key features: the patient’s approximate age, sex, and general anatomic location of the lesion. More specifically, over 98% of patients have age and sex information, and around 80% have the general anatomical location of the lesions documented. Other features, such as lesion size and personal or family history of melanoma, are limited to only 1% to 17% of records, making them less effective for this study. Moreover, only the ISIC 2019 and 2020 datasets provide ground truth of the training and testing sets’ metadata for the patient’s age, sex, and anatomical location. Therefore, age_approx, sex and anatom_site_general features are adequate for fair comparisons in this study.

We applied imputation strategies to handle missing values in the selected three metadata features. The patient’s sex was imputed using the mode, assuming the most frequent value represents the majority distribution. The approximate age was filled with the median to mitigate the impact of skewed distributions. For the general anatomic location, missing values were replaced with ‘unknown’ to preserve dataset consistency without introducing bias. These strategies maintain data completeness while minimising distortion, enhancing the robustness of downstream analyses in multimodal deep-learning models.

4. Methodology

In skin lesion classification, deep learning models necessitate balanced classes to prevent overfitting, even when using categorical cross-entropy to tackle imbalance. The ISIC datasets present a significant class imbalance, necessitating the creation of a balanced dataset maintaining the same number of images in each class. This study introduces a curated balanced dataset, ISIC-DICM-17K Dataset, which encompasses dermoscopic images and clinical metadata with a 1:1 ratio of melanoma and non-melanoma without duplicate records. The subsequent sections elaborate on the curation process of the ISIC-DICM-17K Dataset (Section 4.1) and the description of data modalities for the input of multimodal deep learning models (Section 4.2).

4.1. ISIC-DICM-17K Dataset

We present the ISIC-DICM-17K (ISIC Dermoscopic Images and Clinical Metadata 17K) Dataset, a balanced collection of 17,060 dermoscopic images with clinical metadata curated from the ISIC Archive Gallery. This dataset expands the curated balanced dataset [6] comprising 4,905 melanoma and 4,905 non-melanoma images curated from the 2016 to 2020 ISIC challenge datasets.

We identify 80,504 unique image files and metadata records with unique ISIC image IDs. Binary-similarity images from ISIC 2016 to 2020 datasets were removed us-

ing the Rdfind⁴ and duplicate removal strategy [6]. To reduce the confusion, we define melanoma (MEL) for records labelled ‘malignant’ or ‘indeterminate/malignant’ and non-melanoma (NON-MEL) for the remaining categories. Table 3 summarises the ‘benign_malignant’ defined classes of the additional curated images. We implemented and applied duplicate removal to the remaining 20,311, and the curated set of 56,987 records [6], 371 and 7 duplicates were removed, respectively.

The curation and duplicate removal strategy provides a clean dataset with 19,933 dermoscopic images and clinical metadata (naming as Curated19933 in Table 4), 3,625 MEL and 16,308 NON-MEL classes. Due to the significant class imbalance between MEL and NON-MEL classes (1:4.5 ratio), a clean and balanced dataset has been crafted, comprising 7,250 dermoscopic images (3,625 in each class), including clinical metadata.

The balanced ISIC skin image dataset [6] consists of dermoscopic images only. We retrieved the correspondence clinical metadata from the ISIC Archive using the image filenames. Table 4 summarises the MEL and NON-MEL class distribution across the curated and balanced datasets. This study presents an extended version of balanced datasets named the ISIC-DICM-17K Dataset by employing curation and balance steps. We identified 3,625 NON-MEL records with complete metadata for the selected features and combined them with the MEL records. In total, this dataset includes 8,530 MEL and 8,530 NON-MEL images. Each group contains 3,625 from this study and 4,905 from Cassidy et al. [6], resulting in 17,060 images with metadata.

4.2. Data Modality

The models were trained on the ISIC-DICM-17K dataset using two types of input data: i) dermoscopic images only, and ii) dermoscopic images with clinical metadata. Metadata features were represented as feature vectors and then fused with image vectors as input to multimodality models. We evaluate them using the ISIC 2020 test set with six anatomical site categories. However, the training set has eight categories. The training set’s anterior, lateral, and posterior torso were consolidated into a single torso category to align the categories. Therefore, the meta-feature vector for the model input consists of sex, age_approx, and 8 one-hot encoded anatom_site_general features. We model it as a binary classification problem to categorise benign (NON-MEL) and malignant (MEL) records.

Following prior works [18, 49], we conduct our experiments using CNN backbones, EfficientNets [40] (ENets) and ResNets [20], respectively, followed by late fusion to combine features. Metadata is processed through linear layers and concatenated with CNN-derived image features be-

⁴<https://github.com/pauldreik/rdfind>

Table 3. Class distribution of MEL and NON-MEL after curation. *Duplicates* and *Remaining* denote the number of duplicate image files removed and the remaining images after applying the duplicate removal strategy, respectively.

Lesion Type	#Records	Class	Subtotal	Total	Duplicates	Remaining
benign	16,398	NON-MEL	16,660	20,311	378	19,933
indeterminate	144	NON-MEL				
indeterminate/benign	61	NON-MEL				
indeterminate/malignant	71	MEL	3,651			
malignant	3,580	MEL				

Table 4. Number of MEL, NON-MEL, and total records in each dataset after curation and balance steps. The ISIC-DICM-17K dataset, derived from the curated data, includes dermoscopic images and clinical metadata with a balanced 1:1 ratio of melanoma (MEL) to non-melanoma (NON-MEL) without duplicate records.

Dataset	Curated			Balanced		
	MEL	NON-MEL	Total	MEL	NON-MEL	Total
Cassidy et al. [6]	4,905	52,082	56,987	4,905	4,905	9,810
Curated19933	3,625	16,308	19,933	3,625	3,625	7,250
ISIC-DICM-17K	-	-	-	8,530	8,530	17,060

for the final fully connected layer. Ha et al. [18] retained the original ENet architecture, integrating pre-processed metadata features into the metadata block. In contrast, Yap et al. [49] combined dermoscopic images and metadata, feeding the pre-processed metadata into a modified ResNet-50 to extract image features, which were then concatenated with metadata features and processed through two fully connected layers for classification. This study further investigates MetaFormer [12], a hybrid framework combining convolutional layers for vision encoding and transformer layers for fusing vision and metadata. It uses MB-Conv blocks in the first three stages and Relative transformer blocks in the last two. MetaFormer is adapted to use a pre-processed categorical metadata feature vector instead of a non-linear embedding.

5. Experimental Results

Implementation Details. For this study, the ISIC-DICM-17K dataset includes 1,962 dermoscopic images with clinical metadata in the validation set (981 MEL and 981 NON-MEL) [6] and 15,098 dermoscopic images with clinical metadata in the training set. The training set was used to train the supervised learning models and to fine-tune the models that were pretrained using ImageNet1K for the backbones (transfer learning models). All images were resized, with the shortest side reduced to 224 pixels, followed by center-cropping to 224×224 pixels. Each model was trained with epochs=100 and batch_size=32. Adam [24] (ENet and ResNet) and AdamW [26] (MetaFormer) optimisers and image augmentation [18] were employed. We adopted a cosine annealing schedule with one warm-up epoch [25]. Each model was tuned for the initial learning

rate for the cosine cycle, ranging from 1e-4 to 3e-4. During the warm-up epoch, the learning rate is set to the initial learning rate×0.1. Early stopping was employed with patience=10 to ensure each model converged optimally.

Benchmarks. In Table 5, we evaluated the model performance using dermoscopic images (image-only) with clinical metadata (image+meta) on the ISIC 2020 test set. The results were evaluated on Kaggle⁵ due to lack of ground truth labels in the test set. The baseline from Cassidy et al. [6] (4,905 MEL and 4,905 NON-MEL) was calculated using identical settings, serving as a reference for evaluating the ISIC-DICM-17K Dataset (8,530 MEL and 8,530 NON-MEL). Across all tested model configurations, those trained on the proposed dataset consistently outperformed the ones trained on the baseline dataset, regardless of backbone architecture. Clinical metadata significantly enhances model performance across our experiments. In particular, all supervised ENet backbone variants consistently benefit from metadata integration, achieving a 3%-6% performance gain. Furthermore, transfer learning models outperformed supervised learning counterparts in nearly all scenarios, with MetaFormer demonstrating the most substantial improvement at 12%. Fine-tuned ENet-B3 with metadata achieves the best AUC of 0.8851, outperforming the model without metadata. Integration of clinical metadata proves beneficial, particularly when combined with transfer learning, as demonstrated by consistently higher AUC scores.

Statistical Comparison. We conducted t-SNE analysis on ENet-B2 (image-only) and ENet-B4 (image+meta) back-

⁵<https://www.kaggle.com/competitions/siim-isic-melanoma-classification/>

Table 5. AUC on the ISIC 2020 testing set, the largest dermoscopic image test dataset in ISIC challenges, comprising 10,982 images with clinical metadata. Results obtained using supervised learning and transfer learning models are reported on their best epoch. The best results from each setting are bolded. The results were evaluated using ISIC 2020 Challenge platform due to the absence of ground truth labels in the test set. Specifically, submissions were accessed based on AUC between the predicted probabilities and observed targets.

Backbone	Supervised			Transfer		
	baseline [6]	<i>image</i>	<i>image + meta</i>	baseline [6]	<i>image</i>	<i>image + meta</i>
ENet-B0	0.8211	0.8400	0.8769	0.8479	0.8722	0.8755
ENet-B1	0.8115	0.8358	0.8630	0.8350	0.8651	0.8540
[18] ENet-B2	0.8100	0.8451	0.8457	0.8431	0.8759	0.8500
ENet-B3	0.8116	0.8314	0.8786	0.8618	0.8754	0.8851
ENet-B4	0.7908	0.8361	0.8838	0.8514	0.8339	0.8698
[49] ResNet-50	0.8201	0.8277	0.8534	0.8292	0.8204	0.8521
MetaFormer-0	0.7614	0.7430	0.7564	0.8564	0.8645	0.8531
[12] MetaFormer-1	0.7531	0.7412	0.7869	0.8495	0.8523	0.8568
MetaFormer-2	0.7592	0.7717	0.7305	0.8370	0.8727	0.8463

Table 6. Statistical analysis of MEL vs. NON-MEL separability across three layers: input (Input), feature extraction (FE) and prediction (Pred). The table reports intra-class distances (MEL and NON-MEL), inter-class distances, and three clustering validity metrics: the Calinski-Harabasz (CH) index, which measures cluster compactness and separation (higher is better); the Silhouette index (Si), indicating cluster cohesion and separation (higher is better); and the Davies-Bouldin (DB) index, where lower values signify better clustering.

Modality	Layer	Metrics					
		Intra(MEL)	Intra(NON-MEL)	Inter-class	CH	Si	DB
<i>image</i>	Input	22.7276	23.3148	25.4125	891.0790	0.0806	3.1154
	FE	23.0591	22.2817	24.9537	1488.9255	0.0824	3.2510
	Pred	29.0680	30.7721	31.2098	692.5916	0.0359	4.7788
<i>image + meta</i>	Input	22.6976	23.3018	25.4022	897.0860	0.0811	3.1056
	FE	28.4263	29.6396	29.5993	226.9248	0.0188	8.3152
	Pred	27.3637	27.1797	29.2844	1181.9816	0.0608	3.6490

bones, which achieved the highest AUC scores in supervised learning. Figure 1 represents a progressive separation of MEL and NON-MEL classes across three layers: input (Input), feature extraction (FE) and prediction (Pred). At the Input layer, intra-class distances for MEL and NON-MEL are comparable, while inter-class distances remain low, indicating poor separability. In the Feature Extraction (FE) layer, the image+meta modality outperforms image-only, as intra-class distances increase slightly. In contrast, inter-class distances remain stable, suggesting that additional metadata helps preserve class structure. At the Prediction (Pred) layer, image+meta maintains better inter-class separation while having lower intra-class distances, indicating more compact and well-defined clusters. In the image-only modality (Fig. 1a), the t-SNE plot reveals overlapping clusters with minimal separation between MEL (blue) and NON-MEL (orange). However, the image+meta modality (Fig. 1b) exhibits significantly more distinct clustering, illustrating enhanced separability when integrating clinical metadata. Table 6 shows that incorporating clinical metadata improves the clustering of features and the

class separation between the image-only and image+meta modality. The results demonstrate that including metadata enhances MEL vs. NON-MEL separability, particularly in the final prediction layer, where higher CH, better Si, and lower DB indices a stronger clustering performance. The CH index improves significantly by 597.8465, reflecting more compact clusters at higher layers. However, the DB index increases to 8.3152, highlighting a trade-off between intra-class compactness and inter-class separation, underscoring the complexity of incorporating clinical metadata in multimodal models.

Visual Comparison. We computed an overlapping index, $HeatMap_{index}$ [22], to evaluate the performance on supervised learning and transfer learning models. This measures the extent to which the classification result has better coverage on the focusing area of the skin lesion images. The $HeatMap_{index}$ is computed by the sum of the intensity of pixels in the heatmap within the segmentation area divided by the sum of all pixels in the heatmap. Using the ISIC 2017 test set (Task 1: Lesion Segmentation) with 600 ground truth segmentation masks, we generated Grad-CAM [36]

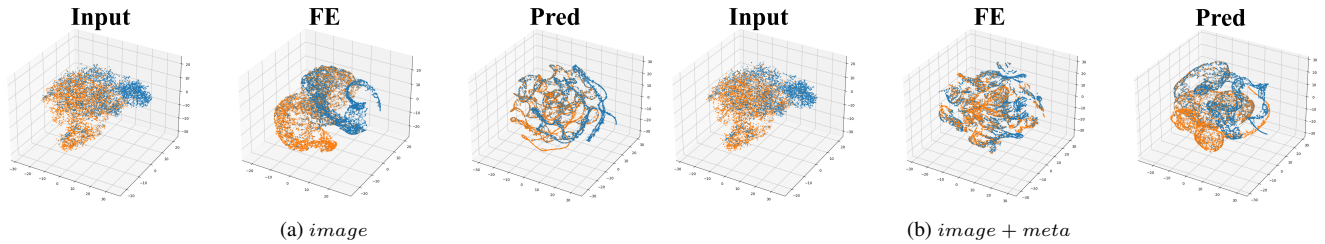


Figure 1. t-SNE visualisation of the curated balanced training set, highlighting the separability between MEL (blue) and NON-MEL (orange) classes across the input (Input), feature extraction (FE), and prediction (Pred) layers. Blue and orange demonstrate the most distinct and robust class distinction by increasing cluster separation and intra-class cohesion at the image+meta modality prediction layer (Pred).

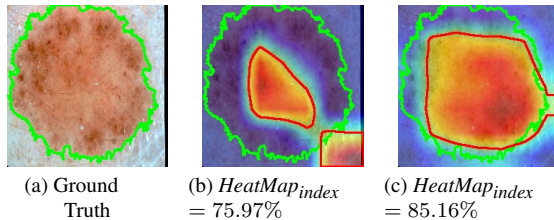


Figure 2. Examples of skin lesions focusing area with corresponding heatmaps created by the ENet-B2 model. Green contours represent the ground truth, and red contours represent the predicted Grad-CAM heatmap result.

heatmaps to visualise image regions with high attention for neural network classification. Our null hypothesis H_0 test for image-only modality demonstrated significant rejection across all EfficientNet backbones (Table 5), with a p-value of $1.57e-3$ for ENet-B2. This shows that transfer learning models (Fig. 2c) focus on the skin lesion area more effectively than supervised learning models (Fig. 2b), thus enhancing feature extraction and classification accuracy.

6. Discussion and Conclusion

This study introduces ISIC-DICM-17K Dataset, a curated and balanced dataset that improves skin lesion classification by incorporating clinical metadata. Our analysis shows that multimodal models utilising clinical metadata outperform image-only models across different architectures. For instance, the transfer learning ENet-B3 model with metadata achieved an AUC of 0.8851, substantially improving over the supervised learning variant’s AUC of 0.8314. This highlights the effectiveness of clinical metadata in enhancing the model’s ability to distinguish between melanoma and non-melanoma classes in skin lesion classification. Incorporating clinical metadata combined with dermoscopic images further augmented model performance in supervised learning. Metadata features, such as anatomical site, patient age and sex, improved the distinguishability of melanoma and non-melanoma classes, leading to better clustering and fea-

ture separation as observed in statistical metrics. Table 6 and Figure 1 further supported the finding, showing more distinct clusters for melanoma and non-melanoma classes in the multimodality (image+meta) compared to the single modality (image-only).

Our study establishes a foundation for future advancements in skin lesion analysis. Focusing on key clinical metadata features, such as the patient’s approximate age, sex, and anatomical site, we recommend incorporating additional features, such as lesion size and a personal or family history of melanoma. Including additional metadata features holds immense potential to push the boundaries of skin lesion classification. Our findings underscore the importance of further integrating more comprehensive metadata with skin lesion images to enhance the accuracy and reliability of classification models, inspiring a transformative direction for skin cancer research.

7. Limitations

The ISIC 2020 test set lacks ground truth labels, preventing the evaluation of additional performance metrics for skin lesion classification. Furthermore, our data access was limited to the last update in November 2023, which may impact the dataset’s completeness. Notably, our primary objective was to propose a curated, balanced dataset rather than to develop new model architectures. Lastly, while we sought to incorporate as much clinical metadata as possible, many features contained missing values, which were not factored into the model’s performance assessment.

References

- [1] Melanoma UK 2020 melanoma skin cancer report. <https://www.melanomauk.org.uk/2020-melanoma-skin-cancer-report.1>
- [2] Farhat Afza, Muhammad Sharif, Mamta Mittal, Muhammad Attique Khan, and D Jude Hemanth. A hierarchical three-step superpixels and deep learning framework for skin lesion classification. *Methods*, 202:88–102, 2022. 2
- [3] Ali H Alzamili and Nur Intan Raihana Ruhaiyem. A comprehensive review of deep learning and machine learning

- techniques for early-stage skin cancer detection: Challenges and research gaps. *Journal of Intelligent Systems*, 34(1):20240381, 2025. 2
- [4] Seokyoung An, Kyungsik Kim, Sungji Moon, Kwang-Pil Ko, Inah Kim, Jung Eun Lee, and Sue K Park. Indoor tanning and the risk of overall and early-onset melanoma and non-melanoma skin cancer: systematic review and meta-analysis. *Cancers*, 13(23):5940, 2021. 1
- [5] Muhammad Azeem, Kaveh Kiani, Taha Mansouri, and Nathan Topping. Skinlesnet: classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional neural network. *Cancers*, 16(1):108, 2023. 2
- [6] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap. Analysis of the isic image datasets: Usage, benchmarks and recommendations. *Medical image analysis*, 75:102305, 2022. 1, 4, 5, 6
- [7] Jyoti Chandra, Nazeer Hasan, Nazim Nasir, Shadma Wahab, Punniyakoti Veeraveedu Thanikachalam, Amirhossein Sahebkar, Farhan Jalees Ahmad, and Prashant Kesharwani. Nanotechnology-empowered strategies in treatment of skin cancer. *Environmental Research*, 235:116649, 2023. 1
- [8] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kallou, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 3
- [9] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kallou, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018.
- [10] Marc Combalia, Noel CF Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Cristina Carrera, Alicia Barreiro, Allan C Halpern, Susana Puig, et al. Bcn20000: Dermoscopic lesions in the wild. *arXiv preprint arXiv:1908.02288*, 2019. 3
- [11] Getamesay Haile Dagnaw, Meryam El Mouhtadi, and Musa Mustapha. Skin cancer classification using vision transformers and explainable artificial intelligence. *Journal of Medical Artificial Intelligence*, 2024. 1
- [12] Qishuai Diao, Yi Jiang, Bin Wen, Jia Sun, and Zehuan Yuan. Metaformer: A unified meta framework for fine-grained recognition. *arXiv preprint arXiv:2203.02751*, 2022. 5, 6
- [13] Caixia Dong, Duwei Dai, Yizhi Zhang, Chunyan Zhang, Zongfang Li, and Songhua Xu. Learning from dermoscopic images in association with clinical metadata for skin lesion segmentation and classification. *Computers in Biology and Medicine*, 152:106321, 2023. 1, 2
- [14] Benlier Erol, Usta Ufuk, Unal Yasin, Cayci Aslihan, and Kir Koray. True hematogenous metastases of melanoma on contralateral skin graft donor site: a case report. *Melanoma Research*, 18(6):443–446, 2008. 1
- [15] Xiaohang Fu, Lei Bi, Ashnil Kumar, Michael Fulham, and Jinman Kim. Graph-based intercategory and intermodality network for multilabel classification and melanoma diagnosis of skin lesions in dermoscopy and clinical images. *IEEE Transactions on Medical Imaging*, 41(11):3266–3277, 2022. 2
- [16] Nils Gessert, Maximilian Nielsen, Mohsin Shaikh, René Werner, and Alexander Schlaefer. Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. *MethodsX*, 7:100864, 2020. 2
- [17] 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. 3
- [18] Qishen Ha, Bo Liu, and Fuxu Liu. Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge. *arXiv preprint arXiv:2010.05351*, 2020. 4, 5, 6
- [19] Nazeer Hasan, Mohammad Imran, Afsana Sheikh, Nidhi Tiwari, Abhinav Jaimini, Prashant Kesharwani, Gaurav Kumar Jain, and Farhan Jalees Ahmad. Advanced multifunctional nano-lipid carrier loaded gel for targeted delivery of 5-fluorouracil and cannabidiol against non-melanoma skin cancer. *Environmental Research*, page 116454, 2023. 1
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4
- [21] Monika Janda, Catherine M Olsen, Victoria J Mard, and Anne E Cust. Early detection of skin cancer in australia—current approaches and new opportunities. *Public health research & practice*, 32(1), 2022. 1
- [22] Joanna Jaworek-Korjakowska, Andrzej Brodzicki, Bill Cassidy, Connah Kendrick, and Moi Hoon Yap. Interpretability of a deep learning based approach for the classification of skin lesions into main anatomic body sites. *Cancers*, 13(23):6048, 2021. 6
- [23] Mario Fernando Jojoa Acosta, Liesle Yail Caballero Tovar, Maria Begonya Garcia-Zapirain, and Winston Spencer Percybrooks. Melanoma diagnosis using deep learning techniques on dermoscopic images. *BMC Medical Imaging*, 21:1–11, 2021. 2
- [24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [25] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 5
- [26] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5
- [27] Arief Kelik Nugroho, Retantyo Wardoyo, Moh Edi Wibowo, and Hardyanto Soebono. Image dermoscopy skin lesion classification using deep learning method: systematic literature

- review. *Bulletin of Electrical Engineering and Informatics*, 13(2):1042–1049, 2024. 2
- [28] Şaban Öztürk and Tolga Çukur. Deep clustering via center-oriented margin free-triplet loss for skin lesion detection in highly imbalanced datasets. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4679–4690, 2022. 2
- [29] Andre GC Pacheco and Renato A Krohling. An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, 25(9):3554–3563, 2021. 1, 2
- [30] PP Pranav and S Sarath. Comparative study of skin lesion classification using dermoscopic images. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–5. IEEE, 2023. 2
- [31] Vankayalapati Radhika and B Sai Chandana. Skin melanoma classification from dermoscopy images using anu-net technique. *International Journal of Advanced Computer Science and Applications*, 13(10), 2022. 2
- [32] G Reshma, Chiai Al-Atroshi, Vinay Kumar Nassa, BT Geetha, Gurram Sunitha, Mohammad Gouse Galety, and S Neelakandan. Deep learning-based skin lesion diagnosis model using dermoscopic images. *Intelligent Automation & Soft Computing*, 31(1), 2022. 2
- [33] Veronica Rotemberg, Nicholas Kurtansky, Brigid Betz-Stablein, Liam Caffery, Emmanouil Chousakos, Noel Codella, Marc Combalia, Stephen Dusza, Pascale Guitera, David Gutman, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific data*, 8(1):34, 2021. 3
- [34] Wessam Salma and Ahmed S Eltrass. Automated deep learning approach for classification of malignant melanoma and benign skin lesions. *Multimedia Tools and Applications*, 81(22):32643–32660, 2022. 2
- [35] Gehad Ismail Sayed, Mona M Soliman, and Aboul Ella Hassanien. A novel melanoma prediction model for imbalanced data using optimized squeezeNet by bald eagle search optimization. *Computers in biology and medicine*, 136:104712, 2021. 2
- [36] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 6
- [37] Jaisakthi SM, Mirunalini P, Chandrabose Aravindan, and Rajagopal Appavu. Classification of skin cancer from dermoscopic images using deep neural network architectures. *Multimedia Tools and Applications*, 82(10):15763–15778, 2023. 2
- [38] Ellak Somfai, Benjamin Baffy, Kristian Fenech, Rita Hosszú, Dorina Korozs, Marcell Polik, Miklos Sardy, and András Lőrincz. Handling dataset dependence with model ensembles for skin lesion classification from dermoscopic and clinical images. *International Journal of Imaging Systems and Technology*, 33(2):556–571, 2023. 2
- [39] Qilin Sun, Chao Huang, Minjie Chen, Hui Xu, and Yali Yang. Skin lesion classification using additional patient information. *BioMed research international*, 2021(1):6673852, 2021. 2
- [40] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 4
- [41] R Thamizhamuthu and D Manjula. Skin melanoma classification system using deep learning. *Computers, Materials & Continua*, 68(1), 2021. 2
- [42] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018. 3
- [43] Turker Tuncer, Prabal Datta Barua, Ilknur Tuncer, Sengul Dogan, and U Rajendra Acharya. A lightweight deep convolutional neural network model for skin cancer image classification. *Applied Soft Computing*, page 111794, 2024. 1
- [44] Sirawich Vachmanus, Thanapon Noraset, Waritsara Piyanonpong, Teerapong Rattananukrom, and Suppawong Tuarob. Deepmetaforge: A deep vision-transformer metadata-fusion network for automatic skin lesion classification. *IEEE Access*, 11:145467–145484, 2023. 2
- [45] David Wen, Andrew Soltan, Emanuele Trucco, and Rubeta N Matin. From data to diagnosis: skin cancer image datasets for artificial intelligence. *Clinical and experimental dermatology*, 49(7):675–685, 2024. 2
- [46] Adrienne D Woods, Daria Gerasimova, Ben Van Dusen, Jayson Nissen, Sierra Bainter, Alex Uzdavines, Pamela E Davis-Kean, Max Halvorson, Kevin M King, Jessica AR Logan, et al. Best practices for addressing missing data through multiple imputation. *Infant and Child Development*, 33(1):e2407, 2024. 2
- [47] Chao Xin, Zhifang Liu, Keyu Zhao, Linlin Miao, Yizhao Ma, Xiaoxia Zhu, Qiongyan Zhou, Songting Wang, Lingzhi Li, Feng Yang, et al. An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, 149:105939, 2022. 1
- [48] Qian Xu, Wenzhao Xie, Bolin Liao, Chao Hu, Lu Qin, Zhengzijing Yang, Huan Xiong, Yi Lyu, Yue Zhou, and Aijing Luo. Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review. *Journal of healthcare engineering*, 2023(1):9919269, 2023. 2
- [49] Jordan Yap, William Yolland, and Philipp Tschandl. Multimodal skin lesion classification using deep learning. *Experimental dermatology*, 27(11):1261–1267, 2018. 4, 5, 6
- [50] Lisu Yu, Yifei Wang, Liyu Zhou, Jinsheng Wu, and Zhenghai Wang. Residual neural network-assisted one-class classification algorithm for melanoma recognition with imbalanced data. *Computational Intelligence*, 39(6):1004–1021, 2023. 2