Towards Robust Skin Lesion Classification: Lesion Segmentation, Mole Collision Simulation and Hierarchical learning *

Hang Nguyen[1,2], Paul Fricker[1,2], Marianne Defresne[1,2], Frederik Pahde[3,4], Serena Bonin[6], Jonathan Wolfe[6,7], Eros Azzalini[6], Iris Zalaudek[5], Skye Tanzmann[2,6], and Zung Nguyen[1,2]

[1] Torus AI, 12 Av. de l'Europe, Ramonville-Saint-Agne, France
[2] Belle.ai, 245 First Street Riverview II, 18th Floor Cambridge, MA, USA
[3] Fraunhofer Heinrich Hertz Institute, Berlin, Germany
[4] Technische Universität Berlin, Germany
[5] Department of Dermatology and Venereology, University of Trieste, Italy
[6] DSM-Department of Medical Sciences, Unit of Dermatology, Univ. of Trieste, Italy
[7] Jefferson-Einstein Montgomery Medical Center, East Norriton, PA, USA
[8] Harvard University

Abstract. Accurate and robust classification of skin lesions remains a critical challenge in dermatological AI due to issues such as visual similarity between lesion types and dataset imbalances. In this work, we propose a comprehensive framework to improve skin lesion classification by integrating three key strategies: lesion segmentation, synthetic mole collision simulation, and hierarchical learning. First, lesion segmentation is used to localize the mole and focus the model on relevant regions, reducing background noise. Second, we introduce a novel synthetic data generation technique that simulates mole collisions by combining two lesions into a single image, improving the accuracy of the model in case of multi lesions appearance. These synthetic images also serve as a form of data augmentation, enhancing model generalization. Finally, we employ hierarchical learning that predicts lesion classes and sub-classes. Experimental results demonstrate that while our multi-label model may be slightly outperformed on class-specific metrics such as sensitivity, it achieves superior or comparable performance on global metrics like AUROC and specificity. This study highlights the potential of combining structural priors, synthetic augmentation, and label hierarchy to advance robust skin lesion classification.

Keywords: Skin cancer detection, Multi-tasking, Hierarchical Learning

* This project is supported by European Union's Horizon 2020 research and innovation program through Intelligent Total Body Scanner for Early Detection of Melanoma (iToBoS, grant agreement No. 965221).

1 Introduction

Skin cancers are among the most common cancers in the world and carry a risk of severe morbidity and, if diagnosis is delayed, mortality. This is especially pertinent for patients with skin of color; although such patients are less likely to be afflicted by skin cancer, they are more likely to die from the disease due to delayed presentation and detection [11]. According to [30], in 2020, melanoma represented 324,000 new cancer cases and 57,000 cancer deaths worldwide, while non-melanoma skin cancer represented approximately 1.2 million new cancer cases and 63,000 cancer deaths. Early detection and prompt treatment reduce morbidity and mortality [27].

Automatic skin cancer detection with image analysis is rapidly improving with the onset of deep learning [19]. Convolutional neural networks (CNN) are deep learning algorithms that are commonly used for image analysis, including image classification, segmentation, and object detection [25, 13]. Current CNN diagnostic accuracy is superior or equal to that of clinicians. The integration of CNNs into healthcare can significantly aid diagnostic accuracy, especially for populations with skin of color, as well as quicken the diagnosis process, improve treatment outcomes, and strengthen workflow efficiency [14, 13].

Lesion segmentation involves defining lesion borders and reducing background noise. The initial, pre-processing step of lesion segmentation makes it possible for the CNN to later extract lesion asymmetry, border, color, and diameter (ABCD) features for classification [1]. Other works have explored various pre-processing techniques to refine lesion segmentation and enhance feature extraction, these include downsampling image size, color space transformations, contrast enhancement, color normalization, cropping, flipping, and rotation [23]. For example, [12], achieved contrast enhancement through fast local Laplacian filtering (FlLpF) along HSV color transformation to reduce noise from irrelevant features. In contrast, this study used lesion segmentation to reduce possible noise by cropping and padding with averaged color around the lesion; averaged color padding was used instead of a blackout strategy to avoid confusing the model that associates black color with particular melanomas.

Although unlikely, an image may contain several adjacent or contiguous, unrelated lesions, known as collision skin lesions (CSL). Due to their rarity, CSLs have yet to be widely studied or characterized [7]. Experienced dermatologists often fail to recognize CSLs, highlighting the need for detection assistance [10]. In this study, we generate a plethora of synthetic CSL images, both for data augmentation and to prepare our model to perform multi-label classification if necessary. To our knowledge, this is the first study to produce and classify AI-generated synthetic CSL images.

This study developed a system for image-based skin cancer detection using hierarchical learning, a technique for machine learning that improves classification performance by predicting sub-type or sub-class information. Traditional algorithms use a flat classification approach in which other than ground truth, all other classes are treated as equally wrong. The flat classification model lacks decision-making transparency and consideration of semantic relationships be-

tween disease classes [15]. In contrast, the hierarchical approach allows for more accessible interpretation of results as humans can trace the sequential decision and prediction pathway [15, 6]. Additionally, inter-class relations can be implemented hierarchically by expert-defined class labels. Thus, the hierarchical approach is superior to the flat classification model based on result interpretability and transparency, which are of the utmost importance for deep learning medical diagnostics insofar the physician can validate the output [5].

Our model was trained with 10 classes and 122 sub-classes of skin lesions. For example, to enable the model to characterize melanoma with greater accuracy, we trained it to identify sub-classes such as *acral melanoma, melanoma in-situ, mucosal melanoma*, etc. In contrast with previous hierarchical research [16, 29], we did not predict the sub-class labels using clustering methods but rather used an adapted loss function to handle potential missing subclass information. In this case, the corresponding sub-class loss was set to zero, and no gradient update was made in the direction of the sub-class.

The contribution of this paper is as follow. We train an AI system for skin cancer detection (DermoscopyAI) capable of multi-label classification and employ a novel loss term designed to handle hierarchical class labels that achieves excellent results on a large test set collected in the Itobos project. We prove that hierarchical class-sub-class learning in DermoscopyAI results in higher performance than the "classical" non-hierarchical learning based on classes alone. We also share some of our pre-processed training data to encourage the reproduction of research.

2 DermoscopyAI Design

The DermoscopyAI system performs two main steps: skin lesion segmentation and image-based lesion classification (see Fig. 1). We first trained a model to segment the lesion. Using the segmentation result, we cropped the image around the lesion and padded the image with averaged color (see Fig. 2 L) to obtain a square image. Then, we trained the classifier on such modified images. This preprocessing is necessary as (i) a lesion might be very small relative to the entire image, creating noise that may confuse the model, and (ii) the lesion might be off-center, making it difficult for the model to localize the lesion. We selected sub-classes based on the following criteria:

- Cell type (*i.e.* squamous, melanocytes, and connective tissue)
- Severity level or stage
  - in situ is the early stage
  - invasive means spreading beyond the primary tissue layer
  - metastatic means spreading to other tissues and organs
- Pigmentation description (*i.e.* level of pigmentation, non-pigmented/amelanotic, and pigment variegation)
- Other geometric features, including the location of the lesion

Moreover, special sub-classes are separated from others to facilitate the estimation of the cancer risk level. For example, *Spitz nevus* is separated from other benign nevi in the class NV. and *Pyogenic granuloma* is separated from other benign vascular lesions in the class VASC. Details about our classification system can be found at https://github.com/AIpourlapeau/skincancer.

2.1 Lesion segmentation

A preliminary segmentation model was trained on ISIC images to segment skin lesions [17, 24]. We used Unet architecture with encoder EfficientNet B4 with weights pre-trained on ImageNet and input size $320 \times 320 \times 3$. This model was used to crop training images for classification model.

Next, we trained a secondary segmentation model for inference. This model reused the encoder from the classification model (with its weights frozen) to improve segmentation accuracy. This led to an increase in the validation Jaccard score from 0.782 to 0.835.

Both models were trained using a binary cross-entropy loss and an Adam optimizer with a decay learning rate (starting from 0.0001 with a rate decay of 0.992).

2.2 Image classification

Model: We used an EfficientNet B4 architecture pre-trained on ImageNet to extract features, on which are applied dropout (25%) and global average pooling. Features are then fed into three classification heads, all composed of a single linear layer and a Sigmoid activation function. Thus, the model has three outputs: oops score, class, and sub-class. The oops score corresponds to image type, valued at 0 for dermoscopic images, 0.5 for macro(clinical close-up) images, and 1 for irrelevant images (without any lesion). For the class and sub-class predictions, the Sigmoid activation was chosen to target multi-label classification (*i.e.* an image can belong to several classes).

To prevent the sub-class score from being higher than the class score and to ensure more consistent results, we introduce a multiplication step. This step also strengthens accurate class prediction as the model is forced to learn important features through sub-class labels. The multiplication step is done as follows:

- Let X be class output, $X \in R_+^{10}$, which is a positive vector of dimension 10
- Let $\hat{Y}$ be pre-sub-class output, $\hat{Y} \in R_+^{10\times 16}$, which is a 10×16 positive matrix, 16 is maximum number of sub-class
- Let Y be sub-class output, we have: $Y = X \otimes \hat{Y}$, where $\otimes$ is vector point-wise product, that is $Y_{ij} = X_i * Y_{ij}$.

The model input size is set to $512 \times 512 \times 3$. We used an NVIDIA GeForce RTX 3060 to train with Adam optimizer and decay learning rate scheduler, whose starting learning rate was 0.001 and rate decay was 0.999.

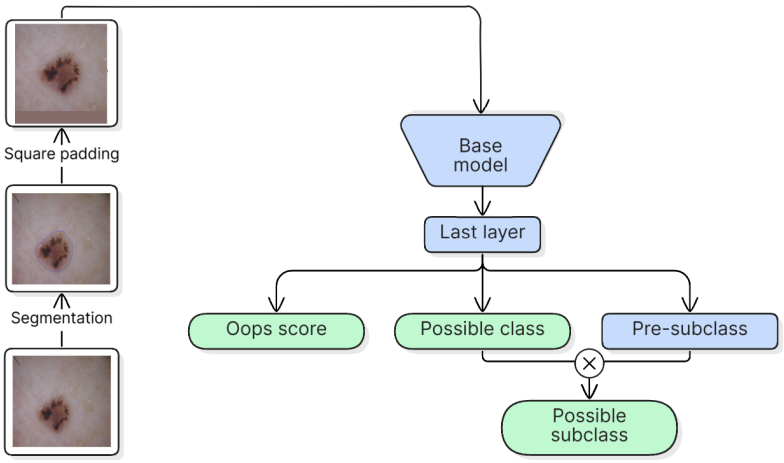

Fig. 1. Training flow and model structure in sub-class-based approach. Green boxes are the model's outputs. Blue boxes are the model's layers.

Loss: We applied an adaptive loss function as follows:

$$L = L_{MSE}(Oops) + L_{focal}(Class) + \mathbb{1}_{info} L_{focal}(sub - class) \tag{1}$$

where L_{MSE}, L_{focal} are Mean Squared Error and Focal losses, $\mathbb{1}_{info}$ is an indicator function which equals 1 if there is sub-class information, otherwise it equals 0. With this setup, the model will not be penalized when sub-class information is missing.

Dataset: We trained the model on a dataset of $104,813$ images, including the ISIC dataset and images obtained by our clinical partners.

Mole collision simulation: On a rare occasion, one image may contain several lesions of different classes, known as collision skin lesions (CSLs). With this possibility in mind, the model must be capable of multi-label classification and not just multi-class classification. Figure 2 shows some examples of such collisions. To enhance the robustness of the classification model, we generated 228,640 synthetic collision images (see Fig. 2 R). To generate such images, we first identify the location of the lesion in a single-lesion-image using lesion segmentation. Then, we crop and merge two selected images so that their lesions appear at two random positions within a single composite image. This process is carefully performed to ensure that both source images belong to the same data fold. Although the total number of synthetic collision images is twice that of the training images, during training, each batch contains 20% synthetic collision images and 80% original training images. By incorporating these synthetic images, we not only improve the model's accuracy in handling collision scenarios, but also benefit from their role as a form of data augmentation, enhancing overall model generalization.

Finally, we applied cross-validation on the whole dataset by splitting it into five folds, training on four folds, and leaving one fold for validation. We trained two models whose outputs were averaged to obtain the final prediction.

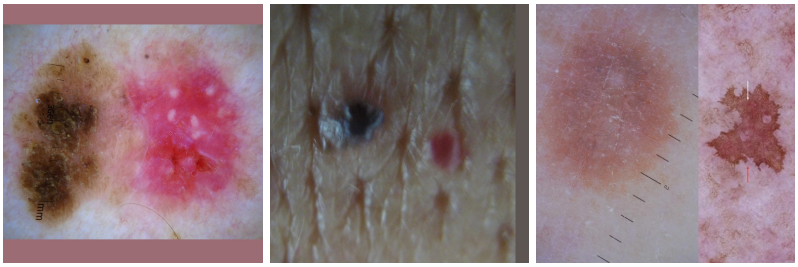

Fig. 2. Example of collision. (L) dermoscopic image of collision of BCC and BKL. (C) macroscopic image of collision of BCC and VASC. (R) synthetic collision of DF and BKL.

Data sharing: It is worth noting that the ISIC dataset contains very little sub-class information. With the help of our medical partners, we were able to obtain sub-class labels for around 40% of the ISIC images[9]. Moreover, there are many duplicated images in the ISIC dataset [8]. If the same images are in both the validation and training sets, the accuracy of the validation step will be reduced, leading to poor monitoring of the training process. To detect duplicated images, we used MobinetV3 to obtain a feature vector for each image, and then used Cosine similarity distance to compute the distance between every image. Images with a distance greater than 0.9 will have the same ID number. Different from [8], we do not delete these images but we put them in the same set during the train-validation split. Therefore, we ensure that the accuracy of similarity search models will not affect the classifier model. In other words, duplicated images act as augmentation; wrong predictions are not deleted.

3 Evaluation Result

3.1 Evaluation on public dataset

ISIC2018. We first tested our model on task 3 of ISIC 2018 [9] (also known as the HAM10000 dataset) as a clear test set is provided. Since images are already centered around the lesion, segmentation-based cropping is not necessary. We first compared our model to a version trained without sub-class information (see Table 1). Even if no such label is available for this dataset, taking sub-classes into account with hierarchical learning leads to improved predictions. We also compared our model with other approaches (by gathering their published results,

[9] Please find the data here: https://github.com/AIpourlapeau/skincancer

so the train/test split may differ). They all use class-level information only. Due to the multi-label approach, our DermoscopyAI is outperformed on best-class-based metrics (accuracy and sensitivity), but it performs better (or equal) on metrics considering prediction over all classes (AUROC, specificity). Multi-label training is beneficial in real-case scenarios, especially in case of collisions between two types of lesions.

Table 1. Results on ISIC 2018 (task 3) test set.

Model	Accuracy	AUROC	Sensitivity	Specificity
DermoscopyAI (no sub-class)	86.4%	0.974	79.0%	96.6%
DermoscopyAI	89.0%	0.982	81.3%	**97.1%**
(Al-Masni, Kim and Kim 2020) [4]	89.3%	-	81.0%	81.3%
(Afza *et al.*2022) [2]	85.8%	0.977	86.1%	86.0%
(Maqsood and Damaševičius 2023) [21]	**98.6%**	**0.988**	**93.9%**	96.4%

PH2 dataset. We assessed the robustness of our model on the public dataset PH2 [22], used as external validation. None of the 200 images (40 melanoma and 160 nevi) were ever seen during training. Despite the distribution shift, the model achieves high classification metrics: 93.5% for accuracy, 82.8% for sensitivity, and 96.2% for specificity. Those results cannot be directly compared with methods [2, 3, 20, 21, 26, 28] training on PH2 dataset as the results provided (up to 100% accuracy [3]) may result from overfitting.

3.2 Result of a blind test set

We evaluated our model and the Kaggle winner's solution [18] on the Itobos test set, which is independent of training and validation data. It contains 4,894 dermoscopic images collected and labeled by the Dermatology Unit at the University of Trieste. To ensure inference is objective and reflects real-world performance, it was run by the Fraunhofer Heinrich Hertz Institute so that the authors could not access the images and the metadata. The test set contains nine classes: MEL, BCC, SCC, AK, NV, BKL, DF, VASC, and Other. The evaluation result is summarized in Table 2. In Kaggle's solution, there were 18 CNN model structures, each trained on five-folds, resulting in up to 90 models to run at inference. We dropped 25 models as they required meta-data or they predicted only four classes. Thus, we used 65 models to predict the Itobos test set. The final result is the average value of the outputs. We note [18] used several very large architectures, such as EfficientNetB6 and B7. In contrast, our solution is much lighter, with only two EfficientNetB4 models, resulting in an inference time $\sim 50\times$ faster using an NVIDIA GeForce RTX 3060.

Ensemble of 65 models has top-1, top-2, and top-3 accuracy of 0.72, 0.88, and 0.94, respectively, while our DermoscopyAI has 0.80, 0.92, and 0.95, respectively. The best improvement (AUROC 0.88 vs 0.71) is for the class Others which contains diverse lesions: *clear cell acanthoma* (class BKL, sub-class BKL CCA), *eccrine poroma* (class BAL, sub-class BAL Poroma), *sebaceous hyperplasia* (class BAL, sub-class BAL SGH), *acral hemorrhage* (class VASC, sub-class VASC Haem) and *nail hemorrhage* (class VASC, sub-class VASC Haem). This result demonstrates the advantage of incorporating hierarchical learning—classes and sub-classes—during training. We do not discuss the accuracy at the sub-class level since we do not have that information for this test set.

Table 2. Summary of evaluation results on Itobos set. There are two MEL columns for two different thresholds. The thresholds were chosen on an independent set to optimal sensitivity and specificity.

Class	MEL	MEL	BCC	SCC	AK	NV	BKL	DF	VASC	Others
AUROC - Kaggle		0.91	**0.97**	**0.96**	0.95	0.93	**0.95**	**0.98**	**0.98**	0.71
AUROC - Ours		**0.92**	0.96	**0.96**	**0.96**	**0.97**	**0.95**	0.96	0.96	**0.88**
Threshold	**0.3**	**0.5**	**0.5**	**0.3**	**0.3**	**0.7**	**0.50**	**0.2**	**0.2**	**0.2**
Sensitivity (%)	95	87	85	91	87	88	80	85	87	80
Specificity (%)	62	81	95	92	93	93	96	98	98	82
# Images	868	868	697	211	148	2422	283	46	85	134

We also report cutoff thresholds in Table 2. The default value is 0.50, but it can be tuned for most diseases to achieve a better sensitivity/specificity trade-off. We set thresholds on an independent set so that malignant lesions (especially melanoma, but also BCC, SCC, and AK) have high sensitivity (95% for MEL) and reasonable specificity (62%). On the contrary, we favored high specificity for benign classes such as nevus. The class "Others" contains very diverse lesions that are difficult to detect, so we used a low threshold of 0.2 to catch most of them.

4 Conclusion

AI-based image analysis can help healthcare providers render a more accurate diagnosis. To better reflect real-case scenarios, we developed a new classification system covering a much broader set of skin lesions than the ISIC system and organized them into classes (10) and sub-classes (122). We trained our DermoscopyAI for multi-label classification, which is beneficial in case of CSLs, and used an adaptive loss function for hierarchical learning. At inference, our system crops the lesion based on segmentation prediction (to reduce background noise) and then classifies the resulting image. The hierarchical class/sub-class approach proved beneficial on public datasets (ISIC2018 and PH2) and blind tests over 4,894 images.

References

1. Riham Abdel Kader, Wassim El Hajj Chehade, and Ali Al-Zaart. Segmenting skin images for cancer detection. *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 392–396, 2018.
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.
3. Ibrahim Abdulrab Ahmed, Ebrahim Mohammed Senan, Hamzeh Salameh Ahmad Shatnawi, Ziad Mohammad Alkhraisha, and Mamoun Mohammad Ali Al-Azzam. Multi-models of analyzing dermoscopy images for early detection of multi-class skin lesions based on fused features. *Processes*, 11(3):910, 2023.
4. Mohammed Al-Masni, Dong-Hyun Kim, and Tae-Seong Kim. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Computer methods and programs in biomedicine*, 190:105351, 2020.
5. Catarina Barata, M. Emre Celebi, and Jorge S. Marques. Explainable skin lesion diagnosis using taxonomies. *Pattern Recognition*, 2021.
6. Catarina Barata, Jorge S. Marques, and M. Emre Celebi. Deep attention model for the hierarchical diagnosis of skin lesions. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2757–2765, 2019.
7. Andreas Blum, Graeme Siggs, and Ashfaq et. al Marghoob. Collision skin lesions—results of a multicenter study of the international dermoscopy society (ids). *Dermatology Practical Conceptual*, 7:51–62, 07 2017.
8. 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.
9. Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, 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.
10. Fikrle et al. Clinical-dermoscopic-histopathological correlations in collision skin tumours. *Indian J Dermatol.*, 2021.
11. Gupta et al. Skin cancer concerns in people of color: Risk factors and prevention. *Asian Pac J Cancer Prev.*, pages 5257–5264, 2016.
12. Saba et al. Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction. *J Med Syst*, 2019.
13. Shah et al. A comprehensive study on skin cancer detection using artificial neural network (ann) and convolutional neural network (cnn). *Clinical eHealth*, pages 76 – 84, 2023.
14. Yee et al. The role of artificial intelligence and convolutional neural networks in the management of melanoma: a clinical, pathological, and radiological perspective. *Melanoma Research*, pages 96 – 104, 2024.
15. Yu et al. Hierarchical skin lesion image classification with prototypical decision tree. *npj Digital Medicine*, 2025.
16. Nikolaos Gkalelis and Vasileios Mezaris. Subclass deep neural networks: Reenabling neglected classes in deep network training for multimedia classification. In *MultiMedia Modeling*, pages 227–238, Cham, 2020.

17. David Gutman, Noel C. F. 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). *eprint arXiv:1605.01397*, 2016.
18. Qishen Ha, Bo Liu, and Fuxu Liu. Identifying melanoma images using efficientnet ensemble: Winning solution to the SIIM-ISIC melanoma classification challenge. *CoRR*, abs/2010.05351, 2020.
19. Mohamed A. Kassem, Khalid M. Hosny, Robertas Damaševičius, and Mohamed Meselhy Eltoukhy. Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review. *Diagnostics*, 11(8), 2021.
20. M Latha, G Manjula, YM Raghavendra, HC Rashmi, et al. Enhancing skin cancer classification on the ph2 dataset through transfer learning technique. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2(03):500–507, 2024.
21. Sarmad Maqsood and Robertas Damaševičius. Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare. *Neural networks*, 160:238–258, 2023.
22. Teresa Mendonça, Pedro M Ferreira, Jorge S Marques, André RS Marcal, and Jorge Rozeira. Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5437–5440. IEEE, 2013.
23. Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M. Emre Celebi, and Ghassan Hamarneh. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, 88:102863, 06 2023.
24. Codella N, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza S, Liopyris K Kalloo A, Mishra N, Kittler H, and Halpern A. 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). *eprint arXiv: 1710.05006*, 2017.
25. Ir Purwono, Alfian Ma'arif, Wahyu Rahmaniar, Haris Imam, Haris Imam Karim Fathurrahman, Aufaclav Frisky, and Qazi Mazhar Ul Haq. Understanding of convolutional neural network (cnn): A review. *International Journal of Robotics and Control Systems*, 2:739–748, 01 2023.
26. Dasari Anantha Reddy, Swarup Roy, Sanjay Kumar, and Rakesh Tripathi. A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification. *Procedia Computer Science*, 2023.
27. Carucci JA. Rigel DS. Malignant melanoma: prevention, early detection, and treatment in the 21st century. *CA Cancer J Clin.*, pages 215–240, 2000.
28. Jonatas Silva, Atécio Alves, Paulo Santos, and Luiz Matioli. A new svm solver applied to skin lesion classification. *Statistics, Optimization & Information Computing*, 12(4):1149–1172, 2024.
29. Jared Dunnmon Geoffrey Angus Albert Gu Sohoni, Nimit and Christopher Ré. No subclass left behind: Fine-grained robustness in coarse-grained classification problems. *Advances in Neural Information Processing Systems*, pages 19339–19352, 2020.
30. Siegel RL et al. Sung H, Ferlay J. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.*, 71(3):209–249, 2021.