TIPTEKNO'24 MEDICAL TECHNOLOGIES CONGRESS
10-12 October 2024
Yasmin Bodrum Resort, Muğla, Türkiye

A Comparative Study on Skin Cancer Detection: Multi-class vs. Binary

Sudenaz Basut, Yagmur Kurtbas, Nilay Guler, Erdem Okur, Mehmet Turkan
Izmir University of Economics
Izmir, Turkiye

***Abstract*—Skin cancer, particularly melanoma, is a major public health concern due to its high fatality rate. Early diagnosis is crucial for improving patient outcomes, and advances in computer-aided diagnostic systems based on deep learning have showed promise in increasing diagnostic accuracy. This study examines two methods for handling the multi-class classification issue in skin cancer diagnosis. The first strategy utilizes a single EfficientNet-b0 model to classify all classes at once, whereas the second approach, that can be thought as "waterfall" method, employs a sequence of binary classifiers, each designed to detect one specific class at a time. Both techniques were evaluated on the ISIC 2018 dataset, and the findings show that the waterfall like strategy improves classification accuracy by around 8%. This study illustrates the potential benefits of sequential binary classification in dealing with complicated multi-class problems in medical image analysis especially for skin cancer; nevertheless, more research with other metrics is required to corroborate these findings and explore different network models.**

***Keywords*—Skin cancer, melanoma, multi-class, convolutional neural networks, EfficientNet**

I. Introduction

Skin cancer, particularly melanoma, is a major and rising public health concern worldwide. Melanoma, which accounts for a lesser fraction of skin cancer diagnoses, is responsible for nearly 75% of skin cancer-related fatalities. In the last decade, the number of new aggressive melanoma cases detected yearly has grown by 32%, emphasizing the need for early identification and care. The number of fatalities from melanoma is anticipated to increase by 3.8% in 2024. Thus, early identification is critical, since the five-year survival rate for melanoma is roughly 99% when identified early, but it reduces rapidly as the illness develops. Further facts and statistics that signifies the importance of melanoma can be found in [1].

Physical examinations and biopsies are commonly used to diagnose melanoma, but they are invasive, time-consuming, and often inaccurate. This has increased the relevance of non-invasive diagnostic procedures such as dermoscopic imaging, which allows for precise renderings of skin layers that are unseen to the human eye. Dermoscopy, despite improving diagnostic sensitivity, still depends primarily on dermatologists' knowledge, making it vulnerable to possible human error.

In recent years, the integration of computer-aided diagnostic (CAD) systems, particularly those that use artificial intelligence (AI) and machine learning (ML), has transformed melanoma detection [2], [3]. Deep learning algorithms and convolutional neural networks (CNNs) have demonstrated great promise for accurately evaluating dermoscopic images, decreasing diagnostic mistakes, and even outperforming experienced dermatologists in certain circumstances. These algorithms can analyze massive volumes of image data, detect minute trends, and deliver quick, reliable diagnoses, making them critical weapons in the battle against skin cancer.

The development and implementation of AI-powered diagnostic techniques provide a non-invasive, efficient, and accessible option for early melanoma identification, which is critical for improving patient outcomes and lowering mortality. A part of this requires differentiating a melanoma case from other possible skin cancer types or some cases where a benign lesion seems like a melanoma case. To emphasize on this differentiation, this paper compares a straightforward approach based on an EfficientNet-b0 model and another approach where more than one network models run in an order to eliminate different types of lesions successively to detect melanoma. Hence, changing the problem from multi-class to binary classification steps.

The remaining of this paper is organized as follows. Section II describes related works in the literature. Section III details the proposed methods in this study. Section IV reports experimental results and provides statistical comparisons. Finally, Sec. V concludes the paper with a brief conclusion.

II. Background

In this section, only the studies which utilize the same dataset as in this paper are mentioned. The choice is made because of the uniqueness of the dataset and also to make the proposed approach comparable in terms of systematic design.

A. The ISIC 2018 Dataset

The ISIC 2018 Dataset is created by the International Skin Imaging Collaboration (ISIC) for the 2018 challenge [4], [5], [6]. The dataset consists of $10,015$ training images and one ground truth response CSV file which has one header row and $10,015$ corresponding response rows. It also has around 100 validation and $1,512$ test images. However, the uniqueness of this dataset comes from the fact that it has more than one disease classes. It includes not just Melanoma, but also Melanocytic nevus, Basal cell carcinoma, Actinic

979-8-3315-2981-9/24/$31.00 ©2024 IEEE

keratosis, Benign keratosis, Dermatofibroma, Vascular lesion and benign nevus. With total of eight classes to separate, the dataset presents a unique problem for melanoma detection. Of course, having various classes with different number of images in the dataset, creates multiple class imbalance and rate of representation problem. In the next subsection, three recent studies are mentioned which utilize this dataset.

B. Literature Review

Gessert *et. al.* [7] identify many major problems, including the large class imbalance in the dataset, which they address using strategies such as loss weighting and balanced batch sampling. They also stress the use of unscaled multi-crop assessment, which comprises training and testing using smaller, unscaled image crops to capture high-resolution characteristics and global context at the same time. They created an ensemble network model, which includes architectures like DenseNet [8], SENet [9], and ResNeXt [10]. Then, they have fine-tuned this model using pretrained models from the ImageNet dataset [11]. Another study by Zhuang *et. al.* [12] uses a collection of sophisticated deep neural network topologies, especially the SENet and the Progressive Neural Architecture Search Network (PNASNet-5-Large). These models were chosen for their superior performance in feature learning, especially in scenarios with little training data. They use a variety of preprocessing approaches, including color constancy algorithms and data augmentation, to improve the model's resilience and accuracy. The dataset's significant imbalance, in which certain categories contained disproportionately more samples than others, is reduced by experimenting with different strategies, such as class-weighted loss functions and focused loss functions, to penalize mistakes in underrepresented groups more severely. They also acknowledged using other strategies, such as clustering and typical machine learning techniques, but the ensemble proved to be the most successful. As a last example, Amro *et. al.* [13] propose two algorithms, similar to ours, for disease classification: a one-step classifier for all seven lesion types and a two-step classifier that first distinguishes between Melanocytic nevus (the dominating class) and other lesions, then classifies the remaining six kinds. The dataset's imbalance was resolved by upsampling the underrepresented classes. The final model used an ensemble of Xception [14] and DenseNet121 networks.

As it can be seen from above mentioned studies which work on the same dataset, all approaches differ greatly. Since the problem in this case is not just a binary classification and the nature of the dataset creates great class imbalance, unusual approaches are possibly required. The next section explicitly provides the details of the two proposed approaches in this study and their architectural comparison.

III. Proposed Methods

In this study, two different approaches are proposed and compared for the multi-class classification problem. The first approach involves using a single network model to classify all classes at once, meaning each test image is classified into one of the eight classes described previously. The second approach is akin to a "waterfall" method, where seven different optimized versions of the same network model are sequentially used to classify each class one-by-one. The very first submodel is specifically trained to classify a single class, labeling all other classes as "other". These "other" labeled images are then passed to the next network model, which attempts to classify them as a second class or as "other" again, continuing this process until no test images remain or only those that cannot be classified are left. This method essentially treats each step as a binary classification problem: distinguishing between a specific disease and all others. It is briefly illustrated in Fig. 1. The final statistical comparison of these two approaches aims to determine whether treating this multi-class problem as a series of binary classifications yields better results.

The choice of network model is crucial for this task. EfficientNet, known for its compound scaling methodology, has significantly influenced the understanding of balancing efficiency and accuracy in deep learning. By scaling width, depth, and resolution in a balanced manner, it provides versatile models that can be adapted to various hardware constraints. Moreover, its lightweight and robust architecture consistently delivers high performance across a range of computer vision tasks. The model also offers several variations with predefined scaling options, such as EfficientNet-b0. Given these advantages, this basic variant "EfficientNet-b0" is selected for use in the proposed approaches. To ensure a coherent and fair comparison, this network is employed in both approaches, with specific optimizations applied to each class in the waterfall approach.

A. Preprocessing

The significant class imbalance in the ISIC 2018 dataset poses a great challenge. To address this problem, the images are augmented to create a balanced training set with $5,000$ images per class. This number is chosen due to the nature of the dataset, where the majority of cases are naturally benign nevi, containing nearly $5,000$ images. To achieve this balance, techniques such as rotation, translation, and reflection are applied, particularly to underrepresented classes, to increase their sample sizes to match the target number.

Another important issue is the variation in the aspect ratios of the images. Since the images were captured in different facilities using various devices, the aspect ratios vary widely across the dataset. This variation poses a problem because the images are needed to be resized to 224×224 pixels to be used as inputs for EfficientNet-b0. To preserve information and avoid distortion, each image is zero padded to form a square image, ensuring that if an image had dimensions of $m \times n$, it is padded to become $t \times t$, where $t > m, n$. This padding ensures that no information was lost when resizing all images to 224×224 pixels for the network.

The dataset was already divided into training, validation, and test sets, so no further preprocessing is required before training the network models.

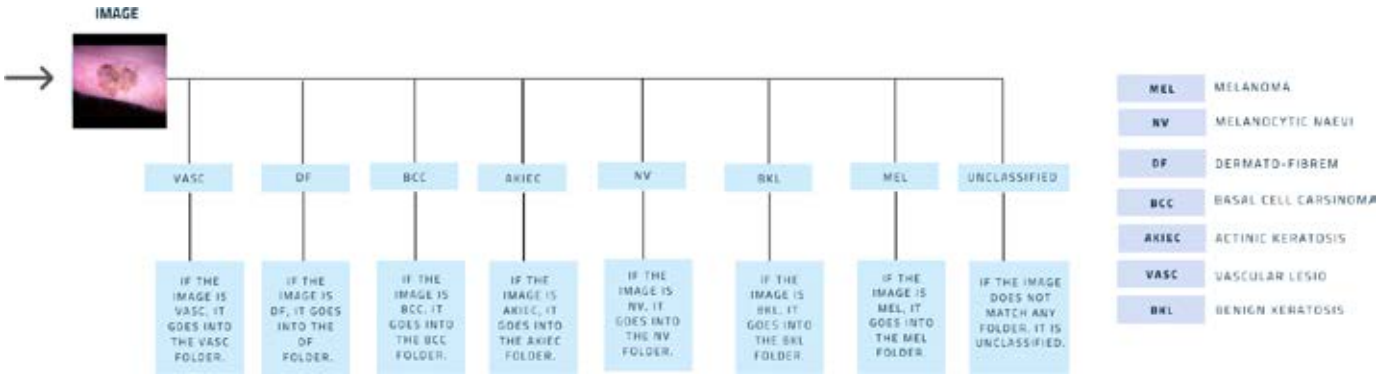


Fig. 1: Waterfall-like approach. Each class is classified with a variant of EfficientNet-b0.

B. Training

The training process is an essential aspect to discuss. Initially, a standard EfficientNet-b0 model is trained, which is readily available in Matlab environment. This represents the first approach, serving as a multi-class classifier for the dataset, and reflects the typical methodology employed by other researchers. The network is trained from scratch on the training set using the correct labels. The network parameters are optimized through experimentation.

For the second approach, the waterfall method, seven EfficientNet-b0 models are trained from scratch on seven distinct sub-datasets. Each subset uses the same augmented training dataset but with different labels. For example, in the melanoma class, the dataset is labeled with only two categories: melanoma and others. The melanoma-specific network (i.e., one-to-all) is trained on this binary-labeled dataset, and the same process is applied to all remaining classes. This approach effectively transforms the problem into a binary classification task at each step (see Fig. 1).

IV. Results and Comparison

After all training sessions, both approaches are tested on $1,512$ test images in the dataset. The first approach employs the multi-class classifier via EfficientNet-b0 model. This test is followed by the second approach, the waterfall method, where seven different EfficientNet-b0 models work sequentially to classify an input image. If an image can not be classified, it is considered as a benign lesion.

In the waterfall approach, the first model aims to classify the "VASC" (Vascular Lesion) class among all classes. The second model focuses on identifying the "DF" (Dermatofibroma) class among the remaining six classes. The third model targets the classification of the "BCC" (Basal Cell Carcinoma) class from the remaining five classes. The fourth model aims to classify the "AKIEC" (Actinic Keratosis) class from the remaining four classes. The fifth model classifies the "NV" (Melanocytic Nevus) class from the remaining three classes. The sixth model classifies the "BKL" (Benign Keratosis) class from the remaining two classes, and the seventh model distinguishes the remaining "MEL" (Melanoma) class. Images that could not be classified through this cascaded process are considered as "OTHER" (Benign, unclassified). The results, including the validation and test accuracies of both approaches, are presented in Table I. To ensure a fair comparison, the second approach is shown twice in the table, including its average accuracy.

TABLE I: Statistical comparison of the proposed approaches. ACC indicates "Accuracy".

Approach	Classifier	Validation ACC	Test ACC
First Approach	Multi-Class Single EfficientNet-b0	88.86%	73.94%
Second Approach (average over all)	Binary Seven Different EfficientNet-b0	-	**81.51%**
Second Approach (individual results)	MEL - EfficientNet-b0	96.25%	77.65%
	NV - EfficientNet-b0	99.60%	39.88%
	VASC - EfficientNet-b0	76.28%	97.62%
	DF - EfficientNet-b0	96.61%	96.89%
	AKIEC - EfficientNet-b0	86.52%	84.39%
	BKL - EfficientNet-b0	70.59%	80.16%
	BCC - EfficientNet-b0	92.36%	93.98%

The table clearly demonstrates an improvement in the second approach, with an almost 8% increase in accuracy. This improvement is anticipated due to the inherent characteristics of skin lesions. Typically, a lesion comprises melanocytes, the cells responsible for producing melanin, the pigment that gives skin its color. Many skin cancers and other related conditions originate from these cells. Each class in the ISIC 2018 dataset technically appears in specific regions of a lesion but not throughout the entire lesion, with the remaining areas being considered healthy. The presence of shared features among these classes can confuse a multi-class classifier, leading to misclassification between the lesion classes and healthy sections. This overlap of features likely hinders the detection capabilities of the first approach, as the experimental results suggest.

Further examination of the individual accuracy results in the second approach provides valuable insights into skin lesions. Notably, the "NV" class stands out, with a 39.88% accuracy, representing the lowest performance in the second approach. A plausible explanation for this lies in the nature of this class, which represents Melanocytic Nevus, commonly known as a mole. While moles are generally benign, they have the potential to develop into melanoma. The issue arises because many lesions in the dataset are moles. A mole with malignant

features includes both malignant and healthy characteristics, whereas benign moles exhibit only healthy features. This mix of features can lead to incorrect classification, highlighting the need for further refinement in the detection system to effectively distinguish moles from malignant skin lesions.

As a final note, while this study suggests that a sequentially designed cascaded binary classification system for a multi-class skin lesion problem performs better than a standard multi-class classifier model, the results are not yet sufficient to confirm and generalize this outcome. Future work should include additional evaluation metrics beyond accuracy to definitively support these findings. Moreover, exploring state-of-the-art network models may help identify other suitable models that could perform better under these conditions.

V. Conclusion

This study demonstrates that when applied to the task of detecting skin lesion problems, a sequential binary classification strategy acts as a waterfall, that classifies one type at a time, outperforms a typical multi-class classification model. Using seven optimized EfficientNet-b0 models trained on different classes, this waterfall-like technique improves accuracy by 8% on the ISIC 2018 dataset compared to the conventional multi-class classification strategy. These findings indicate that addressing a multi-class problem as a sequence of binary classifications helps alleviate some of the difficulties associated with overlapping features in skin lesion images, especially when dealing with unbalanced datasets. However, these findings are not yet definitive, and follow up research should include more assessment measures and investigate various network designs to confirm and generalize on these positive outcomes.

References

[1] "Skin Cancer". Available online: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/ (accessed on 11 August 2024).
[2] Okur, E. and Turkan, M., "A survey on automated melanoma detection", Eng. App. Artif. Intell., 73, 2018, 50-67.
[3] Adegun, A. and Viriri, S., "Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art". Artif. Intell. Rev. 54, 2, 2021, 811–841.
[4] "ISIC 2018". Available online: https://challenge.isic-archive.com/landing/2018/ (accessed on 11 August 2024).
[5] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A., "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018. https://arxiv.org/abs/1902.03368
[6] Tschandl, P., Rosendahl, C., Kittler, H., "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions". Sci. Data 5, 2018, 180161.
[7] Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., Schlaefer, A., "Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting", ISIC 2018 Challenge: Skin Lesion Analysis Towards Melanoma Detection, 2018.
[8] Huang, G., Liu, Z., Weinberger, K.Q., "Densely connected convolutional networks". IEEE Conf. Comp. Vis. Pat. Rec., 2017, 2261-2269.
[9] Cheng, D., Meng, G., Cheng, G., Pan, C., "SeNet: Structured edge network for sea–land segmentation". IEEE Geoscience Remote Sensing Lett., 14, 2017, 247-251.
[10] Xie, S., Girshick, R.B., Dollar, P., Tu, Z., He, K., "Aggregated residual transformations for deep neural networks". IEEE Conf. Comp. Vis. Pat. Rec., 2017, 5987-5995.
[11] Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L., "ImageNet: A large-scale hierarchical image database". IEEE Conf. Comp. Vis. Pat. Rec., 2009, 248-255.
[12] Zhuang, J., Li, W., Manivannan, S., Wang, R., Zhang, J., Liu, J., Pan, J., Jiang, G., Yin, Z. "Skin lesion analysis towards melanoma detection using deep neural network ensemble". ISIC 2018 Challenge: Skin Lesion Analysis Towards Melanoma Detection, 2018.
[13] Amro, M. K., Singh, B., Rizvi, A. "Skin lesion classification and segmentation for imbalanced classes using deep learning". ISIC 2018 Challenge: Skin Lesion Analysis Towards Melanoma Detection, 2018.
[14] Chollet, F. "Xception: Deep learning with depthwise separable convolutions". IEEE Conf. Comp. Vis. Pat. Rec., 2017, 1800-1807.