JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 15, No. 3, September 2021, pp. 115-124

Assessment of Segmentation Impact on Melanoma Classification Using Convolutional Neural Networks

Qikang Deng, Jose Cruz Castelo Beltran, and DoHoon Lee*

Department of Information Convergence Engineering, Pusan National University, Busan, Korea

dengqikang@pusan.ac.kr, jose.castelo@pusan.ac.kr, dohoon@pusan.ac.kr

Abstract

Among the different types of skin cancer, melanoma is the one with the highest death rate. Therefore, the early detection of melanoma and the development of technologies that can assist in this task have become significantly important. Convolutional neural networks are one of the most popular skin cancer classification methods. However, most of the available skin cancer datasets include images with lesions that are hard to differentiate from healthy skin or with a high presence of hair that can occlude the lesion. This characteristics of the images makes it harder to extract lesion features. Therefore, utilizing segmentation to extract the lesion location is an important step to reduce hair noise and improve lesion analysis. In this paper, two combining methods for segmentation and classification were explored: concatenation and multiplication. By utilizing these methods, it was possible to improve the accuracy of different neural network architectures by around 1% when compared to unmodified models without segmentation. The best-performing model was selected for further training. This model in conjunction with the segmentation module allowed for the correct re-classification of around 10% of the total examples in the dataset, indicating that a segmentation phase leads to an overall accuracy improvement and suggested that by improving the segmentation, an improvement on the overall accuracy can be obtained.

Category: Computer Graphics / Image Processing

Keywords: Skin cancer; Melanoma classification; Medical image segmentation

I. INTRODUCTION

According to the American Cancer Society (ACS), skin cancer has become an increasing health problem worldwide [1] and is the most commonly diagnosed cancer in the United States. It is important to note that skin cancer is usually divided into three different groups. Basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma. BCC and SCC are usually referred to as non-melanoma cancer; however, they are more commonly diagnosed compared to melanoma, which also presents a high death rate. The ACS estimated that by 2020, around 100,350 cases would be diagnosed, and 6,850 diseases will occur due to melanoma.

Although melanoma cases are more prevalent in the Caucasian population, evidence indicates that the detection of melanoma in Asian countries occurs at a more advanced stage and with a worse prognosis than their Caucasian counterparts [2]. However, a study conducted by Gyeongsang National University Hospital from 2008 to 2018 [3] has estimated that the number of patients with skin cancer has been increasing with time. Moreover, if melanoma is detected at an early stage, it is highly curable; otherwise, it is more likely than other types of

Open Access http://dx.doi.org/10.5626/JCSE.2021.15.3.115 http://jcse.kiise.org

Received 15 June 2021; **Accepted** 28 August 2021

*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

skin cancer to spread to other parts of the body. Therefore, early detection of skin cancer, specifically melanoma, is extremely important in order to reduce the probability of disease. To accomplish this, the existence of a computer system that can assist in the tracking and detection of skin lesions on patients would be extremely valuable.

Previous researches have widely utilized support vector machine (SVM) and convolutional neural network (CNN) architectures for skin cancer classification. Additionally, CNN architectures like InceptionV3 have reached dermatologist-level classification of skin cancer [4]. However, most skin cancer lesion datasets present images with a high presence of hair. This can affect the ability of CNNs to extract features for classification, making this task harder to solve with a simple CNN. An additional technique that can be utilized is an inpainting algorithm [5] which can remove the hair in lesion images. However, inpainting algorithms will ineluctably remove some useful information for feature extraction and classification. For this reason, other researchers have used a simple combination of segmentation and classification steps to firstly locate and crop the lesion, then use the CNN model to classify the segmented region. By combining segmentation with classification, it is possible to improve the recognition ability of the classifier.

According to the "ABCDEs" rules defined in the work of Tsao et al. [6], "B" which stands for border, the irregularities from a lesion contain valuable information that can impact melanoma classification, and a simple cropping segmentation will lead to a loss of information on ragged, notched, or blurred edges of the lesion. In this paper, two new combinations of segmentation and classification are proposed instead of a simple segmentation and classification method to classify melanoma: concatenation and multiplication. In these proposed methods, the lesion is found and located using MultiResUnet [7]. This will calculate the probability of each pixel in the image to belong to the region of interest.

In the proposed multiplication approach, a probabilistic mask is multiplied by each of the red, green, blue (RGB) channels in the image. Then, a new RGB image with weighted pixel values is created. Finally, the CNN model can classify this example using the generated weighted RGB image. On the other hand, the concatenation approach obtains a probabilistic mask and concatenate it with the RGB image, generating a four-channel image. Utilizing this four-channel image, the CNN model can classify melanoma with extra information from the probabilistic mask channel. In this way, the two proposed combination methods weigh the importance of each pixel and generate a segmentation mask. This allows retaining all the features of the image while focusing on the located region of interest in the lesion.

For this reason, several deep learning neural networks were investigated in this study with the goal of comparing their performance in the melanoma classification task. With these selected architectures, a series of experiments were performed by implementing image augmentation and lesion segmentation techniques to measure and compare the increase in performance. Based on the results of the initial experiments, the best performing model was selected to receive additional training in order to maximize the performance and obtain the final result of this project.

Although the implementation of the proposed model as a computer or mobile application was desirable, it is outside the scope of the study at this stage and is instead considered as a future research opportunity.

II. RELATED WORK

A. Skin Cancer Classification

Over time, the study of skin cancer segmentation and classification has gone through different stages where different techniques and methods were used to address these tasks. The methods utilized during these different time stages can be generally classified as the following: computer vision algorithms, machine learning and deep learning stages. The study of skin cancer classification started with the utilization of pure computer vision mathematical methods such as Black-hat Morphology and Total Variation [8] and simple linear iterative clustering (SLIC) Superpixel [9]. These were some of the first methods utilized in the attempt of segmenting and classifying skin lesion images. Subsequently, with the surge of machine learning methods, researchers started exploring algorithms such as k-means, k-nearest neighbors (KNN), decision tree and SVM with improved performance compared to the computer vision methods utilized before. In the work of Lynn and Kyu [10], a comparison of the mentioned machine learning classifiers was performed having SVM as the best performing method with a 78.2% accuracy. SVM was selected as the classification method to employ for a consumer application for skin cancer assessment due to its promising results [11].

These days the state-of-the-art deep learning architectures are the most common and successful methods utilized for the task. Among the different existing neural network architectures, ResNet and InceptionV3 are the best performing models with a classification and validation accuracy of nearly 0.8900 (ResNet 101) and 0.9000 (InceptionV3) [12], respectively, in skin lesion images.

B. Skin Lesion Datasets

Unfortunately, access to skin lesion datasets has been a limitation for the research and development of automated methods for skin cancer classification. At the beginning of the last decade, publicly available datasets were technically inexistent. Therefore, every group interested in

researching this topic had to collect skin lesion datasets by their own means. However, publication of the Hospital Pedro Hispano (PH^2) database [13], a collection of 200 images annotated and segmented by professional dermatologists, started to change circumstances. Although the release of the PH^2 dataset was a significant step forward in the advancement of research and development of automated classification systems for skin cancer, the size of the set was still small for the needs of some classification methods. With the aim of progressing further in this study field, the International Skin Imaging Collaboration (ISIC) has been collecting dermoscopic images from various clinical centers across the world for the past six years. To this date the ISIC archive contains over 13,000 images making it the largest skin lesion dataset publicly available.

III. METHODOLOGY

A. Data Construction

In this paper, three different datasets were sampled and combined to obtain a more balanced dataset (5,106 melanoma and 6,343 nevus) that was utilized for the training and testing of the presented models. The original datasets that were utilized are as follows:

- 2020 ISIC challenge dataset [14]: 584 melanoma images and 32,542 non-melanoma (or nevus) images from >2,000 patients. Each image was associated with one of these individuals using a unique patient identifier.
- 2019 ISIC challenge dataset [15–17]: 25,331 images were classified into nine types of skin cancer, including 4,522 images of melanoma.
- 2018 ISIC challenge [15, 16] (Task 1. Lesion segmentation dataset): 2,596 lesion images and 2,596 corresponding mask images.

1) Classification Dataset

Since the 2020 ISIC dataset presented a small number of melanoma examples when compared to nevus, the 4,522 melanoma examples from the 2019 ISIC dataset were merged resulting in a total of 5,106 melanoma examples. A selection of nevus examples from patients who also had melanoma was made in order to reduce the class imbalance that still exists. As a consequence of systematically merging the ISIC 2019 and 2020 datasets, a more balanced dataset with 5,106 melanoma images and 6,343 nevus images was obtained.

2) Segmentation Dataset

As part of the proposed method, the segmentation of the lesion plays an important role in the performance of the model. In order to implement and validate the accuracy of the segmentation section of the model, it is essential to utilize a professionally segmented set of medical images. The 2018 ISIC challenge [15, 16] provided the public with a collection of skin lesion images as well as the ground truth images reviewed and curated by practicing dermatologists with expertise in dermoscopy. Additionally, with 2,594 dermoscopic images as well as 12,970 corresponding ground truth masks (five for each image), the ISIC 2018 dataset was, to our best knowledge, the most extensive dataset available to perform skin lesion segmentation tasks. For the aforementioned reasons, ISIC 2018 dataset was selected to develop the segmentation module of the proposed model.

B. Data Preprocessing

Data preprocessing consists of a set of actions performed to the input data before they are used for training a model. The objective of this step was to diversify the training data to expose the model to a wider range of examples to improve the performance. Examples of image augmentation methods that were utilized for this project are image flip, rotation, and zoom. These kinds of augmentation are usually useful for any kind of task performed with neural networks. Additionally, it would be possible to implement some task-specific augmentation that can help to reduce noise or to overcome partial occlusion on images. Finally, the images were resized to 224×224 pixels in order to speed up the training process.

A good rule of thumb is that the inputs should always be small values. Normalization refers to rescaling the input data values from the original range to a smaller one that ranges from zero to one. For the case of colored images similar to the ones from the utilized datasets, the values ranged from 0 to 255. Therefore, to achieve normalization, the value of each pixel in every one of the channels (red, green, blue) was divided by 255.

C. Segmentation

In the medical imaging area, one of the crucial factors lies in the ability to focus the attention on the area of interest. Depending on the type of images, the area of interest can represent organs, tumors, or lesions that often constitute the most critical information to perform a diagnosis. Therefore, different efforts have been made by researchers to produce algorithms capable of identifying and segmenting the area of interest from a picture. Although most of the first segmentation methods relied on geometric [18] or fuzzy algorithms [19], neural networks have been a step forward given the high performance shown while working with images. Therefore, for the purpose of this project, the MultiResUNet [7] model was utilized. This model showed improved performance over its base mode U-Net, one of the most popular and successful models for image segmentation.

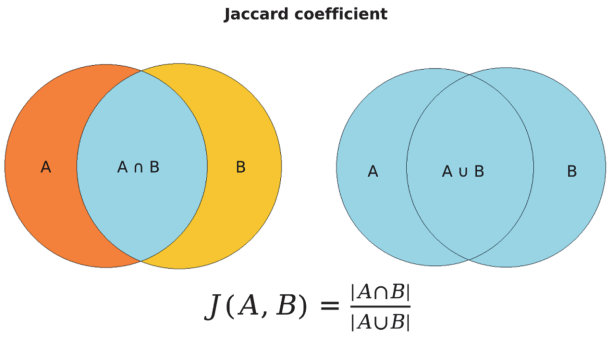


Fig. 1. Venn diagram illustrating Jaccard coefficient.

As a means to evaluate the performance of a segmentation model, the Jaccard index, given by (1), becomes a useful metric.

$$J(A,\ B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

The value of this metric expresses the similarity between two given sets and it is defined as the ratio of the intersection and the union of the given sets. On the imaging segmentation case illustrated in Fig. 1, set A represents the ground truth mask while set B represents the predicted segmentation mask. In this way, the Jaccard index provides an insight into the precision of the segmentation, over-segmentation and under-segmentation in both directions. For the development of the segmentation module, the MultiResUNet model was pre-trained utilizing the ISIC 2018 dataset which contains lesion images as well as the ground truth mask. The training process was done using Adam optimizer and binary cross-entropy. After training the model, the results showed a training accuracy of 0.949 and a validation Jaccard index of 0.796. The weights gained during training were retained for use in the final design when the segmentation and classification components were combined.

Since the ground truth mask was performed by a professional dermatologist, the values of the mask were either zero or one, indicating the pixel belonged to the lesion which was the area of interest for this task. On the other hand, the output of the MultiResUNet module consisted of a pixel-wise matrix ranging from one to zero where each value represents the probability for each pixel to be part of the skin lesion. Fig. 2 shows an example of the output obtained by the MultiResUNet model.

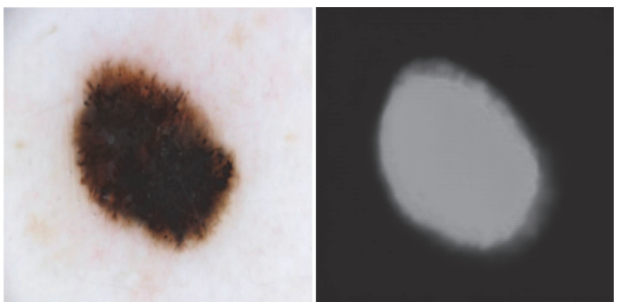


Fig. 2. Original image on the left and generated mask on the right.

D. Classification

The second task that needs to be accomplished by the system was the ability to predict whether a lesion was a melanoma or not. Neural networks have shown a high ability to work with images, specifically for the classification task. Over time, different neural architectures have been proposed by the research community. On some occasions it can be difficult to know which architecture is best for a certain task. Therefore, in order to provide an overview of the performance of some of these model architectures, some of the most popular architectures were implemented to make a comparison specifically in the melanoma classification task.

The model architectures that were considered in the project are as follows:

- LeNet [20]
- EfficientNet B0 [21]
- ResNet 50V2 [22]
- InceptionV3 [23]

Each of these models was trained on the constructed dataset for 30 epochs utilizing sparse categorical cross-entropy as loss function and Adam optimizer. These results would be used as the baseline when comparing the same models merged with the segmentation module. The performance of each model can be found in Table 1.

E. Segmentation and Classification Merge

In order to locate the lesion and keep all the information from the original image, the probabilistic mask was assumed as the important value for every pixel in the image. Classification models can utilize the segmentation mask to enhance the important areas and dismiss the unimportant part of the original image. To utilize the probabilistic mask in a classification model, two different approaches were adopted.

1) Mask concatenation: In this approach, the probabilistic mask obtained from the segmentation module was passed to the classification model as an input. In this way, the classification model received an input of four channels; three channels from the original image, and the other one channel being the probabilistic mask. The concatenated output was then fed to the classification model to learn the image features. Fig. 3 illustrates the architecture of the proposed model by utilizing the mask concatenation approach.

2) Mask multiplication: The second approach consisted

Table 1. Results of combining classification and segmentation

Classification	Approach	Training		Validation	
		Loss	Accuracy	Loss	Accuracy
LeNet	Unmodified	0.2344	0.9045	0.2187	0.9187
	Mask multiplication	0.1741	0.9338	0.2190	0.9174
	Mask concatenation	0.1858	0.9296	0.1948	0.9222
ResNet50V2	Unmodified	0.2498	0.8990	0.2397	0.9082
	Mask multiplication	0.2979	0.8780	0.2660	0.8912
	Mask concatenation	0.2050	0.9206	0.2096	0.9139
EfficientNetB0	Unmodified	0.2303	0.9097	0.2356	0.9069
	Mask multiplication	0.3250	0.8662	0.2877	0.8820
	Mask concatenation	0.2512	0.8997	0.2072	0.9178
InceptionV3	Unmodified	0.2894	0.8813	0.2529	0.8990
	Mask multiplication	0.2440	0.9082	0.2792	0.9060
	Mask concatenation	0.2313	0.9144	0.2437	0.9091

Mask multiplication and mask concatenation utilized a probabilistic mask obtained with MultiResUNet.

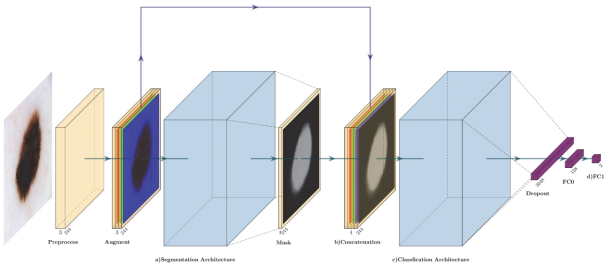


Fig. 3. Mask concatenation approach. (a) Segment the augmented image (3×244×244) to obtain a mask feature (1x244x244) with values ranging from 0 to 1. (b) Concatenate the mask feature (1×244×244) and augmented image (3×244×244) to obtain a 4×244×244 feature matrix. (c) Pass the obtained 4×244×244 feature matrix as the classification model's input. (d) Obtain the predicted class. The blue boxes represent the selected segmentation and classification models.

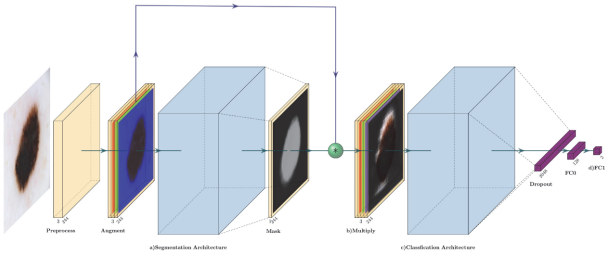


Fig. 4. Mask multiplication approach. (a) Segment the augmented image (3×244×244) to obtain a mask feature (1×244×244) with values ranging from 0 to 1. (b) Perform element-wise multiplication between each channel of augmented image (1×244×244) ×3 with the mask feature (1×244×244) to obtain a new 3×244×244 feature matrix. (c) Pass the obtained 3×244×244 feature matrix to the classification model's input. (d) Obtain the predicted class. The blue boxes represent the selected segmentation and classification models.

of multiplying each channel of the image by the probabilistic mask obtained from the segmentation module. This was intended to directly reduce the impact of pixels outside of the area of interest during the classification process. Since the multiplication output remained as a three channels image, this approach is suitable for general classification models even without the need for modification. Fig. 4 illustrates the architecture of the proposed model by utilizing the mask multiplication approach.

The merged models utilized the same hyperparameters indicated in the segmentation and classification sections. The following results provide an overview of how the different classification models performed when combined with the MultiResUNet segmentation model using concatenation and multiplication.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Results

In addition to testing each of the architectures with both segmentation merging approaches, each one of these models was also tested without any segmentation implemented. In this way, it was possible to observe the impact on the performance generated by segmenting the image. Table 1 presents the results of each of the combination experiments that were performed.

Although after being trained by 30 epochs, the

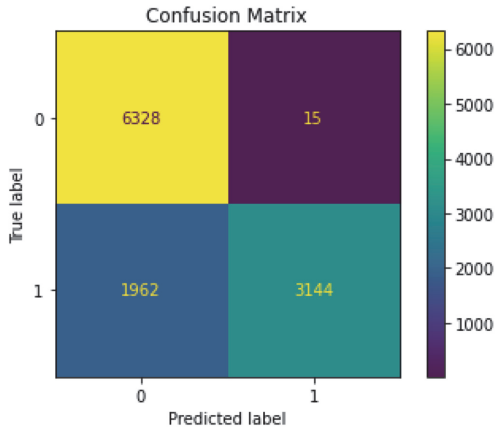


Fig. 5. Confusion matrix.

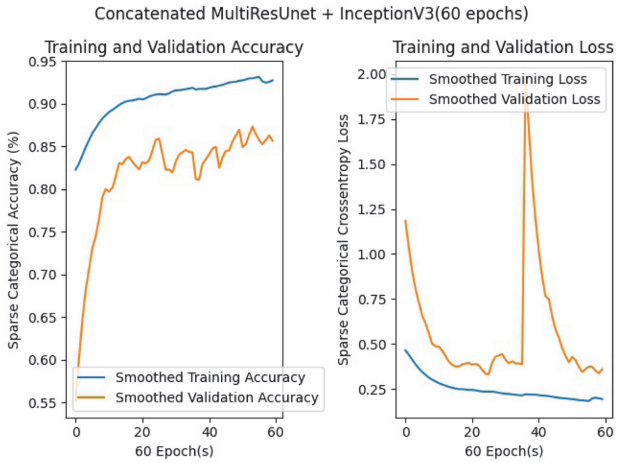


Fig. 7. Smoothed accuracy and loss of InceptionV3 with mask concatenation segmentation.

InceptionV3 was the model with the lowest validation accuracy among the best of each architecture, the other models showed signs of overfitting. Additionally, InceptionV3 showed a promising performance on the skin cancer task; however with the results obtained by the experiments conducted in this paper, it was observed that the overall performance of this model can yet be improved. As it can be seen in Fig. 6, the InceptionV3 model continued to improve after the initial 30 epochs. Therefore, it was trained for additional 30 epochs to demonstrate the improvement obtained by utilizing a segmentation module in conjunction with the InceptionV3 architecture.

After the 60 epochs of total training, the results are as follows: train loss 0.1816, train accuracy 0.9299, validation loss 0.2221, and validation accuracy 0.9113. Fig. 7 shows the validation accuracy and validation loss of the InceptionV3 model with the mask concatenation approach during the 60 epochs training. As it can be observed in the validation loss graph in Fig. 7, a sudden spike appeared in epoch number 36. This behavior was due to the Keras API method to calculate the loss, where it was divided by the mini-batch size. Given that the number of examples in a dataset was unlikely to be an integer multiple of the selected batch size (32 in this case), the last mini-batch contained fewer samples than the batch size, therefore dividing the loss by a lower number causing a spike in the graph.

To assess the benefit of using segmentation before classification we considered the incorrectly classified examples when performing the classification without previously segment the image. Among these misclassified examples, 1,181 were re-classified correctly after utilizing the segmentation module prior to classification, which

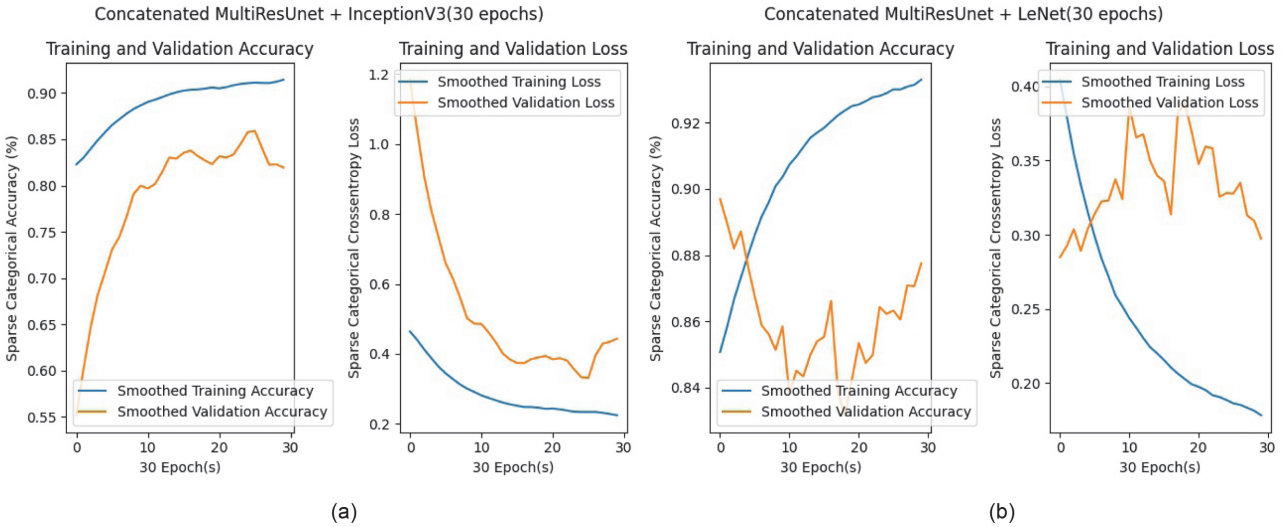


Fig. 6. InceptionV3 (a) and LeNet (b) smoothed accuracy and loss.

Table 2. Performance comparison for melanoma and non-melanoma classification task between the proposed method and related works utilizing different approaches to combine the whole image context with ground truth segmentation mask and automatically generated segmentation mask

Study	Method	AP	ACC	SENS	SPEC	BA	AUC
Codella et al. [24]	Whole Image (WI)	0.596	0.755	0.627	0.796	0.711	0.808
	Crop (CR)	0.618	0.81	0.72	0.832	0.776	0.819
	Crop GT (CRGT)	0.629	0.781	0.707	0.799	0.753	0.827
	Part 3B: VOTE (WI, CRGT)	0.588	0.834	0.533	0.9079	0.720	0.829
	Part 3: VOTE (WI, CR)	0.602	0.834	0.52	0.9112	0.7156	0.828
	Part 3B: AVG (WI, CRGT)	0.649	0.807	0.693	0.836	0.764	0.843
	Part 3: AVG (WI, CR)	0.645	0.805	0.693	0.832	0.762	0.838
Codella et al. [25]	Part 3B: Top Rank	0.624	0.855	0.547	0.931	0.739	0.783
	Part 3: Top Rank	0.637	0.855	0.507	0.941	0.724	0.804
Codella et al. [26]	Part 3B	0.591	0.836	0.253	0.98	0.616	0.815
	Part 3	0.589	0.77	0.72	0.723	0.721	0.815
	Part 3B: LL Ensemble	0.523	0.726	0.693	0.734	0.713	0.603
	Part 3: LL Ensemble	0.506	0.752	0.64	0.78	0.71	0.643
Current study	Proposed	0.784	0.911	0.615	0.997	0.806	0.806

Part 3 and Part 3B represent 2016 ISIC challenge part 3 and part 3B consisting in melanoma and non-melanoma classification without and with the ground truth segmentation mask, respectively.
Proposed method was evaluated on the dataset presented in Section III-A.
AP: average precision; ACC: accuracy; SENS: sensitivity; SPEC: specificity; BA: balanced accuracy; AUC: area under the receiver operating characteristic curve.

corresponded to 10.31% of the total examples in the utilized dataset. It is important to note that the ground truth of 1,173 of the 1,181 correctly re-classified examples belonged to the non-melanoma class while the remaining eight corresponded to melanoma. On the other hand, when implementing segmentation prior to classification, 416 correctly classified examples were incorrectly re-classified. Consequently, 185 belonged to the non-melanoma class while 231 corresponded to the melanoma class.

Based on the numbers mentioned above we could observe that the implementation of a segmentation method has the highest impact on the false positives, where the correct re-classification of 1,173 examples gave substantial help to improve the overall accuracy at the expense of decreasing the true positive rate. Nevertheless, as shown in Table 1, with the utilization of the segmentation module the overall accuracy was improved by around 1% on all of the implemented architectures.

Additional comparison between the proposed method and related works is presented in Table 2. When evaluating a model there is always a trade-off between sensitivity and specificity. Based on the proposed method specificity, we could observe a high confidence when discriminating a benign example while other methods with similar specificity tend to see their sensitivity reduced in a more drastic magnitude. In this way, the proposed method shows a better ability to handle the sensitivity-specificity trade-off obtaining the highest balanced accuracy among the compared methods.

B. Discussion

Based on the results in Table 1, it was possible to observe that for every model, the variant that achieved the highest performance was always the one utilizing for the mask concatenation approach. On the other hand, the mask multiplication approach showed a reduction in performance on most of the evaluated models, except for the InceptionV3 model where it performed better than the plain InceptionV3 but it was still behind the concatenation approach. The reduction in performance with the multiplication approach can be explained by the loss of information when multiplying a segmentation mask with low accuracy for certain examples of the dataset.

Even though good performance has been obtained without implementing skin lesion segmentation [4], it was possible to conclude that image segmentation can be used to improve classification performance. Moreover, if the segmentation model can be improved, better performance on the classification could be expected. When observing different images in the datasets, the most evident difference that exists among the examples in the set was the presence and absence of body hair between images.

This difference became even more relevant in cases where the hair can partially occlude the lesion. As follow-up research, it could be possible to measure the potential impact on the overall model while implementing some augmentations to alleviate the presence of hair. For this, two opposite approaches have been explored by the research community, that is hair removal [27] and hair adding [28] approaches. In the same way as it was done by the segmentation module, it would be possible to perform a comparison of different models and techniques in combination with these two techniques to observe the resulting performance of the different models.

V. CONCLUSION

In this paper, two different approaches for combining segmentation and classification (mask multiplication and concatenation) were utilized to improve the classification accuracy of different CNN architectures. For this, LeNet, ResNet, EfficientNet, and InceptionV3 were trained with the obtained dataset (5,106 melanoma images and 6,343 nevus images) from the systematic merge of the 2019 and 2020 ISIC datasets. The performance of the trained models in the melanoma classification task was compared against themselves when a segmentation module utilizing a mask concatenation or mask multiplication approach was applied. For the segmentation module, a MultiResUNet model trained on the ISIC 2018 dataset was utilized.

After training the models for 30 epochs and analyzing their performance, it was observed that the overall accuracy when applying a segmentation module was improved by around 1% on all the tested architectures. Additionally, the InceptionV3 with mask concatenation was selected for further training given that overfitting was not as severe as the other models. After an additional of 30 epochs for a total of 60 epochs training, this model allowed the reclassification of 10.31% of the total examples used in the database, resulting in the biggest impact on the false positives obtained by the model without segmentation.

With the development of the current project, a broad perspective on the performance of different neural network architectures for the melanoma classification task was provided. Most important, by comparing the performance of different combinations of architectures and segmentation approaches, it is expected to provide the research community with an overview and a starting point for other projects that intend to utilize some of the methods explored in this work. The development of medical-grade applications for diagnosis assistance has always been the goal of this research area. With the rise of skin cancer as a global health problem, and with melanoma being the deadliest type, it is hoped that the knowledge presented in this paper could contribute to the evolution of these technologies.

ACKNOWLEDGEMENTS

This work was supported by BK21 FOUR Program by Pusan National University Research Grant, 2020. The authors would like to thank all members of the VBLab at Pusan National University for their continuous support.

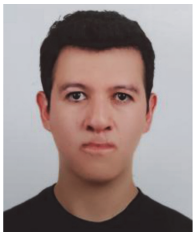
REFERENCES

1. American Cancer Society, "Cancer Fact & Figures 2020," 2020 [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html.
2. J. W. Chang, J. Guo, C. Y. Hung, S. Lu, S. J. Shin, R. Quek, et al., "Sunrise in melanoma management: time to focus on melanoma burden in Asia," *Asia-Pacific Journal of Clinical Oncology*, vol. 13, no. 6, pp. 423-427, 2017.
3. Y. J. Park, G. H. Kwon, J. O. Kim, N. K. Kim, W. S. Ryu, and K. S. Lee, "A retrospective study of changes in skin cancer characteristics over 11 years," *Archives of Craniofacial Surgery*, vol. 21, no. 2, pp. 87-91, 2020.
4. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.
5. J. Koehoorn, A. C. Sobiecki, D. Boda, A. Diaconeasa, S. Doshi, S. Paisey, A. Jalba, and A. Telea, "Automated digital hair removal by threshold decomposition and morphological analysis," in *Mathematical Morphology and Its Applications to Signal and Image Processing*. Cham, Switzerland: Springer, 2015, pp. 15-26.
6. H. Tsao, J. M. Olazagasti, K. M. Cordoro, J. D. Brewer, S. C. Taylor, J. S. Bordeaux, et al., "Early detection of melanoma: reviewing the ABCDEs," *Journal of the American Academy of Dermatology*, vol. 72, no. 4, pp. 717-723, 2015.
7. N. Ibtehaz and M. S. Rahman, "MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74-87, 2020.
8. A. H. Khan, D. N. F. Awang Iskandar, J. F. Al-Asad, and S. El-Nakla, "Classification of skin lesion with hair and artifacts removal using black-hat morphology and total variation," *International Journal of Computing and Digital Systems*, vol. 10, pp. 597-604, 2021.
9. D. Patino, J. Avendano, and J. W. Branch, "Automatic skin lesion segmentation on dermoscopic images by the means of superpixel merging," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2018, pp. 728-736.
10. N. C. Lynn and Z. M. Kyu, "Segmentation and classification of skin cancer melanoma from skin lesion images," in *Proceedings of 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, Taipei, Taiwan, 2017, pp. 117-122.
11. T. M. de Carvalho, E. Noels, M. Wakkee, A. Udrea, and T. Nijsten, "Development of smartphone apps for skin cancer risk assessment: progress and promise," *JMIR Dermatology*, vol. 2, no. 1, article no. e13376, 2019. https://doi.org/10.2196/13376

12. A. Demir, F. Yilmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: ResNet-101 and Inception-v3," in *Proceedings of 2019 Medical Technologies Congress (TIPTEKNO)*, Izmir, Turkey, 2019, pp. 1-4.
13. T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH^2: a dermoscopic image database for research and benchmarking," in *Proceedings of 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, 2013, pp. 5437-5440.
14. V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, et al., "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific Data*, vol. 8, article no. 34, 2021. https://doi.org/10.1038/s41597-021-00815-z
15. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, article no. 180161, 2018. https://doi.org/10.1038/sdata.2018.161
16. N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, 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 *Proceedings of 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, Washington, DC, 2018, pp. 168-172.
17. M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, et al., "BCN20000: dermoscopic lesions in the wild," 2019 [Online]. Available: https://arxiv.org/abs/1908.02288.
18. X. Li, B. Aldridge, L. Ballerini, R. Fisher, and J. Rees, "Depth data improves skin lesion segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Heidelberg, Germany: Springer, 2009, pp. 1100-1107.
19. R. J. Stanley, R. H. Moss, W. Van Stoecker, and C. Aggarwal, "A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images," *Computerized Medical Imaging and Graphics*, vol. 27, no. 5, pp. 387-396, 2003.
20. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
21. M. Tan and Q. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, 2019, pp. 6105-6114.
22. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 770-778.
23. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 2818-2826.
24. N. C. Codella, Q. B. Nguyen, S. Pankanti, D. A. Gutman, B. Helba, A. C. Halpern, and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, no. 4/5, article 5, 2017. https://doi.org/10.1147/JRD.2017.2708299
25. D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. 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)," 2016 [Online]. Available: https://arxiv.org/abs/1605.01397.
26. N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, 2015, pp. 118-126.
27. T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor: a software approach to hair removal from images," *Computers in Biology and Medicine*, vol. 27, no. 6, pp. 533-543, 1997.
28. S. Kitada and H. Iyatomi, "Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning," 2018 [Online]. Available: https://arxiv.org/abs/1809.02568.

Qikang Deng

Qikang Deng received his B.S. degree in the Department of Computer Science from South China Normal University, China, in 2017. In 2020, he joined the VB Laboratory in Department of Computer Engineering for pursuing his M.S. degree at Pusan National University. His research interests include autonomous vehicles, biomedical image analysis, and image generation.



José Cruz Castelo Beltrán

José Cruz Castelo Beltrán received a B.S. degree in Mechatronics Engineering at CETYS University in Mexico. In 2020 he started a master's degree in Computer Engineering at Pusan National University, Korea. His research interests include biomedical images processing, big data applications and blockchain technologies.

DoHoon Lee

DoHoon Lee is a professor in Pusan National University. He received the B.S., M.S., and Ph.D. degrees from Pusan National University, Korea. His research interests are in machine vision, data analysis, biometrics, and video prediction.