AI-BASED LOCALIZATION AND CLASSIFICATION OF SKIN DISEASE WITH ERYTHEMA

Although computer-aided diagnosis (CAD) is used to improve the quality of diagnosis in various medical felds such as mammography and colonography, it is not used in dermatology, wherenoninvasive screening tests are performed only with the naked eye, and avoidable inaccuracies may exist. This study shows that CAD may also be a viable option in dermatology by presenting a novel method to sequentially combine accurate segmentation and classifcation models. Given an image of the skin, we decompose the image to normalize and extract high-level features. Using a neural network-based segmentation model to create a segmented map of the image, we then cluster sections of abnormal skin and pass this information to a classifcation model. We classify each cluster into diferent common skin diseases using another neural network model. Our segmentation model achieves better performance compared to previous studies, and also achieves a near-perfect sensitivity score in unfavorable conditions. Our classifcation model is more accurate than a baseline model trained without segmentation, while also being able to classify multiple diseases within a single image. This improved performance may be sufcient to use CAD in the feld of dermatology.

Computer-aided diagnosis (CAD) is a computer-based system that is used in the medical imaging feld to aid healthcare workers in their diagnoses1. CAD has become a mainstream tool in several medical felds such as mammography and colonography1,2. However, in dermatology, although skin disease is a common disease, one in which early detection and classifcation is crucial for the successful treatment and recovery of patients, dermatolo_gists perform most noninvasive screening tests only with the naked eye. Tis may result in avoidable diagnostic inaccuracies as a result of human error, as the detection of the disease can be easily overlooked. Furthermore, classifcation of a disease is difcult due to the strong similarities between common skin disease symptoms. Terefore, it would be benefcial to exploit the strengths of CAD using artifcial intelligence techniques, in order to improve the accuracy of dermatology diagnosis. Tis paper shows that CAD may be a viable option in the feld of dermatology using state-of-the-art deep learning models.Te segmentation and classifcation of skin diseases has been gaining attention in the feld of artifcial intel_ligence because of its promising results. Two of the more prominent approaches for skin disease segmentation and classifcation are clustering algorithms and support vector machines (SVMs). Clustering algorithms generally have the advantage of being fexible, easy to implement, with the ability to generalize features that have a similar statistical variance. Trabelsi et al.3experimented with various clustering algorithms, such as fuzzy c-means, improved fuzzy c-means, and K-means, achieving approximately 83% true positive rates in segmenting a skin disease. Rajab et al.4 implemented an ISODATA clustering algorithm to find the optimal threshold for the segmentation of skin lesions. An inherent disadvantage of clustering a skin disease is its lack of robustness against noise. Clustering algorithms rely on the identifcation of a centroid that can generalize a cluster of data. Noisy data, or the presence of outliers, can signifcantly degrade the performance of these algorithms. Terefore, with noisy datasets, caused by images with diferent types of lighting, non-clustering algorithms may be preferred; however, Keke et al.5implemented an

improved version of the fuzzy clustering algorithm using the RGB, HSV, and LAB color spaces to create a model that is more robust to noisy data. SVMs have gained attention for their efectiveness in high-dimensional data and their capability to decipher "…subtle patterns in noisy and complex datasets"6. Lu et al.7 segmented erythema in the skin using the radial basis kernel function that allows SVMs to separate nonlinear hyperplanes. Sumithra et al.8 combined a linear SVM with a k-NN classifer to segment and classify fve diferent classes of skin lesions. Maglogiannis et al.9 implemented a threshold on the RGB value for segmentation and used an SVM for classifcation. Although more robust than clustering algorithms, SVMs are more reliant on the preprocessing of data for feature extraction. Without preprocessing that allows a clear defnition of hyperplanes, SVMs may also underperform.

Owing to the disadvantages of these traditional approaches, convolution neural networks (CNNs) have gained popularity because of their ability to extract high-level features with minimal preprocessing10. CNNs can expand the advantages of SVMs, such as robustness in noisy datasets without the need for optimal preprocessing, by capturing image context and extracting high-level features through down-sampling. CNNs can interpret the pixels of an image within its own image-level context, as opposed to viewing each pixel in a dataset-level context. However, although down-sampling allows CNNs to view an image in its own context, it degrades the resolution of the image. Although context is gained, the location of a target is lost through down-sampling. Tis is not a prob_lem for classifcation, but causes some difculty for segmentation, as both the context and location of the target are essential for optimal performance. To solve this, up-sampling is needed, which works in a manner opposite to that of down-sampling, in the sense that it increases the resolution of the image. While down-sampling takes a matrix and decreases it to a smaller feature map, up-sampling takes a feature map and increases it to a larger matrix. By learning to accurately create a higher-resolution image, CNNs can determine the location of the targets to segment. Tus, for segmentation, we use a combination of down-sampling and up-sampling, whereas for classifcation, we use only down-sampling. To further leverage the advantages of CNNs, skip-connections were introduced, which provided a solution to the degradation problem that occurs when CNN models become too large and complex. We implement skip-connections in both segmentation and classifcation models. In the segmentation model, blocks of equal feature numbers are connected between the down and up-sampling sections. In the classifcation model, these skip-connections exist in the form of inverted residual blocks. Tis allows our models to grow in complexity without any performance degradation.In this paper, we present a method to sequentially combine two separate models to solve a larger problem. In the past, skin disease models have been applied to either segmentation or classifcation. In this study, we sequen_tially combine both models by using the output of a segmentation model as input to a classifcation model. In addition, although past studies of non-CNN segmentation models used innovative preprocessing methods, recent CNN developments have focused more on the architecture of the model than on the preprocessing of data. As such, we apply an innovative preprocessing method to the data of our CNN segmentation model. Te methods described above lack the ability to localize and classify multiple diseases within one image; however, we have developed a method to address this problem. Our objective is two-fold. First, we show that CAD can be used in the feld of dermatology. Second, we show that state-of-the-art models can be used with current computing power to solve a wider range of

complex problems than previously imagined. We begin by explaining the results of our experimentation, followed by a discussion of our fndings, a more detailed description of our methodology, and fnally, the conclusions that can be drawn from our study.

RESULT AND DISCUSSION :

Figure 1 shows the schematic fow of our study. We started with the original image. We preprocessed this image by decomposing it into its hemoglobin and melanin constituents. Tese images were then input to the U-Net to generate the segmented output. We drew contours around each cluster and used a convex hull algorithm to draw rectangles around these clusters and crop them as individual images. Tese cropped images were used as input to the EfcientNet, which generated a prediction along with the confdence rate.Table 1 shows the results of the test data for segmentation on our Dermnet dataset. Te K-means clustering algorithm showed sub-optimal performance, owing to its limitations with noisy data. Te SVM method showed a signifcant improvement in performance, that was attributed to the advantages of using SVMs to extract infor_mation from decomposition, rather than clustering algorithms. Even without the extra information, the U-Net trained without decomposition outperformed the previous two methods in terms of sensitivity. Te U-Net model was also trained with decomposition and showed the highest sensitivity rate.In our results, we focused on the sensitivity metric because our objective was to assess the viability of using CAD with skin images. Although our U-Net model was not as good as the SVM model in terms of the specifcity rate, it showed the best sensitivity rate, thus satisfying the objective of our study. In addition, we included the Dice coefcient and Hausdorf distance to demonstrate the performance of our methods with greater transparency. Our method showed clear improvements considering these alternative metrics. A major contributing factor7to the underperformance of other methods is that performance of the SVM algorithm deteriorated when the images contained diferences in lighting and shade. Te K-means clustering method3 was also afected by the lighting and shade in the images. As our data had a signifcant mix of shade and lighting, the CNN was able to generalize the data better by learning to use the context of the image.In any classifcation problem, it is important to set the baseline performance. We set our baseline to be the accuracy rate of the data without segmentation. Te original image was input into the EfcientNet without going through the U-Net to determine the baseline accuracy rate. We compared this to the accuracy rate of the model trained to classify segmented images. Figure 2 shows the accuracy rates for the classifcation of our Dermnet dataset. We observed similar accuracy in the baseline model with and without contextual segmentation. Te performance did not decrease when compared with the baseline. Tus, as we gained knowledge of the location of the disease without degrading the performance, we may say that the classifcation model was successfully

TABLE1:

Method	Sensitivity	Specificity	Dice Coef	Hausdorff distance
K-means method	0.6148	0.6324	0.5165	10.487
SVM method	0.8200	0.8100	0.7123	8.138
U-Net method without decomposition	0.8953	0.7205	0.7215	8.153
U-Net method with decomposition	0.9589	0.7682	0.8126	7.165

Performance metrics for segmentation with dermnet images.

TOP-N ACCURACY METRIC

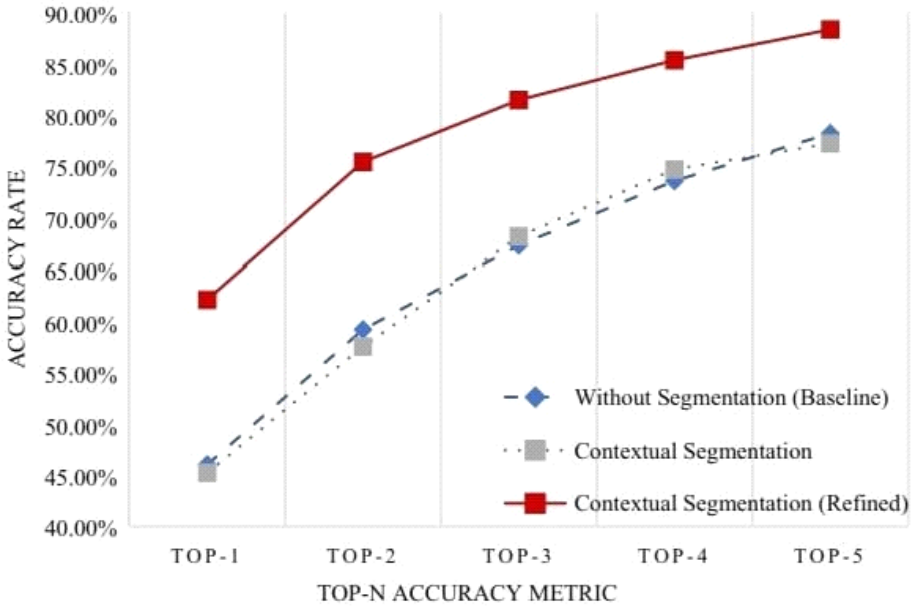


TABLE2 :

Method	AUC	Specificity	Sensitivity	F1-score
Without segmentation	0.8207	0.9642	0.4748	0.4092
Contextual segmentation	0.8104	0.9652	0.4185	0.3876
Refined contextual segmentation	0.8802	0.9513	0.6141	0.6079

Performance matrics for classification with dermnet images.

NOT SEGMENTED & SEGMENTED:

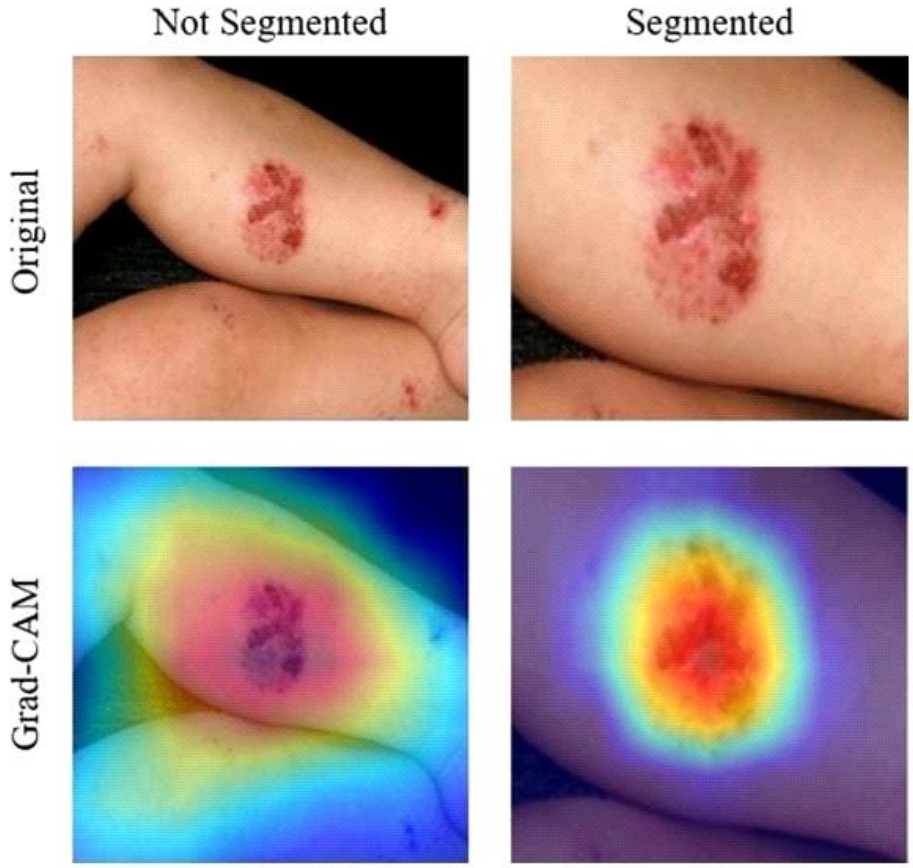


TABLE3 :

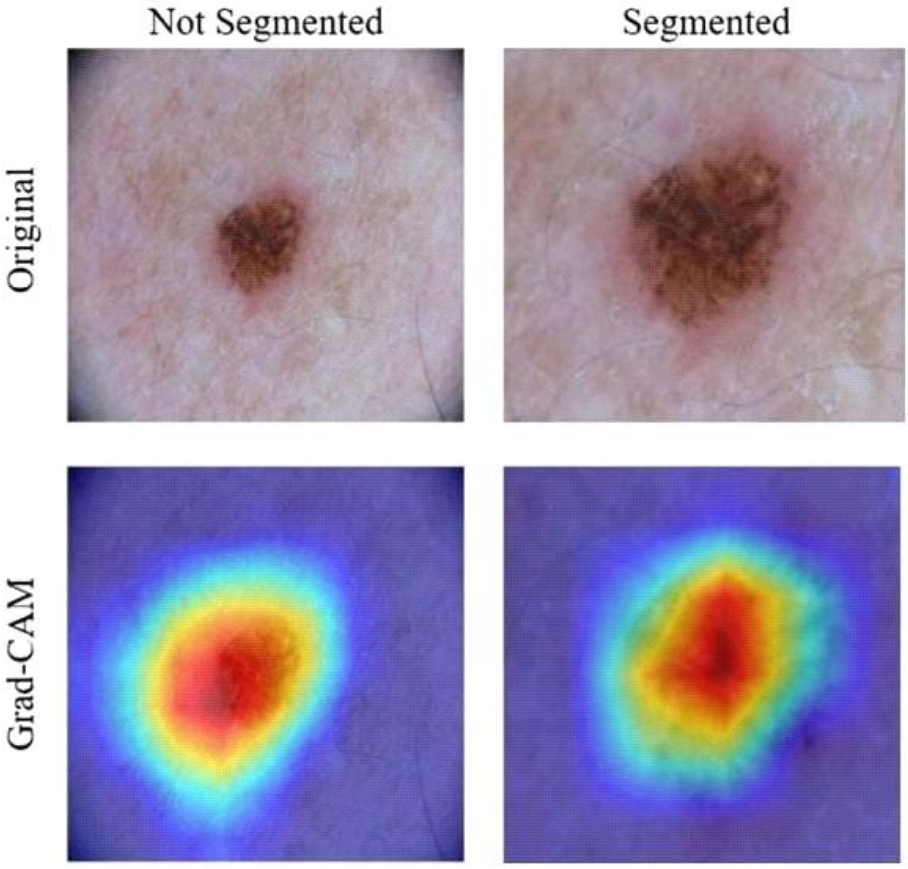
Method	Sensitivity	Specificity	Dice Coef	Hausdorff distance
ISIC2016				
K-means method	0.5422	0.8249	0.5439	9.960
SVM method	0.7229	0.8602	0.6939	8.243
U-Net method without decomposition	**0.9708**	0.9175	0.9060	5.085
U-Net method with decomposition	0.9562	**0.9422**	**0.9198**	**4.764**
ISIC2017				
K-means method	0.5709	0.7734	0.4926	10.567
SVM method	0.7650	0.7576	0.5967	9.388
U-Net method without decomposition	0.8971	0.8969	0.8188	5.392
U-Net method with decomposition	0.9043	0.9076	0.8199	5.338
HAM 10,000				
K-means method	0.5500	0.9300	0.6381	6.807
SVM method	0.7256	0.8389	0.6674	8.381
U-Net method without decomposition	0.9542	**0.9530**	0.9121	4.683
U-Net method with decomposition	**0.9569**	0.9504	**0.9166**	**4.621**

Performance metrics for segmentation with dermatoscopic datasets.

TABLE4 :

Method	AUC	Specificity	Sensitivity	F1-score
ISIC2016				
Without segmentation	**0.765**	**0.726**	**0.860**	**0.864**
Contextual segmentation	0.719	0.641	0.826	0.833
Refined contextual segmentation	0.727	0.698	0.844	0.845
ISIC2017				
Without segmentation	**0.790**	0.741	0.761	0.740
Contextual segmentation	0.750	0.744	0.726	0.723
Refined contextual segmentation	0.774	**0.785**	**0.766**	**0.762**
HAM 10,000				
Without segmentation	**0.891**	**0.933**	0.866	**0.871**
Contextual segmentation	0.831	0.884	0.825	0.810
Refined contextual segmentation	0.871	0.919	**0.873**	0.866

Performance metrics for classifcation with dermatoscopic datasets.


Not Segmented
Segmented
Original
Grad-CAM

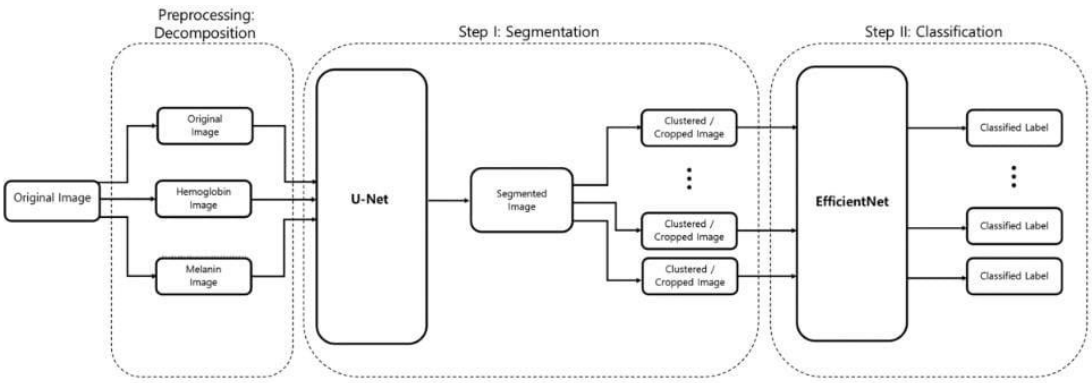


TABLE5 :

Top-level categories		
1. Acne and Rosacea	7. Eczema	13. Psoriasis
2. Actinic keratosis	8. Exanthems	14. Scabies
3. Atopic dermatitis	9. Fungal infections	15. Systemic disease
4. Bullous disease	10. Herpes	16. Urticaria
5. Cellulitis	11. Light chain disease	17. Vasculitis
6. Contact dermatitis	12. Lupus erythematosus	18. Viral infections

Categories for classifcation.

TABLE6 :

Dataset: Dermnet	Number of data					
	Segmentation			Classification		
Class	Train	Test	Total	Train	Test	Total
Acne and Rosacea	4	2	6	746	131	877
Actinic keratosis	4	2	6	1193	181	1374
Atopic dermatitis	3	2	5	642	120	762
Bullous disease	3	2	5	393	92	485
Cellulitis	3	2	5	223	73	296
Contact dermatitis	3	2	5	231	74	305
Eczema	4	3	7	1667	234	1901
Exanthems	3	2	5	354	87	441
Fungal infections	4	3	7	1601	227	1828
Herpes	3	2	5	397	94	491
Light chain disease	3	2	5	538	117	655
Lupus erythematosus	3	2	5	371	90	461
Psoriasis	4	3	7	2044	275	2319
Scabies	3	2	5	448	98	546
Systemic disease	3	2	5	633	119	752
Urticaria	3	2	5	138	63	201
Vasculitis	3	2	5	411	94	505
Viral infections	4	3	7	1443	209	1652
Total	60	40	100	13,473	2378	15,851

Distribution of data in dermnet dataset.

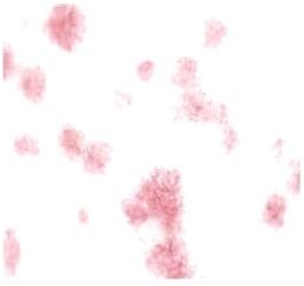
TABLE7 :

Class	Number of data		
	Train	Test	Total
Dataset: ISIC 2016			
Benign	727	303	1030
Malignant	173	75	248
Total	900	378	1278
Dataset: ISIC 2017			
Benign	1372	393	1843
Melanoma	374	117	386
Seborrheic keratosis	254	90	521
Total	2000	600	2750
Dataset: HAM 10000			
Actinic keratosis	164	163	327
Basal cell carcinoma	257	257	514
Benign	549	550	1099
Dermatofibroma	58	57	115
Melanoma	556	557	1113
Melanocytic nevi	3352	3353	6705
Vascular lesion	71	71	142
Total	5007	5008	10,015

Distribution of data in dermatoscopic datasets.

Algorithm 1 AnalyzeSkin
1: **procedure** SEGMENT(x)
2: h, m = DECOMPOSE(x)
3: *mask* = U-NET([x, h, m]) 4: CLASSIFY(*mask*)
5: **end procedure**
6: **procedure** CLASSIFY(*mask*)
7: *clusters* = FINDCLUSTERS(*mask*)
8: **for** *cluster* **in** *clusters* **do**
9: *cluster* = FIXRATIO(*cluster*)
10: *cluster* = RESIZE(*cluster*)
11: *class* = EFFICIENTNET(*cluster*) 12. *top_prediction* = GETHIGHESTCONFIDENCE(*class*)
12: print(*top_prediction*)
13: **end for**
14: **end procedure**

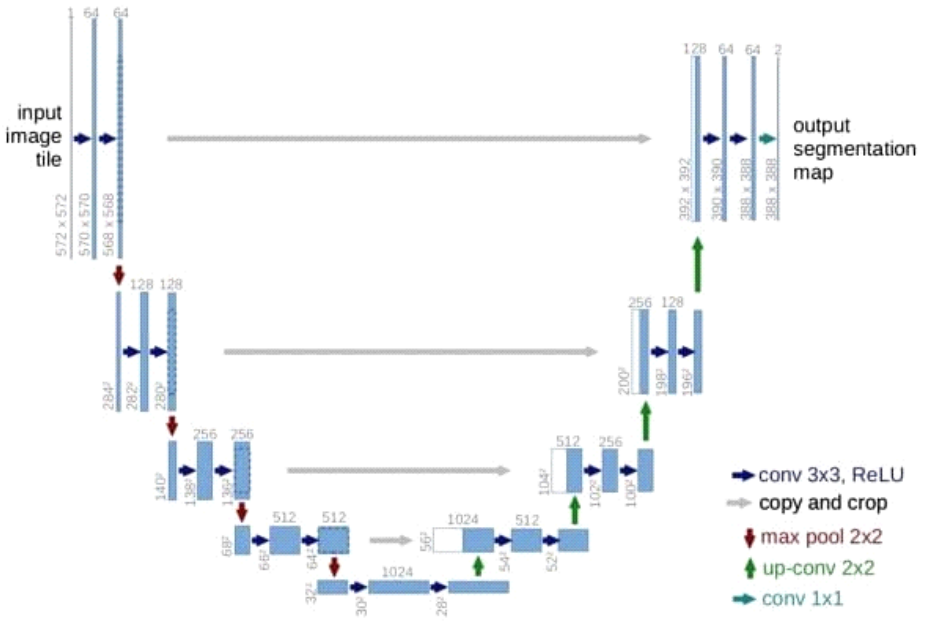
a. Original Image



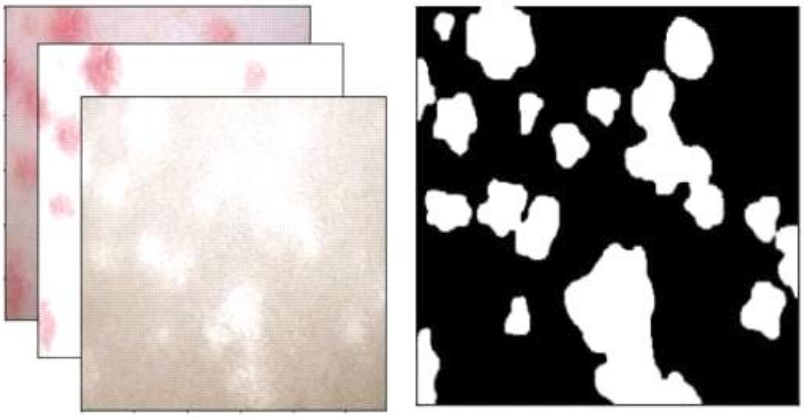
b. Hemoglobin Image

c. Melanin Image

SEGMENTATION :



.



a. Input b. Output

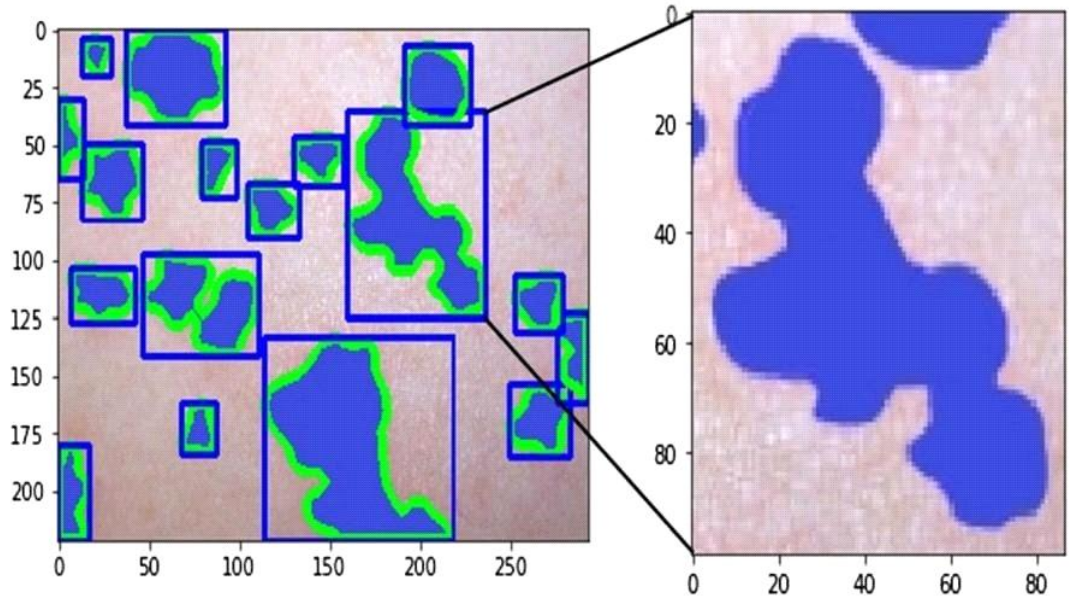


TABLE8 :

Model	Top-1 accuracy (%)	Training time per epoch (s)
EfficientNet-B0	39.71	187.965
EfficientNet-B1	43.15	250.170
EfficientNet-B2	44.46	255.180
EfficientNet-B3	43.30	309.375
EfficientNet-B4	**45.77**	**392.925**
EfficientNet-B5	45.54	522.975
EfficientNet-B6	45.83	643.965
EfficientNet-B7	47.54	942.720

ETHICS DECLARATIONS:

This study was exempted from the approval by the Institutional Review Board of Seoul National University Boramae Medical Center (No. 07-2020-148). The informed consent was waived by the Insti_tutional Review Board of Seoul National University Boramae Medical Center because patient records Information was anonymized and de-identifed prior to analysis. All experiments were performed in accordance with the relevant guidelines and regulations.

CONCLUSION:

We have shown that even without a large dataset and high-quality images, it is possible to achieve sufcient accuracy rates. In addition, we have shown that current state-of-the-art CNN models can outperform models created by previous research, through proper data preprocessing, self-supervised learning, transfer learning, and special CNN architecture techniques. Furthermore, with accurate segmentation, we gain knowledge of the location of the disease, which is useful in the preprocessing of data used in classifcation, as it allows the CNN model to focus on the area of interest. Lastly, unlike previous studies, our method provides a solution to classify multiple diseases within a single image. With higher quality and a larger quantity of data, it will be viable to use state-of-the-art models to enable the use of CAD in the feld of dermatology.