**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 fields such as mammography and colonography, it is not used in dermatology, where noninvasive 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 classification 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 classification model. We classify each cluster into different 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 classification 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 sufficient to use CAD in the field of dermatology.

Computer-aided diagnosis (CAD) is a computer-based system that is used in the medical imaging field to aid healthcare workers in their diagnoses1. CAD has become a mainstream tool in several medical fields such as mammography and colonography1,2. However, in dermatology, although skin disease is a common disease, one in which early detection and classification is crucial for the successful treatment and recovery of patients, dermatologists 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, classification of a disease is difficult due to the strong similarities between common skin disease symptoms. Therefore, it would be beneficial to exploit the strengths of CAD using artificial intelligence techniques, in order to improve the accuracy of dermatology diagnosis. Tis paper shows that CAD may be a viable option in the field of dermatology using state-of-the-art deep learning models. The segmentation and classification of skin diseases has been gaining attention in the field of artificial intelligence because of its promising results. Two of the more prominent approaches for skin disease segmentation and classification are clustering algorithms and support vector machines (SVMs). Clustering algorithms generally have the advantage of being flexible, easy to implement, with the ability to generalize features that have a similar statistical variance. Table at 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. Raj abet 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 identification of a centroid that can generalize a cluster of data. Noisy data, or the presence of outliers, can significantly degrade the performance of these algorithms. Therefore, with noisy datasets, caused by images with different types of lighting, non-clustering algorithms may be preferred; however, at 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 effectiveness in high-dimensional data and their capability to decipher “…subtle patterns in noisy and complex datasets”6. Level at al.7 segmented erythema in the skin using the radial basis kernel function that allows SVMs to separate nonlinear hyper planes. Sumithr at al.8 combined a linear SVM with a k-NN classifier to segment and classify five different classes of skin lesions. Maglogiannis at al.9 implemented a threshold on the RGB value for segmentation and used an SVM for classification. Although more robust than clustering algorithms, SVMs are more reliant on the preprocessing of data for feature extraction. Without preprocessing that allows a clear definition of hyper planes, 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. This is not a problem for classification, but causes some difficulty 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. Thus, for segmentation, we use a combination of down-sampling and up-sampling, whereas for classification, 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 classification models. In the segmentation model, blocks of equal feature numbers are connected between the down and up-sampling sections. In the classification model, these skip-connections exist in the form of inverted residual blocks. This allows our models to grow in complexity without any performance degradiation.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 classification. In this study, we sequentially combine both models by using the output of a segmentation model as input to a classification 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. The 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 field 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 findings, a more detailed description of our methodology, and finally, the conclusions that can be drawn from our study.

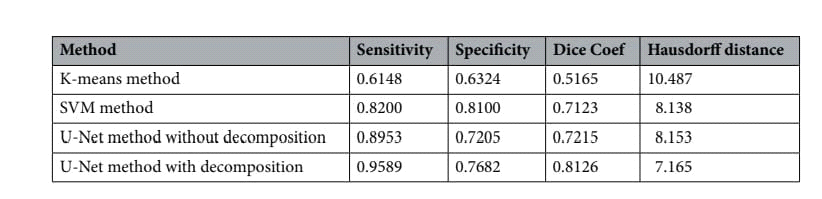
**RESULT AND DISCUSSION :**

Figure 1 shows the schematic flow of our study. We started with the original image. We preprocessed this image by decomposing it into its hemoglobin and melanin constituents. These 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. These cropped images were used as input to the Efficient Net, which generated a prediction along with the confidence rate. Table 1 shows the results of the test data for segmentation on our dataset. The K-means clustering algorithm showed sub-optimal performance, owing to its limitations with noisy data. The SVM method showed a significant improvement in performance, was attributed to the advantages of using SVMs to extract information 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. The 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 specificity rate, it showed the best sensitivity rate, thus satisfying the objective of our study. In addition, we included the Dice coefficient and or 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 differences in lighting and shade. The K-means clustering method3 was also affected by the lighting and shade in the images. As our data had a significant 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 classification problem, it is important to set the baseline performance. We set our baseline to be the accuracy rate of the data without segmentation. The original image was input into the Efficient Net 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 classification of our dataset. We observed similar accuracy in the baseline model with and without contextual segmentation. The performance did not decrease when compared with the baseline. Thus, as we gained knowledge of the location of the disease without degrading the performance, we may say that the classification model was successfully

**FIGURE1 :**

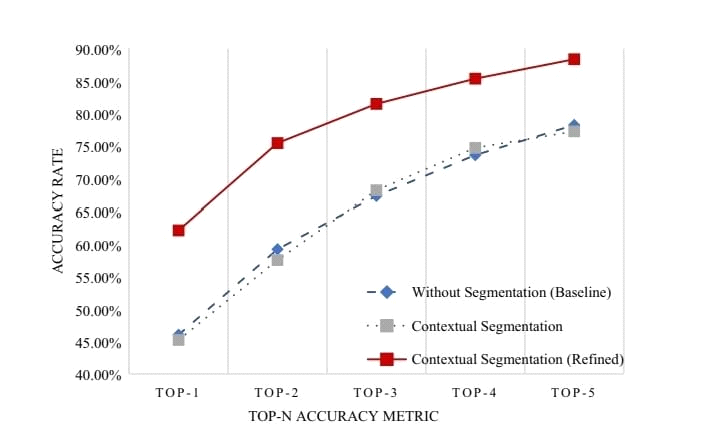
Schematic flow. From left to right, the original is first decomposed into hemoglobin and melanin images. All three images are input to the U-Net which outputs a black-and-white mask image. Tis mask image is used to draw contours each cluster. A convex hull algorithm is applied to crop each cluste

**TABLE1:**



Performance metrics for segmentation with dermnet images.

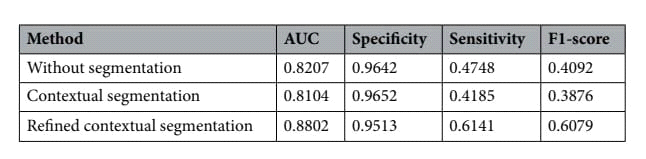
**TOP-N ACCURACY METRIC**



**FIGURE2 :**

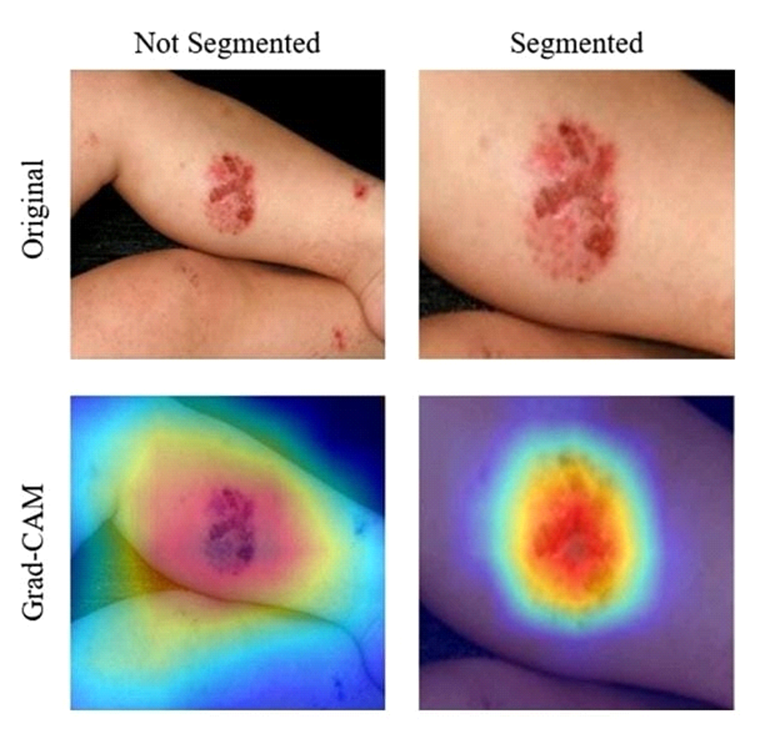
Accuracy rate for classification. The x-axis represents the Top-n accuracy metric, while the y-axis represents accuracy. The blue line is the accuracy of the model trained without segmentation. Images did not enter the U-Net before entering the Efficient Net. The gray line represents the accuracy of the model trained with segmentation. Images were segmented and cropped through the U-Net before entering the Efficient Net. The red line represents the accuracy of the model trained with segmentation and refined data. Images were segmented, cropped, and verified to ensure that segmentation had been done correctly before entering the Efficient Net.

**TABLE2 :**



Performance matrices for classification with term net images.

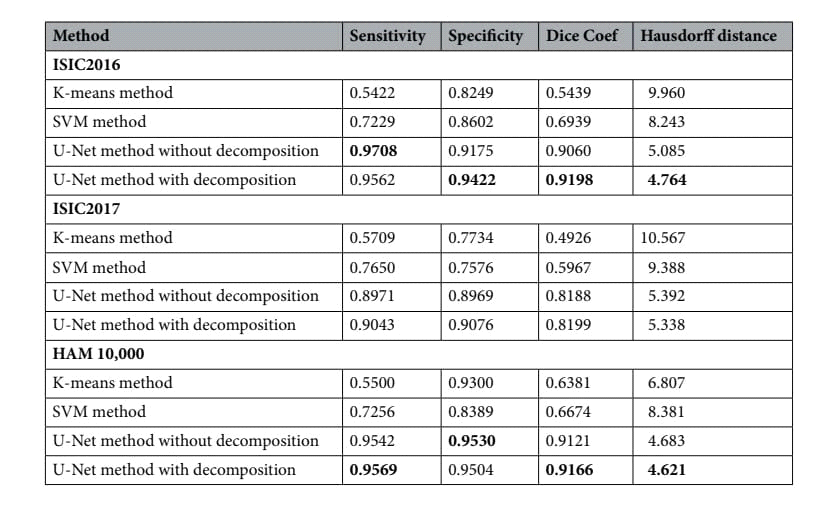
**NOT SEGMENTED & SEGMENTED**:

****

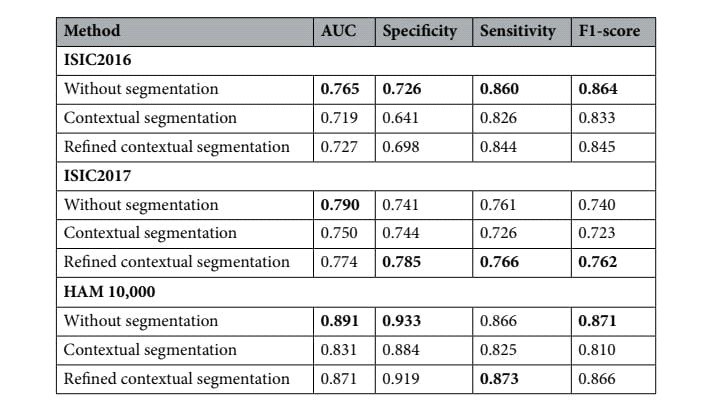
**FIGURE3 :**

Grad-CAM results for un segmented and segmented images in our term net dataset. The top row shows the original input images. The left image shows the un segmented image and the right image shows the segmented image. The bottom row shows the result of Grad-CAM11. The left image of Grad-CAM for the un segmented image shows that the Efficient Net model focused on a larger surface other than erythema. The right image of Grad-CAM for the segmented image shows than that the Efficient Net model correctly focused mostly on erythema.

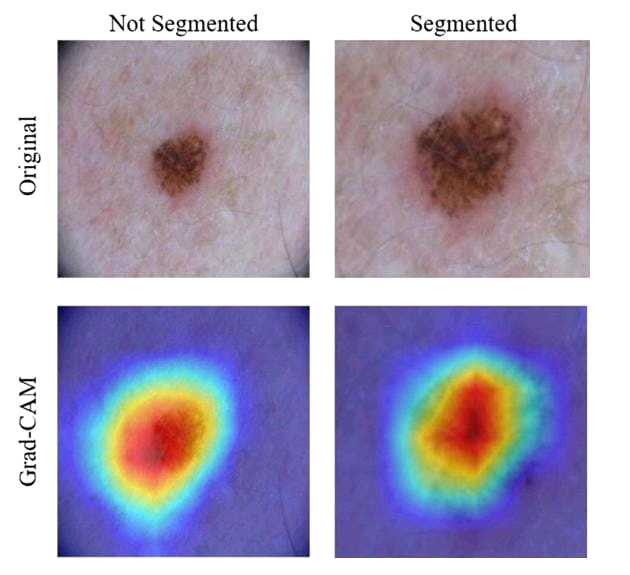
**TABLE3 :**

 Performance metrics for segmentation with termatoscopic datasets.

**TABLE4 :**



Performance metrics for classification with dermatoscopic datasets.

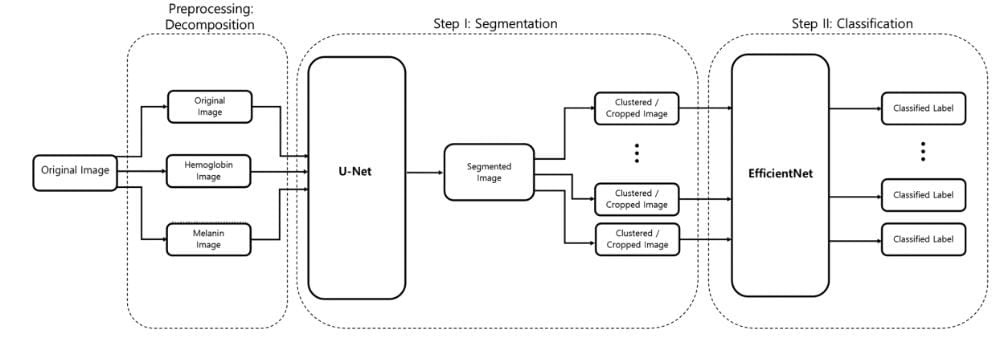


**FIGURE4 :**

Grad-CAM results for un segmented and segmented images in the ISIC2017 dataset. For both images of Grad-CAM, the Efficient Net model correctly focused mostly on erythema.in all but one category. In short, with dermatoscopic images, models trained without segmentation learned to generalize skin lesions most effectively. This was a result of an improved performance when the location of the skin lesion is mostly fixed. The segmentation phase aids models to ignore areas of normal skin and to focus on areas of disease. With dermatoscopic images, this information is in significant, as the location of the disease is static. Figure 4 shows a visual representation of this. The Grad-CAM images show that with both non-segmented and segmented images, the models correctly focused on the skin disease. Because of this, the segmentation phase only decreased the resolution of the image without providing useful information, thus decreasing the performance of the model.

**Analysis methodology :**

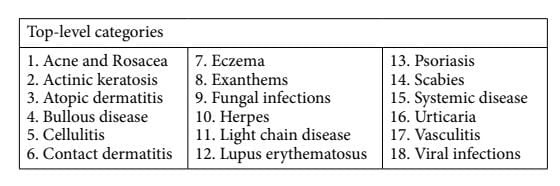
Our 2-phase analysis model for localization and classification is shown via the pseudo code in Algorithm 1 and visually in Fig. 5. We decomposed the original image into its hemoglobin and melanin constituents using preprocessing, to help our model extract valuable information from data that would have been otherwise unavailable. We provide theseimages as input to our segmentation model, the U-Net, which generated a segmented image. This segmented image was then analyzed for clusters, which were subsequently cropped and input to our classification model, the Efficient Net, which then produced a classified label, thus completing our analysis model.



**FIGURE5 :**

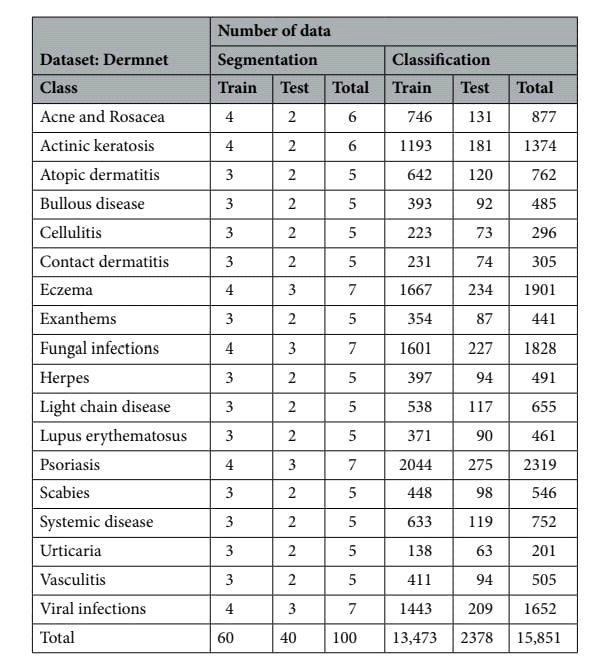
Two-phase analysis model. The original image primarily enters a preprocessing stage, where normalization and decomposition occur. Afterwards, the first step is segmentation, where cluster of abnormal skin are segmented and cropped. The second step is classification, where each cluster is classified into its corresponding class.

**TABLE5 :**



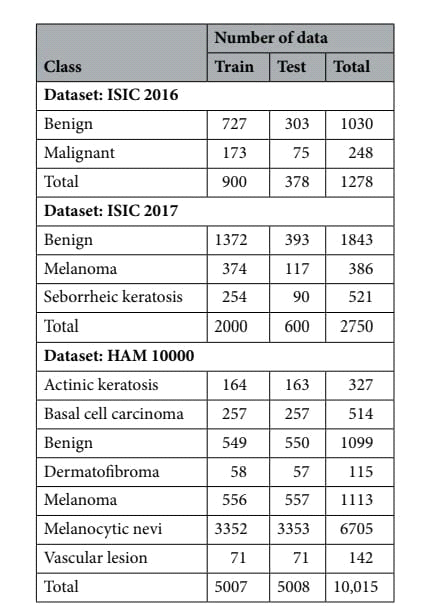
Categories for classification.

**TABLE6 :**

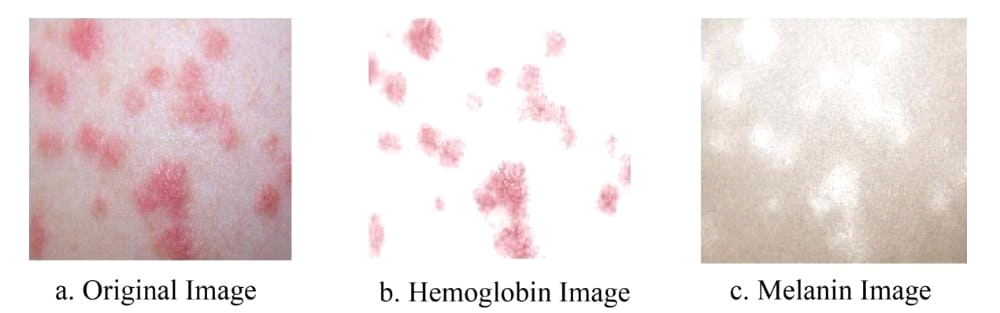


Distribution of data in term net dataset.

**TABLE7 :**



Distribution of data in dermatoscopic datasets



**FIGURE6 :**

Decomposed result of skin. The original image is decomposed into its hemoglobin and melanin constituents through ICA.

**PREPROSSING:DECOMPOSITION :**

The main constituents of the skin that are visible to humans are melanin and hemoglobin. These constituents provide valuable information for the segmentation of abnormal skin. To ensure that our model can learn to use these features, we used independent component analysis (ICA) to extract the melanin and hemoglobin constituents7,15,16. Assuming that these components are linearly separable, the separated linear vectors can be represented by the following formula7:

Lx,y = dmqmx,y + dhqhx,y +del \_\_

where dm and dh represent the density vectors of melanin and hemoglobin, respectively, qmx,y and qhx,y represent the quantity of these components, and \_ represents values that are caused by other colors. As shown in7, by applying ICA, we can decompose skin as

[qmx,y, qhx,y]= ?D?1L(x,y) ? E

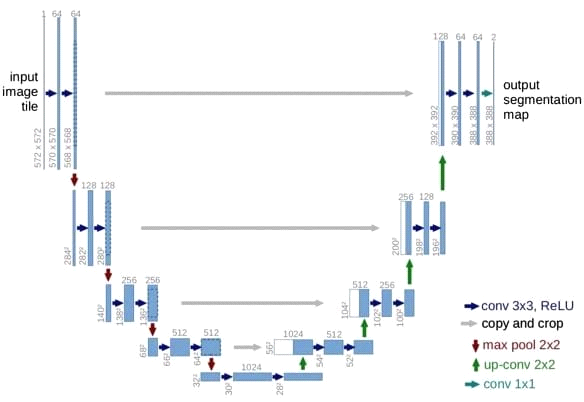
E = minx,y(?D?1L(x,y))

Ix,y = exp(?L?x,y)

where D represents the estimated values of dm and dh, and Ix,y represents the decomposed result. Figure 6 shows an example of one of these decompositions.

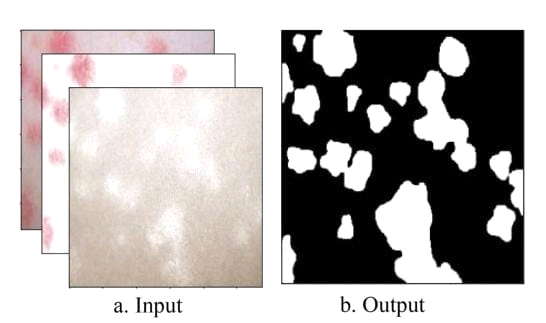
**SEGMENTATION :**

The U-Net17, as shown in Fig. 7, is an architecture created by CNNs, that has attracted attention for accurate biomedical image segmentation through the combination of down-sampling, up-sampling, and skip connections. Its name is attributed to the shape of its architecture, the first half of the ‘U’ representing down-sampling. Here, the context and key features of the input images are gained at the cost of a decrease in resolution. Te second half of the ‘U’ represents up-sampling. Here, the resolution is increased to gain knowledge of the location of the target segment. To combat degradation due to the complexity of the model, skip connections are added to each up-sampling block.



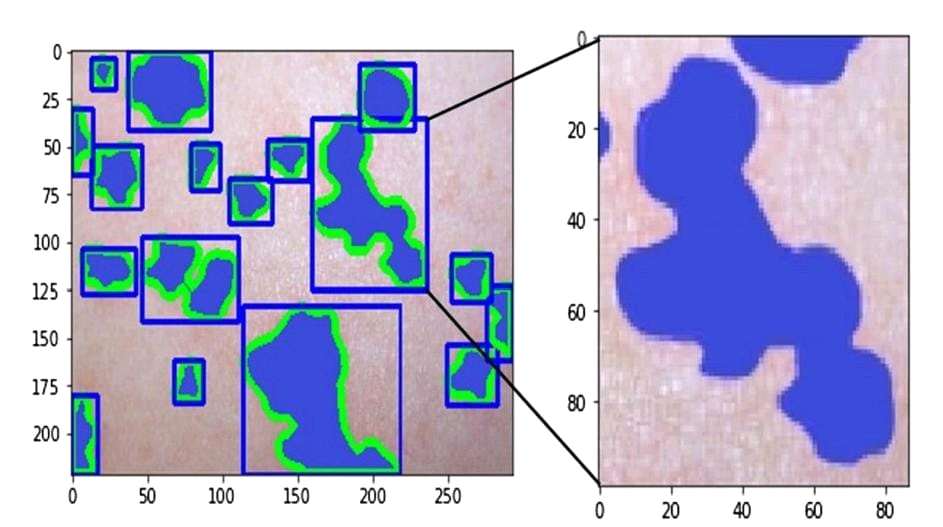
**FIGURE7 :**

U-Net architecture. A fully CNN network, comprised of down-sampling, up-sampling, and skip connections17.



**FIGURE8 :**

Input and output of the U-Net. The inputs of the U-Net are the original,hemoglobin,and melanin images obtained from the preprocessing step. The output of the U-Net is a single masked image.



**FIGURE9 :**

Contour finding algorithm applied to output of U-Net. Clusters of abnormal skin are identified through a contour finding algorithm. Each cluster is cropped in the shape of a rectangle through a convex hull algorithm used to surround each contour.

Sensitivity = TP/TP + FN

Specificity = TN/TN + FP

DiceCoef . = 2 × TP/(TP + FP) + (TP + FN)

The Hausdorff distance (HD) is used to measure the dissimilarity between the predicted segmentation masks the and ground truth. Te Hausdorf distance can be calculated by the formula18;

SetX = {x1, ... xn}andY = {y1, ... , yn}

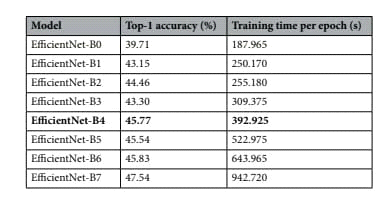
H(X, Y) = max(h(X, Y), h(Y, X)),

where h(X, Y) = max min||x\_y||.

x?X y?Y

We use an implementation of the method presented18 to calculate the Hausdorf distance between the output and ground truth.

**CLASSIFICATION:**



**TABLE8 :**

Training time required for efcientnet-B0 through B7.

such that : N(d, w, r) = ?i=1...sF\_d•Li (X<r•H\_i,r•W\_ i,w•C\_i>)

Memory(N) ? targetmemory

FLOPS(N) ? targetflops

Specificity = TN/TN + FP

Sensitivity = TP/TP + FN

F1 ? score = 2TP/2TP + FP + FN

For all performance metrics, scores are calculated individually for each class present in the dataset. The scores are then weighted and averaged according to the number of data points in a class corresponding to the entire dataset.

**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 Institutional Review Board of Seoul National University Boramae Medical Center because patient records Information was anonymized and de-identified 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 sufficient 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 classification, 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.