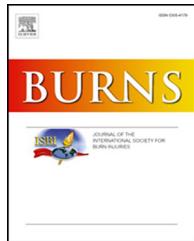




Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/burns



Application of multiple deep learning models for automatic burn wound assessment

Che Wei Chang^{a,b,*}, Chun Yee Ho^b, Feipei Lai^{a,c}, Mesakh Christian^c, Shih Chen Huang^b, Dun Hao Chang^{b,d}, Yo Shen Chen^b

^a Graduate Institute of Biomedical Electronics & Bioinformatics, National Taiwan University, Taipei, Taiwan

^b Division of Plastic and Reconstructive Surgery, Department of Surgery, Far Eastern Memorial Hospital, New Taipei, Taiwan

^c Department of Computer Science & Information Engineering, National Taiwan University, Taipei, Taiwan

^d Department of Information Management, Yuan Ze University, Taiwan

ARTICLE INFO

Article history:

Accepted 14 July 2022

Keywords:

Machine learning
Deep learning
Burn wound segmentation
Deep burn segmentation
% TBSA burned
Rule of palm

ABSTRACT

Purpose: Accurate assessment of the percentage of total body surface area (%TBSA) burned is crucial in managing burn injuries. It is difficult to estimate the size of an irregular shape by inspection. Many articles reported the discrepancy of estimating %TBSA burned by different doctors. We set up a system with multiple deep learning (DL) models for %TBSA estimation, as well as the segmentation of possibly poor-perfused deep burn regions from the entire wound.

Methods: We proposed boundary-based labeling for datasets of total burn wound and palm, whereas region-based labeling for the dataset of deep burn wound. Several powerful DL models (U-Net, PSPNet, DeeplabV3+, Mask R-CNN) with encoders ResNet101 had been trained and tested from the above datasets. With the subject distances, the %TBSA burned could be calculated by the segmentation of total burn wound area with respect to the palm size. The percentage of deep burn area could be obtained from the segmentation of deep burn area from the entire wound.

Results: A total of 4991 images of early burn wounds and 1050 images of palms were boundary-based labeled. 1565 out of 4994 images with deep burn were preprocessed with superpixel segmentation into small regions before labeling. DeeplabV3+ had slightly better performance in three tasks with precision: 0.90767, recall: 0.90065 for total burn wound segmentation; precision: 0.98987, recall: 0.99036 for palm segmentation; and precision: 0.90152, recall: 0.90219 for deep burn segmentation.

Conclusion: Combining the segmentation results and clinical data, %TBSA burned, the volume of fluid for resuscitation, and the percentage of deep burn area can be automatically

* Corresponding author at: Graduate Institute of Biomedical Electronics & Bioinformatics, National Taiwan University, Taipei, Taiwan.
E-mail address: b92401068@gmail.com (C.W. Chang).

diagnosed by DL models with a pixel-to-pixel method. Artificial intelligence provides consistent, accurate and rapid assessments of burn wounds.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The incidence of burn injuries varies significantly across different countries. Though the incidence of burns trended lower in the highly developed countries, it is still a concern in the less developed countries [1]. The successful management of acute burn injuries relies on timely fluid resuscitation, which is related to accurate estimation of %TBSA burned. For the past decade, there were articles reporting the inaccurate estimation of %TBSA burned by first line medical staff [2–4]. Yet, even experienced burn surgeons make inaccurate %TBSA estimation. An interesting research conducted by Parvizi et al. reported that experienced burn specialists had a high deviation of %TBSA estimation when assessing the same images of burn wounds [5].

In recent years, machine learning (ML) has been widely used in chronic wound diagnosis [6], while the utilization in burn assessment was relatively scarce [7,8]. Early works of ML in burn wounds focus on burn depth classification [9,10]. The results were unsatisfactory. Since most burn injuries had mixed burn depths, classifying any image of burn wound into a single burn depth is not likely. To improve the results, some studies applied the following processing: 1. transferring to other color space, such as CIELAB or HSV, 2. choosing the channel with the highest intensity, 3. cropping on only small areas [11–13]. Later works deal with burn depth segmentation and burn wound segmentation [14,15]. Combining the above approaches and thresholding the above data for new images, the burn depth segmentation can be achieved.

There are still some challenges with these methods. First, the segmentation of burn depth only presented the most homogenous part but not all information of a given burn depth. Secondly, after those approaches, the ML models usually overfit to the training dataset, and may have poor results in new coming images. Third, most studies did not mention a standard labeling method for deep learning model training. Deep learning (DL) is an emerging subset of machine learning. The architectures of DL comprised multiple fully connected or convolution layers (CNN) with activation functions to calculate the possibility of input data. Compared with traditional ML, DL model can comprehend all unfavored factors rather than excluding them to get better results. For example, the luminance component was usually eliminated to minimize the effect of lighting. The elimination of the luminance component, though producing better validation, makes the model difficult to apply in actual clinical conditions. For DL models, the luminance component could be reserved as another input to learn various lighting in different clinical situations.

Our previous study utilized the segmentation result of the burn wounds and palms [16]. Based on the Rule of palm, the %TBSA burnt could be calculated. In this study, we proposed superpixel segmentation preprocessing before

image labeling to segment the deep burn area. We trained DL models with various architectures and combined multiple models outputting %TBSA estimation and deep burn segmentation.

2. Methods

2.1. Image acquisition

This study was approved by the research ethics review committee of the Far Eastern Memorial Hospital (number 109037-F). We collected the images of acute burn wounds in the database and EMR from January 2016 to December 2020. All images were taken under different conditions, such as outpatient clinics, operation theaters or wards. These images were also taken by various devices, including cell phones and digital cameras.

An image was excluded if the wound was coated with ointments or dressing; if it involved body parts that can identify a patient, such as a face, adjacent to tattoo and birthmarks; if it appeared to have undergone an intervention, such as debridement or skin graft; and if the images were taken within 48 h of burn injuries. If the long axis of an image was less than 1000 pixels, the image was excluded due to low image quality. All images were randomly assigned numbers to be de-identified and were then presented to burn surgeons for labeling. To estimate the %TBSA by the Rule of palm, we collected images of the volar side of hands from our colleagues with nearly equal numbers of male to female. These images were also taken in various conditions by different devices.

2.2. Boundary-based labeling

Two labeling methods, boundary-based labeling and region-based labeling, were applied. The burn wounds on images were labeled with *Labelme* (python3.7) using polygons to delineate all borders of the wound. The task was co-labeled by 2 out of 3 burn surgeons. One surgeon labeled the images, and the others re-checked the results. If there was no consensus on the borders after discussion, usually due to poor light conditions or blurring images, the images were also excluded. The dataset of total burn area (Dataset 1) contained original images and json files. The images of open hands were also labeled with *Labelme* using polygons. The borders of palms (hand without digits) were delineated and saved as json file as Dataset 2. Fig. 1 shows the workflows of boundary-based labeling for Dataset 1 and Dataset 2 and models training from these two datasets.

2.3. Region-based labeling

The characters and textures of deep burn varied, which could be pink, dark red, white, or brown. The clinical burn depths,

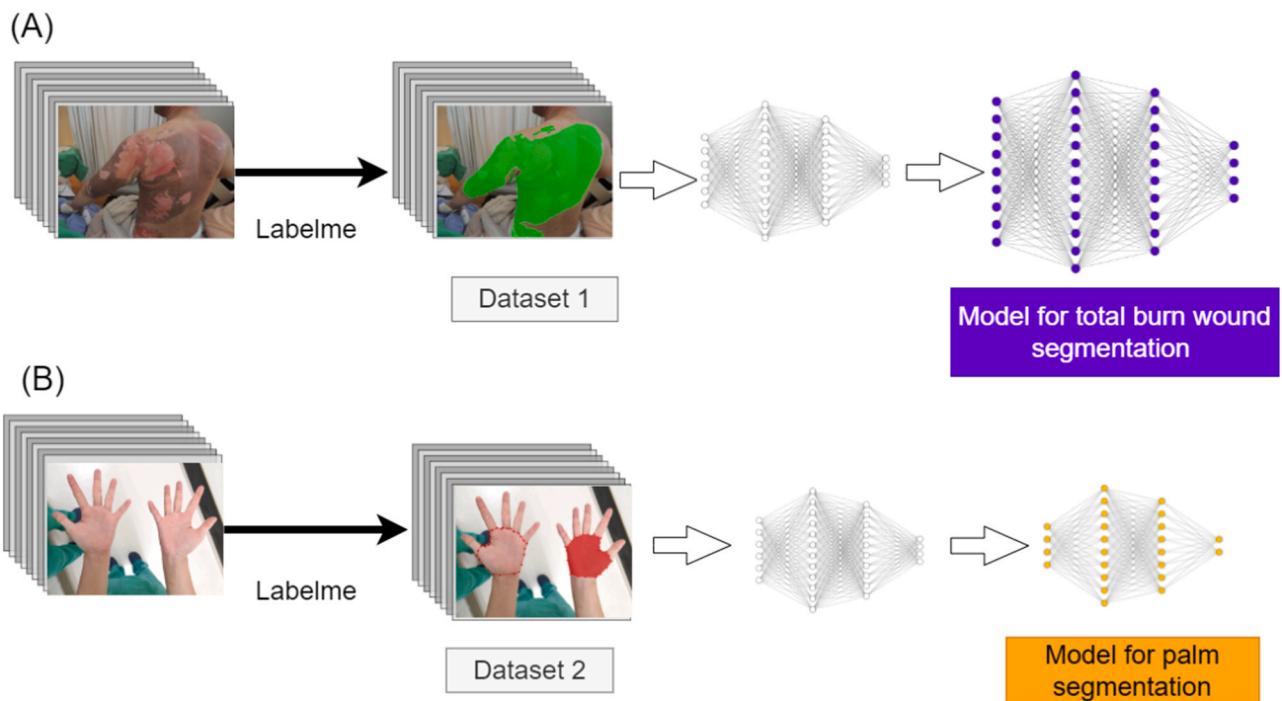


Fig. 1 – Boundary-based labeling and models training. (A) Borders of burn wounds were labeled and saved as Dataset 1. Models training from Dataset 1 were applied for total burn wound segmentation. (B) Borders of hands without digits were labeled and saved as Dataset 2. Models training from Dataset 2 were used for palm segmentation.

first degree, superficial second, deep second, and full thickness were pathological definitions. In practice, the burn depths were diagnosed by the inspection of burn specialists but not by biopsy. We labeled the deep burn area with poor perfusion but not differentiated the burn depths pathologically. The deep burn with poor perfusion is usually white or brown in color, and leathery in texture. However, the borders of deep burn areas were not easy to outline manually. To deal with the problem, we preprocessed the images with superpixel segmentation before labeling.

After selecting the images with deep burn, these images were segmented by simple linear iterative clustering (SLIC) method [17]. SLIC is the most popular method of superpixel segmentation. It clusters the pixels by separating color vectors (l, a, b) in the CIELAB color space from the pixel positions (x, y) rather than using a five-dimension vector (l, a, b, x, y). The desired number of equally sized superpixels (K) is chosen. For an images with N pixels, the superpixel center at every grid interval ($S = \sqrt{N/K}$) is defined. Compactness (m) is a variable that controls the weight of the spatial term. The center of K superpixel C_k is $(l_k, a_k, b_k, x_k, y_k)$. Each pixel I is represented as $(l_i, a_i, b_i, x_i, y_i)$. The color distance (D_{lab}) and Spatial distance (D_{xy}) between center of superpixels and each pixel are defined as follows:

$$D_{lab} = \sqrt{(l_i - l_k)^2 + (a_i - a_k)^2 + (b_i - b_k)^2},$$

$$D_{xy} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2},$$

$$D = D_{lab} + \frac{m}{S} * D_{xy},$$

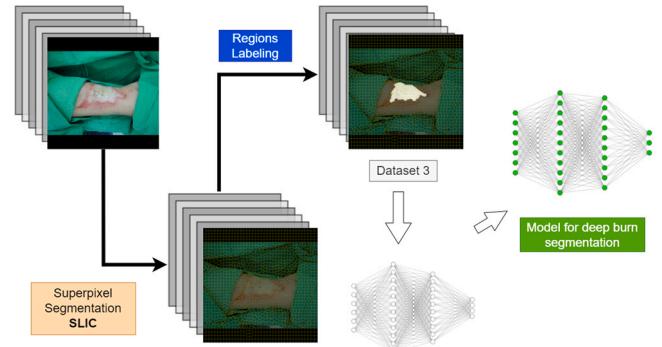


Fig. 2 – Region-based labeling and model training. The images were preprocessed with SLIC to form many superpixels. Burn surgeons could label any superpixels, small homogenous regions, as deep burn. Model training from Dataset 3 aimed to segment deep burns from the whole.

where D is the sum of color distance and the spatial distance normalized by the grid interval. The borders of superpixels to minimize all distances (D) could be calculated after many iterations.

All images were resized to long axis of 1000 pixels filled with a black background to 1000*1000 pixels to speed up model training. The parameters are set as K from 1000 to 1500 and $m = 10$, to form superpixels. After preprocessing with SLIC, images were presented to plastic surgeons to select the regions of deep burns. The labeled images and original

images were saved as the dataset of deep burn (Dataset 3). Fig. 2 shows the workflow of superpixel preprocessing, region-based labeling for Dataset 3 followed with models training.

2.4. Deep learning models

The three datasets were input into four well-developed DL models (U-Net, DeeplabV3+, PSPNet, and Mask R-CNN) to compare their performances on each task. U-Net proposed by Ronneberger et al., is the most popular semantic segmentation model in the medical field [18]. Its architecture, which involves a series of encodings followed by a decoding process, resembles the letter U. DeeplabV3+ proposed by Chen et al. is an encoder-like architecture [19]. DeeplabV3+ adopts atrous convolutions to solve the loss of details when going through multiple convolution layers. PSPNet proposed by Zhao et al., utilizes multiple sub-regional features extraction from large size to small size [20]. The model up-samples all the sub-regional features and then concatenates the original feature map to form a single layer. Mask R-CNN is proposed by researchers of Facebook [21]. It is an advanced model of Faster R-CNN and a tool for instance segmentation. The standard Dice loss was chosen as the loss function for U-Net, DeeplabV3+, and PSPNet. The loss function (L) could be defined by true positive (TP), false positive (FP), and false negative (FN), which is given by:

$$L(TP, FP, FN) = 1 - \frac{2TP + \epsilon}{2TP + FP + FN + \epsilon}$$

The ϵ term is used to avoid the issue of dividing by 0 when precision and recall are empty. Mask R-CNN uses a multi-task loss function given by $L = L_{\text{class}} + L_{\text{box}} + L_{\text{mask}}$. The L_{class} component contains the Region Proposal Network loss. The L_{box} measures failure of object localization or bounding. The last component L_{mask} constitutes the loss from failure to correctly predict the segmentation mask.

All models were combined with ResNet101 as their encoder [22]. ResNet101 is a powerful algorithm that can be compatible with any deep learning model and is easy to optimize. We also initialized them with pre-trained model weights derived from large-scale object detection, segmentation, and captioning datasets, such as ImageNet and COCO. The standard image augmentations we used were rotating, shifting, scaling, gaussian blurring, and contrast normalization. All the models were trained in the Taiwan Computing Cloud, using their container service, on a server configured with four NVIDIA TESLA V100 GPU cards, a 16 core Intel XEON Gold 61 CPU and 360 GB RAM. The parameters of each model are listed in Appendix Table 1.

2.5. Evaluation metrics

We used five metrics to evaluate the performance of DL models: Dice's coefficient (F1 score), intersection over union (IoU), precision, recall and accuracy. These metrics could be defined by true positive (TP), false positive (FP), true negative (TN) and false negative (FN) for any input image, as follows:

DC (F1 score) is twice the area of the intersection of the ground truth and prediction divided by the sum of their areas. It is given by:

$$DC = \frac{2 |Area(Predict) \cap Area(Ground truth)|}{|Area(Predict)| + |Area(Ground truth)|}$$

or

$$\frac{2TP}{2TP + FP + FN}$$

The intersection over union (IoU) denotes the area of the intersection of the ground truth and prediction divided by the area of their union. It is given by:

$$IoU = \frac{|Area(Predict) \cap Area(Ground truth)|}{|Area(Predict) \cup Area(Gound truth)|} \quad \text{or} \quad \frac{TP}{TP + FP + FN}$$

Precision is defined as the ratio of the actual wound pixels that the model correctly classified to all predicted wound pixels. It is also called the positive predicted value (PPV) and given by:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall is defined as the ratio of the actual wound pixels that are correctly classified to all actual wound pixels. It is also called sensitivity and given by:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Accuracy denotes the percentage of correctly classified pixels. It is given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

3. Results

3.1. Total burn wound segmentation

Initially, 6993 images of burn wounds were collected. All images were presented to burn surgeons to label. Final 4991 images were co-labeled with *labelme* and processed as the dataset of total burn area (Dataset 1). Dataset 1 was randomly split at a ratio of 7:2:1 into three datasets for training, validation and testing and then fed into four DL models for training. Three-fold validation was applied to prevent bias of splitting the dataset. Table 1 shows the segmentation results of total burn wounds. DeeplabV3+ had the best performance (DC: 0.8938, IoU: 0.8289) and was chosen as Model 1. Mask R-CNN had a performance with a DC of 0.8913 and an IoU of 0.8214. Interestingly, though U-Net was the oldest architecture among all, it outperformed the newer PSPNet with a DC of 0.8487 and an IoU of 0.8179.

In Fig. 3, the testing images of burn wounds comprised burn wounds with different burn depths. Four DL models achieved decent results, and DeeplabV3+ showed the best performance.

3.2. Palm segmentation

To convert the burn wound to %TBSA burned, we adopt the Rule of palm as a guide of 0.5 %TBSA. We collected 1050 images of hands from our colleagues. The dataset of palm comprised

Table 1 – Segmentation results of the total burn wound

	F1 score	IoU	Precision	Recall	Accuracy	Loss
U-Net	0.8847	0.8178	0.9149	0.8872	0.9836	0.1152
PSPNet	0.8487	0.7667	0.8839	0.8570	0.9783	0.1519
Mask R-CNN	0.8913	0.8213	0.8869	0.8957	0.9841	0.1086
DeeplabV3+	0.8938	0.8288	0.9076	0.9006	0.9846	0.1063

535 images of female hands and 515 images of male hands. The border of hands without digits were labeled and saved as Dataset 2. Dataset 2 was split into training, validation, and testing set, and then was applied with three-fold validation as Dataset 1. Dataset 2 was input into DL models for training. Dataset 2 is more homogenous than the dataset of burn wounds. Four models had similar results on every evaluation metrics. DeeplabV3+ had a marginally better performance than

all with a DC of 0.9901 and an IoU of 0.9804 (Table 2). DeeplabV3+ was chosen as Model 2 for palm segmentation.

After that the burn wound and palm were correctly segmented, the %TBSA burned can be calculated with respect to the distance of taking pictures. The formula of the %TBSA was given by:

$$\%TBSA = \sum \frac{M_{burn} * D_{burn}^2 * 0.5}{M_{palm} * D_{palm}^2},$$

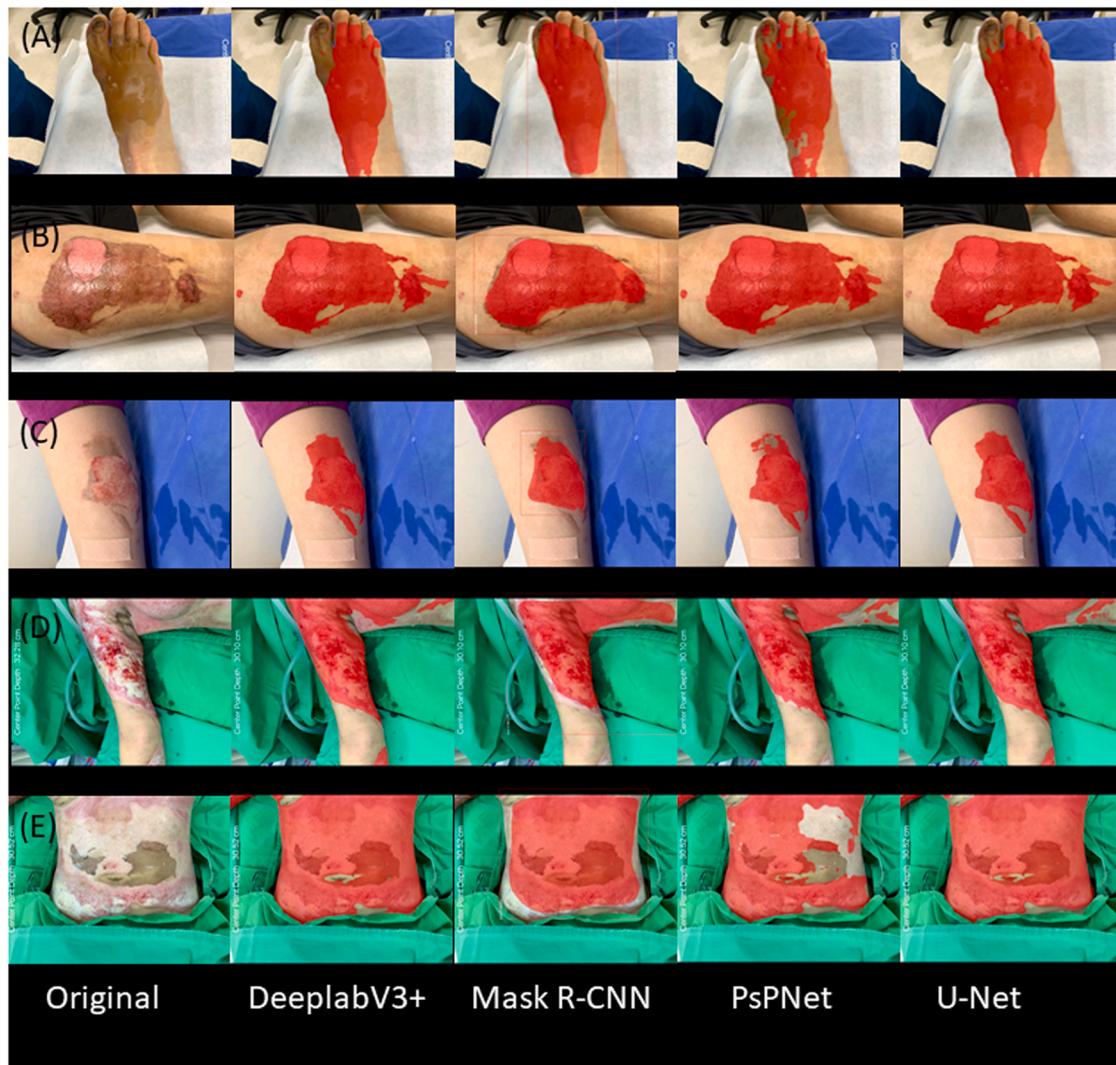


Fig. 3 – Total burn wound segmentation by different models (from left to right, Original, DeeplabV3+, Mask R-CNN, PSPNet, U-Net). A. intact bullae, degree unknown, B. partial bullae rupture with superficial burn, C. partial bullae rupture with superficial burn and deep burn, D. Scatter of superficial burn, deep burn including full thickness burn, E. Large and confluent of deep burn including full thickness burn.

Table 2 – Segmentation results of palm.

	F1 score	IoU	Precision	Recall	Accuracy	Loss
U-Net	0.9887	0.9778	0.9896	0.9879	0.9974	0.0215
PSPNet	0.9821	0.9700	0.9878	0.9818	0.9966	0.0180
Mask R-CNN	0.9858	0.9790	0.9821	0.9896	0.9975	0.0138
DeeplabV3+	0.9901	0.9804	0.9898	0.9903	0.9977	0.0099

where M_{burn} = the number of pixels of the masked burn area, M_{palm} = the number of pixels of the masked palm area (0.5 %TBSA), D_{burn} = the filming distance of the image of the patient's burn wound and D_{palm} = the filming distance of the image of the patient's hand. In our own practice, we use several tools to gauge the distance of taking photos, including tape measure and selfie sticks. In recent two years, the iPhone with lidar (12 or 13 pro) can render images with distances of taking pictures in every pixel, which can build 3D images at the same time.

3.3. Deep burn segmentation

Among the 4991 images, 1565 images of burn wounds have deep burns and were selected out. These images were processed with superpixel segmentation to define the borders of meaningful regions. Then, burn surgeons could select any regions of deep burns. The labeled images and original images were saved as the dataset of deep burn (Dataset 3). Dataset 3 was also split into training, validation, and testing set and then was applied with three-fold validation, which is similar to the total burn area segmentation. DeeplabV3+ performed with the best segmentation among all four architectures with a DC of 0.9018 and an IoU of 0.8124. Mask R-CNN had the second-best performance, followed by U-Net and PSPNet (Table 3). DeeplabV3+ was chosen as our Model 3 for deep burn segmentation. Fig. 4 shows that all the DL models had successfully segmented deep burn from whole burn wounds with various results in testing images.

3.4. Web-based server

According to the above results, DeeplabV3+ had better performances than other models for each task. We set up the web-based server comprising three DeeplabV3+. The %TBSA burned could be obtained from the segmentation results of Model 1 and Model 2 with the distance of taking the pictures.

Combining the %TBSA burned, age, body weight, and body height, the suggested volume for resuscitation was calculated by the modified Parkland formula [23]. Comparing the segmentation results of Model 1 and Model 3, the percentage of deep burn area could be estimated. Fig. 5 shows the workflow of the automatic burn diagnosis system. Fig. 6 shows a print screen of our web-based diagnosis system from the cell phone of our colleague.

4. Discussion

4.1. Wound segmentation and tissue classification

The application of ML in wound diagnosis had two major tasks: wound segmentation and tissue classification (or segmentation). Wound segmentation requires ML models to define the wound edges from normal skin, whereas tissue classification segments different types of tissue from the whole wound. The concepts are similar but require different approaches to obtain the best results. Veredas et al. utilized color histogram for wound segmentation and the Bayes rule for tissue classification with an accuracy of 88.08 % [24]. García-Zapirain et al. employed HSI color space for wound segmentation and a Linear combination of discrete Gaussians for tissue classification with a DC of 92 % [25]. Elmogy et al. used YCbCr color space for wound segmentation and RGB color space for tissue classification with a DC of 96 % [26]. Rajathi et al. employed simple gradient descent for wound segmentation and DL for tissue classification with an accuracy of 99.55 % [27].

The concept of different approaches to the two tasks was not proposed for acute burn wounds. In our study, we proposed boundary-based labeling for models to be utilized in total burn wound and palm segmentation, whereas region-based labeling and superpixel segmentation preprocessing for models to segment deep burn. The areas of deep burn

Table 3 – Segmentation results of the deep burn area.

	F1 score	IoU	Precision	Recall	Accuracy	Loss
U-Net	0.8949	0.7919	0.8950	0.8948	0.9854	0.1050
PSPNet	0.8870	0.7725	0.9125	0.8628	0.9798	0.1129
Mask R-CNN	0.9005	0.8095	0.9105	0.8907	0.9864	0.0994
DeeplabV3+	0.9018	0.8124	0.9015	0.9021	0.9888	0.0981

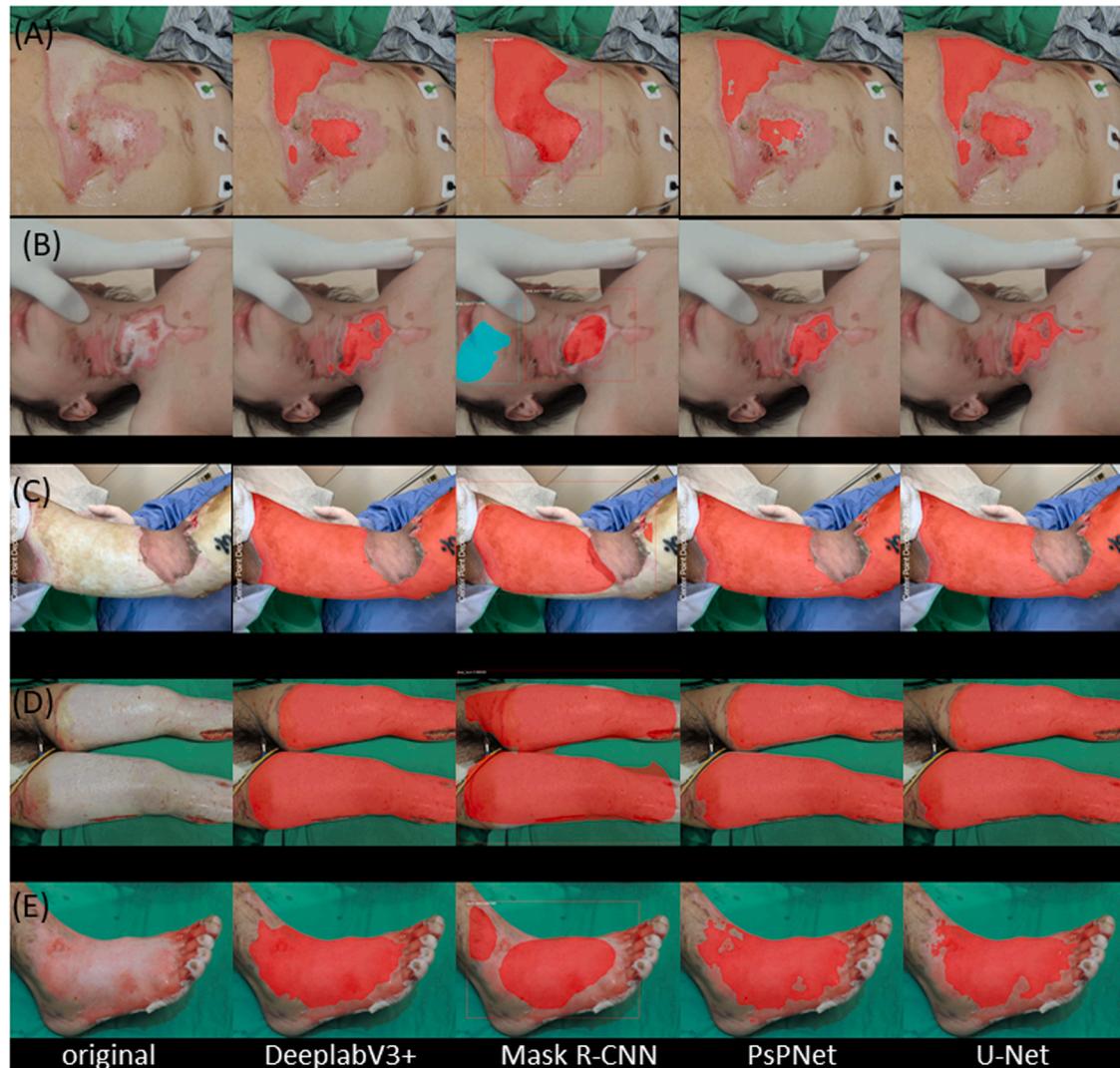


Fig. 4 – Deep burn segmentation by different models from left to right, Original, DeeplabV3+, Mask R-CNN, PSPNet, U-Net.

were viewed as a type of tissue in the whole burn wound. We used different approaches for total burn wound and deep burn segmentation because of the following reasons. First, the model for entire burn wound segmentation is directly related to the calculation of %TBSA burned. It is too important to require an independent model for this task. The model needs to segment wounds with various burn depths, which comprised wide ranges of colors and textures.

Secondly, the borders of deep burn with poor perfusion are hard to define. They are usually presented with a whitish to yellow, shiny, and leathery appearance. By inspecting images only, they may be misdiagnosed as normal skin. That makes them difficult to be accurately labeled even by burn experts. When we compared these areas with normal skin in different color spaces and histograms, both of their features

are highly overlapped (Fig. 7). To classify them from normal skin by thresholding methods is less likely, especially the dataset is big. To address this challenge, we applied superpixel segmentation to burn areas to pre-define meaningful areas. Then, we can easily choose small regions of deep burns from this whole burn area and label them as deep burns. Preprocessing with superpixel segmentation is the approach of Bayes' theorem.

Superpixel segmentation is first proposed by Ren et al. in 2003 [28]. The concept of superpixel segmentation is the grouping of adjacent pixels according to their features, including colors, intensities, and textures into small regions called superpixels. Several algorithms have been proposed in recent years, including Felzenszwalb's efficient graph in 2004, Quickshift in 2008, simple linear iterative clustering (SLIC) in

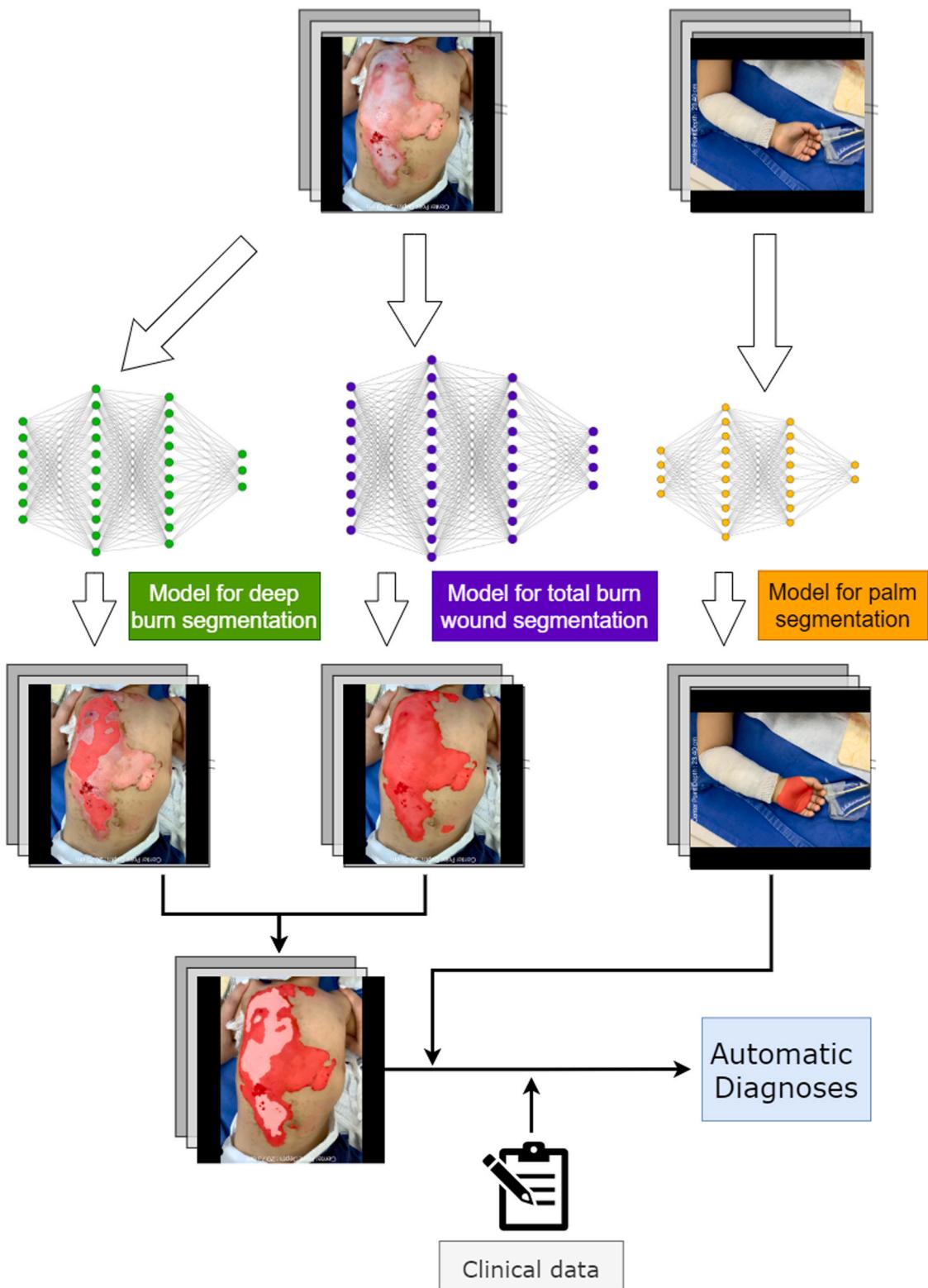


Fig. 5 – Workflow of the automatic burn diagnosis system.

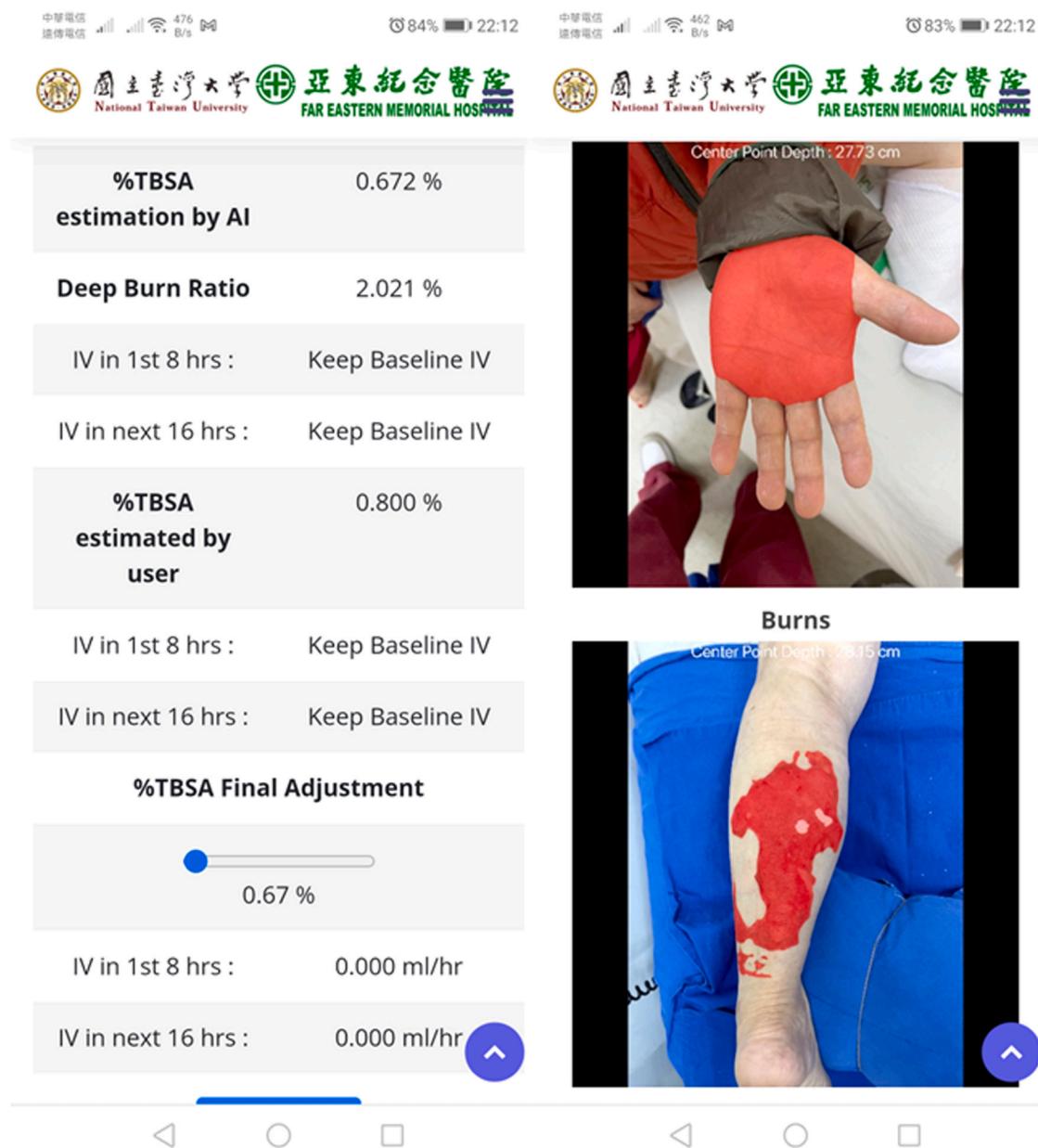


Fig. 6 – The print screen of the web-based automatic burn diagnosis system from the cell phone of our colleague.

2012. Linear spectral clustering (LSC), a 10-dimensional space version of SLIC, was proposed in 2015 to get a better boundary [29]. Superpixel-based edge detection (SBED), another modified version of SLIC, was proposed in 2020 to minimize the noise by a centroid updating approach [30]. Utilizing different segmentation algorithms would produce slightly different borders. We do not know which algorithms produce less labeling bias since they develop their ground truths. We applied SLIC to our data because it provides 1. good adhesion to

object boundaries, 2. regular shape and similar size, 3. fast and requiring less computing resource [31].

4.2. %TBSA burned

To calculate the fluid required for the resuscitation of burn patients, clinicians have to estimate the %TBSA of burn patients. There are three standard methods for the estimation of %TBSA. The most straightforward method is by the

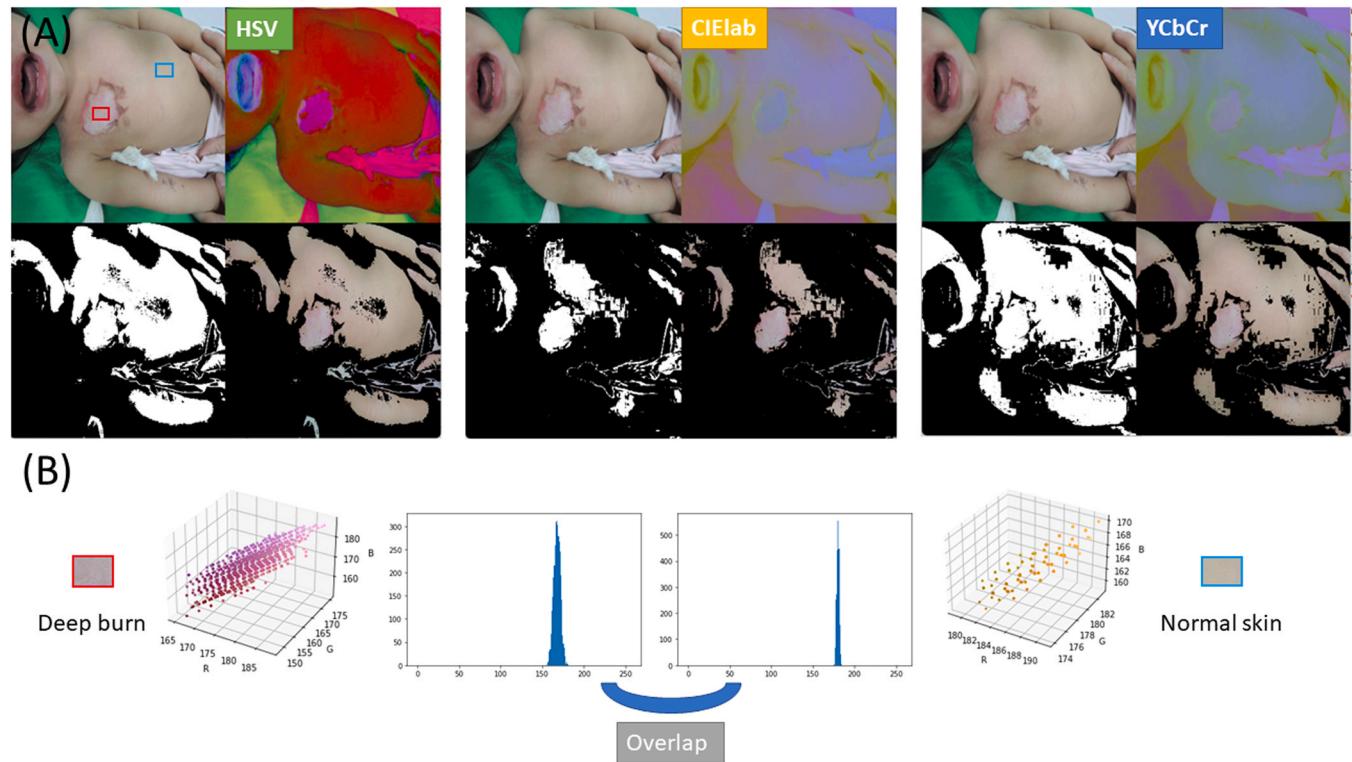


Fig. 7 – (A) It is hard to set a threshold of the deep burn areas without including normal skin in different color spaces. **(B)** The features of small regions of deep burn and normal skin are highly overlapped in the histogram, although the range of deep burn is more diverse.

definition of %TBSA burned. When a DL model accurately segments burn wounds, the absolute burn area (cm^2) can be calculated, combining the subject distance, focal length, and pixel numbers. Given the body height, body weight, age, and gender, the body surface area (BSA) can be estimated. The %TBSA burned is acquired by dividing these two numbers. The challenges for this method are that there are more than 25 formulae to estimate an adult BSA, and the accuracy of them was doubtful [32–34]. In addition to this, these formulae did not properly apply to children, which deferred the clinical application [35,36].

The Wallace Rule of Nines was the most well-known tool for the %TBSA estimation [11]. By dividing the body parts into 11 parts of 9 %TBSA, we can estimate the proportion of injured skin in different positions and add up to a total %TBSA burned. In recent studies, most body parts did not follow the Rule of nine. For example, an upper extremity is around 7 %TBSA, a lower extremity is more than 20 %TBSA, and the percentage was related to BMI [37–41]. Moreover, Choi et al. reported a study with the use of 3D camera that the %TBSA of body parts is largely related to body contour in a person with a similar BMI [42]. The body contour is hard to evaluate, suggesting that the Rule of nine is not a precise method for estimating the %TBSA burned. The same challenge is encountered when calculating the %TBSA burned of children even with the Lund-Browder chart.

We adopted the Rule of palm to convert the burn segmentation into %TBSA burned. The palm method is another tool usually for the estimation of %TBSA burned with a

smaller area. The palm is the ruler of our body. Recently studies showed that the area of palm excluding digits is approximately 0.5 %TBSA [43–45]. The relation of ratio to genders, races, BW, BH and body contour is neglectable. In addition, the palm method can also be applied to children [46,47]. The only clinical limitation is that when the Rule of palm was applied to a burn area greater than 20 %TBSA, it will cause significant deviations in different raters. Fortunately, AI models calculate burn areas and palm areas in a pixel-to-pixel manner. They do not make errors as human inspection.

4.3. Limitations

We excluded the images of burn wounds occurring within 48 h of injury. The DL models had various results when inputting images of immediate burn injury with only erythema but not bullae formation. This is often the case of first-degree burn or some second-degree burn without bullae formation, such as sunburn. Lucky, the first-degree burn was not required fluid resuscitation for the burn patient. In addition, the dataset was acquired in a medical center in Northern Taiwan, and 98.3 % of the patient population was Chinese. Our models do not train with images from darker or lighter skin tones. We, therefore cannot guarantee promising results when our DL models applies to burn victims with other skin tones.

The estimation of %TBSA is calculated from the segmentation result of burn wounds and palm. When the palms of patients were also burnt, the segmentation of palm might also

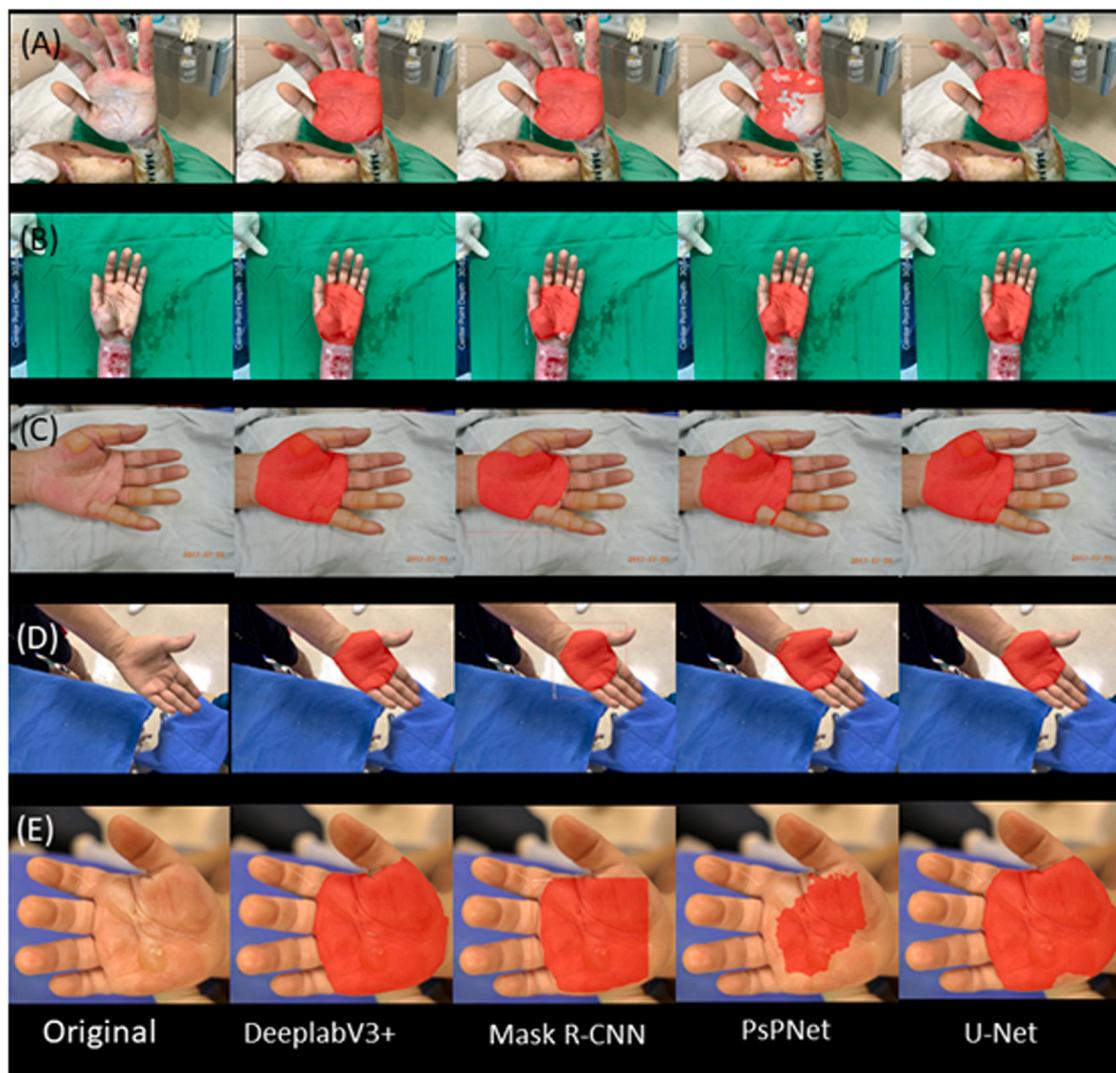


Fig. 8 – Segmentation of burnt palm. A. full thickness burn, B. Adjacent forearm deep burn, C. bullae formation on palm, D. hand with digits closed, E. Child burned hand.

be affected. Fig. 8 shows the segmentation results on burnt palm. DeepLabV3+ and U-Net showed satisfactory segmentation, whereas the Mask R-CNN and PSPNet were excluded the areas with bullae formation. To overcome this problem, we need more images of burn palm for training and testing.

Though we could take photos parallel to the wound bed, the wound could not be accurately depicted by a 2-dimensional (2D) image since our body surface was not on a flat plane. Thanks to the development of LiDAR camera on cell phones, 3D laser scanning was more popular and easily available than in the past. We are working on the DL models training and testing from 3D image datasets.

5. Conclusion

In this study, we proposed boundary-based labeling for the total burn wounds and palm dataset, whereas region-based labeling for the dataset deep burn area. Multiple deep learning models were trained and tested from these three datasets with promising results. Combining the segmentation of total burn wounds and palms, the %TBSA burned could be calculated. The segmentation of deep burn indicated areas possibly requiring more time to heal. We set up a web-based server including the outputs from multiple DL models to provide automatic burn wound diagnoses.

Ethical approval

This study was approved by the research ethics review committee of the Far Eastern Memorial Hospital (number 109037-F).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We appreciate our colleagues in the Division of plastic surgery of the Far Eastern Memorial Hospital to collect the images of burn wounds and hands. We also thank our colleagues in the Graduate Institute of Biomedical Electronics & Bioinformatics of National Taiwan University for the collection of images of hands. the Grant is “Innovation Project of the Far Eastern Memorial Hospital”, “Taiwan”, Grant No. PI20200002.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.burns.2022.07.006](https://doi.org/10.1016/j.burns.2022.07.006).

REFERENCES

- [1] Smolle C, Cambiaso-Daniel J, Forbes AA, Wurzer P, Hundeshagen G, Branski LK, et al. Recent trends in burn epidemiology worldwide: a systematic review. *Burns* 2017;43:249–57.
- [2] Harish V, Raymond AP, Issler AC, Lajevardi SS, Chang L-Y, Maitz PK, et al. Accuracy of burn size estimation in patients transferred to adult Burn Units in Sydney, Australia: an audit of 698 patients. *Burns* 2015;41:91–9.
- [3] Baartmans MG, Van Baar M, Boxma H, Dokter J, Tibboel D, Nieuwenhuis MK. Accuracy of burn size assessment prior to arrival in Dutch burn centres and its consequences in children: a nationwide evaluation. *Injury* 2012;43:1451–6.
- [4] Sadidine H, D'Asta F, Moiemen N, Wilson Y. Does overestimation of burn size in children requiring fluid resuscitation cause any harm? *J Burn Care Res* 2017;38:e546–51.
- [5] Parviz D, Kamolz L-P, Giretzlehner M, Haller HL, Trop M, Selig H, et al. The potential impact of wrong TBSA estimations on fluid resuscitation in patients suffering from burns: things to keep in mind. *Burns* 2014;40:241–5.
- [6] Zahia S, Zapirain MBG, Sevillano X, González A, Kim PJ, Elmaghriby A. Pressure injury image analysis with machine learning techniques: a systematic review on previous and possible future methods. *Artif Intell Med* 2020;102:101742.
- [7] Marijanović D, Filko D. A systematic overview of recent methods for non-contact chronic wound analysis. *Appl Sci* 2020;10:7613.
- [8] Anisuzzaman D., Wang C., Rostami B., Gopalakrishnan S., Niezgoda J., Yu Z. Image-Based Artificial Intelligence in Wound Assessment: A Systematic Review. *Advances in Wound Care*. 2021.
- [9] Suvarna M, Nirajan U. Classification methods of skin burn images. *AIRCC'S Int J Comput Sci Inf Technol* 2013;5:109–18.
- [10] Tran H, Le T, Le T, Nguyen T. Burn image classification using one-class support vector machine. *ICCASA*. Springer; 2015. p. 233–42.
- [11] Kuan P, Chua S, Safawi E, Wang H, Tiong W. A comparative study of the classification of skin burn depth in human. *J Telecommun Electron Comput Eng (JTEC)* 2017;9:15–23.
- [12] Wang Y, Ke Z, He Z, Chen X, Zhang Y, Xie P, et al. Real-time burn depth assessment using artificial networks: a large-scale, multicentre study. *Burns* 2020;46:1829–38.
- [13] Yadav D, Sharma A, Singh M, Goyal A. Feature extraction based machine learning for human burn diagnosis from burn images. *IEEE J Transl Eng Health Med* 2019;7:1–7.
- [14] Liu NT, Salinas J. Machine learning in burn care and research: a systematic review of the literature. *Burns* 2015;41:1636–41.
- [15] Huang S, Dang J, Scheckter CC, Yenikomshian HA, Gillenwater J. A systematic review of machine learning and automation in burn wound evaluation: a promising but developing frontier. *Burns* 2021;47:1691–704.
- [16] Chang CW, Lai F, Christian M, Chen YC, Hsu C, Chen YS, et al. Deep learning-assisted burn wound diagnosis: diagnostic model development study. *JMIR Med Inform* 2021;9:e22798.
- [17] Achanta R, Shaji A, Smith K., Lucchi A., Fua P., Süsstrunk S. Slic superpixels. 2010.
- [18] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer; 2015. p. 234–41.
- [19] Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proc Eur Conf Comput Vis (ECCV)* 2018:801–18.
- [20] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. *Proc IEEE Conf Comput Vis Pattern Recognit* 2017:2881–90.
- [21] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. *Proc IEEE Int Conf Comput Vis* 2017:2961–9.
- [22] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vis Pattern Recognit* 2016:770–8.
- [23] Henry S. ATLS 10th edition offers new insights into managing trauma patients. *Bull Am Coll Surg* 2018.
- [24] Veredas FJ, Luque-Baena RM, Martín-Santos FJ, Morilla-Herrera JC, Morente L. Wound image evaluation with machine learning. *Neurocomputing* 2015;164:112–22.
- [25] García-Zapirain B, Elmogy M, El-Baz A, Elmaghriby AS. Classification of pressure ulcer tissues with 3D convolutional neural network. *Med Biol Eng Comput* 2018;56:2245–58.
- [26] Elmogy M, García-Zapirain B, Elmaghriby AS, El-Baz A. An automated classification framework for pressure ulcer tissues based on 3d convolutional neural network. *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE; 2018. p. 2356–61.
- [27] Rajathi V, Bhavani R, Wiselin Jiji G. Varicose ulcer (C6) wound image tissue classification using multidimensional convolutional neural networks. *Imaging Sci J* 2019;67:374–84.
- [28] Ren X, Malik J. Learning a classification model for segmentation. *Computer Vision IEEE International Conference*. IEEE Computer Society; 2003. p. 10.
- [29] Li Z, Chen J. Superpixel segmentation using linear spectral clustering. *Proc IEEE Conf Comput Vis Pattern Recognit* 2015:1356–63.
- [30] Zhang H, Wu C, Zhang L, Zheng H. A novel centroid update approach for clustering-based superpixel methods and superpixel-based edge detection. *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE; 2020. p. 693–7.

- [31] Zhang J, Aviles-Rivero AI, Heydecker D, Zhuang X, Chan R, Schönlieb C-B. Dynamic spectral residual superpixels. *Pattern Recognit* 2021;112:107705.
- [32] Shuter B, Aslani A. Body surface area: Du bois and Du bois revisited. *Eur J Appl Physiol* 2000;82:250–4.
- [33] Redlarski G, Palkowski A, Krawczuk M. Body surface area formulae: an alarming ambiguity. *Sci Rep* 2016;6:1–8.
- [34] Yu C-Y, Lin C-H, Yang Y-H. Human body surface area database and estimation formula. *Burns* 2010;36:616–29.
- [35] Amit LM, Song Y-W. Formulae evaluation for estimating body surface area of Korean children. *J UOEH* 2018;40:19–32.
- [36] Sigurdsson TS, Lindberg L. Six commonly used empirical body surface area formulas disagreed in young children undergoing corrective heart surgery. *Acta Paediatr* 2020;109:1838–46.
- [37] Neaman KC, Andres LA, McClure AM, Burton ME, Kemmeter PR, Ford RD. A new method for estimation of involved BSAs for obese and normal-weight patients with burn injury. *J Burn Care Res* 2011;32:421–8.
- [38] Williams RY, Wohlgemuth SD. Does the "rule of nines" apply to morbidly obese burn victims? *J Burn Care Res* 2013;34:447–52.
- [39] Borhani-Khomani K, Partoft S, Holmgaard R. Assessment of burn size in obese adults; a literature review. *J Plast Surg hand Surg* 2017;51:375–80.
- [40] Moore R.A., Waheed A., Burns B. Rule of nines. StatPearls [Internet]. 2021.
- [41] Livingston EH, Lee S. Percentage of burned body surface area determination in obese and nonobese patients. *J Surg Res* 2000;91:106–10.
- [42] Choi J., Patil A., Vendrow E., Touponse G., Aboukhater L., Forrester J.D., et al. Practical Computer Vision Application to Compute Total Body Surface Area Burn: Reappraising a Fundamental Burn Injury Formula in the Modern Era. *JAMA surgery*. 2021.
- [43] Rhodes J, Clay C, Phillips M. The surface area of the hand and the palm for estimating percentage of total body surface area: results of a meta-analysis. *Br J Dermatol* 2013;169:76–84.
- [44] Thom D. Appraising current methods for preclinical calculation of burn size—a pre-hospital perspective. *Burns* 2017;43:127–36.
- [45] Jose RM, Roy DK, Wright PK, Erdmann M. Hand surface area—do racial differences exist? *Burns* 2006;32:216–7.
- [46] Cox S, Kriho K, De Klerk S, van Dijk M, Rode H. Total body and hand surface area: measurements, calculations, and comparisons in ethnically diverse children in South Africa. *Burns* 2017;43:1567–74.
- [47] Choi H, Park MS, Lee H-M. Hand surface area as a percentage of body surface area in Asian children: a pilot study. *Burns* 2011;37:1062–6.