**Deep Learning Based on Superpixel Segmentation Assisted Labeling for Automatic Pressure Ulcer Diagnosis**

Che-Wei Chang1,2, Mesakh Christian3, Dun-Hao Chang2,4 ,Wei-Jen Chen1, Dun-Hao Chang2,4, Yo-Shen Chen2, Yen-Che Yu2, Tyng-luen Roan2, Feipei Lai1,3

1 Graduate Institute of Biomedical Electronics & Bioinformatics, National Taiwan University, Taipei, Taiwan.

2 Division of Plastic and Reconstructive Surgery, Department of Surgery, Far Eastern Memorial Hospital, New Taipei, Taiwan.

3 Department of Computer Science & Information Engineering, National Taiwan University, Taipei, Taiwan.

4 Department of Information Management, Yuan Ze University, Taiwan.

Corresponding Author: Feipei Lai

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

Phone: +886-2-3366-4924

Email: flai@ntu.edu.tw

Introduction

A pressure ulcer or pressure sore is a pressure-induced injury upon skin inside out from bony eminence. These injuries appear most commonly on the sacral area, followed by the heel and occipital area. The mechanism is to apply on the skin more than end capillary pressure, which is 20-30 mmHg. Since no blood can flow in, the tissues were gradually necrosis. However, simply applying that level of pressure to the skin may not necessarily result in a pressure ulcer. It was found that pres­sure should persist for more than 2 hours to cause irreversible ischemic damage1. Patients who developed pressure ulcers have different etiologies to prevent their protective mechanism of the spontaneous position change. It could be medical conditions, such as stroke, spine cord injury, shock, malnutrition, antidepressant, and catheter use2. Major surgery or trauma are also contributing factors.

In the United States, pressure ulcers affected around 3 million adults annually3. In Europe, the prevalence rates ranged from 8.3% to 23% among different countries4. The prevalence rate is higher in long-term care facilities than in hospitals2,5. Although this tissue had got attention from the public, the prevalence of pressure ulcers has remained unchanged in this decade 3,6. As the life span of humans increased, the affected populations are expected to rise. Not only the increasing affected population is problems. Patients with pressure ulcer also had difficulty attending medical care because of their underlying diseases. This condition has further deteriorated by the Covid-19 pandemic.

Machine learning (ML) has many applications in the field of medicine, such as in drug development and disease diagnosis7-10. Automatic wound diagnosis based on ML can provide an objective assessment of wound statuses, such as wound size and healing stages. They filled the unmet need by sharing the medical provider's loading and providing timely suggestions for patients when medical care was not easily accessed.

Prior work

There were two major tasks for ML to do for pressure ulcers. The first is wound segmentation; the other is tissue classification11-13. Wound segmentation is to separate the wound area from the non-wound area. There were three standard methods to achieve this result：threshold, edge-base and region-based segmentation. Threshold segmentations applied algorithms to separate pixels of the wound from pixels of normal skin, which are more homogenous in intensity14,15. Edge-based segmentation initializes an approximate shape as a border, such as elliptic or circle. Each iteration, the border is moving to minimize the function energy. The border final will fit into the edge of wound16,17. Region-based segmentation is to select a small region on the wound and gradually recruits adjacent pixels with similar color energy18,19. The above methods can be further combinated to acquire better results20,21. After successful wound segmentation, the size of a wound can be calculated by adding focal length, pixel number, and distance to wound. Although the size of a wound is the most direct information about wound healing, it is less clinical usage than the tissues composition of a wound. A pressure ulcer usually took weeks to months to be observed decreasing in size.

Tissue classification, as anthor assessment of wound healing becomes the second important goal for ML models. Before performing tissue classification, images of wounds usually be converted from RGB to other color spaces. These three primary colors (RGB) are correlated to each other. The common format of color spaces were HSV, follow by CIELab and YCbCr22. Usually, the luminance components such as Y channel or L channel were excluded to decrease the light effect in various clinical situations. The next step is to calculate the probability map of pixels of different tissues. ML methods, such as k-means, SVM, RF or Bayesian were used to learn the threshold and classify tissues.

There were articles that addressed both tasks of wound segmentation and tissue classification. Although two taskd seemed to be similar, some authors used different preprocessed methods to get better results. For example, García-Zapirain et al converted the images to HSI color space for wound segmentation and used linear combinations of discrete Gaussians for tissue classification23. Elmogy et al suggested YCbCr images for wound segmentation, while RGB images for tissue classification24. Veredas et al used color histograms for wound segmentation and Bayes rule for tissue classification24. Rajathi et al used deep learninng for tissue classification but a simple gradient descent for wound segmentation12,13,25. These studies demonstrated important skills in data preprocessing and features extraction. The problem is that some approaches, even with good results, can only be applied to their own datasets. That will diminish their clinical application. Secondly, most studies of ML in wound diagnosis used the datasets of chronic wounds not focusing on pressure ulcer. There was not much data showing that models training from mix etiology of ulcer can equally be used for pressure ulcer.

Deep learning (DL) is a field for ML with multiple convolution layers. The structures of DL models have the abilities for embracing more data. These data could be intendedly trimmed out for ML to get better results. For example, the lumination channels of color spaces (L channel of CIELab and Y channel of YCbCr) were usually excluded for ML model to decrease the effect under different light conditions. For DL models, instead, lumination channels could be input with other channels in order to comprehend various light conditions. DL models, on the other hand, require a large amount of labeled data to achieve good results and prevent overfitting. How to label data systemically become very critical.

Very few articles addressed the detail of labeling images but stated as “ labeled by medical experts”. In our study, we reveal our step-by-step approach of data labeling and the revision we made during labeling. This method could be suitable for any images of pressure ulcers, not just for our datasets. In the second part, we input more than 2,800 images of pressure ulcers to some most popular and influential DL models. We combined the results from different models to provide automatic wound diagnosis.

Method

1-1 Evaluation metrics

Before we described our method of data preprocessing, we need to define the evaluation metrics we used. There were five metrics：

Dice Coefficient (F1 score), intersection over union (IoU), Precision, Recall and Accuracy. The Dice Coefficient (DC) and IoU are two common metrics to assess segmentation performance, whereas precision, recall and accuracy are the metrics of assessing classification results. These metrics are defined using true positive (TP), false positive (FP), true negative (TN) and false negative (FN) predictions for any input image considered as follows:

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

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:

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

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

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

1-2 Image acquisition

This study was approved by the research ethics review committee of Far Eastern Hospital (No. 109145-E). We reviewed the medical records of patients from Jan. 2016 to Dec. 2020 with ICD9 codes from 707.00 to 707.10. These codes present a pressure ulcer and its location, such as ulcer, pressure, sacrum 707.03. We collected the images of pressure ulcers from progress notes, operation notes, discharge notes, and ER electric charts. All images were 3600\*2700 pixels and saved as jpg files. These images were assigned random numbers for de-identification and presented to plastic surgeons for labeling.

An image was initially excluded if the wound was dressed with ointment, wound was partial covered with dressing, active bleeding in wound or coating with hematoma, wound presented with obvious, images were out of focus and under poor light conditions.

2-1 Boundary base labeling

Our initial method to label a pressure ulcer is similar to pizza-making (Figure 1.) The first step is to make the round cake and then putting on all the toppings, such as sausage, mushroom, and olive. So as to make a pizza, the border of the whole pressure ulcer was first drawn as ulceration. The boundaries of different tissues, such as granulation, eschar, slough inside ulceration, were labeled orderly. The label software is *Labelme* on python 3.6 with a mouse or pen tablet. All images were co-labeled by two of five plastic surgeons to yield a single consensus labeled image. Any image was excluded when two surgeons had no agreement on the labeled result.

***Inside ulceration***

Granulation is pink to red, watery tissue, and it is the healthy tissue in wounds. The presentation of granulation suggested the wound was healing. During the process of ulceration healing, the ratio of granulation to ulceration would increase. Eschar is black, dry and nonviable skin and subcutaneous tissue. They usually have clear margins from other tissues. That made them most easy to be labeled. Slough has the most various appearance and could be whitish, yellow or brown. They may comprise the partial dead tissues, such as dermis or fascia infiltrating with inflammation cells. Slough could also be complete necrotic tissues with exudate coating.

Some pressure ulcer is a 3D basin-like structure. There could be some part of ulceration under the shadow of edges on the image. These pixels were hard to define which types of tissue they were even by experienced plastic surgeons. The relation of pressure ulcer and different tissues can be described as：Ulceration = (Granulation ∩ Eschar∩ Slough ∩ Undefined). A pressure ulcer may not have all types of tissues.

***Outside ulceration***

Re-epithelialization (Re-ep) is a term to describe the resurfacing of a wound with new epithelium 26. It is the essential sign of wound healing outside the ulceration. The process includes the formation of wound bed matrix, the migration of keratinocytes from edges, the reformation of basement membrane and stratification and differentiation of the neo-epithelium. It starts approximately 24 hours after injury and continues all wound healing process. The re-ep skin is always pink and will turn white when there are skin maceration or hyperkeratosis formation.

|  |
| --- |
|  |
| Figure 1. Pizza-making labeling method. The border of ulceration was first drawn, followed by borders of different tissues inside ulceration. Re-ep was drawn if the wound showed. |

2-2 Different combination of classes

The above label methods create a detailed annotation with different types of tissues inside and outside of ulceration. However, the informative label data generate a challenge of repeat annotation on the same pixels. Different classes need to be trained together, and at least two modes were required. There were some possible combinations of different classes into two groups. To decide which combination of groups not only depends on final segmentation results, but which tissue is more important for diagnosis. Granulation and re-ep were the two most important tissue to detect wound healing. However, re-ep had fewer total numbers of pixels compared with other tissues among all labeled images. The shape of re-ep is usually ring or crescent, making it difficult to predict accurately.

We tried two practical class combinations to test which models had better segmentation performance for re-ep. One is that ulceration was trained in the first model (model 1) and all types of tissues included re-ep, granulation, slough, and eschar were trained in second model (model 2). The other combination is that re-ep and ulceration were trained in the first model (model 3) and tissues inside ulceration (granulation, slough, and eschar) were trained in the second model (model 4).

We collected 755 labeled images of pressure ulcer as dataset. The dataset was split at a ratio of 7:2:1 for training, validation and testing. We selected the class combination as mentioned and input U-Net (with ResNet101) for training. The comparison was made between model 2 (re-ep with other tissues) and model 3 (re-ep with ulceration). Figure 2 showed confusion matrix of model 2 and model 3. The first row of both matrix was total pixels of re-ep and is the same in both. When re-ep was trained with ulceration, model 3 resulted in more true positive for re-ep. Model 3 had better performance (precision: 0.9974, recall: 0.6291, DC: 0.7716) to detect re-ep than model 2 (precision: 0.9981, recall: 0.5685, DC: 0.7244). Based on this result, re-ep and ulceration were trained in a model, and other tissues inside ulceration were trained in the other model.

|  |
| --- |
|  |
| Figure 2. The Confusion Matrix of model 2 and model 3. In the first matrix, re-ep was trained with granulation, slough, and eschar (Others) in model2. In the second matrix, re-ep was trained with ulceration only in model 3. |

2-3 Region-based labeling

Pizza-making is a straight-forward, boundary base label method. We found it challenging for plastic surgeons to acquire consensus when they labeled tissues (granulation, slough, eschar) inside ulceration. The composition of tissues on wound bed is a subtle transition. It is sometimes hard to draw boundaries between different tissues. It was also difficult for two plastic surgeons to have consensus on these borders.

We introduced superpixel segmentation to define meaningful regions among tissues. We tried to reduce the tedious hand draw methods for plastic surgeons and hoped to improve the inter-rater agreement. There are different algorithms to achieve this goal. Simple linear iterative clustering (SLIC) is the most popular algorithm27. It comprises in five dimensions (l,a,b,x,y) ,where(l,a,b) are color vectors in CIELAB color space and (x,y) are pixel position. A desired number of equally-size superpixel K is defined, and the grid interval is S. M, compactness is a variable that controls the weight of spatial term. The center of K superpixel Ck is ( lk,ak,bk,xk,yk). Each pixel Ii can be present as ( li,ai,bi,xi,yi). Euclidean distances D is as followed：

Dcolor=

Dxy=

D= Dcolor+Dxy

The process of associating pixels with the nearest cluster center computed the boundary repeatedly until converge.

To test whether the dataset preprocessed by SLIC had a positive effect on the models to segment different tissues (granulation, slough and eschar). The same 755 images which had been labeled with *labelme* were used. All images were resized to 1000\*750 pixels but filled with a black background to 1000\*1000. The downsized images can be processed with SLIC in laptops without powerful GPU. Images were processed by SLIC with K=800 to 1000, m=10, to form superpixels (regions). Figure 3 showed the different number of superpixels (K). After preprocessing with SLIC, images were presented to the same two plastic surgeons for sorting these regions to their right classes.

Dataset labeled with *labelme* and dataset labeled with SLIC preprocessing were both input to U-Net with the same finetune parameters to yield model 5 and model 6.

The performance of tissue classification from two models was showed in Table 1. The model trained by the down-size images with SLIC preprocessing (Model 6) had better results in all evaluation metrics. The results did not necessarily mean that model 6 has better segmentation performance in real-wound images since the two datasets' ground truth was different. We used the other 75 new testing images of pressure ulcers to make a direct comparison. In Figure 3, model 6 still had more satisfactory results and predicted fewer undefined pixels than model 5.

|  |
| --- |
|  |
| Figure 3. Different number of superpixels (K). (a) Resize image. (b) K=500. (c) K=1000. (d) K=2000. |

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1. Tissue classification from models with different label methods | | | |
|  | Model 5 (*Labelme*) | Model 6 (SLIC) |
| IoU score | 0.3921 | 0.4635 |
| F1 score | 0.4336 | 0.5129 |
| Precision | 0.5036 | 0.5622 |
| Recall | 0.7979 | 0.8233 |
| Accuracy | 0.9897 | 0.9927 |
| Loss | 0.5664 | 0.4870 | |

|  |
| --- |
|  |
| Figure 3. Segmentation results of different tissues\* on testing image. First row: original images of pressure ulcer. Second row: result of model 6. Third row: result of model 5. (\*Granulation: red; Slough: yellow; Eschar: blue) |

Based on the above results, we revised the boundary base labeling for all classes (pizza-making method) to boundary base labeling for re-ep and ulceration; regional base labeling (SLIC preprocessing) for granulation, slough and eschar. One dataset (boundary-based) was used to train models for wound segmentation. The other dataset (region-based) was applied to trained models for tissue classification, which came from different labeled methods (Figure 4).

|  |
| --- |
|  |
| Figure 4. On the left side of the workflow, it showed different label methods for wound segmentation and tissue classification. On the right side, two models predict the input images, and combine the clinic data to provide automatic diagnosis. |

3. Deep learning models

The two datasets were input to five well-developing DL models to compare their performance on two tasks. All models were combined with the ResNet101 as their encoder. 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 augmentations of images we used were rotations, shifts, scale, gaussian blur, and contrast normalization.

U-Net

U-Net proposed by Ronneberger et al is the most popular semantic segmentation model in the medical field28. It involved a series encoder then a decoder process. When visualized the architecture, it is similar to the letter U and has its name. Many models were derived from U-Net structures, such as FCNet, SegNet, DeconNet, V-Net, U-Net++. The standard Dice loss was chosen as the loss function. The formula is given by:

The *ϵ* term is used to avoid the issue of dividing by 0 when precision and recall are empty.

DeeplabV3

DeeplabV3 was proposed by chen et al29. Multiple convolution layers were the structure to get high dimension features, but they loss the details when layers go deeper. Atrous convolution is the solution to solve this problem. Deeplab V3 composed multiple scales of atrous convolutions (spatial pyramid pooling) to preserve the details and acquire high dimension features. The standard Dice loss was chosen as the loss function of U-Net.

Pyramid Scene Parsing (PsPnet)

PsPNet was proposed by [H Zhao](https://scholar.google.com.tw/citations?user=4uE10I0AAAAJ&hl=zh-TW&oi=sra) et al30. After through CNN layers, such as ResNet pretrained with dilated network for feature extraction, the model adopts whole and multiple sub-regional features from large size to smaller size, which calls pyramid pooling module. Then, the model up-sampled all the sub-regional then concatenated the original feature map to form a single layer. The standard Dice loss was chosen as the loss function of U-Net.

Feature Pyramid Network (FPN)

FPN, proposed by Lin et al. had a modified convolution pathway31. It not only has the high level of features of convolutional layers but also combine the features from different levels of up-sampling layers. It works faster than featurized image pyramid which had features from different scales of images. It acquired more detail than single feature map. FPN is the hybrid and modified model of featurized image pyramid and single feature map. The standard Dice loss was chosen as the loss function as U-Net.

Mask R-CNN

Mask R-CNN was proposed by team of facebook32. It was the advanced model from Faster R-CNN and a tool of instance segmentation. Mask R-CNN uses a multi-task loss function given by L = Lclass + Lbox +Lmask (Figure 1). The Lclass component contains the RPN class loss (failure of the Region Proposal Network to separate object prediction from background. The Lbox means failure of object localization or bounding by RPN. The last component Lmask loss constitutes the failure of Mask R-CNN object mask segmentation.

Result

Wound segmentation & Tissue classification

The first dataset contented 2893 images labeled with boundary-based method and used for wound segmentation and re-ep detection. The second dataset included 2836 images labeled with SLIC preprocessing, region-based method and used for tissue classification inside ulceration. The first dataset had more labeled images than the second, because it is challenging to have consensus when labeling tissues.

The performance of wound segmentation and re-ep detection from five models were showed in Table 2. Five models had satisfactory results on this task, where DeeplabV3 had an edge against other algorithms. In table 3, DeeplabV3 also had a leading performance of tissue classification. We applied two DeeplabV3 as our main models for automatic diagnosis (Figure 4).

However, it's worth noting that Mask R-CNN is a powerful model but had worse results on tissue classification. Mask R-CNN outputs instance segmentation by defining the bounding box for objects then posing segmentation masks. If the bounding box could not define correctly, the mask would be wrong segmented. In many ulcerations, granulation tissues encased (Intermingle?) the slough tissues or versus, which posed wrong range of bounding boxes as well as masks (supplement1). The actual wound pixels were underestimated and recall was affected most. That will not cause many issues for model outputting semantic segmentation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 2. Performance of wound segmentation + re-epithelialization detection | | | | | | |
|  | IoU score | F1 score | Precision | Recall | Accuracy | Loss |
| U-Net | 0.9745 | 0.9867 | 0.9868 | 0.9867 | 0.9911 | 0.0132 |
| DeepLabV3 | 0.9782 | 0.9887 | 0.9888 | 0.9887 | 0.9925 | 0.0112 |
| PsPnet | 0.9211 | 0.9404 | 0.9317 | 0.9494 | 0.9780 | 0.0595 |
| Mask R-CNN |  |  |  |  |  |  |
| FPN | 0.8196 | 0.8939 | 0.8556 | 0.9492 | 0.9346 | 0.1061 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 3. Performance of tissue classification | | | | | | |
|  | IoU score | F1 score | Precision | Recall | Accuracy | Loss |
| U-Net | 0.9830 | 0.9905 | 0.9913 | 0.9897 | 0.9899 | 0.0094 |
| DeepLabV3 | 0.9834 | 0.9915 | 0.9915 | 0.9915 | 0.9957 | 0.0084 |
| PsPnet | 0.9390 | 0.9753 | 0.9614 | 0.9897 | 0.9899 | 0.0094 |
| Mask R-CNN | 0.7886 | 0.8480 | 0.9191 | 0.7871 | 0.8903 | 0.2032 |
| FPN |  |  |  |  |  |  |

|  |
| --- |
|  |
| Supplement 1. (a) The granulation tissues intermingled with slough and eschar. (b)Some Eschar was encased by slough and slough was encased by granulation. (c) and (d) showed poor segmentation results due to above phenomenon. |

Four major outputs of automatic diagnosis

Combining segmentation results and clinical data, the algorithms provided four major outputs, from the sign of wound healing：detection of re-ep, the progress of wound healing：ratio of granulation to ulceration area and trend of ratio, the estimated size of the wound, and whether surgical debridement is required. The detection of re-ep came from DeeplabV3 to do wound segmentation and also segment the area of re-ep adjacent to ulceration. The ratio of granulation to ulceration area needs results from both DeeplabV3 of wound segmentation and tissue classification. The increasing covering of granulation tissues provided the strength of wound contracture and indicated the wound was healing. The estimated size of the wound combined the wound segmentation with parameters extracting from input images. Assuming that the image was taken at distance D. The area of a wound can be given by：

As the fourth output, Whether the pressure ulcer requires surgical debridement depends on the presentation of infection or/and necrotic tissue33. The decision tree consists of three checkboxes in Figure 5. The necrotic tissue can be assessed from the results of our DL models (blue box). The infection (gray box) and potential infection of necrotic tissue (orange box) need the input of clinic data and other wound conditions to evaluate. We set up above all algorithms and models in web servers. Medical staff and caregivers can upload the images of pressure ulcers and input clinical information from their phones or laptops to obtain the results of the automatic diagnosis.

|  |
| --- |
|  |
| Figure 5. Decision tree of surgical debridement. Combining the DL segmentation result(blue) with clinical data(grey) and other wound conditions (grey and orange). |

Discussion

Compared with traditional ML for wound segmentation, DL does not need to transfer images to other color spaces, not setting a threshold to define a class. They can learn from potential important but unfavorable factors for performance to deal with any testing images. However, DL relies on large, organized labeled, and domain-specific data to extract features. How to label images of wounds with few variations and bias is the key. The most challenging task is to label various tissues of ulceration. Besides the boundary hand drawing method, the painting method was commonly used. But the problem remained unsolved; there were no pre-defined regions or boundaries. The inter-rater variability and even intra-rater variation (the same image, different round) were still considerable34. Algorithms to define the meaningful areas of wounds, such as superpixel segmentation, are required before being labeled.

The concept of superpixel segmentation was first proposed by Ren at al **35** to group adjacent pixels with similar features, such as intensities, colors and textures. After processing by algorithms, an image will be segmented into many small regions called superpixels. In the last decade, different superpixel algorithms had been proposed, such as Felzenszwalb’s efficient graph in 2004, Quickshift in 2008, and simple linear iterative clustering (SLIC) in 2012. SLIC, a k means clustering algorithm, became the most popular superpixel algorithm27. There were updated versions of SLIC. For example, linear spectral clustering (LSC) took ten-dimensional space to get a better boundary36. Superpixel based edge detection algorithm (SBED) was a centroid updating approach to decrease the effect from noise37. They demand a higher level of computer calculation. We still adopt the original SLIC to draw boundaries for tissues, because it has three crucial properties：fast and simple38; regular shape and similar size; good adhesion to object border.

To perform superpixel segmentation before labeling, we downsized the images to 1000\*750 pixels. The total number of labeled pixels of each class in superpixel datasets was far less than the number of pixels labeled with *labelme* (supplement 2). Why the model used less data but had better performance on testing images ?(Figure 3) Zhang et al reported an interesting study 39. When they input randomly labeled objects or random pixels, after 10 thousand steps, neural network models still converged to fit the training set perfectly. The neural networks were rich enough to memorize wrong training data, but their results on testing datasets were poor. It indicated that if the model trained from dataset with large inter or intra rater variability, the segmentation results of real-world images may not be satisfactory. Superpixel segmentation decreased the inter and intra rater variability from drawing borders between different tissues as steps on *labelme*.

Before the showing up of machine learning, radiation oncologists had already faced this challenge. Radiation therapy had been applied to almost every cancer in the human body. A successful treatment relied on the precision hand draw tumor contour in CT or MRI images. The targeted tumor contour came from the agreement of multi-experts (relative ground truth). However, the inter-rater agreement ranges from 50 to 90% depending on the tasks and image modalities40. To increase agreement, algorithms or models were used to define meaningful regions for medical experts. They only have to finetune than to hand-draw border on their own41. The models are usually trained from a small, but highly reliably labeled dataset, which is comprised by textbook-like medical images41. Otherwise, they could be algorithms that can properly segment images related to ground truth, such as SLIC42.

It is obvious that increasing the number of superpixels (K) could segment images more accurately (boundary recall)43,44, but it also increased the computation time. Interestingly, when superpixel segmentation combined with other ML models, the increasing K was not always positive correlative with performance45. If the number is too high, the texture information of each superpixel would be reduced. In our datasets, the average wound area is around 41.8% of the entire image, and we resized all images to 1000\*750 pixels. Our experienced plastic surgeons could easily label all types of tissues by K 800 to 1000 without compromising the details (Figure 3). If the number of superpixels kept rising, the consensus of labeling is hard to achieve as boundary label method.

|  |
| --- |
|  |
| Supplement 2. The total number of pixels of different tissues among *labelme* dataset (orange bar) and superpixel dataset (blue bar). |

As for class combination, we found that when re-ep training with ulceration had better results than training with other tissues (granulation, eschar, and slough). It can’t be easily explained by imbalance data. The total pixels of ulceration were slight more than the sum of three types of tissues (supplement 3). When re-ep was trained with ulceration, we actually put it toward more minority, which will theoretically get poor results of re-ep46,47. Class overlapping will deteriorate classification result as well. Neither can it be explained by class overlapping because re-ep is less likely to overlap with tissues than edge of ulceration48.

The other possible factor may be the number of classes. However, the related articles were limited and has controversial conclusions. In a study of ImageNet(large data、large classes、small individual objects), Abramovich et al found that more classes will improve the accuracy of model classifiers49. It is because sorting objects into more classes is an operation of supervised feature extraction. In other articles, merge classes can improve classifiers' accuracy because that reduces the inter-class label errors of dataset50. For example, if we sorted Giant pandas to *Ursus* (bear), it is wrong because they belong to *Ailuropoda*. If we merged *Ursus and Ailuropoda* into *Ursidae,* Giant pandas could be classed as *Ursidae* correctly.However, our status of datasets is not fit the above theories completely. First, our dataset has different distributions from ImageNet with small classes and large individual pixels. Secondly, Re-ep is not a class can be merge into other classes. We believed that re-ep 1.outside the wounds 2. incontinence of other tissues 3.fewer pixels may be classified as background when training with other tissues.

|  |
| --- |
|  |
| Supplement 3. The number of pixels of each class. The Pixels of ulceration are slightly more than the sum of pixels of granulation, slough, and eschar. |

Limitation

1. Tissues of pressure ulcer were more complicated than what we labeled. Pressure ulcers may expose ligaments, periosteum or bone cortex. Though some articles adopted more types of tissues51,52, we used used three primary tissues (granulation, slough, and eschar) for reasons. More types of tissue increase the difficulty of labeling and augment inter-rater variability. To classify more tissues may not change final assessment of wound healing. Granulation is the most critical tissue of wound healing to be observed. However, detection of only three types of tissues has the limitation, if we want the DL model to output more advanced suggestions, such as the need for reconstruction surgery.
2. The estimation of wound size was a projection of 3 dimension (3D) wound surface on a image. The angle of a camera to the wound bed will affect the result of the projection area. We suggested that the camera should be parallel to the wound bed to get accurate results. Moreover, our segmentation results didn’t include the depth information of pressure ulcers. To solve above problems, 3D wound surface images were needed to be constructed. Still, special devices are required to gather more images and information, such as stereo images, multiview by structure light guided or Lidar guided (ToF array).

Conclusion

We proposed systemic label methods to create pressure ulcer datasets for deep learning. Superpixel preprocessing, a region-based method was applied for various tissues inside ulceration while boundary-based method was used for re-epithelialization and ulceration. Several powerful DL models had trained from the datasets, and had promising results of wound segmentation and tissue classification. Combining the segmentation results and other clinic information, our algorithm can detect the sign of wound healing, monitor the progress of wound healing, estimate the size of a wound and suggest the need for surgical intervention.

Reference

1. Dinsdale S. Decubitus ulcers: role of pressure and friction in causation. *Archives of physical medicine and rehabilitation.* 1974;55(4):147-152.

2. Horn SD, Bender SA, Ferguson ML, et al. The National Pressure Ulcer Long‐Term Care Study: pressure ulcer development in long‐term care residents. *Journal of the American Geriatrics Society.* 2004;52(3):359-367.

3. Mervis JS, Phillips TJ. Pressure ulcers: Pathophysiology, epidemiology, risk factors, and presentation. *Journal of the American Academy of Dermatology.* 2019;81(4):881-890.

4. Vanderwee K, Clark M, Dealey C, Gunningberg L, Defloor T. Pressure ulcer prevalence in Europe: a pilot study. *Journal of evaluation in clinical practice.* 2007;13(2):227-235.

5. Baumgarten M, Margolis D, Gruber-Baldini AL, et al. Pressure ulcers and the transition to long-term care. *Advances in Skin & Wound Care.* 2003;16(6):299-304.

6. Bauer K, Rock K, Nazzal M, Jones O, Qu W. Pressure Ulcers in the United States' Inpatient Population From 2008 to 2012: Results of a Retrospective Nationwide Study. *Ostomy/wound management.* 2016;62(11):30-38.

7. Kwon S, Hong J, Choi E-K, et al. Detection of Atrial Fibrillation Using a Ring-Type Wearable Device (CardioTracker) and Deep Learning Analysis of Photoplethysmography Signals: Prospective Observational Proof-of-Concept Study. *Journal of medical Internet research.* 2020;22(5):e16443.

8. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ precision oncology.* 2020;4(1):1-10.

9. Kanavati F, Toyokawa G, Momosaki S, et al. Weakly-supervised learning for lung carcinoma classification using deep learning. *Scientific reports.* 2020;10(1):1-11.

10. Han W, Johnson C, Gaed M, et al. Histologic tissue components provide major cues for machine learning-based prostate cancer detection and grading on prostatectomy specimens. *Scientific reports.* 2020;10(1):1-12.

11. Marijanović D, Filko D. A Systematic Overview of Recent Methods for Non-Contact Chronic Wound Analysis. *Applied Sciences.* 2020;10(21):7613.

12. Zahia S, Zapirain MBG, Sevillano X, González A, Kim PJ, Elmaghraby A. Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. *Artificial intelligence in medicine.* 2020;102:101742.

13. Anisuzzaman D, Wang C, Rostami B, Gopalakrishnan S, Niezgoda J, Yu Z. Image Based Artificial Intelligence in Wound Assessment: A Systematic Review. *arXiv preprint arXiv:200907141.* 2020.

14. Sungkrityayan K, Swarupa P, Sambu S, Adithya K. Segmenting skin ulcers based on thresholding and watershed segmentation. Paper presented at: IEEE Communications and Signal Processing (ICCSP) International Conference2015.

15. Veredas FJ, Luque-Baena RM, Martín-Santos FJ, Morilla-Herrera JC, Morente L. Wound image evaluation with machine learning. *Neurocomputing.* 2015;164:112-122.

16. Rani P, Aliahmad B, Kumar DK. A novel approach for quantification of contour irregularities of diabetic foot ulcers and its association with ischemic heart disease. Paper presented at: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)2017.

17. Trabelsi O, Tlig L, Sayadi M, Fnaiech F. Skin disease analysis and tracking based on image segmentation. Paper presented at: 2013 International Conference on Electrical Engineering and Software Applications2013.

18. Seixas JL, Barbon S, Siqueira CM, Dias IFL, Castaldin AG, Felinto AS. Color energy as a seed descriptor for image segmentation with region growing algorithms on skin wound images. Paper presented at: 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)2014.

19. Mesa H, Veredas FJ, Morente L. A hybrid approach for tissue recognition on wound images. Paper presented at: 2008 Eighth International Conference on Hybrid Intelligent Systems2008.

20. Silveira M, Nascimento JC, Marques JS, et al. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing.* 2009;3(1):35-45.

21. Gholami P, Ahmadi-Pajouh MA, Abolftahi N, Hamarneh G, Kayvanrad M. Segmentation and measurement of chronic wounds for bioprinting. *IEEE journal of biomedical and health informatics.* 2017;22(4):1269-1277.

22. Li D, Mathews C. Automated measurement of pressure injury through image processing. *Journal of clinical nursing.* 2017;26(21-22):3564-3575.

23. García-Zapirain B, Elmogy M, El-Baz A, Elmaghraby AS. Classification of pressure ulcer tissues with 3D convolutional neural network. *Medical & biological engineering & computing.* 2018;56(12):2245-2258.

24. Elmogy M, García-Zapirain B, Elmaghraby AS, El-Baz A. An automated classification framework for pressure ulcer tissues based on 3d convolutional neural network. Paper presented at: 2018 24th International Conference on Pattern Recognition (ICPR)2018.

25. Rajathi V, Bhavani R, Wiselin Jiji G. Varicose ulcer (C6) wound image tissue classification using multidimensional convolutional neural networks. *The Imaging Science Journal.* 2019;67(7):374-384.

26. Rousselle P, Braye F, Dayan G. Re-epithelialization of adult skin wounds: Cellular mechanisms and therapeutic strategies. *Adv Drug Deliv Rev.* 2019;146:344-365.

27. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. *Slic superpixels.* 2010.

28. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention2015.

29. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv 2017. *arXiv preprint arXiv:170605587.* 2019.

30. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2017.

31. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2017.

32. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. Paper presented at: Proceedings of the IEEE international conference on computer vision2017.

33. Tanaka M, Noda Y, Sekine Y, et al. JSPU Guidelines for the Prevention and Management of Pressure Ulcers.

34. Wannous H, Treuillet S, Lucas Y. Supervised tissue classification from color images for a complete wound assessment tool. Paper presented at: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society2007.

35. Ren X, Malik J. Learning a classification model for segmentation. Paper presented at: Computer Vision, IEEE International Conference on2003.

36. Li Z, Chen J. Superpixel segmentation using linear spectral clustering. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition2015.

37. Zhang H, Wu C, Zhang L, Zheng H. A novel centroid update approach for clustering-based superpixel methods and superpixel-based edge detection. Paper presented at: 2020 IEEE International Conference on Image Processing (ICIP)2020.

38. Zhang J, Aviles-Rivero AI, Heydecker D, Zhuang X, Chan R, Schönlieb C-B. Dynamic spectral residual superpixels. *Pattern Recognition.* 2021;112:107705.

39. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:161103530.* 2016.

40. Cabitza F, Campagner A, Albano D, et al. The elephant in the machine: Proposing a new metric of data reliability and its application to a medical case to assess classification reliability. *Applied Sciences.* 2020;10(11):4014.

41. Lu S-L, Xiao F-R, Cheng JC-H, et al. Randomized Multi-Reader Evaluation of Automated Detection and Segmentation of Brain Tumors in Stereotactic Radiosurgery with Deep Neural Networks. *Neuro-oncology.* 2021.

42. Irving B, Franklin JM, Papież BW, et al. Pieces-of-parts for supervoxel segmentation with global context: Application to DCE-MRI tumour delineation. *Medical image analysis.* 2016;32:69-83.

43. Gong Y-J, Zhou Y. Differential evolutionary superpixel segmentation. *IEEE Transactions on Image Processing.* 2017;27(3):1390-1404.

44. Zhang Y, Guo Q, Zhang C. Simple and fast image superpixels generation with color and boundary probability. *The Visual Computer.* 2021;37:1061-1074.

45. Maghsoudi OH. Superpixel based segmentation and classification of polyps in wireless capsule endoscopy. Paper presented at: 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)2017.

46. Ali A, Shamsuddin SM, Ralescu AL. Classification with class imbalance problem. *Int J Advance Soft Compu Appl.* 2013;5(3).

47. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *Journal of Big Data.* 2019;6(1):1-54.

48. García V, Alejo R, Sánchez JS, Sotoca JM, Mollineda RA. Combined effects of class imbalance and class overlap on instance-based classification. Paper presented at: International Conference on Intelligent Data Engineering and Automated Learning2006.

49. Abramovich F, Pensky M. Classification with many classes: challenges and pluses. *Journal of Multivariate Analysis.* 2019;174:104536.

50. Hashemi S, Yang Y, Mirzamomen Z, Kangavari M. Adapted one-versus-all decision trees for data stream classification. *IEEE Transactions on Knowledge and Data Engineering.* 2008;21(5):624-637.

51. Veredas F, Mesa H, Morente L. Binary tissue classification on wound images with neural networks and bayesian classifiers. *IEEE transactions on medical imaging.* 2009;29(2):410-427.

52. Nizam K, Fauzi MFA, Ahmad NN, Nair HK. Characterization of Tissues in Chronic Wound Images. Paper presented at: 2018 IEEE Student Conference on Research and Development (SCOReD)2018.

53. Rostami B, Anisuzzaman D, Wang C, Gopalakrishnan S, Niezgoda J, Yu Z. Multiclass Wound Image Classification using an Ensemble Deep CNN-based Classifier. *arXiv preprint arXiv:201009593.* 2020.

## Abstract

**Background:**

pressure ulcer is a damage to skin and underlying tissues from bony eminence. Patients who suffered from this disease may have difficulty attending medical care. Recently. Covid pandemic has deteriorated this situation. Automatic diagnosis based on machine learning (ML) brings promising solutions.

**Objective:** Traditional ML required complicated preprocessing steps for feature extraction. The clinical applications were limited to particular datasets. Deep learning (DL) extracting features from convolution layers can embrace more data, which might be deliberately excluded in traditional algorithms. However, DL requires large, systemic labeled data for training. It is a challenge for plastic surgeons to label tissues of pressure ulcers.

**Methods:** We proposed a superpixel-assisted, region-based method to label the dataset for tissue classification. The boundary-based method was used to label the other datasets for wound and re-epithelialization (re-ep) segmentation. Five popular DL models (U-Net, DeeplabV3, PsPNet, FPN, and Mask R-CNN) with encoder (ResNet101) were trained on the two datasets.

**Results:** Total 2836 images of pressure ulcers were labeled for tissue classification, and 2893 images were labeled for wound and re-ep segmentation. DeeplabV3 has the best performances among both tasks with precision: 0.9915, recall: 0.9915 and accuracy: 0.9957 for tissue classification; precision: 0.9888, recall: 0.9887 and accuracy: 0.9925 for wound and re-ep segmentation.

**Conclusions:** Combining segmentation results with clinical data, our algorithm could detect the sign of healing, monitor the progress of the wound, estimate the wound size, and suggest the need for surgical debridement.

**Keywords:** deep learning; semantic segmentation; instance segmentation; burn wounds; percentage total body surface area; %TBSA

補充用

https://scikitimage.org/docs/0.13.x/auto\_examples/segmentation/plot\_segmentations.html

<https://davidstutz.de/superpixel-algorithms-overview-comparison/>

<https://blog.csdn.net/u012931582/article/details/70314859>

https://stackoverflow.com/questions/35293468/does-reducing-classes-in-a-classification-method-improve-accuracy/35297915

https://stats.stackexchange.com/questions/348584/do-more-object-classes-increase-or-decrease-the-accuracy-of-object-detection

* 為何是用在data processing 而不是架在其他的model
* 是否有其他的 tissue type

Area=

Area = \* (42\*31.5)^

Area = \* (42\*31.5)^

|  |  |
| --- | --- |
|  |  |
| 1686191  Re-ep precision : 0.9981  Re-ep recall : 0.5685  Re-ep iou : 0.4482  f1-score : 0.7244 | 1686191  Re-ep precision : 0.9974  Re-ep recall : 0.6291  Re-ep iou : 0.4639  f1-score : 0.7716 |
| Others precision : 0.9971  Others recall : 0.6920 | Ulceration precision : 0.9977  Ulceration recall : 0.9172 |
| Loss: 0.61338  mean iou\_score: 0.69275  mean f1-score: 0.72431  mean recall: 0.82691  mean precision: 0.83809  mean accuracy: 0.99266 | Loss: 0.33023  mean iou\_score: 0.83316  mean f1-score: 0.86231  mean recall: 0.91913  mean precision: 0.90534  mean accuracy: 0.99314 |

* Dataset 是chronic wound or pressure ulcer 的差別

Though some of the articles had focus their work on one type of wound, such as DM ulcer, most articles had trained their models with chronic wounds, which means mix etiology of chronic wounds. Rostami et al. reported ensemble DNNs to classify different types of chronic wounds53. The classification results were good, but the clinical application was questionable. First, chronic ulcers that define as ulceration healing less than 20% of wound area in 4 weeks, always have multiple etiologies. 47% of patients with DM ulcer on foot had suffered from peripheral artery occlusion. Patients with DM ulcer have more likely to suffer from venous insufficiency as well. Second,