Comprehensive Assessment of Fine-Grained Wound Images Using a Patch-Based CNN With Context-Preserving Attention

Ziyang Liu, Emmanuel Agu, Peder Pedersen, Clifford Lindsay, Bengisu Tulu, and Diane Strong

Abstract*—*Goal: **Chronic wounds affect 6.5 million Americans. Wound assessment via algorithmic analysis of smartphone images has emerged as a viable option for remote assessment. *Methods:* We comprehensively score wounds based on the clinically-validated Photographic Wound Assessment Tool (PWAT), which comprehensively assesses clinically important ranges of eight wound attributes: Size, Depth, Necrotic Tissue Type, Necrotic Tissue Amount, Granulation Tissue type, Granulation Tissue Amount, Edges, Periulcer Skin Viability. We proposed a DenseNet Convolutional Neural Network (CNN) framework with patch-based context-preserving attention to assess the 8 PWAT attributes of four wound types: diabetic ulcers, pressure ulcers, vascular ulcers and surgical wounds. *Results:* In an evaluation on our dataset of 1639 wound images, our model estimated all 8 PWAT sub-scores with classification accuracies and F1 scores of over 80%. *Conclusions:* Our work is the first intelligent system that autonomously grades wounds comprehensively based on criteria in the PWAT rubric, alleviating the significant burden that manual wound grading imposes on wound care nurses.**

***Index Terms*—Chronic wounds, deep learning, medical imaging, smartphone assessment, transfer learning.**

***Impact Statement*— We proposed a CNN Densenet with context-preserving attention mechanism that assess all 8 PWAT sub-scores for chronic wound images and achieves classification accuracies and F1 scores of over 0.8.**

I. Introduction

Motivation: Over 6.5 million people in the US (or approximately 2% of the population [1]) have chronic wounds with a high prevalence in the elderly population [2], [3] and cost the healthcare system over $25 billion annually [4]. Chronic wounds are often painful and require proper treatment in order to heal properly. Care of chronic wounds involves cleaning, debridement, changing of dressings and using antibiotics [5]. Without proper care, such wounds may become infected [6] or cause limbs to be amputated. Due to the large and growing number of chronic wounds, there is an increasing demand for more efficient chronic wound care, especially information technology solutions that assist the work of medical personnel and reduces the cost of care. Some background and detailed descriptions of the various types of chronic wounds can be found in the Supplementary Materials.

Smartphone-based image analyses has emerged as a viable option for remote wound assessment [7]–[10]. Since 2011, our group has been researching and developing the Smartphone Wound Analysis and Decision-Support (SmartWAnDS) [11]–[14]. SmartWAnDS autonomously analyzes chronic wound images captured using smartphone cameras and provide wound care recommendations to patients and their caregivers. The envisioned SmartWAnDS system will allow patients to receive standardized feedback on their wounds in their homes between hospital visits, engaging patients in the care of their wounds. It can also provide care recommendations to assist wound nurses in remote locations when wound doctors are temporarily unavailable. The best care and recommended treatment for a wound are based on its healing progress and condition. Thus at any point in time, before deciding on the recommended treatment, the wound's healing progress since the preceding examination has to be graded (scored). The research presented in this work describes the SmartWAnDS module that autonomously grades wound healing status based on its visual appearance in a smartphone photograph.

In collaboration with wound experts, our group created a *WoundNet*, a large chronic wound image dataset that contains 1639 chronic wound images in total. *WoundNet* contains four types of wounds: diabetic foot ulcers, pressure ulcers, vascular ulcers and surgical wounds. Venous ulcers and arterial ulcers are the two main types of vascular ulcers. These types of chronic wounds were included in our *WoundNet* dataset because they

Manuscript received March 26, 2021; revised June 18, 2021 and June 21, 2021; accepted June 21, 2021. Date of publication June 24, 2021; date of current version July 21, 2021. This work was supported by the National Institutes of Health under Grant 1R01EB025801. *(Corresponding author: Emmanuel Agu.)*

Ziyang Liu and Emmanuel Agu are with the Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: zliu10@wpi.edu; emmanuel@cs.wpi.edu).

Peder Pedersen is with the Electrical and Computer Engineering Department, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: pedersen@wpi.edu).

Clifford Lindsay is with the Department of Radiology, University of Massachusetts Medical School, Worcester, MA 01655 USA (e-mail: Clifford.Lindsay@umassmed.edu).

Bengisu Tulu and Diane Strong are with the Foisie Business School, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: bengisu@wpi.edu; dstrong@wpi.edu).

Digital Object Identifier 10.1109/OJEMB.2021.3092207

TABLE I
PRIOR WORK ON ASSESSING WOUND ATTRIBUTES RELATED TO THE 8 PWAT SUB-SCORES

PWAT subscore	Related research	method	task	wound type	dataset size	results
1. Size	Chino et al. 2020 [15]	deep neural network	segment the wound and estimate the size	venous and arterial ulcer	446	estimate wound area in cm^2 with error of 14%
	Spinczyk et al. 2017 [16]	triangulation technique	wound 3D surface reconstruction	not specified	10 patient	measure wound area with error of 11%
2. Depth	None	N. A.	N. A.	N. A.	N. A.	N. A.
3. Nec Type and 4. Nec Amount	Hsu et al. 2017 [17]	clustering method and SVM	detect necrotic tissue	postsurgical wound	42	detection accuracy of 95.23%
	Blanco et al. 2020 [18]	superpixel driven deep learning approach	segmenting tissue including necrotic and wounded area quantification	arterial and venous ulcers	217	spot wounded tissues with AUC = 0.986
	Godeiro et al. 2018 [19]	deep neural network	classifying tissue including necrotic	chronic wounds	30	tissue classification accuracy 96%
	Nejati et al. 2018 [20]	deep neural network, SVM	classifying tissue including necrotic	chronic wounds	350	tissue classification accuracy 86.4%
	Hsu et al. 2019 [21]	robust image segmentation, SVM	wound segmentation and detect tissue including necrotic	chronic wounds	293	tissue classification accuracy 83.58%
	Maity et al. 2018 [22]	deep neural network	classifying tissue including necrotic	chronic wounds	68	tissue classification accuracy 99%
	Babu et al. 2018 [23]	Naive bayes and Hoeffding tree	wound segmentation and classifying tissues including necrotic	diabetic wound	N. A.	tissue classification accuracy 90.9%
	Rajathi et al. 2019 [24]	deep neural network	classifying tissue including necrotic	varicose ulcer	1250	tissue classification accuracy 99.55%
5. Gran Type and 6. Gran Amount	Blanco et al. 2020 [18]	superpixel driven deep learning approach	segmenting tissue including granulation and wounded area quantification	arterial and venous ulcers	217	spot wounded tissues with AUC = 0.986
	Godeiro et al. 2018 [19]	deep neural network	classifying tissue including necrotic	chronic wounds	30	tissue classification accuracy 96%
	Nejati et al. 2018 [20]	deep neural network, SVM	classifying tissue including necrotic	chronic wounds	350	tissue classification accuracy 86.4%
	Hsu et al. 2019 [21]	robust image segmentation, SVM	wound segmentation and detect tissue including granulation	chronic wounds	293	tissue classification accuracy 83.58%
	Maity et al. 2018 [22]	deep neural network	classifying tissue including granulation	chronic wounds	68	tissue classification accuracy 99%
	Babu et al. 2018 [23]	Naive bayes and Hoeffding tree	wound segmentation and classifying tissues including granulation	diabetic wound	N. A.	tissue classification accuracy 90.9%
	Rajathi et al. 2019 [24]	deep neural network	classifying tissue including granulation	varicose ulcer	1250	tissue classification accuracy 99.55%
Edge	None	N. A.	N. A.	N. A.	N. A.	N. A.
Skin	None	N. A.	N. A.	N. A.	N. A.	N. A.

are the most common types seen by wound experts at hospitals. Table II(a) summarizes the statistics of the wound types in our *WoundNet* repository. Fig. 2(a) shows example diabetic, venous, arterial and pressure ulcers.

In this paper, we propose a patch-based neural networks-based method with context-preserving attention [25] to assess four different wound types (diabetic foot ulcers, pressure ulcers, vascular ulcers and surgical wounds), which analyzed a smartphone image of a wound to assess its healing status. To comprehensively assess wounds in WoundNet and generate ground truth labels for training our neural networks model, each wound image was assessed using the Photographic Wound Assessment Tool (PWAT), a clinically validated wound grading rubric [26]–[28]. The PWAT wound rubric evaluates eight attributes of wounds [28] that can be extracted visually from an image: 1) Size 2) Depth 3) Necrotic Tissue Type 4) Necrotic Tissue Amount 5) Granulation Tissue Type 6) Granulation Tissue Amount 7) Edges and 8) Skin viability. Each PWAT sub-score grades a single wound attribute with a score of 0 (best), 1, 2, 3 or 4 (worst). PWAT grading yields 8 wound sub-scores and all 8 PWAT sub-scores are summed to generate a total PWAT wound score (max = 32). Higher PWAT scores indicate a worse wound condition. In the rest of this paper the PWAT sub-scores for Necrotic Tissue Type, Necrotic Tissue Amount, Granulation Tissue Type and Granulation Tissue Amount are abbreviated as Nec Type, Nec Amount, Gran Type and Gran Amount respectively. A table that presents detail descriptions of each PWAT sub-score and their corresponding grading criteria can be found in the Supplementary Materials.

Novelty of our work: Our work is addresses limitations of prior work include that they have 1) not assessed wounds comprehensively based on a validated wound rubric and instead only a few selected wound attributes or assessed a single wound type, or 2) Focused on the simpler binary classification task of detecting the presence (yes/no) of specific tissue types. We advance the state-of-the-art by comprehensively scoring wounds based on the clinically-validated Photographic Wound Assessment Tool (PWAT) grading rubric. We are also the first to formulate wound image classification as a fine-grained image classification problem due to the high inter-class similarity of wound images. Table I summarizes related work demonstrating that they typically assessed only a few wound attributes and did not comprehensively assess wounds. A detailed analysis and comparison of the limitation of related work is presented in Table I in the supplementary materials. A brief description of the fine-grained image classification problem in computer vision

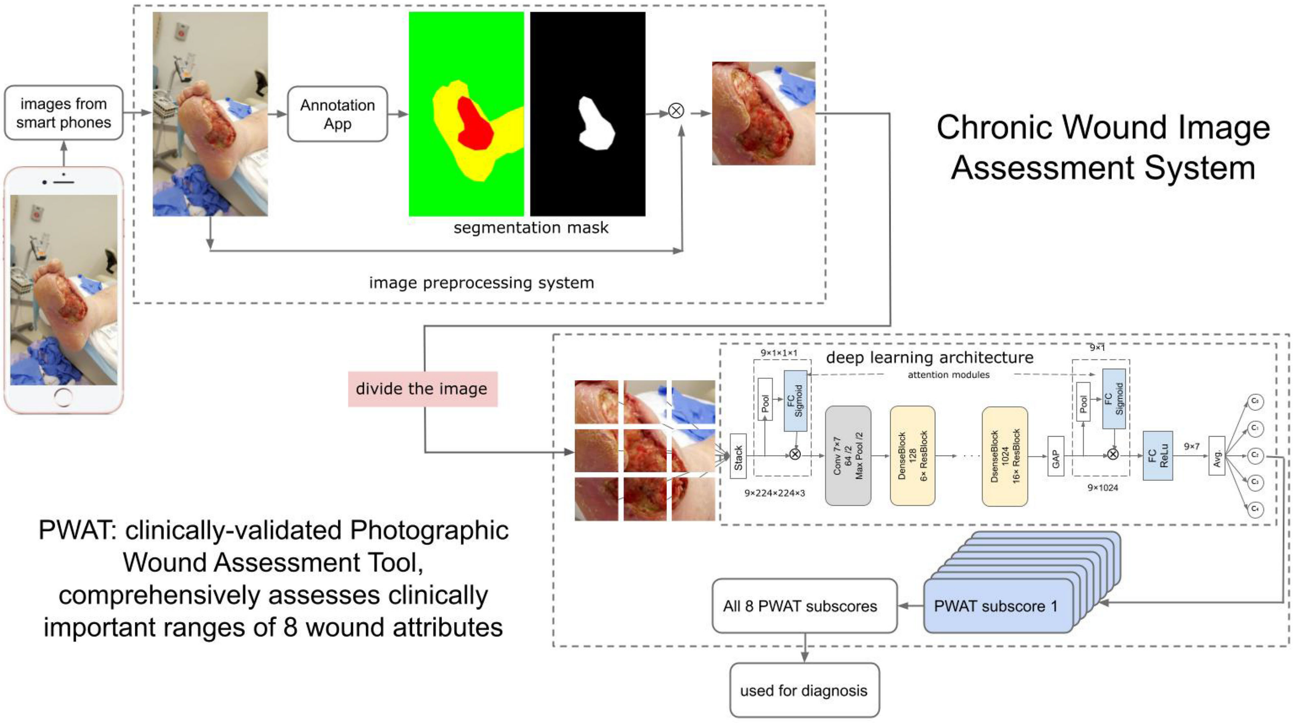

images from smart phones
Annotation App
segmentation mask
image preprocessing system
Chronic Wound Image Assessment System
divide the image
deep learning architecture
attention modules
9×1×1×1
9×224×224×3
Stack
Pool
FC Sigmoid
Conv 7×7 64 /2 Max Pool /2
DenseBlock 128 6× ResBlock
DseneBlock 1024 16× ResBlock
GAP
9×1
9×1024
FC ReLu
9×7
Avg.
C0
C1
C2
C3
C4
PWAT: clinically-validated Photographic Wound Assessment Tool, comprehensively assesses clinically important ranges of 8 wound attributes
All 8 PWAT subscores
PWAT subscore 1
used for diagnosis

Fig. 1. Our chronic wound image assessment system comprehensively assesses chronic wound healing status based on its visual appearance in a smartphone photograph, which uses a deep learning framework with patch-based context-preserving attention mechanism.

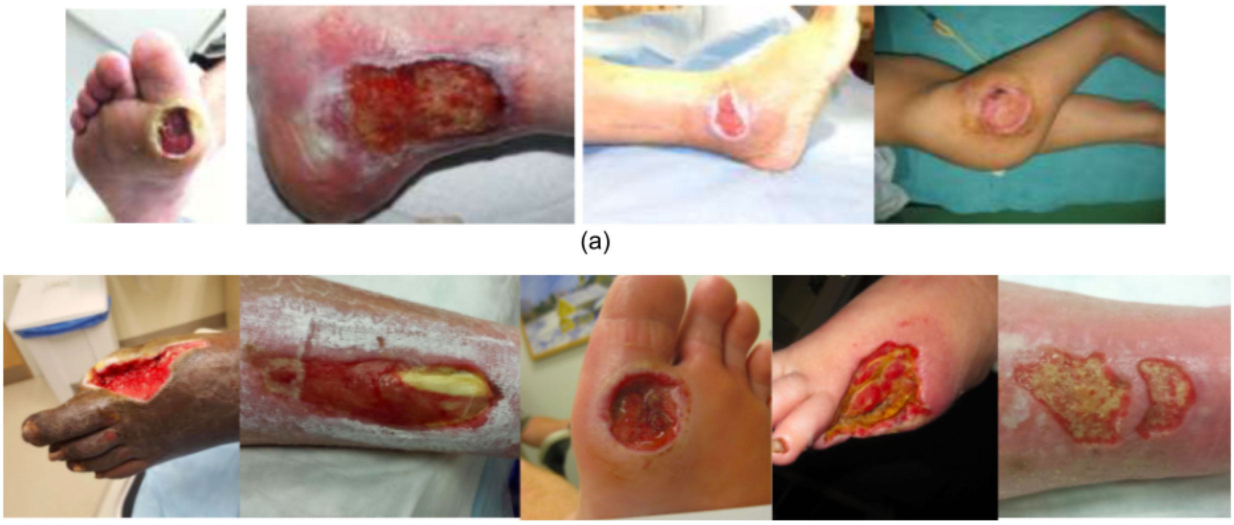


Fig. 2. (a) Examples of diabetic, venous, arterial and pressure ulcers wound types (left to right) (b) Example wound images corresponding to PWAT Necrotic Amount scores 0 (left) to 4 (right). The target sub-classes (Necrotic amount scores) can appear quite similar visually, posing a fine-grained image classification problem.

is also shown in the Supplementary Materials. Fig. 2 illustrates the fine-grained nature of PWAT sub-score prediction and (b) shows example images corresponding to scores 0 to 4 for the subscore Nec Amount (left to right) based on the PWAT necrotic tissue wound grading rubric. As shown in Fig. 2(b) wounds that fall into different PWAT wound classes are quite challenging to distinguish visually based on the type or amount of a specific types of tissue shown in the wound images.

Our contributions: There are three main contributions in this paper:

1) In collaboration with wound experts, we created *WoundNet*, a large dataset of 1639 chronic wound images of four types

TABLE II
STATISTICS OF TYPES OF WOUNDS AND PWAT SUB-SCORES OF IMAGES IN TEXTITWOUNDNET DATASET

(a) Statistics of types of wounds in *WoundNet* dataset

Wound Types	Numbers of Images
1. Diabetic Foot Ulcers	121
2. Pressure Ulcers	13
3. Vascular Ulcers	1349
4. Surgical Wounds	156

(b) Statistics of PWAT sub-scores of images in *WoundNet* dataset

PWAT subscore	0	1	2	3	4
1. Size	141	393	537	337	231
2. Depth	134	135	814	378	178
3. Necrotic Type	509	472	273	76	309
4. Necrotic Amount	347	160	150	290	692
5. Granulation Type	143	319	626	95	456
6. Granulation Amount	137	77	194	342	889
7. Edges	142	311	1035	131	20
8. Skin	435	1061	143	0	0

(diabetic foot ulcers, pressure ulcers, vascular ulcers and surgical wounds). Each image in *WoundNet* was annotated with the 8 sub-scores in the PWAT wound grading rubric. This large wound image repository made deep learning methods feasible.

2) We innovatively proposed a deep learning framework with patch-based context-preserving attention mechanism [25] that utilized the DenseNet neural network [29] to comprehensively assess all 8 PWAT sub-scores and also dealt with the challenging fine-grained image task of recognizing clinically-important grades of wound.

3) We performed rigorous evaluations of our proposed model and its components. Our results show that our proposed patch-based Convolution Neural Networks (CNNs) with context-preserving attention mechanism that utilized DenseNet performed well, preserving the global context between the patches. Our model achieves classification accuracies and F1 scores of over 0.8 for fine-grained classification of all 8 PWAT sub-scores.

II. Materials and Methods

This section introduces our proposed neural networks-based method for estimating PWAT subscores of wounds. Sub-section II-A describes *WoundNet*, our chronic wound image dataset. Sub-section II-B gives a system overview of the deep learning architecture that includes the context-preserving patch attention mechanism and DenseNet. Sub-section II-C describes the patch context-preserving attention mechanism. Sub-section II-D defines our evaluation metrics and describes our experiment. The supplementary materials introduces the DenseNet121 neural network architecture [29] used in our current chronic wound assessment system, discusses its advantages and explains our reason for selecting DenseNet121. The supplementary materials also describes other neural networks models utilized as baselines in this study, including ResNet [30] and Bi-linear Convolutional Neural Network (Bi-CNN) [31].

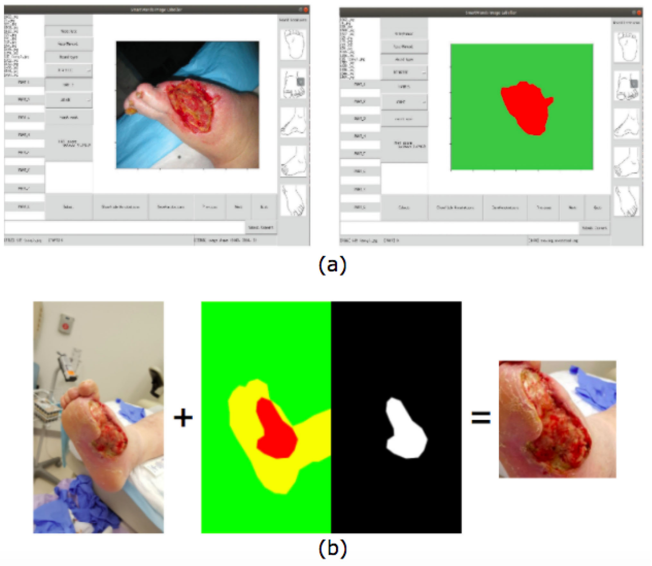


Fig. 3. (a) Example of our wound annotation app (b) Example of wound segmentation mask.

A. WoundNet *Dataset and Preprocessing*

Our *WoundNet* chronic wound image dataset contains 1639 images in total. 1323 of these images were provided by the University of Massachusetts Memorial Medical Center from their archives. 114 wound images were captured by our research group using a mechanical wound imaging box that ensured consistent imaging distance, angle and lighting. 202 wound images were collected from the Internet using an image search. All images in *WoundNet* were labeled with all 8 PWAT sub-scores based on the PWAT subscore scoring instruction. PWAT sub-scores 1 through 7 were assigned values 0, 1, 2, 3 and 4 and were modeled as 5-class classification problems. As PWAT sub-score 8 can only be assigned values 0, 1 and 2, it was modeled as a 3-class classification problem. The number of *WoundNet* images that were assigned each of the 8 PWAT sub-scores are summarized in Table II(b).

Some of our original wound images were poorly captured, posing challenges for image analyses. For instance, some images had very small wounds with large background areas. Some images had wounds that were too large and were mostly occupied by wound and skin area. To make the wound images more consistent in our dataset, we pre-processed the images using the following steps. First, we segmented the wound and the skin out of the whole image using a wound annotation app developed in our previous research [32], which is shown in Fig. 3(a). This segmentation app implemented the deep extreme cuts algorithm proposed by Mananis *et al.* [33] that ensured consistent, systematic wound image segmentation. A detailed description about deep extreme cuts can be found in the Supplementary Materials. Then segmentation mask of the wound yielded by the extreme cuts algorithm was then used as a bounding box to crop the skin and the wound area out from the original wound image. The cropped wound images were resized

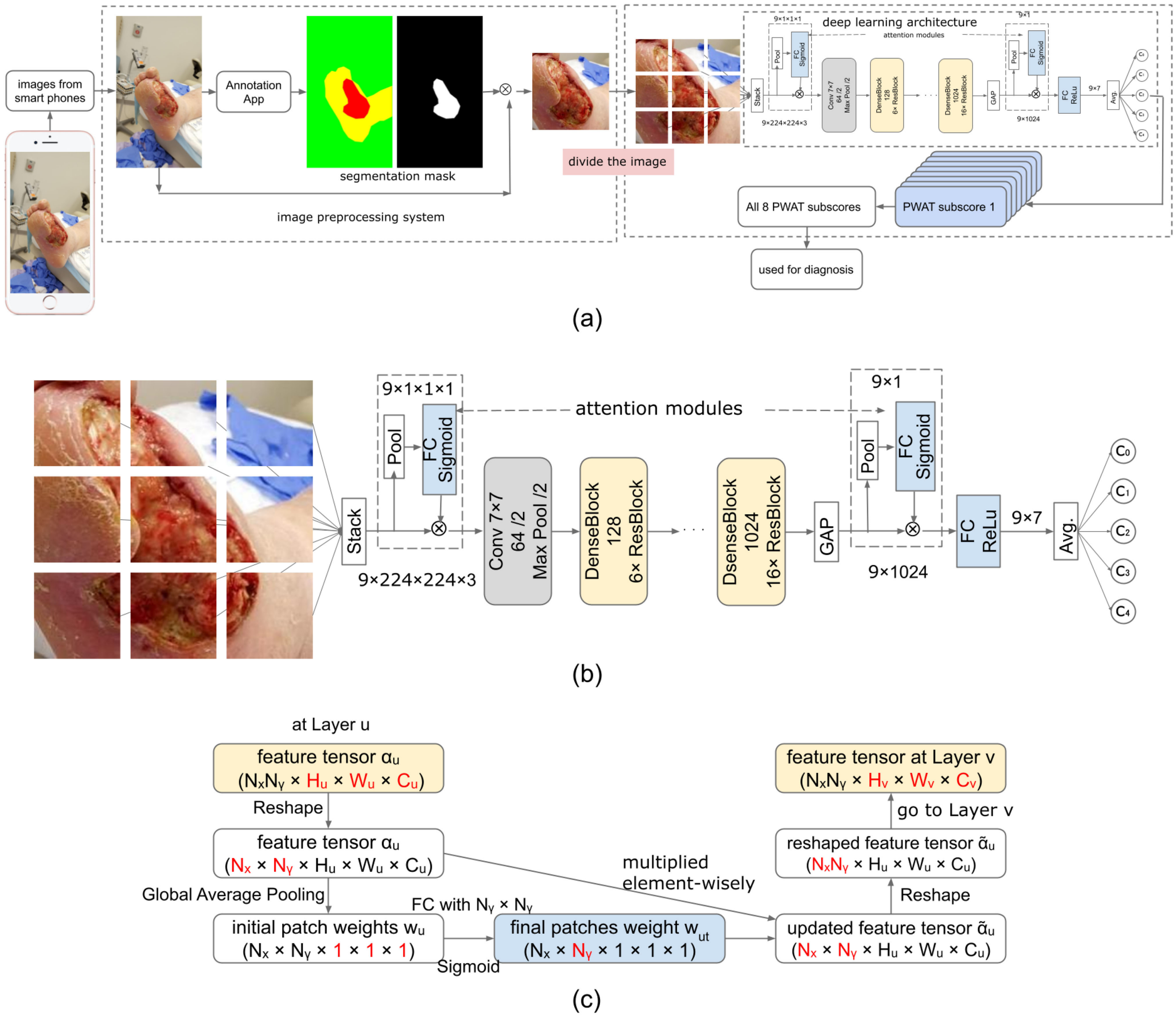

images from smart phones
Annotation App
segmentation mask
image preprocessing system
divide the image
deep learning architecture
attention modules
All 8 PWAT subscores
PWAT subscore 1
used for diagnosis
(a)
9×1×1×1
attention modules
9×1
Pool
FC Sigmoid
Stack
9×224×224×3
Conv 7×7 64 /2 Max Pool /2
DenseBlock 128 6× ResBlock
DsenseBlock 1024 16× ResBlock
GAP
9×1024
FC ReLu
9×7
Avg.
C0
C1
C2
C3
C4
(b)
at Layer u
feature tensor αu (NxNy × Hu × Wu × Cu)
Reshape
feature tensor αu (Nx × Ny × Hu × Wu × Cu)
Global Average Pooling
initial patch weights wu (Nx × Ny × 1 × 1 × 1)
FC with Ny × Ny
Sigmoid
final patches weight wut (Nx × Ny × 1 × 1 × 1)
multiplied element-wisely
updated feature tensor α̃u (Nx × Ny × Hu × Wu × Cu)
Reshape
reshaped feature tensor α̃u (NxNy × Hu × Wu × Cu)
go to Layer v
feature tensor at Layer v (NxNy × Hv × Wv × Cv)
(c)

Fig. 4. (a) Our chronic wound image analysis system including our annotation app, segmenatation, patch generation and assessment using our novel patch-based CNN with context-preserving attention architecture (b) Patch Attention DenseNet architecture (c) Context-Preserving Patch Attention Mechanism.

to a dimension of $512 \times 512 \times 3$. An example of an original wound, its segmentation mask, and cropped image is shown in Fig. 3(b).

B. Overview of Our Wound Assessment System

The diagram of our entire chronic wound image analysis system is shown in Fig. 4(a). It describes the whole system from the first step of capturing wound images using smartphone cameras to finally estimating all 8 PWAT sub-scores. Our deep learning architecture for PWAT sub-score prediction is shown in Fig. 4(b). We adopted a trainable end-to-end deep learning architecture that we call Patch Attention DenseNet [25]. It combines DenseNet121 [29] with a context-preserving patch attention mechanism that preserves the global context between the wound's patches and combines both local and global image information. Local information such as wound tissue types in each patches should be combined with global context information such as the relative positions of parts of the wound region to provide a robust assessment of the overall condition of the wound. For images as complex as wound images, simple methods of combining the results of predictions on patches such as averaging or maximum are simple local operations that fail to integrate global context information from different patches that is required to comprehensively assess the overall wound condition. The context-preserving patch attention block was inserted into the pre-trained standard DenseNet121 architecture as shown in Fig. 4(b). This block was inspired by channel-wise attention [34], which uses attention across patches. Using a wound image with dimension $512 \times 512 \times 3$ as input, our Patch Attention DenseNet architecture predicts the sub-class (e.g 0, 1, 2, 3 or 4) of the input image for one of the PWAT sub-scores.

C. Context-Preserving Patch Attention Mechanism

The calculation process of each step in a convolutional neural network can be interpreted as a mapping function F_{tr} taking an input feature tensor α_{in} and generates an output feature tensor α_{out} expressed as $\alpha_{out} = F_{tr}(\alpha_{in})$. With patches as input from the original image, the output feature tensor of the mapping can be interpreted as a collection of the local descriptors that are expressive for the whole image. Instead of simply averaging or applying other methods to the prediction for each patch from the output of the network, our context-preserving attention mechanisms maps each patch's output in a way that leverages the contextual information from the outputs of other patches. It is necessary to capture patch dependencies by providing them with access to global information. Modelling patch interdependencies enhances the learning process of the convolutional features and increases the network's sensitivity to informative features that can be exploited by the subsequent mapping in the network. The patch attention mechanism can provide the mapping for each patch with access to global information using 2 steps that are similar to the squeeze and excitation in Squeeze-and-Excitation Networks [34]. The squeeze and excitation steps allow the Squeeze-and-Excitation Networks to gain access to global spatial information and perform dynamic channel-wise feature recalibration by modelling channel interdependencies. This mechanism recalibrates filter responses before they are input into the next mapping.

The calculation process of the attention mechanism is shown in Fig. 4(c). The original model input α has a dimension $N_X \times N_Y \times H \times W \times C$, where N_X is the batch size, N_Y is the number of crops and $H \times W \times C$ are the crop dimensions. All crops in the batch dimension are stacked so that the model input's dimension becomes $N_X N_Y \times H \times W \times C$. When the attention module is at layer u, N_X and N_Y were separated to change the feature tensor α_u into dimension $N_X \times N_Y \times H_u \times W_u \times C_u$.

To apply the attention module, the initial patch weights w_u with the size $N_X \times N_Y$ is created using global average pooling over the feature tensor's last three dimensions. It is the global average pooling over $H_u \times W_u \times C_u$ and w_{u_c} is the c-th element of w_u:

$$w_{u_c} = \frac{1}{H_u \times W_u \times C_u} \sum_{j_h=1}^{H_u} \sum_{j_w=1}^{W_u} \sum_{j_c=1}^{C_u} u_c \tag{1}$$

This step is inspired from the squeeze step in the Squeeze-and-Excitation Networks [34], in which it squeezes global spatial information into a channel descriptor to generate channel-wise statistics. On the other hand, the patch attention mechanism we applied generates patch-wise statistics over the 9 wound's patches. The statistics are expressive for the whole image and will be used for calculating the final patches weight in next step.

After aggregating information in the squeeze step, Squeeze-and-Excitation Networks [34] apply a simple gating mechanism with a sigmoid activation, which aims to fully capture channel-wise dependencies. Inspired from the Squeeze-and-Excitation Networks, the patch attention mechanism uses a similar but simplified step to capture patch-wise dependencies. The fully-connected attention layer f_u^a that has the learnable parameters $\boldsymbol{\theta}_u^a$ with size $N_Y \times N_Y$ transforms the patch weights. Then, the transformed weights are substituted into a sigmoid activation function $\boldsymbol{\sigma}$ as $\boldsymbol{\sigma}(f_u^a(w_u)) \in [0, 1]$ and the final patches weight w_{ut} are generated.

$$w_{ut} = \sigma(f_u^a(w_u)) \tag{2}$$

The final patches weights w_{ut} with dimension $N_X \times N_Y$ are then multiplied by the original feature tensor α_u with size $N_X \times N_Y \times H_u \times W_u \times C_u$ element-wisely. Thus, the patches' weights are learned. Here w_{ut} and α_u multiplied element-wisely and obtain the final output of the block $\tilde{\alpha}_u$.

$$\tilde{\alpha}_{u_c} = w_{ut_c} \alpha_{u_c} \tag{3}$$

$$\tilde{\alpha}_u = [\tilde{\alpha}_{u_1}, \tilde{\alpha}_{u_2}, \ldots, \tilde{\alpha}_{u_c}] \tag{4}$$

The dimension of the tensor $\tilde{\alpha}_u$ is then changed to $N_X N_Y \times H_u \times W_u \times C_u$ for subsequent steps in the network. The deep learning model's predictions on crops are averaged at the output. Although we inserted this attention mechanism at the beginning and end of our architecture, it can flexibly be placed at a different location in the network.

D. Evaluation

1) Evaluation Metrics: Our deep learning architecture was evaluated using 4 performance measures: test set accuracy, weighted F1 score, multi-class sensitivity and multi-class specificity. Test set accuracy is defined as the ratio of the number of images in the test set that were correctly classified by the model divided by the total number of images in the test set. The weighted F1 score is defined as:

$$F1_{score} = \sum_{i=1}^{5} w_i \frac{2 \times P_i \times R_i}{P_i + R_i} \tag{5}$$

where P_i and R_i can be calculated as:

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{6}$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{7}$$

In the above equations, TP_i represents the number of True Positives of class i, FP_i is the number of False Positives of class i, TN_i for True Negative of class i and FN_i for False Negative of class i. wi is the weight, which represents the proportion of class i images in the test dataset. We note that due to class imbalance in our wound dataset, it was important to also evaluate the model using the F1 score, sensitivity and specificity in addition to accuracy. Using only testing accuracy as our metric would encourage the correct classification of classes with more instances and neglect classes with fewer instances. We utilized multi-class sensitivity and multi-class specificity to evaluate our architecture, defined as:

$$\text{multiclass sensitivity} = \frac{1}{5} \sum_{i=1}^{5} \frac{TP_i}{TP_i + FN_i} \tag{8}$$

$$\text{multiclass specificity} = \frac{1}{5}\sum_{i=1}^{5}\frac{TN_i}{TN_i + FP_i} \tag{9}$$

2) Experiments: First, as a pilot experiment, we explored whether our model could benefit from data augmentation, which has been utilized to boost model performance in some prior work [35]–[37]. We trained the ResNet50 model on our *WoundNet* chronic wound dataset with and without data augmentation. We applied rotation, flipping and translation augmentations to each image.

Second, we studied the performance of our context-preserving patch-based attention method on *WoundNet*, our chronic wound dataset. We conducted the following experiments. For each PWAT sub-score, we trained the model using 5-fold cross validation on our data with a test set of 10% of the entire dataset. We also ensured that the distribution of images of each PWAT sub-score ([0,1,2,3,4] per bin for sub-scores 1–7 and ([0,1,2] for sub-score 8, Skin) in the test set was the same as their distribution in the entire dataset. Due to various sources of randomness in the deep learning model such as random weight initialization and the randomness in batch gradient descent during training, it was possible that the model could converge to different local minimums. To ensure that our results were stable, we trained each fold 3 times with the same training and test sets. Thus, there were 15 training results per PWAT sub-score. We adopted a two-step model training strategy, which involved transfer learning. We froze the model's weights on the early layers and fine-tuned its last layer.

Third, we investigated the performance boost attributable to the context-preserving patch mechanism on our chronic wound image dataset. The Patch Attention DenseNet121 architecture we adopted was trained with an image of size $512 \times 512 \times 3$. Each input image was cropped to 9 patches. Using a fewer number of patches such as 4 patches will not fully utilize the patch attention mechanism. Using a higher number of patches such as 16 patches will make the cropped images too small. To explore the effectiveness of context-preserving patch attention mechanism, we compared the model's performance in predicting sub-score 2 (Depth) with and without the context-preserving patch attention mechanism. To explore the effects of varying patch size, we compared the Patch Attention DenseNet121 model's performance with image sizes $512 \times 512 \times 3$ and $256 \times 256 \times 3$. We also compared our model to standard DenseNet121 with patches generated from original images but no attention mechanism, and also to DenseNet121 and ResNet50 with no patches or attention mechanism.

III. Results

A. Training and Testing Accuracy Trajectories

Fig. 5(a) and 5(b) show a sample of the trajectories of the training and testing set accuracies for all 8 PWAT sub-scores. The training and testing accuracy trajectories plotted are for the same best set results of the 5 folds that are shown above. In these two figures, the number index i corresponds to the ith PWAT sub-score. For example, train1 is the training accuracy for sub-score 1 (Size). The training and testing accuracies converged and stabilized after about 75 epochs. The differences between training and testing accuracies were small indicating that the model generalized well to the test set.

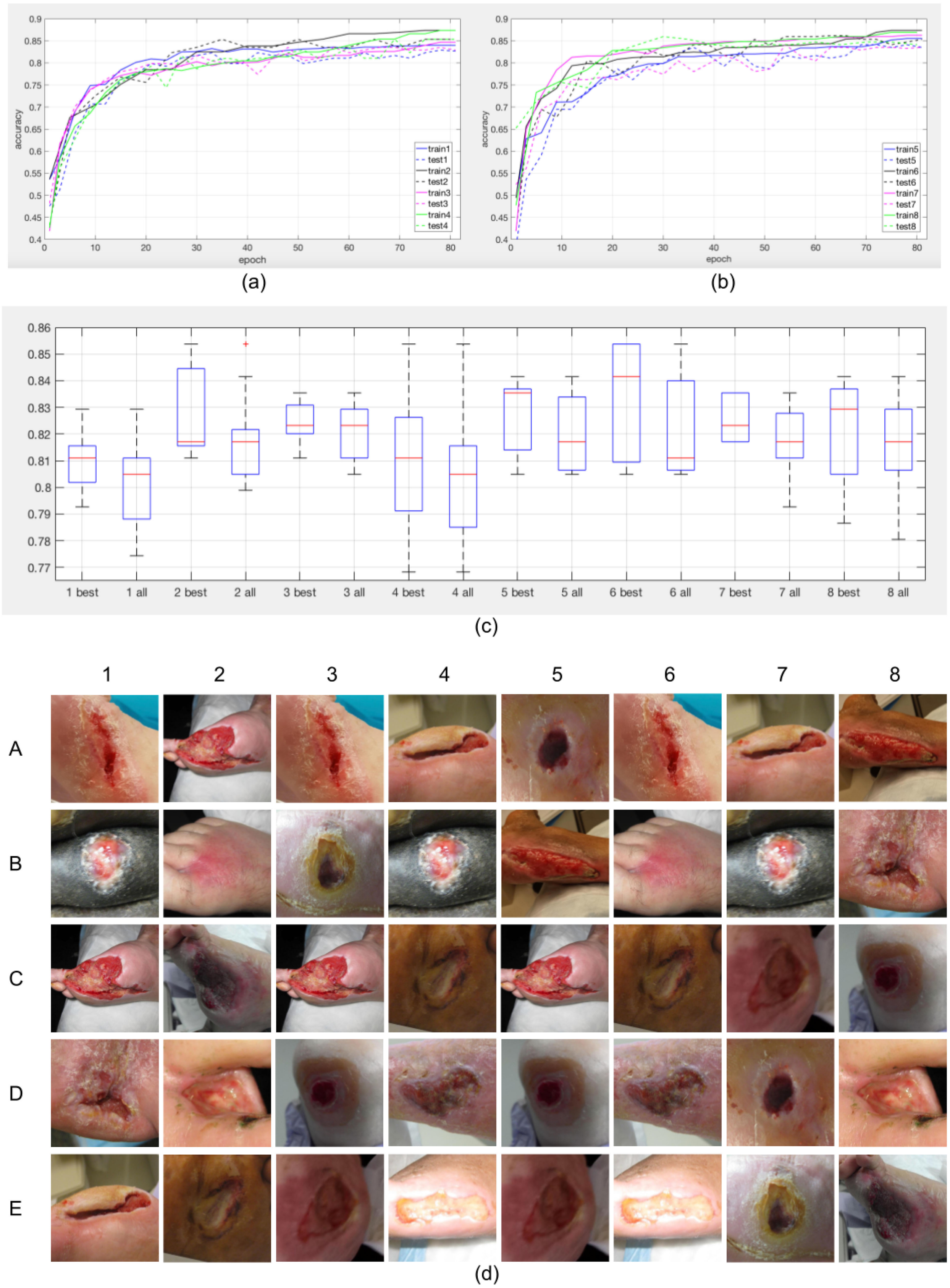
B. Model Performance for Predicting All 8 PWAT Sub-Scores

The mean and standard deviation of the test accuracies, weighted F1 scores, multi-class sensitivity and multi-class specificity from the 5 folds' best results are shown in Table III(a). The calculation of metrics including the weighted F1 scores, multi-class sensitivity and multi-class specificity are defined in II-D1. Our experimental design using these metrics is explained in the second part of II-D2. All mean test accuracies and F1 scores were above 80% while the Patch Attention DenseNet model performed better on some sub-scores than the others. The Patch Attention DenseNet Model performed well for classifying 3 PWAT subscores: 3 (Nec Type), 5 (Gran Type) and 7 (Edges). The mean test accuracies and F1 scores for these sub-scores are high and relatively stable. Table III(a) also shows the mean and standard deviation for the multi-class sensitivity and specificity. The multi-class sensitivity and specificity are the mean sensitivity and specificity of all classes, which provides insight into how well the model performs on each class. Although the dataset is imbalanced, the means of all 8 multi-class sensitivity are above 75% and the means of all 8 multi-class specificity are above 87%, which shows that data imbalanced problem did not significantly affect our models' performance.

The mean for sub-scores 3 (Nec Type), 5 (Gran Type) and 7 (Edges) are relatively high and the standard deviation are low, which means that these three sub-scores are acceptable. While their means are all over 80%, methods to improve the results of 5 of the PWAT sub-scores need exploring in future work. Specifically, looking at the mean and variance of the boxplots, the following are reasons to improve specific PWAT sub-scores. 1) The mean for sub-score 1 (Size) is low but relatively stable. 2) The mean for sub-score 2 (Depth) is relatively high but not stable enough. 3) The mean for sub-score 4 (Nec Amount) is low and not stable. This sub-score needs the most improvement. 4) The mean for sub-score 6 (Gran Amount) is high relatively but not so stable. 5) The mean for sub-score 8 (Skin) is relatively high but not so stable.

C. Box Plot Showing k-Fold Cross-Validation Results of All 8 PWAT Sub-Scores

For each of the 5 folds, 3 results were generated, yielding a total of 15 test results per PWAT sub-score. We studied the best of the 3 iterations as well as all test accuracy results observed for the 5 folds. Fig. 5(c) contains boxplots of the best result among the 3 result iterations, as well as all 15 test results for each PWAT sub-score. For each PWAT subscore, boxplots are shown: 1) the best results and 2) all results. This yields a total of 16 boxplots for the 8 PWAT sub-scores. The index number i below each boxplot indicates that the boxplot corresponds to the ith PWAT sub-score. For example, the first and second boxplots are for the 5 best results and all 15 results respectively for PWAT sub-score 1 (Size). Looking at Fig. 5(c), we can see that the



(a)

(b)

(c)

(d)

Fig. 5. (a) Training and testing accuracy trajectory (Example 1) (b) Training and testing accuracy trajectory (Example 2) (c) Boxplot result for all 8 PWAT subscore (d) column 1 to 8 is PWAT sub-score 1 to 8; row A, B and C predicted correctly while row D and E predicted incorrectly.

TABLE III
OUR MODEL'S RESULTS AND COMPARISON BETWEEN DIFFERENT MODELS

(a) Results of our method for predicting all 8 PWAT sub-scores

	accuracy		F1 score		Sensitivity		Specificity	
PWAT subscore	mean	std	mean	std	mean	std	mean	std
1. Size	0.8098	0.0132	0.8085	0.0137	0.8154	0.0244	0.9503	0.0034
2. Depth	0.8281	0.0185	0.8285	0.0190	0.8156	0.0440	0.9509	0.0056
3. Nec Type	0.8244	0.0090	0.8226	0.0096	0.7711	0.0336	0.9533	0.0028
4. Nec Amount	0.8098	0.0309	0.8072	0.0307	0.7539	0.0447	0.9498	0.0080
5. Gran Type	0.8269	0.0153	0.8263	0.0150	0.8354	0.0306	0.9534	0.0042
6. Gran Amount	0.8330	0.0235	0.8324	0.0229	0.8137	0.0227	0.9512	0.0062
7. Edges	0.8256	0.0093	0.8269	0.0107	0.8157	0.0617	0.9421	0.0062
8. Skin	0.8208	0.0222	0.8189	0.0270	0.7548	0.0771	0.8783	0.0243

(b) Comparing our proposed architecture with baselines for predicting the PWAT Depth Sub-score

deep learning architecture	input image	accuracy	F1 score	sensitivity	specificity
Our proposed method with various patch dimension					
1. Patch attention DenseNet (9 patches, context preserving attention)	512×512	**0.8537**	**0.8554**	**0.8739**	**0.9578**
2. Patch attention DenseNet (9 patches, context preserving attention)	256×256	0.8293	0.8250	0.8120	0.9459
DenseNet with other baseline attention mechanisms					
3. DenseNet (no patch, no attention)	256×256	0.8049	0.8060	0.8021	0.9447
4. DenseNet (9 patches, no attention) (average over 9 patches)	512×512	0.8110	0.8170	0.8046	0.9490
5. DenseNet (9 patches, no attention) (max over 9 patches)	512×512	0.8293	0.8298	0.8148	0.9514
6. DenseNet (9 patches, no attention) (mode over 9 patches)	512×512	0.8232	0.8249	0.8182	0.9498
Baseline CNN image classification architecture					
7. ResNet (no patch, no attention)	256×256	0.8110	0.8127	0.8043	0.9428

Patch Attention DenseNet model performed achieved a mean accuracy >80% for all PWAT sub-scores. The distribution of the 5 best and all results for all 8 PWAT sub-scores are mostly above 80%. For PWAT sub-scores 3. Nec Type and 5. Gran Type, the distribution range of the 5 best results and all 15 results were small and close to each other. For sub-scores 6 (Gran Amount) and 8 (Skin), the distribution range of the 5 best results and all 15 results were close to each other but large. For sub-score 7 (Edges), the distribution range of the 5 best results is small while the distribution range for its all 15 results were a little larger.

D. Comparison of Variations of Our Proposed Architecture With Baseline Methods

We also explored the effects of 1) varying the input image size on our proposed model's accuracy 2) removing the context-preserving attention mechanism from the model and 3) Using alternate methods of combining results of patches instead of our context-preserving method. Finally, Table III(b) shows the training results of these experiments for PWAT sub-score 2 (Depth). In experiment 1 shown in Table III(b), we used the standard Patch Attention DenseNet121 with 9 patches on 512×512 images. In experiment 2, we used the standard Patch Attention DenseNet121 with 9 patches on 256×256 images. In experiment 3, a standard DenseNet121 was used without patches and attention mechanism on 256×256 images. In experiment 4, 5 and 6, a standard DenseNet121 took 9 patches from the original images as input and estimated 9 results. Then we took the average, maximum and mode respectively from these 9 results as the final result for the original images, where:

$$\text{Average} = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n} \quad (10)$$

$$\text{Maximum} = \max\left(a_1, \ldots, a_n\right) \quad (11)$$

and the mode is the value that appears most frequently in the set of value. In experiment 7, a standard ResNet50 was used without patches and attention mechanism on 256×256 images. The results shows that the standard Patch Attention DenseNet121 with 9 patches on 512×512 images performed best when compared to other methods. DenseNet121 that used 9 patches from original images improved the performance a little when comparing to DenseNet121 without using patches. However, the standard Patch Attention DenseNet121 outperformed the DenseNet121 that used 9 patches from original images but no attention mechanism. These results demonstrate that by leveraging contextual information from other patches, our context-preserving attention further improves the model's performance. Table III(b) shows that larger size images and the patch attention mechanism both enhanced the Patch Attention DenseNet model's performance.

E. Investigating Data Augmentation

We found that data augmentation did not consistently improve our model's performance (See Table IV(a)). ResNet50's performance on some sub-scores improved (e.g. granulation amount) while others worsened (e.g. depth). This was consistent with our findings in our prior work [38], where we also found that data augmentation did not consistently improve Bi-CNN's [31]

TABLE IV
DATA AUGMENTATION RESULTS AND CONFUSION MATRICES

(a) Model comparison with and without data augmentation

Architecture	PWAT subscore	training accuracy	testing accuracy
1. ResNet50 no data augmentation	Depth	0.8153	0.8110
2. ResNet50 data augmentation	Depth	0.8648	0.7927
3. ResNet50 no data augmentation	Gran Amount	0.8275	0.8171
4. ResNet50 data augmentation	Gran Amount	0.8206	0.8476

(b) Confusion Matrices1

1. Size		actual class				
		0	1	2	3	4
prediction class	0	14	2	0	0	0
	1	1	35	6	0	0
	2	0	3	42	4	1
	3	0	0	3	25	3
	4	0	0	0	5	20

2. Depth		actual class				
		0	1	2	3	4
prediction class	0	14	0	0	1	0
	1	0	11	6	1	0
	2	0	3	68	5	0
	3	0	0	6	30	1
	4	0	0	0	1	17

3. Nec Type		actual class				
		0	1	2	3	4
prediction class	0	41	4	0	1	0
	1	6	41	0	1	5
	2	1	1	25	0	2
	3	0	0	1	6	0
	4	1	2	2	0	24

4. Nec Amount		actual class				
		0	1	2	3	4
prediction class	0	33	1	1	3	1
	1	1	13	1	1	1
	2	1	0	11	1	1
	3	0	1	0	19	2
	4	0	1	2	5	64

(c) Confusion Matrices2

5. Gran Type		actual class				
		0	1	2	3	4
prediction class	0	15	1	2	0	1
	1	0	29	5	2	2
	2	0	2	49	2	1
	3	0	0	2	5	1
	4	0	0	4	1	40

6. Gran Amount		actual class				
		0	1	2	3	4
prediction class	0	13	0	0	0	2
	1	0	6	0	1	2
	2	0	0	18	3	3
	3	0	2	0	26	4
	4	1	0	2	4	77

7. Edges		actual class				
		0	1	2	3	4
prediction class	0	14	1	2	0	0
	1	0	26	7	0	0
	2	1	5	84	3	0
	3	0	0	5	10	0
	4	0	0	2	1	3

8. Skin		actual class				
		0	1	2	3	4
prediction class	0	37	14	0		
	1	7	88	2		
	2	0	3	13		
	3					
	4					

performance on two PWAT sub-scores. Thus, we concluded that data augmentation may not effectively enhance our models' performance on our dataset and did not use it further.

F. Confusion Matrices

Table IV(b) and Table IV(c) show a sample of the confusion matrices of the test set results for all 8 PWAT sub-scores. Although there were 5 results from 5 folds in Table III(a), due to space constraints, we only show the confusion matrices of the best results of the 5 folds here. The numbers on the diagonal position represent images classified correctly in the confusion matrices. It is easily observed that the majority of test images are on the diagonals of the confusion matrices. The misclassified images mainly distributed beside the diagonal position. The confusion matrices and the fact that the F1 score, which accounts for imbalance does not vary significantly from accuracy show that the imbalanced data does not significantly affect our models' performance.

IV. DISCUSSION

Overall our proposed neural networks model was able to assess all 8 clinically important wound attributes with accuracies and F1 scores of over 0.8. This is encouraging and implies that at the click of a smartphone camera button, a non-expert wound nurse can receive clinically-valid, evidence-based wound assessments (scores) on which to base treatment. Our ablation rigorous study also shows that all aspects of the final model were useful including the context-preserving attention and the DenseNet backbone model. Our model performed well for PWAT sub-scores 3 (Nec Type) and 5 (Gran Type), which involved detecting a tissue type. This task is a texture pattern matching problem, which prior work focused on and performed well on. It also performed well for detecting the edges of wounds. Although edge detection task is well studied in the literature and neural networks have performed it well [39], our research focused on analyzing clinically-important ranges of the shape and tissues of the wound's edges. Our model found sub-scores in which the appearance of images within each sub-score bin can be quite varied challenging to produce consistent, stable results. These include Depth and surrounding skin viability sub-scores. Sub-score 1 (size) is also challenging because it involves depth perception from a 2D image, which is a somewhat ill-posed problem.

High-resolution wound images can facilitate the detection of fine-grained differences between wounds and should be utilized when designing high-performance models. However, utilizing high-resolution wound images with deep learning methods is challenging due to limitations on memory and computational resources. A simple alternative approach is to use small patches with multi-crop evaluation. Small, high-resolution crops are fed into standard models and the target label predictions from all crops are averaged. However, preserving the global context between individual patches is also important. We have extended the simple multi-crop, patch approach by also incorporating the global context between the local patches with a patch-based attention mechanism.

Our model performed well for larger image sizes, which could imply that we could achieve higher performance using larger patches as smartphone cameras get more precise and available GPU memory gets larger. The context-preserving patch-based attention mechanism improved results, which implies that our attention mechanism preserves global context, focuses on the most predictive parts of a wound image and enables effective global information exchange. However, data augmentation, a widely-used technique to enhance performance did not consistently improve our results.

V. CONCLUSION

We have proposed a deep learning photo-based method for comprehensive wound assessment that was based on a comprehensive, clinically-valid wound grading rubric. Prior work was limited because they 1) Assessed a single wound type, 2) Explored a few wound attributes, or 3) Focussed on the simpler binary classification task of detecting the presence (yes/no) of specific tissue types. We advanced the state of the art by classifying which clinically-important tissue sub-types (e.g. type of necrotic tissue) that occur in a wound, which is a challenging fine-grained image classification task. We proposed a Patch

Attention DenseNet deep learning architecture that is able to estimate all 8 PWAT sub-scores. We applied transfer learning to the Patch Attention DenseNet model to learn each of the 8 PWAT subscores separately The accuracy and the F1 score achieved by our Patch Attention DenseNet architecture shows that it generally works well on diverse wound types including diabetic foot ulcers, pressure ulcers, vascular ulcers and surgical wounds.

In future work, we will research methods to improve the PWAT-sub-scores our model did not perform well on, particularly Necrotic Amount. We will also explore other emerging deep learning architectures for predicting PWAT wound sub-scores.

References

[1] K. Järbrink *et al.*, "The humanistic and economic burden of chronic wounds: A protocol for a systematic review, "*Syst. Rev.*, vol. 6, no. 1, 2017, Art. no. 15.

[2] S. R. Nussbaum *et al.*, "An economic evaluation of the impact, cost, and medicare policy implications of chronic nonhealing wounds," *Value Health*, vol. 21, no. 1, pp. 27–32, 2018.

[3] L. Gould *et al.*, "Chronic wound repair and healing in older adults: Current status and future research," *Wound Repair Regeneration*, vol. 23, no. 1, pp. 1–13, 2015.

[4] C. K. Sen *et al.*, "Human skin wounds: A major and snowballing threat to public health and the economy," *Wound Repair Regeneration*, vol. 17, no. 6, pp. 763–771, 2009.

[5] M. A. Fonder, G. S. Lazarus, D. A. Cowan, B. Aronson-Cook, A. R. Kohli, and A. J. Mamelak, "Treating the chronic wound: A practical approach to the care of nonhealing wounds and wound care dressings," *J. Amer. Acad. Dermatol.*, vol. 58, no. 2, pp. 185–206, 2008.

[6] S. J. Landis, "Chronic wound infection and antimicrobial use," *Adv. skin wound care*, vol. 21, no. 11, pp. 531–540, 2008.

[7] H. Nejati, V. Pomponiu, T.-T. Do, Y. Zhou, S. Iravani, and N.-M. Cheung, "Smartphone and mobile image processing for assisted living: Health-monitoring apps powered by advanced mobile imaging algorithms," *IEEE Signal Process. Mag.*, vol. 33, no. 4, pp. 30–48, Jul. 2016.

[8] E. Sirazitdinova and T. M. Deserno, "System design for 3D wound imaging using low-cost mobile devices," in *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10138. International Society for Optics and Photonics, 2017, Art. no. 1013810.

[9] V. N. Shenoy, E. Foster, L. Aalami, B. Majeed, and O. Aalami, "Deepwound: Automated postoperative wound assessment and surgical site surveillance through convolutional neural networks," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2018, pp. 1017–1021.

[10] R. Niri, Y. Lucas, S. Treuillet, and H. Douzi, "Smartphone-based thermal imaging system for diabetic foot ulcer assessment," *Journées d'Etude sur la TéléSanté*, 2019.

[11] L. Wang, P. C. Pedersen, D. Strong, B. Tulu, and E. Agu, "Wound image analysis system for diabetics," in *Medical Imaging 2013: Image Processing*, vol. 8669, 2013, Art. no. 866924.

[12] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2098–2109, Sep. 2016.

[13] L. Wang, "System designs for diabetic foot ulcer image assessment," Doctoral dissertation, Worcester Polytechnic Inst., Worcester, MA, USA, 2016.

[14] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Boundary determination of foot ulcer images by applying the associative hierarchical random field framework," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 024002.

[15] D. Y. Chino, L. C. Scabora, M. T. Cazzolato, A. E. Jorge, C. Traina-Jr, and A. J. Traina, "Segmenting skin ulcers and measuring the wound area using deep convolutional networks," *Comput. Methods Programs Biomed.*, vol. 191, 2020, Art. no. 105376.

[16] D. Spinczyk and M. Wideł, "Surface area estimation for application of wound care," *Injury*, vol. 48, no. 3, pp. 653–658, 2017.

[17] J.-T. Hsu, T.-W. Ho, H.-F. Shih, C.-C. Chang, F. Lai, and J.-M. Wu, "Automatic wound infection interpretation for postoperative wound image," in *Proc. 8th Int. Conf. Graphic Image Process.*, vol. 10225. International Society for Optics and Photonics, 2017, Art. no. 1022526.

[18] G. Blanco *et al.*, "A superpixel-driven deep learning approach for the analysis of dermatological wounds," *Comput. Methods Programs Biomed.*, vol. 183, 2020, Art. no. 105079.

[19] V. Godeiro, J. S. Neto, B. Carvalho, B. Santana, J. Ferraz, and R. Gama, "Chronic wound tissue classification using convolutional networks and color space reduction," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process.*, 2018, pp. 1–6.

[20] H. Nejati *et al.*, "Fine-grained wound tissue analysis using deep neural network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 1010–1014.

[21] J.-T. Hsu *et al.*, "Chronic wound assessment and infection detection method," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, 2019, Art. no. 99.

[22] M. Maity, D. Dhane, C. Bar, C. Chakraborty, and J. Chatterjee, "Pixel-based supervised tissue classification of chronic wound images with deep autoencoder," in *Proc. Adv. Comput. Commun. Paradigms*. 2018, pp. 727–735.

[23] K. Babu, S. Sabut, and D. Nithya, "Efficient detection and classification of diabetic foot ulcer tissue using PSO technique," *Int. J. Eng. Technol.*, vol. 7, no, 3.12, pp. 1006–1010, 2018.

[24] V. Rajathi, R. Bhavani, and G. Wiselin Jiji, "Varicose ulcer (C6) wound image tissue classification using multidimensional convolutional neural networks," *Imag. Sci. J.*, vol. 67, no. 7, pp. 374–384, 2019.

[25] N. Gessert *et al.*, "Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 495–503, Feb. 2019.

[26] P. E. Houghton, C. B. Kincaid, K. E. Campbell, M. Woodbury, and D. Keast, "Photographic assessment of the appearance of chronic pressure and leg ulcers," *Ostomy Wound Manage.*, vol. 46, no. 4, pp. 20–35, 2000.

[27] P. E. Houghton *et al.*, "Effect of electrical stimulation on chronic leg ulcer size and appearance," *Phys. Ther.*, vol. 83, no. 1, pp. 17–28, 2003.

[28] N. Thompson, L. Gordey, H. Bowles, N. Parslow, and P. Houghton, "Reliability and validity of the revised photographic wound assessment tool on digital images taken of various types of chronic wounds," *Adv. Skin Wound Care*, vol. 26, no. 8, pp. 360–373, 2013.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[31] T.-Y. Lin, A. Roy Chowdhury, and S. Maji, "Bilinear cnn models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1449–1457.

[32] A. Wagh, "Semantic segmentation of smartphone wound images: Comparative analysis of AHRF and CNN-based approaches Wpi technical report Wpi-cs-tr-19-02," *IEEE Access*, vol. 8, pp. 181590–181604, 2020.

[33] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 616–625.

[34] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[35] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.

[36] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *Proc. Int. Conf. Artif. Neural Netw.*, 2017, pp. 626–634.

[37] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Berlin, Germany: Springer, 2018, pp. 303–311.

[38] X. Zhao *et al.*, "Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network," *IEEE Access*, vol. 7, pp. 179151–179162, 2019.

[39] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3000–3009.