Design of a Smartphone Application for Automated Wound Measurements in Home Care

Joshua Budman, Kevin Keenahan, Soumyadipta Acharya, Gabriel A Brat

Connected Health Symposium 2015

www.iproc.org/2015/1/e16

DOI: http://doi.org/10.2196/iproc.4703

Abstract

There are 8.5 million Americans who suffer from a chronic wound. Due to the lack of an objective system to measure and characterize wounds, the current standard of care relies highly on provider guesswork. This leads to misinformed care decisions, prolonged healing times, and high healthcare expenditures. Further complicating this process is the need for technologies that can readily be deployed among the thousands of nurses who visit patients in the home environment—the most common site of wound care. This study describes the design and validation of a smartphone image-based system for accurately measuring and characterizing chronic wounds in an automated and objective fashion.

Photos (n=81) were collected by the study team from patients (n=25) at the Johns Hopkins Bayview Wound Clinic in an IRB-approved study. Photos were taken using a variety of smartphones such that our training data set would include nuances of different smartphone cameras. We combined supervised image classification and computer vision to detect wound edges and segment the tissues within the wound. Fifteen individual raters with various levels of training were instructed to trace wound regions in a diverse subset of the wound images arbitrarily selected by the study team (n=10). The ensemble wound edge and tissue segmentation algorithms were compared against an 80% inter-rater gold standard.

The automated method resulted in a sensitivity = 98.31 ± 2.18 and specificity = 92.06 ± 7.86). In contrast, the ruler-based measurement resulted in sensitivity = 1 ± 0 and specificity = 0.57 ± 0.30. A normalized area measurement for the automated method resulted in a normalized area of 1.14 ± 0.17. In comparison, the standard of care method resulted in a normalized area of 1.86 ± 0.30 relative to gold standard. With respect to tissue segmentation, the overall average tissue classification accuracy on k-fold cross validation using the sparse neural network method is 93.6% ± 3.3%.

The result illustrates the large overestimation of wound size that occurred when the wounds were

measured using the ruler measurement. It also corroborates the literature-reported value of measurement inaccuracy by standard methods. Our study shows the ability of an easily deployed smartphone system to classify wounds in an automated manner with high accuracy. Such a system could be used to objectify measurements by nurses in the home care environment.

Introduction

A chronic wound is defined as an open sore that remains unhealed for more than 6 weeks (Figure 1) [1]. Each year, 8.5 million Americans suffer from this condition [4]. Of these patients, approximately 3-4 million have diabetic ulcers, 2.5-3 million have pressure ulcers and 2-2.5 million patients will have venous ulcers or arterial ulcers [5]. Pressure and diabetic ulcers form the most common subtype. Surgical wounds, venous stasis ulcers, and arterial ulcers also have an important effect [2]. Diabetics, obese individuals, or smokers are most at-risk [3]. It is essential that these wounds be monitored and treated effectively to prevent readmissions and significant morbidity.

The U.S. healthcare system spends approximately $30 billion each year to treat chronic wounds [4]. Treatment of these conditions is so expensive because, in part, tracking treatment efficacy over time is difficult. The main metrics used to track healing are change in wound area and tissue composition over time. Existing literature states that the current standard of care measurement technique to gather these metrics, which involves the use of a ruler and naked eye approximation, has a 44% error rate [6]. Moreover, literature reports that the inter- rater measurement error, that is the deviation that takes place when two separate providers measure the same wound, ranges between 16 and 50% [7]. To estimate tissue composition, providers currently rely on naked eye estimates of wound bed composition. In one study, nurses working with paralyzed patients frequently characterized pressure ulcer surface composition incorrectly (kappa = 0.33) [8]. This erroneous classification often led to improper staging of these pressure ulcers.

To combat the inaccuracy of ruler-based measurements, other methods have been developed to measure wound surface area, namely manual planimetry using acetate film and digital planimetry using digital photography. The former involves placing acetate film with a grid on top of the wound and having a clinician trace the boundary while the latter involves acquiring a digital image of a wound with a reference and having a clinician trace the boundary of the wound using software programs. In addition to being time consuming and impractical for use in a home care setting, literature has shown that there is inaccuracy and variability associated with both of these methods. One study conducted showed that using acetate planimetry can miscalculate wound area by 22% [9]. While digital planimetry has been established as the most accurate determinant of wound boundaries, there is still 10-35% error associated with its ability to calculate wound area due to camera angle skew [10, 11]. Additionally, due to its subjective nature, there is still a 3.8% inter-rater variability associated with digital planimetry [11].

There is thus a general need for a standardized method to measure wound area and tissue composition in a simple, automated fashion. Protocols established by the American Professional Wound Care Association (APWCA) state that providers must evaluate their patients' treatments every 4 weeks based on the change in wound area and tissue composition [12]. This suggests that the lack of an objective and accurate method for measuring and documenting changes in wound area and tissue composition longitudinally leads to delayed or premature changes. Considering the ubiquity of smartphone technology in the field of healthcare, a high-fidelity image-based method for performing wound analysis would thus be very valuable.

The study team hypothesized that a combined machine learning and computer vision approach would be required to form a high-accuracy, low-variability wound analysis system. This approach would involve collecting sufficient wound image data to classify wounds, forming smaller homogenized data sets, before applying edge detection methods. The study team conducted an IRB-approved study at the Johns Hopkins Bayview Wound Center, which involved collecting photographs of patients' wounds without any identifiable information. Overall, 81 photographs were collected from 25 patients. Using this dataset as a training set for segmentation and classification, the study team produced an automated method that could classify wound edges and measure wound area with a very high level of accuracy and low variability. This automated method also has the ability to segment and classify wound tissues with a high degree of accuracy. This paper will discuss these methods and propose an application to a telehealth system for wound patients.

Methods

A. Data Set

Photos (n=81) were collected by the study team from patients (n=25) at the Johns Hopkins Bayview Wound Clinic in an IRB-approved study (IRB00039125). Patients enrolled in the study had a variety of chronic wound etiologies. Photos were taken using a variety of smartphones such that our training data set would include nuances of different smartphone cameras. The smartphones used for the study were an iPhone4, a Galaxy S4 and a Galaxy S4 Zoom. As part of the study protocol, patients were consented to have their photographs taken prior to their appointment. If positive consent was received, the study team collected general patient information and also collected patients' actual wound measurements using the current standard of care. Before images were taken, the study team placed a small green dot of a known radius of 0.375 in^2 next to the patients' wounds (Figure 2). This dot established a reference object of known size such that wound and tissue areas could be calculated with a high level of accuracy. The dot would thus allow the study team to normalize for environmental conditions such as irregular room lighting, distance from the wound, and camera skew angle.

B. Photo Training

After the study was completed, the study team built a Matlab 2014a(Mathworks, Natick, MA) platform that allowed individuals to manually classify regions using digital planimetry. Specifically, 15 raters were instructed to trace wound regions in a diverse subset of the wound images selected by the study team (n=10). Images were selected based on the following criteria:

1). A single wound area in the frame of the image

2). A mixture of different wound shapes and surface compositions

Raters varied in experience from physicians that treated wounds in routine practice to members of the study team trained to analyze wounds. The large number of raters (15) was used in order to minimize the inter-rater variability associated with digital planimetry. Pixels that were enclosed by the trace of 80% of these individuals or higher were classified as reference wound pixels. In other words,

$$p_{wound} = ceil\left(\frac{\sum_{m=1}^{s=15} p_{ij}}{s} - T + \epsilon\right) \qquad (1)$$

In (1), it is assumed that a p_{wound} value of 1 is classified as a pixel belonging to the wound while a value of 0 is one not belonging to the wound. Further, in this equation T has a value of 0.8 based on the 80% threshold selected.

This expert system was created to establish a gold standard measurement technique to account for the intrinsic flaws with the current standard. After the photo collection was completed, regions of the green dot in each photo were manually traced by the study team. For future classification purposes, the study team also classified each wound as being either "Dark" or "Light" based on the wound bed composition and also stated whether the wound bed contained one tissue type or more than one tissue type.

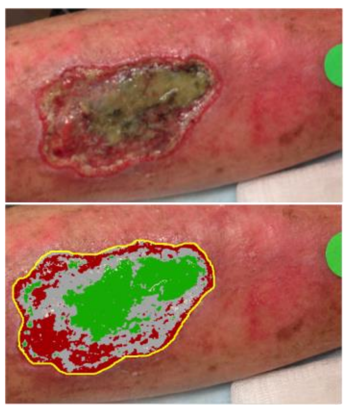
The study team also manually traced specific wound regions in order to enable supervised tissue segmentation and classification. Specifically, the study team traced regions of granulation, necrosis, slough, epithelial tissue, healthy tissue and tendon.

C. Image Analysis for Wound Boundary Detection

The study team used Matlab to formulate and optimize an image analysis system for accurately identifying the boundaries of the wound--segmenting the wound area from the background. Classification algorithms involved an ensemble of state of the art edge detection and boundary segmentation techniques. As the segmentation techniques used in this study were purely bottom-up with no reliance on a training set, there was no risk of training-test data contamination in the analysis. Before applying our algorithms, the 10 images that were analyzed by the raters as described above were pre-processed using a standard image pre-processing routine. This involved erosion, dilation, smoothing, median filtering and color correction in order to remove artifacts and prime the images for analysis. After pre-processing, the boundary detection algorithm was applied to the image set and the sensitivity and specificity of the wound detection was determined. Sensitivity was defined as the percentage of pixels inside the wound that were appropriately classified by the automated algorithms while specificity was the percentage of pixels outside the wound that were correctly classified. The experimental ground truth was considered to be the 80% agreement of the 15 raters as described above. The automated method was compared to the nurses' ruler measurements, which were also collected for the study. Figure 1 shows a sample result of the automated analysis.

D. Image Analysis for Tissue Segmentation

Using the tracing module described previously, the study team traced and classified specific tissue regions within the wound images. Tissue types were classified as one of 5 categories: slough, necrosis, granulation, healthy epithelium, and intact skin. Although the wound boundary segmentation was intended to segment wound tissue from intact skin, the intact skin class was included in the classifier with the idea of correcting for cases where the segmentation did not perform correctly. A large feature set was extracted from each of these tissue regions (n=404) and a sparse neural network-based classification system was constructed. An 80-20 training/cross validation split was performed a total of 10 times, and accuracy is reported as the average percentage of tissue regions in the cross validation set that were classified correctly. Figure 1 below also shows a sample result of successful tissue segmentation.



Tissue	Area (cm^2)	%
Necrosis	0.0	0.0
Granulation	5.55	21.1
Slough	7.03	28.8
Other	13.68	52.1
Total	**26.25**	**100.0**

Figure 1. Sample analysis of lower extremity ulcer on a diabetic patient using the automated image-based wound analysis tool.

Results

A. Wound Boundary Segmentation

For this part of the study, the automated wound boundary segmentation method was compared to the nurses' standard of care, ruler-based measurements on the ten wounds that were rated as described above. Using the definition of sensitivity and specificity described above, the automated method greatly outperformed the standard of care ruler measurements. The comparison was made using F-scores for each of the methods using the experimentally determined gold standard as described previously. The automated method resulted in an average F-score of 0.95 ± 0.041 (Sensitivity = 98.31 ± 2.18, Specificity = 92.06 ± 7.86). In contrast, the ruler-based measurement resulted in an F-score of 0.73 ± 0.066 (Sensitivity = 1 ± 0, Specificity = 0.57 ± 0.30). This number corroborates the literature-reported value of 44% measurement inaccuracy associated with ruler-based area measurements.

Additionally, a normalized area measurement was calculated in order to compare the automated method and the ruler measurement. The automated method resulted in a normalized area of 1.14 ± 0.17. In comparison, the standard of care method resulted in a normalized area of 1.86 ± 0.30. This result is summarized in the boxplot of Figure 2.

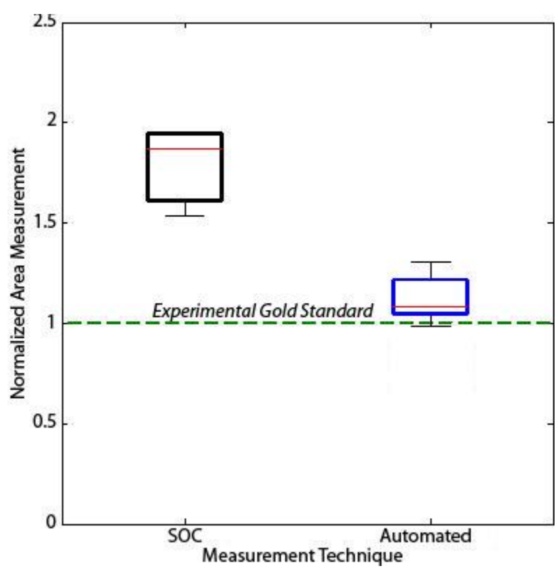


Figure 2. Boxplot comparing normalized area measurement of standard of care result to automated method.

B. Tissue Segmentation

As described in the methods section, the study team did not have a method for validating results of the tissue segmentation using the rater-based system. The average accuracy is reported as the correct classification of tissue segments in the k-fold cross-validation. The overall average tissue classification accuracy on k-fold cross validation using the sparse neural network method is 93.6% ± 3.3%. The class-by-class accuracy is summarized in Table 1.

Table 1. Breakdown of the results of the tissue classification in the k-fold cross-validation.

Class	**Mean Accuracy**	**Standard Deviation**
Granulation	95.68	2.19
Necrosis	97.78	1.72
Epithelium	93.59	3.23
Healthy	89.39	3.82
Slough	91.58	3.97
Overall	**93.60**	**3.30**

Discussion

Wound size and composition are the main characteristics that clinicians use to determine wound healing. Thus, it is essential that the measurement tool used is highly accurate and repeatable. Literature has shown that the current standards of care, ruler measurements for wound size and eyesight for tissue composition, provide unreliable levels of accuracy. This study discusses an automated method for determining wound size and composition with demonstrated accuracy far superior to that of the current standard of care. It delineates the potential clinical relevance of automated, mobile software-based methods in a clinical setting. With this technology, clinician can receive reliable objective data about the state of a patient's wound and track longitudinal changes. Additionally, it creates the potential to objectively analyze treatment effects and modifications. Future work will include further refinement of the algorithms and ensembles. The study team has recently acquired a set of 70,000 additional labeled photographs for the purpose of improving the quality of analysis and validating the accuracy of the automated system.

References

1. Mustoe, T A., K O'Shaughnessy, and O Kloeters. "Chronic wound pathogenesis and current treatment strategies: a unifying hypothesis." *Plastic Reconstructive Surgery* 117.7 (2006): 35-41. Print.
2. Werdin, Frank, Mayer Tennenhaus, Hans-Eberhart Schaller, and Hans-Oliver Rennekampf. "Evidence-based Management Strategies for Treatment of Chronic Wounds." *Open Access Journal of Plastic Surgery* 9.19 (2009). Print.
3. Berlowitz, Dan et al. "Preventing Pressure Ulcers in Hospitals: A Toolkit for Improving Quality of Care." *AHRQ* : 1-156. Print.
4. Sen, C et al. "Human skin wounds: a major and snowballing threat to public health and the economy." *Wound Repair Regeneration* 17.6 (2009): 763-71. Print.
5. "Markets for Advanced Wound Care Technologies." *BCC Research* (2011). Print.
6. Measuring wound length, width, and area: which technique? Langemo D, Anderson J, Hanson D, Hunter S, Thompson P. Advances in Skin & Wound Care, January 2008, 21(1): 42-45 1879-1882.
7. Koel, Gerard, and Frits Oosterveld. ”Reproducibility of Current Wound Surface Measurement.” European Wound Management Conference Pro- ceeding (2008). Print.
8. Beeckman, D et al. “EPUAP Classification System for Pressure Ulcers: European Reliability Study”. *Journal of Advanced Nursing* 60.6 (2007): 682-691. Print.
9. Shaw, J, Bell, P. Wound Measurement in diabetic foot ulceration, from Global Perspective on Diabetic Foot Ulcerations”, edited by Thanh Dinh, ISBN 978-953-307-727-7 December, 2011: under CC BY 3.0 license
10. Mayrovitz, H. Documenting wound areas via digital photography: effect of camera angle on area accuracy, Presented to WOCN Society 40th Annual conference, June 2008.
11. Rennert, R, Golinko, M, Kaplan, D, et al. Standardization of wound photography using the wound electronic medical record. Advances in Skin and Wound Care 22(1).