Diabetic Complications Consortium

Application Title: Image Processing and Machine Learning for the Prediction of Wound Healing

Principal Investigator: James Wrobel, DPM, MS

1. Project Accomplishments:

We have developed a pipeline for segmenting and processing diabetic foot ulcer (DFU) images to predict the binary outcome of wound healing using both the support vector machine (SVM) and random forest (RF) models, trained with electronic health records (EHR) and the features generated from the images.

Imaging Records:

Preprocessing: First, we reduced the noise due to light reflection by applying a mean filter, since the degree of illumination is heterogeneous across the DFU images. A threshold was used to select the pixels with the highest brightness. Then, the kernel K = [1; 1; 1; 1; 0; 1; 1; 1; 1] was generated and convolved over all of the selected pixels, such that the "whitest" spots were substituted with the mean value from all neighboring pixels. In addition, color constancy was used to normalize the color scheme of the images as well [1].

Segmentation: Three methods were compared: a RGB-Support Vector Machine (RGB-SVM) [2], a patch-based convolutional neural network (CNN) [3] and U-Net [4]. The output probability map of the convolutional neural network was postprocessed into a binary mask. For further details on segmentation, please refer to our publication [5].

Feature Generation: From the segmented images of the ulcers, the following features were extracted: mean and standard deviation of each color channel in the RGB, HSV, and LAB color spaces, Gray-level Co-occurrence Matrix (GLCM) [6], Gray-level Run Length Matrix (GLRLM) [6], Gray-level Size Zone Matrix (GLSZM) [6], Neighborhood Grey Tone Difference Matrix (NGTDM) [6], and Gabor filters of four different orientations and wavelengths each, resulting in 16 different filters.

Electronic Health Records:

The EHR of 2291 visits for 381 ulcers from 155 patients who visited the Michigan Medicine Podiatry clinic for diabetic foot ulcers and/or complications thereof were collected from its EHR system, MiChart. Each ulcer was then tracked to see if it had healed eventually at some point, with those lost to follow up being excluded from the study. If an ulcer previously marked as healed re-opened at the same location it was excluded from the study. Any wound without an image at the initial visit was also excluded.

Clinical data processing: The following clinical features from the initial visit of each patient were collected from the EHR: *wound length, width and depth, foot, age, race, gender,*

ethnicity, Charlson comorbidity index, diabetic retinopathy, absolute lymphocyte count, HbA1c, albumin, prealbumin, eGFR, C-reactive protein, erythrocyte sedimentation rate, wound infected, probe to bone test result, X-ray, MRI and/or Tc99 scan signs of osteomyelitis, dorsalis pedis and posterior tibial pulse, ankle and toe systolic pressure, TcPO2, body surface area, body mass index, total contact cast use, offload use, UTSA score, and medication use (immunosuppressants, oral steroids, antihypertensive, oral hypoglycemics, canagliflozin, insulin, heparin, allopurinol, NSAIDS, aspirin, warfarin, and Xa inhibitors). All of the non-numeric features were discretized – for example, categorical features (e.g., medication use) were binarized with a one-hot encoding, and the UTSA grade was separated into the score and grade. eGFR was discretized into different CKD stages. Most of the patient records had portions of clinical parameters missing. The missing clinical parameters were imputed as the Euclidean distance-weighted mean of the three most similar data points (n=3) using the *k*-nearest neighbor (kNN) algorithm [7].

Model building: Random forest (RF) and support vector machine (SVM) with an RBF kernel models were trained using a total of 133 features (48 clinical and 85 image features). The dimension of these features was reduced by principal component analysis (PCA). Grid search was then performed with 5-fold cross validation to find the model with the best area under the receiver operating characteristic (AUROC) curve. For the RF model, 5760 combinations of hyperparameters were evaluated, including the use of bootstrapping, tree selection criteria (Gini impurity or entropy), the maximum depth, the minimum number of samples in a leaf, the minimum number of samples required to split a node, the number of trees, and the number of principal components selected from PCA. For the SVM model, 676 combinations of hyperparameters were evaluated, including C, the penalty parameter of the error term; gamma, the kernel coefficient; and the number of principal components from PCA.

2. Specific Aims:

Develop image processing models that extract surface area, wound base color and surrounding skin to predict healing and infection in DFU treated with TCC.
a. Hypothesis 1: We postulate that the image processing and machine learning model will be capable of accurately predicting wound infection and healing by 20 weeks by integrating deep learning, computer vision, and image processing techniques.

b. Hypothesis 2: We hypothesize our test set image processing and machine learning models will provide more predictive information in terms of healing over surface area changes alone.

Results:

Segmentation: Figure 1 qualitatively depicts the developed pipeline. The preprocessing reduced the bright spots and normalized illumination in the image. The probability map generated by CNN detects the rough outline of the wound region. The postprocessing method further refines segmentation into a binary mask. The performances of each method were compared with respect to precision, sensitivity, specificity, accuracy, mean intersection over union (IoU), Dice similarity coefficient (Dice), and Matthews Correlation Coefficient (MCC) (Table 1, 2).

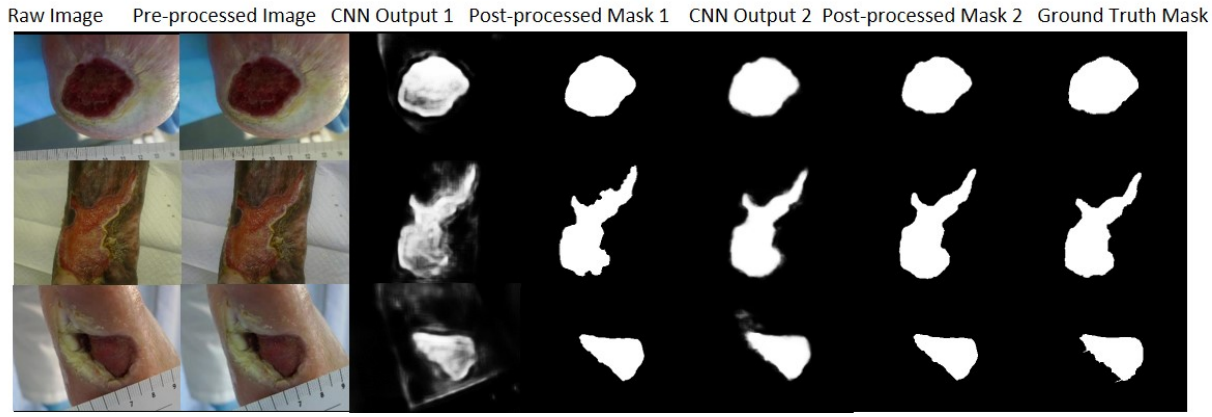

Figure 1. Segmentation of three images as examples. CNN Output 1 is the result of the patch-based CNN and CNN Output 2 the result of U-Net.

Table 1. Comparison of wound segmentation models before post-processing.

Method	Precision	Sensitivity	Specificity	Accuracy	Mean IoU	Dice	MCC
SVM(RGB)	0.564	0.806	0.896	0.877	0.472	0.596	0.594
Patch-based CNN	0.646	0.854	0.932	0.913	0.569	0.700	0.675
U-Net	0.768	0.937	0.960	0.949	0.723	0.823	0.840

Table 2. Comparison of wound segmentation models after post-processing.

Method	Precision	Sensitivity	Specificity	Accuracy	Mean IoU	Dice	MCC
Patch-based CNN	0.722	0.9	0.947	0.934	0.660	0.770	0.753
U-Net	0.830	0.917	0.973	0.966	0.761	0.845	0.839

Prediction of healing with image features: Due to the paucity of records containing the exact date on which individual ulcers healed completely, whether or not the ulcer had eventually healed at any given point was used as our prediction endpoint of healing instead of 20 weeks. If an ulcer previously marked as healed re-opened at the same location it was excluded from the study. Using the EHR and images from the baseline visits of 208 ulcers from 113 patients, SVM and RF predictive models were trained. Of these patients, 20% of the wounds were held out as the test set, and 5-fold cross validation was performed on the remaining 80%. Using image features alone, the 5-fold cross validation resulted in an AUROC of 0.727 +/- 0.087 for the SVM model and 0.667 +/- 0.054 for the RF model.

On the other hand, prediction of healing based on surface area alone (i.e., wound length, width, and depth parameters from the clinical records) resulted in an AUROC of 0.728 +/- 0.045 for the SVM model and 0.668 +/- 0.082 for the RF model, which are very comparable to using the image features alone.

When we combined all of the clinical parameters and image features, the AUROC was 0.734 +/- 0.024 for the SVM model and 0.675 +/- 0.056 for RF, which were better than using either the surface area changes or image features alone.

2. Analyze predictive ability of image processing model + clinical features to predict healing and infection in DFU treated with TCC.

c. Hypothesis 3: We hypothesize our machine learning models will provide significant predictive information by integrating the known clinical features (Brownrigg, Hinchliffe et al. 2016) and surface area changes with advanced image processing factors (e.g. texture measures).

Results:

c. Hypothesis 3: In our dataset, 22 unique wounds from 16 different patients who used the total contact cast at any given point were identified. Given this small dataset, we performed 5-fold cross validation on the entire 22 wounds (rather than excluding 20% as the hold-out set). Since the start of TCC use varies by patient, the features from the earliest visit marked as TCC were used for each individual wound. The time points at which the TCC use started varied from 0 (initial visit) to 385 days after the initial visit (mean 101.5 +/- 128.4 days). Using every available feature, the AUROC for the SVM model was 0.720 +/- 0.117 and that for the RF model was 0.864 +/- 0.164. Despite the scarcity of the data, models using the same features demonstrated a remarkably comparable or better performance in this subset of patients who utilized TCC as part of their treatment routine.

To determine whether or not the TCC group's outcome significantly differed from the non-TCC treatment group with respect to the healing outcome, the contingency table between the TCC and non-TCC group was created (Table 3). A total of 137 plantar ulcers were identified, with 20 of those treated with TCC. Overall, the TCC treatment group had a 65% healing rate and the non-TCC treatment group had a 73% healing rate. The P-value of the Fisher's exact test was 0.5922.

Table 3. Contingency table of the healing outcome in the TCC-treatment group and the control group. P = 0.5922 on Fisher's exact test.

	TCC	Control
Healed	13	85
Not Healed	7	32

3. Publications:

Cui C, Thurnhofer-hemsi K, Soroushmehr R, Mishra A, Gryak J, Dom E, et al. Diabetic Wound Segmentation using Convolutional Neural Networks. EMBC 2019. (In Print)

4. Bibliography:

[1] Fidalgo Barata A, Celebi E, Marques J. Improving Dermoscopy Image Classification Using Color Constancy. IEEE J Biomed Heal Informatics 2014:1–1. doi:10.1109/JBHI.2014.2336473.

[2] Changhan Wang, Xinchen Yan, Smith M, Kochhar K, Rubin M, Warren SM, et al. A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.,

IEEE; 2015, p. 2415–8. doi:10.1109/EMBC.2015.7318881.
[3] Jafari MH, Karimi N, Nasr-Esfahani E, Samavi S, Soroushmehr SMR, Ward K, et al. Skin lesion segmentation in clinical images using deep learning. 2016 23rd Int. Conf. Pattern Recognit., IEEE; 2016, p. 337–42. doi:10.1109/ICPR.2016.7899656.
[4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, Springer, Cham; 2015, p. 234–41. doi:10.1007/978-3-319-24574-4_28.
[5] Cui C, Thurnhofer-hemsi K, Soroushmehr R, Mishra A, Gryak J, Dom E, et al. Diabetic Wound Segmentation using Convolutional Neural Networks. EMBC 2019.
[6] Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 2015;60:5471–96. doi:10.1088/0031-9155/60/14/5471.
[7] Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28:112–8. doi:10.1093/bioinformatics/btr597.