Use of Artificial Intelligence to Improve the Calculation of Percent Adhesion for Transdermal and Topical Delivery Systems

Chao Wang*, Caroline Strasinger^, Yu-Ting Weng*, Xutong Zhao*

FDA/CDER/OTS/OB/DBVI,

^ FDA/CDER/OPQ/OPQAII/DPQAVII

Please visit https://doi.org/10.1007/s10916-023-02027-x for the published paper and more details.

Abstract

- **Adhesion is a critical quality attribute and performance characteristic for transdermal and topical delivery systems (TDS). Regulatory agencies recommend in vivo skin adhesion studies to support the approval of TDS in both new drug applications and abbreviated new drug applications.**
- **The current assessment approach in such studies is based on the visual observation of the *percent adhesion*, defined as *the ratio of the area of TDS attached to the skin to the total area of the TDS*.**
- **Visually estimated percent adhesion by trained clinicians or trial participants creates variability and bias. In addition, trial participants are typically confined to clinical centers during the entire product wear period, which may lead to challenges when translating adhesion performance to the real-world setting.**
- **In this work we propose to use artificial intelligence and mobile technologies to aid and automate the collection of photographic evidence and estimation of percent adhesion.**
- **We trained state-of-art deep learning models with advanced techniques and in-house curated data. Results indicate good performance from the trained models.**

Introduction

- Skin adhesion of TDS is essential for safe and effective use of the respective drug products. Existing evaluation methods for skin adhesion are typically visual assessment by trained technicians or subjects. In in-clinic studies, subjects may not move in the same way as in real life, thus the results may not represent the real-world performance. The visual assessment may be biased and can be subject to perception errors.
- We propose to aid/automate the collection and calculation of percent adhesion during product wear using mobile technology and artificial intelligence. In this to be designed mobile app, a TDS user would be guided to capture high quality photographs of the product adhered to the body at various timepoints during the prescribed wear period. With the given photos, the app can then employ a deep learning model to segment the TDS region in the image, identifying the attached and detached parts of the TDS. Under some assumptions, the percent adhesion can be estimated by the ratio of the total number of pixels representing the part of the TDS that adheres to skin to the total number of pixels of the TDS.
- As development of a mobile app to capture and store photographs is a relatively simple and well-established task, this work focuses on the development of the segmentation model.

Materials and methods

- 796 images were extracted from regulatory submissions of in vivo adhesion studies and manually annotated using CVAT.
- All pixels were classified with the following labels: Background, All Other, Record Card/Sticker, Skin, TDS Adhered to Skin, TDS Detached Above More TDS, TDS Detached Above Skin.
- Data preprocessing:
  1) resizing;
  2) data augmentation (during training and/or testing);
  3) focus crop.
- Models used:
  1) Fully-Convolutional Network (FCN) model with ResNet-50 or ResNet-101 backbones;
  2) U-net, with ResNet-50, ResNet-101 backbones;
  3) Unet++, with ResNet-50, ResNet-101 backbones;
  4) DeeplabV3+, with ResNet-101 backbone.
- Loss functions:
  1) Cross-entropy (with and without weights);
  2) Focal loss;
  3) Focal Tversky loss
- The models were trained with weights initialized by pretrained weights from the ImageNet dataset.
- All models were optimized via the stochastic gradient descent with learning rate scheduled by one cycle learning rate with cosine anneal strategy.
- The whole dataset was split into training (80%), validation (10%), and test (10%) datasets.
- The maximum number of epochs were set to be 200 and a training would be stopped early if the mIoU for validation data did not improve for 30 epochs.
- Both accuracy and mean intersection over union (mIoU) were reported as performance metrics for all training configurations. However, the mIoU is the preferred metric due to class imbalance.

Results and discussion

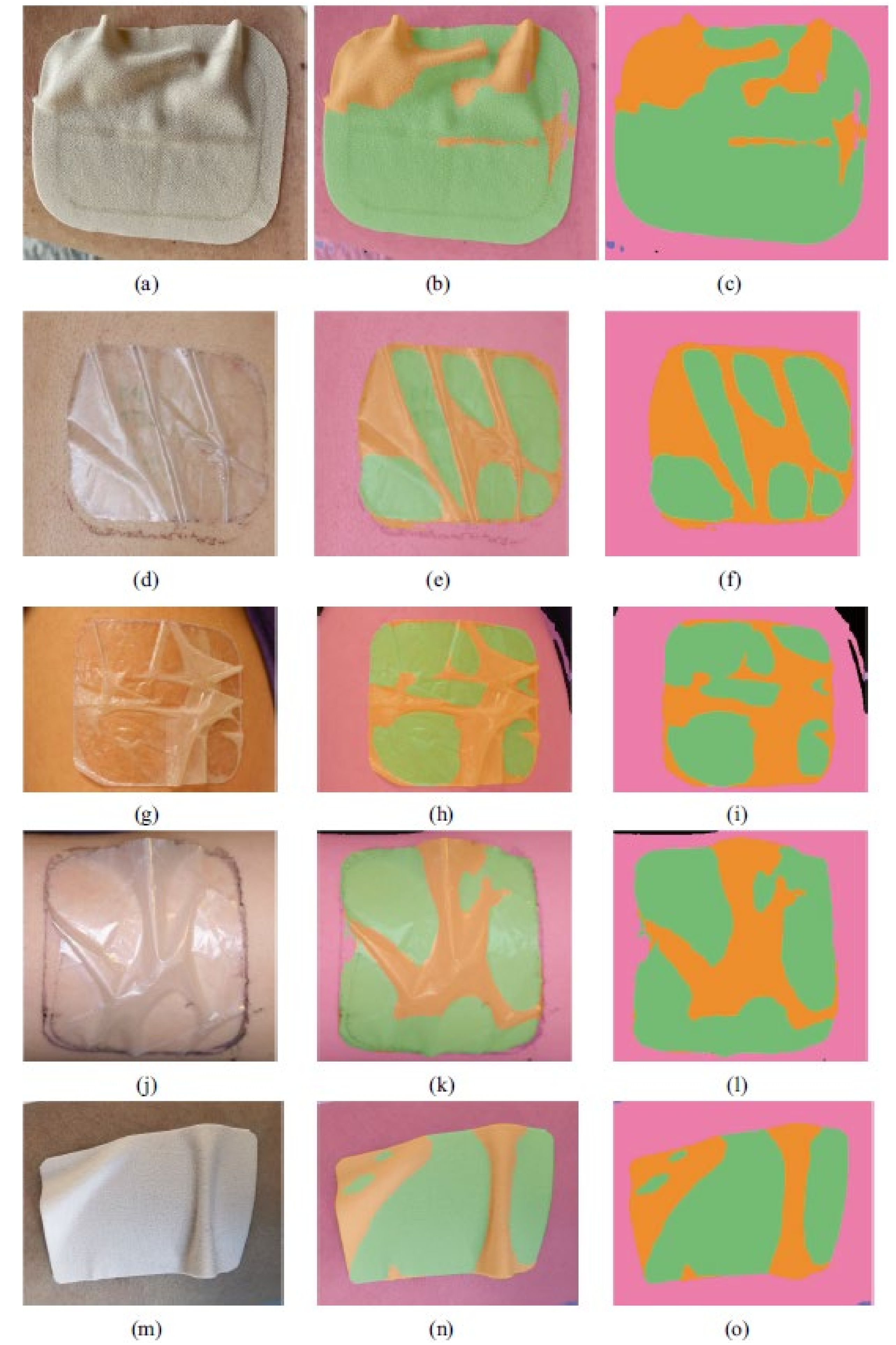
- Various configurations were used to train different models. The results are summarized in Table 1. Each configuration was replicated 10 times with different random seeds and the reported values are the average of the corresponding values.
- The training began with the Resnet-50 FCN model with the unweighted CE loss function and no TDA, TTA, or WTTA. Then more complex models, and/or training techniques were added to the training process to improve performance. The Resnet-50 FCN model without any augmentation and trained with unweighted CE loss yields 0.643 mIoU for test data.
- Further trained with TDA, TTA, WTTA, and weighted CE loss with more weight on detached and attached TDS parts improved the mIoU for test data to 0.670. Replacing the Resnet-50 FCN model by the U-net structure with a Resnet-50 backbone improved the mIoU to 0.675. However, replacing the weighted CE loss by either focal loss or focal Tversky loss did not improve the mIoU. Similar observations for models based on Resnet-101 can be made.
- The U-net with the Resnet-101 backbone trained with TDA, TTA, WTTA, and weighted CE loss yields the highest mIoU for test data, 0.675, which is slightly more than the U-net with Resnet-50 backbone but not significant up to three decimal digits. More complex model such as Unet++ and DeepLabV3+ did not show further improvement over U-net with the Resnet-101 backbone.
- The segmentation performance for the best model (U-net with the Resnet-101 back-bone trained with TDA, TTA, WTTA, and weighted CE loss) is illustrated in the figure on the right. Each row shows the original image (left), original image with segmentation overlay (middle) and segmentation map (right).

Table 1 Summary of training results. For each combination of model and training configuration (TDA, TTA, WTTA, and/or loss function), the training process were replicated 10 times using different random seeds. The reported number of epochs, time (in hours) needed for training, performance measures (accuracy and mIoU) for training, validation, and test data were averages over the 10 replicated runs.

Model	TDA	TTA	WTTA	Loss function	# epochs	Time	Performance					
							Training data		Validation data		Test data	
							Accurary	mIoU	Accurary	mIoU	Accurary	mIoU
resnet50_fcn	No	No	No	ce_unweighted	105.1	1.3	0.987	0.671	0.956	0.691	0.954	0.643
resnet50_fcn	No	No	No	ce_weighted	119.8	1.4	0.984	0.661	0.956	0.691	0.952	0.658
resnet50_fcn	Yes	Yes	Yes	ce_weighted	117.8	23.0	0.972	0.659	0.950	0.694	0.948	0.670
unet resnet50	No	No	No	ce_weighted	117.4	1.0	0.985	0.666	0.955	0.672	0.952	0.657
unet resnet50	Yes	Yes	Yes	ce_weighted	116.0	13.7	0.973	0.663	0.951	0.697	0.949	0.675
unet resnet50	Yes	Yes	Yes	focal_loss	116.0	15.5	0.974	0.597	0.949	0.685	0.950	0.657
unet resnet50	Yes	Yes	Yes	focal_tversky	127.7	14.4	0.929	0.512	0.911	0.528	0.926	0.509
resnet101_fcn	Yes	Yes	Yes	ce_weighted	80.8	16.5	0.969	0.581	0.945	0.673	0.947	0.660
unet resnet101	Yes	Yes	Yes	ce_weighted	105.6	14.2	0.973	0.668	0.949	0.695	0.948	0.675
unet resnet101	Yes	Yes	Yes	focal_loss	111.4	16.6	0.975	0.595	0.950	0.681	0.950	0.660
unet++ resnet101	Yes	Yes	Yes	ce_weighted	109.1	24.5	0.967	0.587	0.939	0.664	0.939	0.653
DeepLabV3+ resnet101	Yes	Yes	Yes	ce_weighted	109.1	24.5	0.967	0.587	0.939	0.664	0.939	0.653

Note:

- Sample standard deviations were not reported due to space limitation and less than 0.03 for accuary and mIoU.
- Each training run was done with a single Tesla V100-PCIE-32GB GPU.
- Abbreviations: TDA, training data augmentation; TTA, testing-time augmentation; WTTA, wrap TTA on top of model during training; mIoU, mean intersection over union.



Conclusion

We used state-of-art models and techniques to train models capable of segmenting a TDS image to estimate percent adhesion, a critical attribute in evaluating TDS performance. The trained models appear to provide good performance on the test data, suggesting that the model may be applied in the following ways:

- The model may be used to cross-check percent adhesion data reported by clinicians or subjects in a clinical trial. Data with large discrepancy between percent adhesion calculated from the segmentation model and that from the clinical report could be subject to further manual review, ultimately leading to a better prediction of adhesion performance of drug products.
- The trained model, if deployed in a future app for mobile devices, could facilitate the collection of TDS images and aid clinicians or trial participants to report the percent adhesion. Given a captured TDS image, the model could provide a real-time segmentation and allow the clinician/participant to revise the segmentation generated by the model based on his/her visual inspection of the TDS. As previously mentioned, this higher quality image capture and potential real time evaluation, could provide a more accurate annotation and percent adhesion than in current practice and even the current model proposed in this paper.