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

Chao Wang (Email Chao.Wang@fda.hhs.gov)
United States Food and Drug Administration

Caroline Strasinger
United States Food and Drug Administration

Yu-Ting Weng
United States Food and Drug Administration

Xutong Zhao
United States Food and Drug Administration

Research Article

Keywords: image processing, percent area adhesion, transdermal delivery system, mobile technology, deep learning

Posted Date: August 22nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3273161/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.
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§
17 August, 2023

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 and the potential use of such models in clinical practice is further explored.

Keywords: image processing; percent area adhesion; transdermal delivery system; mobile technology; deep learning

∗Food and Drug Administration, Silver Spring, Maryland; Chao.Wang@fda.hhs.gov; 0000-0001-8976-3773; Corresponding author
†Food and Drug Administration, Silver Spring, Maryland; Caroline.Strasinger@fda.hhs.gov; 0000-0003-1693-103X
‡Food and Drug Administration, Silver Spring, Maryland; Yu-Ting.Weng@fda.hhs.gov; 0009-0001-9192-510X
§Food and Drug Administration, Silver Spring, Maryland; Xutong.Zhao@fda.hhs.gov; 0000-0003-3179-7369
Statements and Declarations

Disclaimer

Views expressed in this manuscript do not necessarily reflect official policies of the US Food and Drug Administration nor does any mention of trade names, commercial practices, or organization imply endorsement by the US Government.

This article reflects the views of the authors and should not be construed to represent FDA’s views or policies.

Funding

This work was supported in part by the Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration which allows for the use of the Betsy High Performance Computing at the U.S. Food and Drug Administration, and an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Chao Wang, Yu-Ting Weng, and Caroline Strasinger contributed to the study conception and design. Images used in the study were annotated by Chao Wang and reviewed by Caroline Strasinger, Yu-Ting Weng, and Xutong Zhao. Data analysis was performed by Chao Wang. The first draft of the manuscript was written by Chao Wang and Caroline Strasinger, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration for the support of this project, the High Performance Computing at U.S. Food and Drug Administration for computing support, and Jonathan Medley for his assistance with extracting and annotating images.
1 Introduction

Skin adhesion of transdermal and topical systems (collectively referred to as TDS) is essential for safe and effective use of the respective drug products. As such, it is no surprise regulatory agencies such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency both recommend assessment of in vivo adhesion performance for application submission. In more detail, the FDA has issued two guidance for industry specifying their recommendations for clinical study design as well as statistical approaches to assess the in vivo adhesion for products submitted for both New Drug Applications (NDAs) and Abbreviated New Drug Applications (ANDAs) [1, 2]. Regardless of the application types and the different statistical methods used to evaluate adhesion data, all evaluations are based on data comprised of repeated visual assessments of the percent adhesion, i.e., the estimated proportion of the area that adheres to the skin to the total area of the product during the wear period. Thus, the accurate and precise measurement of percent adhesion is critical to subsequent statistical analysis and characterization of the adhesion and ultimately, the product’s performance.

The 2023 Draft Guidance for Industry Assessing Adhesion with Transdermal and Topical Delivery Systems for ANDAs [2] states that when recording measurements of TDS adhesion, applicants may use appropriate methods such as a trained visual assessment and/or dot matrix templates to estimate percent adhesion. Although visual assessment is typically performed by trained clinician(s), observations with the naked eye offers little to no assurance of precision or accuracy. In other methods, a transparent dot matrix or grid template is positioned above the TDS adhered to the subject, and the detached areas are marked to more precisely estimate percent adhesion. While using a template may increase the precision, accuracy still depends on the markings by the clinician and subsequent estimation from the template. Further, depending on application site and size of the product, a TDS that encounters significant torsional strains or curvature of the body can be difficult to match the applied area of the TDS to a two-dimensional template. In all cases, the measurement of percent adhesion, even by trained clinicians, adds observer variation and subjectivity to the measurement. This can lead to reported values significantly inconsistent with the true adhesion performance of the product.

Accurately assessing the adhesion performance of a product can further be influenced by the clinical study design itself. The Draft Guidance for Industry, Assessment of Adhesion for Topical and Transdermal Systems Submitted in NDAs [1] loosely speaks to these challenges by recommending both an inpatient clinical wear study to inform to the adhering capability of the TDS and, when feasible, an outpatient clinical trial to capture the actual user experience. In brief, while an inpatient study offers a greater level of control in study conduct and assessment, it often does not represent the actual product use environment of the TDS (e.g., subjects performing routine daily life). In such “real world” environments multiple visual observations by clinicians are impractical and the subjects themselves must be trained to evaluate and report their own percent adhesion. Naturally, estimations by a large number of marginally trained trial subjects will exhibit substantially greater variability and/or bias than by a small number of experienced clinicians.

Of importance, a recent review of NDA adhesion data notes significant disparity in adhesion performance of a product when the product is assessed in the context of a study with small sample size (N<100) where subjects were confined to a clinic and in
the context of a study with large sample size (N>1000) and subjects were not confined to clinical centers (i.e. “real world”) [3]. It is natural to conjecture that the larger sample size “real world” study would be more representative of the actual performance of the product, even though the data would exhibit greater variability in the larger study. As the FDA moves towards an approach emphasizing clinically meaningful study design (e.g., Real World Evidence Program [4]) to better demonstrate a product’s expected performance, it can be anticipated that accurate and precise visual assessment of the percent adhesion will be challenging and potentially an obstacle to utilizing real world evidence (such as that obtained in large populations and lengthy efficacy trials).

In this paper we propose to aid and automate the collection and calculation of percent adhesion during product wear using mobile technology and artificial intelligence. In this to be designed mobile phone 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.

In brief, we extracted existing images from selected in vivo adhesion studies that were submitted in NDAs and ANDAs and annotated the images. Then we trained state-of-the-art deep learning semantic segmentation models using various techniques. We present the results based on performance measures such as accuracy and mean intersection over union, and finally, discuss how to use the model in practice with the overall goal of achieving a more consistent and accurate assessment of TDS adhesion performance.

2 Material and methods

2.1 Some assumptions

Although TDS applied to the body are in a three-dimensional (3D) space, given the application site(s) of most products, a two-dimensional (2D) plane can be assumed when TDS are fully adhered. Relative to the surface area of the human body, transdermal systems are small in size, and as such the application sites selected by many drug product developers are areas which are relatively flat or a modestly curving surface (e.g., common application sites include the abdomen, back, or deltoid).

When assuming a 2D plane, it is also important to understand varying forms of TDS detachment as illustrated in Figure 1.

a) The detached part may raise off the surface in a planar manner from the skin (i.e., an air gap is formed between the TDS and the underlying skin).

1Topical systems, which target the local and surrounding tissue, must be applied to the site of action, as such these systems may require alternative assumptions when applied to areas with significant curvatures (e.g., a sprained ankle or wrist).
b) The detached part may lift and fold above another part of the TDS that is still adhered to the skin.
c) The detached part may lift and fold under another part of the TDS that is also detached.
d) The detached part may lift off and adhere to itself creating a ridge in the system.

In the detachment scenario where a ridge is formed, the projection of the detached ridge onto a 2D plane may appear as a thin line. If the projection of the ridge onto the 2D image is used as an estimate of the area detached, the percent adhesion may or may not be underestimated, because both the detached area and total area are underestimated. To deal with this issue, more advanced methods are needed to stretch the detached area flat, which is left for future research. In all other detachment cases, the 2D approximation should be relatively accurate.

Figure 1: Illustration of different ways of TDS detachment.
2.2 Dataset

2.2.1 Data source

The images were extracted from regulatory submissions of in vivo adhesion studies, for which FDA recommends that a trained clinician take a digital photograph at the time of assessing percent adhesion or adhesion score and the digital photos be included in the submission. Not all data from all existing submissions were extracted and annotated. However, efforts were made to select representative images from a variety of product designs (e.g., size, shape, color, opacity). If many TDS images were available for a given product, a subset of images with detachment were selected since a multitude of images with TDS near perfect adherence provide little value for the purposes of this project.

The images were manually annotated in CVAT [5] based on the interpretation of the image. In general it is easy to separate the TDS as a whole from the skin. However, for some products, it can be difficult to distinguish between detached areas and attached area based on the image only. Should the image be annotated at the time when the photo was taken, the quality of annotation could be improved. The following labels were used to classify the image pixels at the beginning of the study:

- Background,
- All Other,
- Record Card/Sticker,
- Skin,
- TDS Adhered to Skin,
- TDS Detached Above More TDS,
- TDS Detached Above Skin

Because case d) is very rare in the current dataset, detachment cases c) and d), as described in 1, were not distinguishable and were grouped simply as “TDS Detached Above More TDS.” Such distinguishment will be made in future work as necessary.

In total 796 images were annotated in the following manner: 1) with increased attention to detail, the TDS region was annotated to separate detached area from attached area; 2) a reasonably large region surrounding the TDS, consisting of skin, clothing, and study record card or sticker was also annotated. Due to limited resource, the regions further away from the TDS in photographs were left as unannotated. During training, unannotated regions were masked by black pixels and the corresponding labels set as background.

2.2.2 Data Preprocessing

2.2.2.1 Initial resizing

The image dimensions ranged from 364x248 to 5184x3456. To avoid computationally intensive manipulation during augmentation, when the minimum of height and width was greater than 1024, the image was resized so that the minimum of height and width was 1024. Note that all models were set to take input of size 320x320 pixels.

2.2.2.2 Data augmentation

Data augmentation has been shown to prevent overfitting and achieve better performance [6, 7]. A variety of data augmentation techniques [8] were applied to the training data, including horizontal flip, random rotate, affine transformation, etc.
In addition to training data augmentation (TDA), studies [9, 10] showed that test-time augmentation (TTA), where the prediction for an image is based on merged predictions of multiple augmented copies of the original image, can also improve prediction performance. Furthermore, TTA can be not only applied to an existing trained model, but also wrapped on top of any model during training (WTTA). For our study, the WTTA were considered with the D4 group for both validation and test datasets [11, 8].

2.2.2.3 Focus crop Based on guidance and clinically, the detached portion of a sufficiently adhered TDS is expected to be no more than 20% of the total TDS surface area [1]. Further, when sufficient effort is taken to photographically capture a high quality image, the TDS can be expected to occupy a relatively large portion and be centered within the frame of the photograph. For example, when a rectangular TDS is positioned in the middle of the image with each side occupying 80% of the respective length of the image side, the detached part of a sufficiently adhered TDS is expected to occupy no more than 12.8% of the total photographic image. However, for many photographs in the dataset, the TDS occupied a relatively small portion of the entire frame and was not always centered. Thus, all images of the dataset were cropped with the TDS centered and occupying between 49% and 90% of the total image area.

2.3 Models and training

Many image segmentation models have been developed in recent years [12]. Given the limited data and impracticability of tracking all recent models, we explored the following models which can be expected to perform sufficiently well:

- U-net [16], with ResNet-50, ResNet-101 backbones [17],
- Unet++ [18], with ResNet-50, ResNet-101 backbones [17],
- DeelabV3+ [19], with ResNet-101 backbone [17].

To improve model performance and/or deal with imbalanced data, the briefly described loss functions below were utilized, see also [20]:

Given an array \( x = (x_{i,j,k}, i = 1, \ldots, h; j = 1, \ldots, w; k = 1, 2, 3) \) representing a 3-channel (usually RGB) image with height \( h \) and width \( w \), assuming its ground truth segmentation label is presented by a two-dimensional matrix \( y = (y_{i,j}, i = 1, \ldots, h; j = 1, \ldots, w) \) with values in \( 0, 1, \ldots, n-1 \), where \( n \) is the number of classes including background, which is usually denoted by 0. For a given model, let \( \hat{p} = (\hat{p}_{i,j,k}, i = 1, \ldots, h; j = 1, \ldots, w; k = 1, \ldots, n) \) be the probability array such that \( \hat{p}_{i,j,k} = \text{prob}(y_{i,j} = k | x) \). The predicted labels are usually given by \( \hat{y}_{i,j} = \text{arg} \max_{k} \hat{p}_{i,j,k} \).

Let \( \hat{p}_{i,j,t} = \hat{p}_{i,j,y_{i,j}} \) (think \( t \) as for true label). The log likelihood is given by

\[
\ell_{llk} = \sum_{i,j} \log \hat{p}_{i,j,y_{i,j}} = \sum_{i,j} \hat{p}_{i,j,t}.
\]

The cross-entropy loss function is the same as the negative log-likelihood,

\[
\ell_{CE} = -\ell_{llk}.
\]
The focal loss \cite{21} applies down-weight on easy examples (i.e., data with high probability at ground truth label) and focus on training on hard examples (i.e., data with low probability at ground truth label) via a weight function,

\[ \ell_{\text{focal}} = -\sum_{i,j} \alpha_{i,j,t}(1 - \hat{p}_{i,j,t})^\gamma \log(\hat{p}_{i,j,t}), \]

where \( \gamma > 0 \), \((1 - \hat{p}_{i,j,t})^\gamma \) is a monotone function of \( \hat{p}_{i,j,t} \), which can be treated as weights on the original negative log-likelihood function \(-\log(\hat{p}_{i,j,t})\), \( \alpha_{i,j,t} \) is another layer of weights, and can be defined as proportional to inverse of the class frequency, or as a hyper-parameter.

Focal Tversky loss \cite{22} generalizes the focal loss and Tversky loss \cite{23}, which is defined as

\[ \ell_T = (1 - S_T)^\gamma, \]

where \( \gamma > 0 \) and \( S_T \) is the Tversky similarity measure and defined as

\[ S_T = \sum_k \frac{\sum_{i,j:y_{i,j}=k;\hat{y}_{i,j}=k} \hat{p}_{i,j,k} + \epsilon \sum_{i,j:y_{i,j}=k;\hat{y}_{i,j} \neq k} \hat{p}_{i,j,k} + \beta \sum_{i,j:y_{i,j} \neq k;\hat{y}_{i,j}=k} (1 - \hat{p}_{i,j,k}) + \epsilon}{\sum_{i,j:y_{i,j}=k;\hat{y}_{i,j}=k} \hat{p}_{i,j,k} + \alpha \sum_{i,j:y_{i,j}=k;\hat{y}_{i,j} \neq k} \hat{p}_{i,j,k} + \beta \sum_{i,j:y_{i,j} \neq k;\hat{y}_{i,j}=k} (1 - \hat{p}_{i,j,k}) + \epsilon}. \]

The Tversky similarity measure \( S_T \) generalizes the Dice similarity with two additional parameters via \( \alpha \) and \( \beta \). It reduces to Dice similarity when \( \alpha = \beta = 1/2 \). As higher \( \beta \) may lead to higher generalization and improved performance for imbalanced data \cite{23}, \( \beta = 0.75 \) and \( \alpha = 0.25 \) were used. The hyper-parameter \( \gamma \) was set to be 0.75 \cite{22}.

The models were trained with weights initialized by pretrained weights from the ImageNet dataset \cite{24} provided in \cite{17}. All models were optimized via the stochastic gradient descent with learning rate scheduled by one cycle learning rate with cosine anneal strategy \cite{25}. 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. Each configuration was run 10 times with different random seeds to calculate sample mean and standard deviation of the performance metrics.

### 3 Results

Given the many choices outlined above, various configurations were used to train different models. The results are summarized in Table 1. To mitigate randomness due to seed for random number generator, each run configuration was replicated 10 times with different random seeds. Each numeric value in Table 1 is the average of the corresponding values for 10 runs. 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 of the best model (U-net with the Resnet-101 backbone trained with TDA, TTA, WTTA, and weighted CE loss) are illustrated in Figure 2 with 5 images that were not used for training. In general, the model drew the boundaries between background, skin, and TDS accurately. Many regions of detachment were accurately segmented by the model, however, some boundaries between detached regions and attached regions clearly can be further improved.
<table>
<thead>
<tr>
<th>Model</th>
<th>TDA</th>
<th>TTA</th>
<th>WTTA</th>
<th>Loss function</th>
<th># epochs</th>
<th>Time</th>
<th>Training data</th>
<th>Validation data</th>
<th>Test data</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Accuracy</td>
<td>mIoU</td>
<td>Accuracy</td>
<td>mIoU</td>
</tr>
<tr>
<td>resnet50_fcn</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ce_unweighted</td>
<td>105.1</td>
<td>1.3</td>
<td>0.987</td>
<td>0.671</td>
<td>0.956</td>
<td>0.691</td>
</tr>
<tr>
<td>resnet50_fcn</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ce_weighted</td>
<td>119.8</td>
<td>1.4</td>
<td>0.984</td>
<td>0.661</td>
<td>0.956</td>
<td>0.691</td>
</tr>
<tr>
<td>resnet50_fcn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>117.8</td>
<td>23.0</td>
<td>0.972</td>
<td>0.659</td>
<td>0.950</td>
<td>0.694</td>
</tr>
<tr>
<td>unet resnet50</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ce_weighted</td>
<td>117.4</td>
<td>1.0</td>
<td>0.985</td>
<td>0.666</td>
<td>0.955</td>
<td>0.672</td>
</tr>
<tr>
<td>unet resnet50</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>116.0</td>
<td>13.7</td>
<td>0.973</td>
<td>0.663</td>
<td>0.951</td>
<td>0.697</td>
</tr>
<tr>
<td>unet resnet50</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>focal_loss</td>
<td>116.0</td>
<td>15.5</td>
<td>0.974</td>
<td>0.597</td>
<td>0.949</td>
<td>0.685</td>
</tr>
<tr>
<td>unet resnet50</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>focal_tversky</td>
<td>127.7</td>
<td>14.4</td>
<td>0.929</td>
<td>0.512</td>
<td>0.911</td>
<td>0.528</td>
</tr>
<tr>
<td>resnet101_fcn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>80.8</td>
<td>16.5</td>
<td>0.969</td>
<td>0.581</td>
<td>0.945</td>
<td>0.673</td>
</tr>
<tr>
<td>unet resnet101</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>105.6</td>
<td>14.2</td>
<td>0.973</td>
<td>0.668</td>
<td>0.949</td>
<td>0.695</td>
</tr>
<tr>
<td>unet resnet101</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>focal_loss</td>
<td>111.4</td>
<td>16.6</td>
<td>0.975</td>
<td>0.595</td>
<td>0.950</td>
<td>0.681</td>
</tr>
<tr>
<td>unet++ resnet101</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>109.1</td>
<td>24.5</td>
<td>0.967</td>
<td>0.587</td>
<td>0.939</td>
<td>0.664</td>
</tr>
<tr>
<td>DeepLabV3+ resnet101</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ce_weighted</td>
<td>109.1</td>
<td>24.5</td>
<td>0.967</td>
<td>0.587</td>
<td>0.939</td>
<td>0.664</td>
</tr>
</tbody>
</table>

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 and reported number of epochs, time (in hours) needed for training, performance measures (accuracy and mIoU) for training, validation, and test data are reported as averages over the 10 replicated runs. Sample standard deviations are not reported due to space limitation and less than 0.03 for accuracy 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.
4 Conclusion and discussion

We used state-of-art models and techniques such as training data augmentation, loss functions, and test time augmentation, 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.

The model performance may be improved by several approaches, including the use of more advanced models, collection of additional high-quality data, and federated learning. Efforts to improve the current model continue, while the development of a mobile tool to facilitate the automatic collection and segmentation of quality images can expand the depth and variety of training data. To accelerate the approach, outside sources may consider the creation of a repository for high-quality training images and federated learning from various sources including industry, academia, and clinical centers, all while protecting sensitive health and proprietary data [26, 27, 28].
Figure 2: Illustration of segmentation performance for the best model (U-net with the Resnet-101 back-bone trained with TDA, TTA, WTTA, and weighted CE loss). Each row shows the original image (left), original image with segmentation overlay (middle) and segmentation map (right).
References


deep learning library. Advances in neural information processing systems, 32, 2019


