Clinical Investigation of a Rapid Non-invasive Multispectral Imaging Device Utilizing an Artificial Intelligence Algorithm for Improved Burn Assessment

Jeffrey E. Thatcher, PhD[*], Faliu Yi, PhD[*,], Amy E. Nussbaum, PhD[*], John Michael DiMaio, MD[*,†], Jason Dwight, PhD[*], Kevin Plant, BS[*], Jeffrey E. Carter, MD[‡,||], and James H. Holmes IV, MD[‡]

Currently, the incorrect judgment of burn depth remains common even among experienced surgeons. Contributing to this problem are change in burn appearance throughout the first week requiring periodic evaluation until a confident diagnosis can be made. To overcome these issues, we investigated the feasibility of an artificial intelligence algorithm trained with multispectral images of burn injuries to predict burn depth rapidly and accurately, including burns of indeterminate depth. In a feasibility study, 406 multispectral images of burns were collected within 72 hours of injury and then serially for up to 7 days. Simultaneously, the subject's clinician indicated whether the burn was of indeterminate depth. The final depth of burned regions within images were agreed upon by a panel of burn practitioners using biopsies and 21-day healing assessments as reference standards. We compared three convolutional neural network architectures and an ensemble in their capability to automatically highlight areas of nonhealing burn regions within images. The top algorithm was the ensemble with 81% sensitivity, 100% specificity, and 97% positive predictive value (PPV). Its sensitivity and PPV were found to increase in a sigmoid shape during the first week postburn, with the inflection point at day 2.5. Additionally, when burns were labeled as indeterminate, the algorithm's sensitivity, specificity, PPV, and negative predictive value were: 70%, 100%, 97%, and 100%. These results suggest multispectral imaging combined with artificial intelligence is feasible for detecting nonhealing burn tissue and could play an important role in aiding the earlier diagnosis of indeterminate burns.

[*]*Spectral MD, Inc., Dallas, TX, USA*[†]*Baylor Scott and White, The Heart Hospital, Baylor Scott and White Research Institute, Dallas, TX, USA*[‡]*Atrium Health Wake Forest Baptist Medical Center Burn Center, Winston-Salem, NC, USA*[||]*The Burn Center at University Medical Center New Orleans, LA, USA*

FUNDING

Biomedical Advanced Research and Development Authority (BARDA) contract number: HHSO100201300022C. Clinical investigation of a rapid non-invasive multispectral imaging device utilizing an artificial intelligence algorithm for improved burn assessment. The deidentified analytic dataset that supports the findings of this study is available from the corresponding author upon reasonable request. Images obtained from the investigational device are not publicly available due to reasons of sensitivity.

CONFLICTS OF INTEREST

Jeffrey E. Thatcher receives salary from and has ownership/equity in Spectral MD, Inc. Faliu Yi receives salary from and has ownership/equity in Spectral MD, Inc. Brian McCall received salary from Spectral MD, Inc. at the time of this work. Amy E. Nussbaum received salary from Spectral MD, Inc. at the time of this work. J. Michael DiMaio receives consulting fees from and has ownership/equity in Spectral MD, Inc. Jason Dwight receives salary from and has ownership/equity in Spectral MD, Inc. Kevin Plant receives salary from and has ownership/equity in Spectral MD, Inc. Jeffrey Carter receives consulting fees from and has ownership/equity in Spectral MD, Inc. James H. Holmes receives none.

Address correspondence to Jeffrey Thatcher, PhD, 2515 McKinney Ave., Suite 1,000, Dallas, TX 75201. Email: thatcher@spectralmd.com

https://doi.org/10.1093/jbcr/irad051

Burn care is a critical component of the U.S. health care system that annually manages over 400,000 primary burn diagnoses and 40,000 hospitalizations from individual and disaster scenarios.[1] Survival in modern burn care is improving from advancements in early excision and grafting, antibiotics, fluid resuscitation protocols, and bioengineered skin substitutes.[2] Improvement to burn surface area and depth estimations are also being addressed, especially with optical imaging modalities.[3] Since the introduction of laser Doppler imaging as a clinical adjunct to burn assessment, there has been strong support for imaging to supplement clinical burn depth assessment.[4–6] However, in the U.S. burn depth assessment is typically performed by visual inspection alone.[7] This method is only 60 to 80% accurate for experienced burn clinicians and not much better than a random chance for nonburn trained clinicians.[4,8,9] Furthermore, specialized burn training for new doctors is becoming less common. Before 2008, all residents going through general surgery training were required to rotate on a burn service to gain essential experience and skills in diagnosing and treating burn wounds; however, burn training was removed by American College of Graduate Medical Education Residency Review Committee. As a result, over a decade of surgeons who have not been trained in burn care are currently in practice or about to begin practicing independently.

In burn assessment, it is important to identify regions where the skin's regenerative capacity has been reduced or eliminated. These areas, referred to as "non-healing" burns, are clinically classified as deep partial-thickness (DPT), or full-thickness (FT) burns. The standard treatment for nonhealing burns is excision and grafting. While FT burns are more

straightforward to diagnose, the subtle difference in dermal damage between superficial partial-thickness (SPT) and DPT burns makes them difficult to distinguish. For partial-thickness burns, clinicians often conduct periodic evaluations over the first week before declaring them as superficial or deep. Incorrect burn diagnosis of SPT burns as DPT may result in unnecessary surgical treatment, whereas failing to identify burns that require surgery (DPT and FT) can result in scar formation and contractures.[10] While waiting to make a diagnosis of partial-thickness burns increases the accuracy of diagnosis,[11] it may also increase the length-of-stay and the risk of infection if early excision is warranted.[12]

One promising adjunct to clinician burn assessment is multispectral imaging (MSI). MSI measures the reflectance of visible and near-infrared light from tissues. Using MSI data, diagnosis of burn depth can be achieved by comparing the reflectance of an unknown burn to a reference library of burns of known depth. Early studies using MSI to supplement burn assessment measured the absorbance of burn tissue in red, green, and near-infrared light. These investigators developed linear models to discriminate healing from nonhealing burn areas using ratios of reflectance from these three colors.[13] Eventually, digital imaging and more advanced computational methods were applied.[14,15] Throughout these investigations, it was found that MSI data was highly correlated to burn depth and severity, while at the same time being uninfluenced by patient factors such as age, sex, and total body surface area (TBSA). These studies indicated that the analysis of the reflectance spectrum from partial-thickness burns had 86% sensitivity and 79% specificity and was more accurate than the attending physicians.[16]

Previous studies advanced preclinical experiments with the evaluation of appropriate wavelengths for improved performance of spectral imaging in partial-thickness burns.[17] Additionally, other studies have investigated a variety of machine learning models for accurate discrimination of burn tissue.[18,19] In the current study, we focused on three objectives: (1) implementation of deep learning (DL) as the discriminatory algorithm applied to the MSI data, (2) the feasibility of an observational clinical study design wherein accurate identification of burn healing potential could be made regardless of the patient undergoing surgical or nonsurgical management, and (3) the performance of this technology within indeterminate depth burns (IDBs).

To apply DL to MSI of burns, we investigate a class of artificial neural networks called convolutional neural networks (CNNs). A CNN consists of image filters that work to quantify and then classify abstract features of the original image.[16] These filters adapt to the desired image analysis task when the programmer executes a process of repeated trial-and-error, called training, that uses large numbers of labeled example images. DL has achieved higher accuracies in image classification compared to traditional machine learning approaches and even surpassed humans on specific tasks.[15] An advantage of DL for evaluating MSI data is that information beyond the reflectance measures can be found, including textures that are known to be important indicators of burn depth.[20,21]

One critical factor of DL algorithm development is the accurate labeling of the example images used in training. These labels are referred to as "ground truth." Using the unaided judgment of clinicians to label images would not be sufficient to train an algorithm that outperforms clinicians. Therefore, this study investigates the feasibility of collecting clinical reference standards to augment image labeling by clinicians including burn healing assessments taken at 21-days post injury, and histological assessment of biopsies taken immediately prior to excision.

Evaluating burn healing at 21-days postburn directly informs the viability of the regenerative structures of the skin. If a burn heals by day 21, it is defined as a healing burn.[22] Whereas histological analysis is frequently considered the "gold standard" of burn depth assessment, serving as the basis for comparison of other diagnostic modalities.[4,22–24] Using histology, a pathologist can determine changes to cellular viability, blood vessel patency, and collagen structure caused by the burn injury.[25,26] Prior to the initiation of this study, a board-certified pathologist was consulted to develop the clinical collection of biopsies and their assessment.

One final objective of this study was to estimate the performance of MSI on IDBs. The visual appearance of burns changes with the body's pathophysiologic response to the burn injury.[27] Superficial and deep burns are often straightforward to identify visually, but mid to DPT burns present a challenge to diagnose and treat.[9] These burns are referred to as IDBs.[28] When encountered with partial-thickness burns, it is common to evaluate them periodically for about 3 to 7 days before categorizing them as DPT or SPT. Previous studies indicate that the diagnosis of partial-thickness burns improves over the first week for both clinical visual assessment and optical techniques of laser Doppler and spectral imaging.[11] In this study, we called burns undergoing periodic evaluation by the clinician "indeterminate" and investigated the performance of MSI within this subset.

METHODS

Investigational Device

MSI data were collected using a filter-wheel camera equipped with eight optical band-pass filters. Each image yielded eight greyscale digital images, one for each bandpass filter. The following peak transmission were selected for the filters that were positioned radially on the filter wheel: 420, 581, 601, 620, 669, 725, 860, and 855 nm (filter widths were ±10 nm; Ocean Thin Films; Largo, FL). The filter wheel was rotated between the sensor and lens with a stepper motor. Behind the filter, wheel was the imaging sensor, a complementary metal oxide semiconductor sensor (Sony Inc.) with dimensions of 1044 × 1408 pixels. A telescopic lens was mounted in front of the filter wheel (Distagon T* 2.8/25 ZF-IR; Zeiss Inc.). The light source used was a 4-panel light emitting diode array, where each panel was equipped with a frosted diffuser to create a more even illumination profile within the imager's field-of-view (FOV). The system was calibrated using a square 95% reflectance standard (Spectralon SG3151; LabSphere Inc.; North Sutton, NH) to compensate for the different spectral response of the imaging sensor.

The image was positioned at the end of an adjustable articulating arm mounted on a cart that contained a computer and display (**Figure 1**). The working distance of the camera was 40 cm from the target burn resulting in a 15 × 20

Figure 1. The multispectral imaging device utilized for this study. The system include: (1) a touch-screen display; (2) mobile cart containing the processing unit; (3) an articulating arm to position the camera; and (4) the MSI subsystem. MSI, multispectral imaging.

cm^2 FOV. Green guidance beams projected onto the subject's target burn facilitated positioning the camera and maintaining the 40 cm working distance.

Study Design

We selected a single center for this pilot study designed to evaluate the feasibility of the imaging technology and study methods. The study was approved by the institutional review board of Wake Forest Baptist Medical Center-Burn Center in Winston-Salem, NC, and informed consent was obtained from all subjects prior to enrollment. Adult subjects greater than 18 years of age with flame, scald, or contact burns were candidates. Subjects were enrolled within 72 hours of their initial burn injury and excluded from the study if: their burns were less than 25 cm^2 or isolated to regions other than the arms, legs, or torso; they had an inhalation injury; or their burns were greater than 30% TBSA.

Imaging Procedure.

Upon enrollment, up to three 15 × 20 cm areas of the body that contained at least a portion of burn injured tissue were identified for imaging. These sites were referred to as "study burns." Each study burn was imaged serially up to six separate times in the first 10 days post-injury. Serial imaging of each study burn was performed during routine dressing changes until the patient was discharged from the hospital or the study burn underwent surgical excision and grafting.

In each imaging session, two MSI images were obtained from each study burn. The first image was taken with the sensor directly facing the study burn, and the second image was taken with the sensor offset by approximately 30° from the first image. This variation was chosen to account for potential differences in the camera positioning between users, and to augment the data used for CNN algorithm training.

Reference Standards and Ground Truth Images

To implement a CNN that could highlight the area of nonhealing burn within an image, a labeled image of the correct (ie, - true) area of nonhealing burn was required. This image representing the true area of nonhealing burn was termed the "ground truth" image. Obtaining accurate ground truth images for each MSI image collected in the study was a two-step process.

- First, the healing status of a study burn was obtained using gold standard methods of burn depth determination. These gold standards were either a 21-day healing assessment or biopsies of the burn.
- Second, the gold standard information was used by a panel of three burn practitioners to manually draw the ground truth image.

21-day Healing Assessment.

For burns undergoing nonoperative management, the reference standard was a 21-day healing assessment. This assessment was performed by the primary surgeon at 21 ± 3 days following the initial injury. Healing assessments were documented in the subject's case report form and included a color photograph. Regions within the study burn were designated as healing if they demonstrated over 95% epithelialization by the following characteristics: no longer readily transmitting water; no longer requiring dressings or bandages; dry to the touch; and more pink or opalescent than red or transparent in visual appearance.

Biopsy Collection and Evaluation.

For burns undergoing operative management, the reference standard was obtained through multiple biopsies of the study burn at the time of surgery. All biopsies were obtained during surgery from study burn areas declared to be nonhealing by the burn practitioner, and only from within the area of the burn that was to be excised. A 4.0 mm diameter dermal punch was used. To guide placement of the biopsies, physicians were provided a thin polycarbonate sheet precut with an array of holes evenly spaced at 5.0 cm intervals. To capture a robust sample, clinicians performing the operation were instructed to obtain biopsies from areas of the study burn that were visually distinct with one biopsy for every 25 cm^2 of wound area.

Biopsies were immediately stored in 10% formalin and sent for processing at an independent dermatopathology center (Cockerell Dermatopathology, 2110 Research Row #100, Dallas, TX). Each biopsy was fixed in paraffin, sectioned, mounted on slides, and stained with hematoxylin and eosin. Whole slide scanning was performed to enable investigators to review histologic findings with the dermatopathologist. The evaluation was performed by a blinded, board-certified, dermatopathologist using the following criteria:

- Biopsies of FT burns, or full-thickness burns, were identified by nonviable papillary and reticular dermis (Figure 2).
- Biopsies of DPT, or deep partial-thickness burns, were characterized by nonviable papillary dermis, nonviable

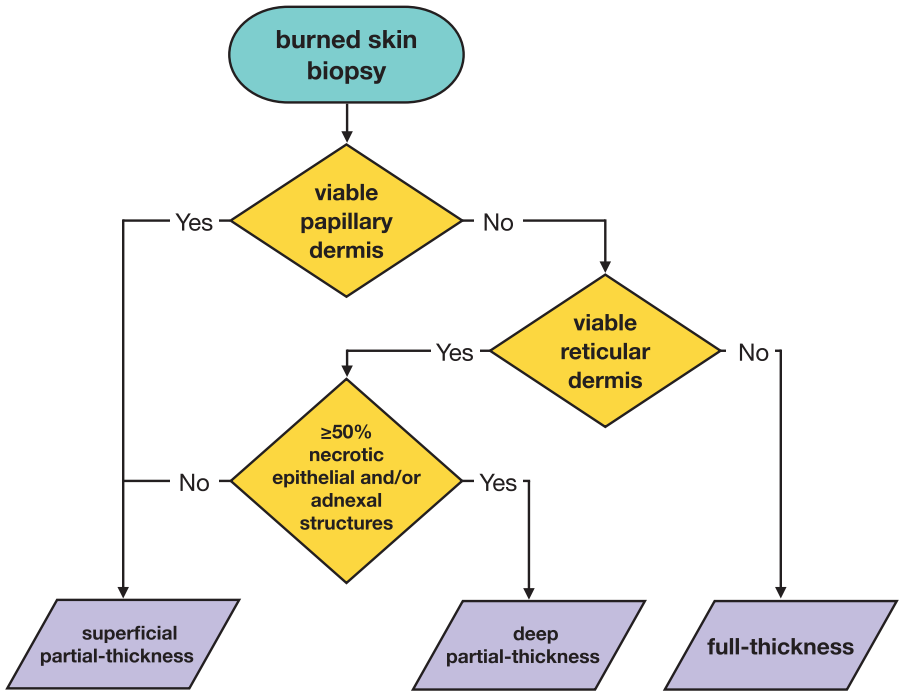

Figure 2. Decision tree used by pathologists to determine burn depth. Superficial burns are unlikely to be sent for excision and biopsy and not included in this decision tree.

epithelial structures of the reticular dermis, and/or less than 50% viability of adnexal structures of the reticular dermis.
- Finally, SPT burns, or superficial partial-thickness burn, was characterized in two ways: (1) a viable papillary dermis; or (2) a nonviable papillary dermis but viable epithelial structures, and/or greater than 50% viability of adnexal structures of the reticular dermis.
- Biopsies that contained **superficial**, or superficial burns, were not obtained in this study design.

Ground Truth Images.
A panel of burn practitioners called the Truthing Panel, used the data from the gold standards to create ground truth images. Truthing Panels always consisted of two burn surgeons, including the subject's primary surgeon, and one physician assistant from the burn center. During the Truthing Panel meetings, a subject's 21-day healing assessment or biopsy results were reviewed by all panel members prior to creating the ground truth images for that subject.

The panel generated one consensus ground truth image for every MSI image collected. Initially, the panel created a color-coded image with the margins of superficial, SPT, DPT, and FT burns. Subsequently, The DPT and FT burn areas from this image were combined to generate the ground truth image of a nonhealing burn (**Figure 3**).

Algorithm Development

Algorithm Architectures and Training.
The CNN architectures evaluated in this study were trained to automatically highlight the area of nonhealing burn tissue within the image, centimeter-by-centimeter. This technique is called image segmentation.

Three independent CNN architectures for image segmentation as well as an ensemble of these CNNs were employed to automatically identify the area of nonhealing burn tissue within an image. The algorithm architectures were the following:

U-Net
U-Net is an encoder-decoder DL semantic segmentation approach that works with very few training images. The U-Net algorithm's decoder up-samples lower resolution input feature maps using skip connections to keep high-resolution features and improve localization.[15,29]

SegNet
Similar to U-Net, SegNet is an encoder-decoder fully CNN for semantic segmentation. However, its decoder up-samples lower resolution input feature maps using pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear up-sampling.[30]

Dilated fully connected neural network (dFCN)
dFCN It is a deep full CNN for semantic segmentation based on dilated convolution. In this scheme, the dilated convolutions allow the receptive field of each convolutional kernel to be increased, and at the same time not reduce the input resolution. This network can produce pixel-level labeling without the typical encoder-decoder "hourglass" structure.[31,32]

Voting Ensemble
Ensemble techniques involve the combination of multiple underlying algorithms. The voting ensemble was a simple

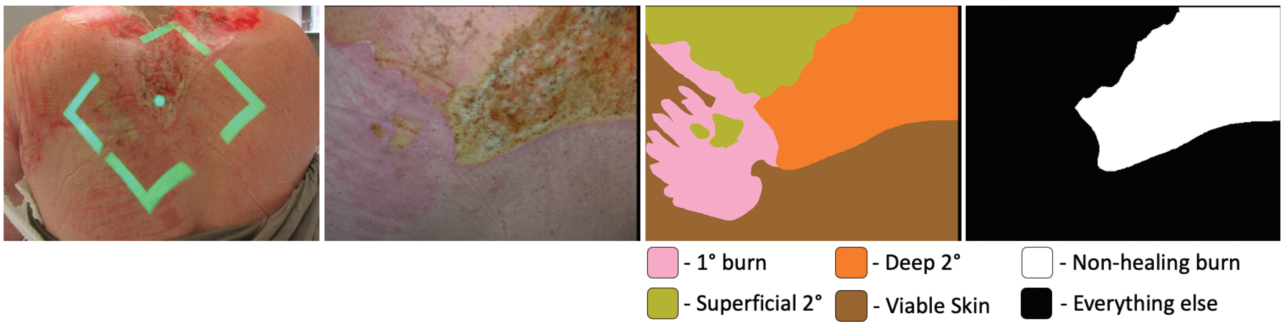

Figure 3. Imaging and ground truth masks from a heterogeneous burn on the dorsal aspect of a subject. Green guiding beams indicate the location and distance of the MSI image; color image of the study burn generated from the MSI data; detailed ground truth provided by expert truthing panel; binary ground truth where all nonhealing burn have been labeled as the target pixels in white. MSI, multispectral imaging.

averaging of the probabilities of the corresponding pixels predicted by the three CNNs described above.[33,34]

CNNs were trained using stochastic gradient descent with a momentum optimizer. The hyperparameters of learning rate (0.0001), momentum (0.9), number of epochs (60), weight decay (0.0005), and batch size (4) were the same for each algorithm. The target pixels, labeled as the nonhealing burn class, represented only about 20% of all the pixels in the data set, and many images contained no target nonhealing burn pixels. To address this class imbalance, the loss function was weighted according to median frequency balancing.[34]

The CNN output was a map displaying the probability of each pixel belonging to the nonhealing burn class, $P(pixel_{ij} = non-healing \mid \lambda_1, \lambda_2, \ldots, \lambda_8, \Phi)$. From this probability map, a binary image was generated, where each pixel was categorized as positive or negative for nonhealing burns. This categorization was determined by applying a threshold, τ, to the probability of each pixel in the probability map (equation 1).

$$1_A := \begin{cases} 1 \ if \ P\left(pixel_{ij} = 1\right) \geq \tau \\ 0 \ if \ otherwise \end{cases} \quad (1)$$

Algorithm Scoring Metrics.
Algorithm scoring was estimated using the leave-one-out cross-validation (LOOCV). For each fold of LOOCV, the held-out set was defined at the level of the subject, because data within a subject was correlated. All pixels classified by the algorithm on the left-out images were compared to their corresponding pixels in the ground truth image. True positives were defined as pixels in the algorithm's output image classified as nonhealing burn that were also labeled as nonhealing burn in the ground truth image. In the same manner, we defined other pixels as a false positive, true negative, and false negative.

We chose to focus our analysis on larger burn areas, because surgeons working on the study agreed that an imaging device would be used for larger burns on the body locations included in the study. 20 cm^2 (approx. 10^4 pixels) was chosen as a cutoff point for nonhealing burn areas to be included in the analysis. When the nonhealing burn area in the ground truth image was less than 20 cm^2, the nonhealing burn was treated as a healing burn. Similarly, when the algorithm predicted a total area of nonhealing burn within an image less than 20 cm^2, all pixel values in that predicted image were set to 0.

Algorithms were compared using the area under the precision and recall curve (PAR-AUC). Plotting the precision and recall (PAR) curve was accomplished by incrementing the threshold value, τ, from 0.0 to 1.0. Note that recall is equivalent to sensitivity, and precision is equivalent to positive predictive value (PPV) (see **Table 1**: equations 2 and 4). From this comparison, the top performing algorithm was selected. We reported the metrics of sensitivity, specificity, PPV, negative predictive value (NPV), and PAR-AUC for the top performing algorithm (**Table 1**).

Prior to reporting performance metrics for the top algorithm, an optimum threshold, τ, was selected by maximizing the F_1-score (equation 7).

$$F_1 = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (7)$$

This was equivalent to selecting the point on the PAR curve closest to the optimal value of [1,1], and it ensured that we obtained the highest combination of sensitivity and PPV, assuming that these metrics should be weighted equally. The F_1-score and PAR curves were used over receiver operating characteristic (ROC) curves because there was a ratio of approximately 1:4 for nonhealing burn pixels to all the other pixels in the images. Therefore, the receiver operating characteristic curve would result in an overly optimistic impression of algorithm performance whereas the PAR curve is better suited for this type of unbalanced data.

Modeling of DL Algorithm Performance Metrics.
To evaluate the performance of image segmentation, we implemented a statistical model to estimate the probability of correctly finding nonhealing burn within an image using standard measures of performance assessment including, sensitivity, specificity, PPV, and NPV. With this model, we also evaluated these metrics throughout the first week following burn injury.

Having collected repeated measurements within clusters (ie, patients and study burns across multiple days) the study design involved correlated observations within patients and study burns. We used *generalized linear mixed models* (GLMMs), a flexible class of models commonly used for hierarchical data, to account for correlated observations. In

Table 1. Metrics used to evaluate segmentation algorithm performance

Metric	Computation	
Sensitivity (Also known as Recall)	$Recall = \text{Sensitivity} = \frac{TP}{TP+FN}$	equation (2)
Specificity	$Specificity = \frac{TN}{TN+FP}$	equation (3)
Positive predictive value (PPV) (also known as precision)	$Precision = PPV = \frac{TP}{TP+FP}$	equation (4)
Negative predictive value (NPV)	$NPV = \frac{TN}{TN+FN}$	equation (5)
Area under the precision and recall curve (PAR-AUC)	$PAR - AUC = \int_{-\infty}^{\infty} Precision(Recall(\tau))\ d\tau$ where *Precision* and *Recall* are probability density functions with respect to τ, the classifier threshold.	equation (6)

addition, this method allowed for analysis of the effect of "time-since-injury" on the scoring metrics described above.

Models were fit using restricted maximum likelihood implemented in the lme4 package for R (R version 4.0.3; LME4 version 1.1-27.1). The following model was used to estimate the DL algorithm's sensitivity:

$$n_{TP,\,i} \sim Binomial\,(n_{GTP,\,i},\ p_i)$$
$$logit\,(p_i) = \beta_0 + \beta_{0,j} + (\beta_1 + \beta_{1,j}) * t_{i,k} + \gamma_j + \epsilon_i \quad (8)$$

where $n_{TP,\,i}$ is the number of True Positive pixels in the DL algorithm result for image i, and $n_{GTP,\,i}$ is the number of nonhealing burn pixels in the Ground Truth Image i. The likelihood of $n_{TP,\,i}$ is modeled with a binomial distribution that requires one parameter, p_i. The parameter p_i represents the sensitivity of the DL algorithm— in other words, the probability that a GTP (ie, ground truth positive) pixel will be predicted true positive by the DL algorithm.

The logit of the sensitivity, or $logit(p_i)$, was modeled as a linear function of an intercept and the time-since-injury, $t_{i,k}$. The correlation of data within each subject was accounted for by modeling both the intercept and time-since-injury as a fixed effect plus a random effect. The intercept was modeled as the sum of a fixed parameter, β_0, and random parameter that varied at the level of the subject, $\beta_{0,j}$. The slope was the sum of a fixed parameter, β_1, and a random parameter that varied at the level of the subject $\beta_{1,j}$.

Specificity, PPV, and NPV were modeled using the same design. Point estimates for these metrics were computed by marginalizing the predictor time-since-injury. The 95% prediction intervals (PI-95) for all metrics used for scoring algorithm performance were computed by random simulations of the coefficients and residual standard deviation of the GLMM.[35]

Algorithm Performance on Burns of Indeterminate Depth

The clinician present at the time of each imaging session was asked if they could make a diagnosis of burn severity. When they could not make a definitive diagnosis (ie, they preferred to wait for the burn to develop further before declaring a diagnosis) we noted that their assessment was "indeterminate" using software on the imaging device. The clinicians involved in making these determinations included burn surgeons and physician assistants working in the burn unit. The performance of the artificial intelligence (AI) algorithm was computed on the study burns that were classified as indeterminate using the model described in equation (8) with an additional term for "indeterminate" added that had two levels indicating whether a burn was assessed as indeterminate.

Table 2. Enrollment summary statistics

Subject Characteristic	Total, n(%)*
Age (years)	47.4 ± 17.2
Burn TBSA (%)	14 ± 7.1
Gender	
Female	8 (21.1%)
Male	30 (78.9%)
Race/ethnicity	
Black/non-Hispanic	5 (13.2%)
White/Hispanic	1 (2.6%)
White/non-Hispanic	32 (84.2%)
Mechanism of injury	
Contact	5 (13.2%)
Flame	31 (81.6%)
Scald	2 (5.3%)
Location of injury†	
Shoulder and arm‡	22 (37.9%)
Thigh and leg‡	18 (31.0%)
Abdomen and chest‡	18 (31.0%)

TBSA, total body surface area.
**Except age, and burn TBSA reported as mean +/- st. dev.
‡Location of injury reported at the burn level rather than subject level.
†Includes both anterior and posterior locations.

RESULTS

Clinical Study

Subject Demographics and Burn Characteristics.
Forty (40) subjects were enrolled in the study from June 2017 to November 2018 with 38 completing the study, one mortality, and one withdrawal (**Table 2**). The study subjects had a mean TBSA of 14.0% ± 7.1. Demographics indicated a mean age of 47.4 years ± 17.2, a majority male population (78.9%), and race-ethnicity makeup of Black-non-Hispanic (13.2%), White-Hispanic (2.6%), and White non-Hispanic (84.2%).

A total of 58 study burns were imaged across all 38 subjects. The cause of study burn injuries varied with 81.6% being

flame burns, followed by 13.2% contact, and 5.3% scald. Study burns were evenly distributed to one of the following regions: arms and shoulders, abdomen and chest, and legs and thighs.

Imaging and Ground Truth.
A total of 406 MSI images were obtained from the 58 study burns. The selected study burn areas were imaged over time with an average of 3.4 ± 1.4 imaging sessions per study burn and a maximum of six.

Twenty-five study burns underwent surgical excision and grafting from which biopsies were obtained, while the remaining 33 were evaluated for burn depth using the 21-day healing assessment. From the 25 study burns undergoing excision and grafting, 178 biopsies were collected, averaging 7.12 biopsies per burn site. The average day of biopsy collection was 8.1 ± 3.9 days following burn injury. Of these 178 biopsies, 23% were found to be SPT, 61% DPT, and the remaining 16% were FT. In 18 of the 25 excised burns (ie, 74%), there was at least one biopsy indicating the presence of a DPT or FT burn. Lastly, 31% of study burns were heterogeneous in-depth, having both healing (ie, SPT) and nonhealing (ie, DPT and/or FT) biopsies.

The resulting set of ground truth masks generated by the truthing panel from all subjects revealed that, as a proportion of pixels, 21.5% of burn areas imaged were SPT, 11.9% were DPT, 8.3% were FT, and 1.9% were superficial. The remainder of the pixels in the ground truth data were either uninjured skin (31.4%) or background objects (25.0%). Study burns were often found to be heterogeneous in-depth, as noted by multiple labels on the ground truth image.[29] burn wounds contained at least some area of nonhealing burn and 27 were at least 20 cm^2 in area. These 27 nonhealing burns represented 19.7% of the total number of pixels across all study burn images.

Algorithm Results

Output images were generated to demonstrate the automated identification of nonhealing burns within an MSI image (**Figure 4**). Resulting images are created through the LOOCV process to estimate the image results that would be provided to the physician in a real-world setting. Comparison of these images to the ground truth created by the truthing panel indicated the success or failure of each area in the output image at the unit of the pixel (**Figure 5**).

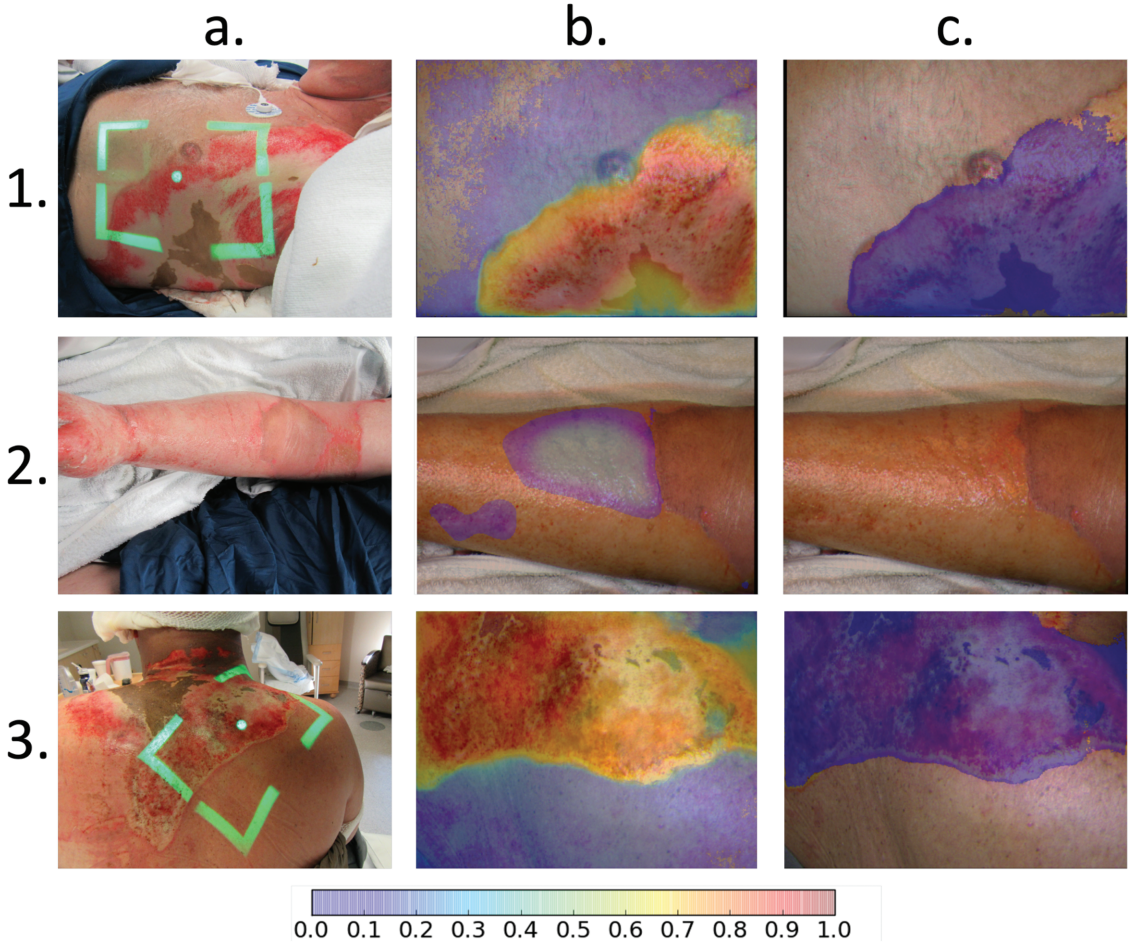

Figure 4. CNN results from three subjects. Columns represent: (a) the reference photo; (b) map of severe burn generated by the AI algorithm with color bar indicating probabilities between 0.0 and 1.0 (probabilities < .05 not shown in the image); and (c) the segmented images resulting from the application of a threshold to the probability map. Rows include: (1) a 71-year-old male with a severe flame burn indicated by the highlighted region in column C; (2) a 44-year-old male with a superficial flame burn indicated by a lack of highlighted region in column C; and (3) a 56-year-old male with a severe flame burn. AI, artificial intelligence; CNN, convolutional neural network.

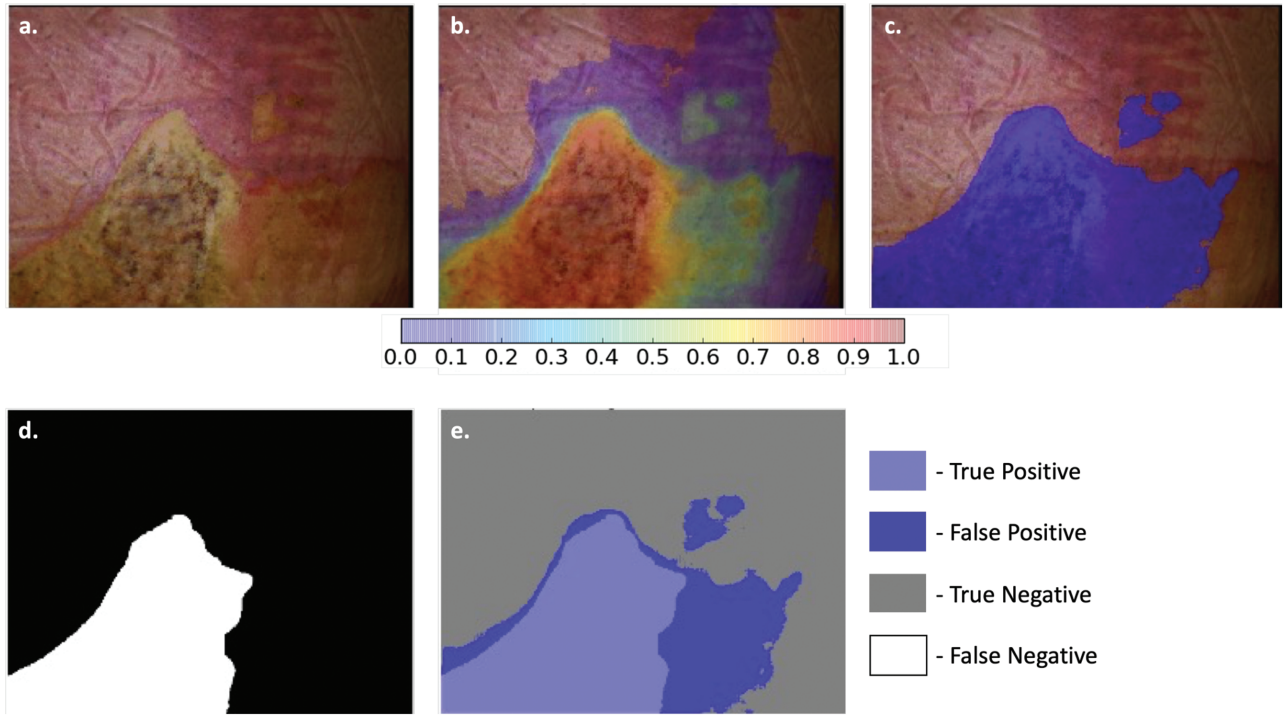

Figure 5. Results of the Voting Ensemble DL algorithm demonstrating the segmentation of nonhealing burn (ie, deep partial-thickness and full-thickness burn) within the image. (a) Color image of a heterogeneous burn on the back of a study subject. (b) Probability "heat map" of the Voting Ensemble DL algorithm. (c) Predicted area of nonhealing burn after threshold has been applied to the probability heat map in image "b". (d) Ground truth location of nonhealing burn in the image. (e) Comparison of the ground truth to the DL algorithm indicating the four outcome types for every pixel in the image. DL, deep learning.

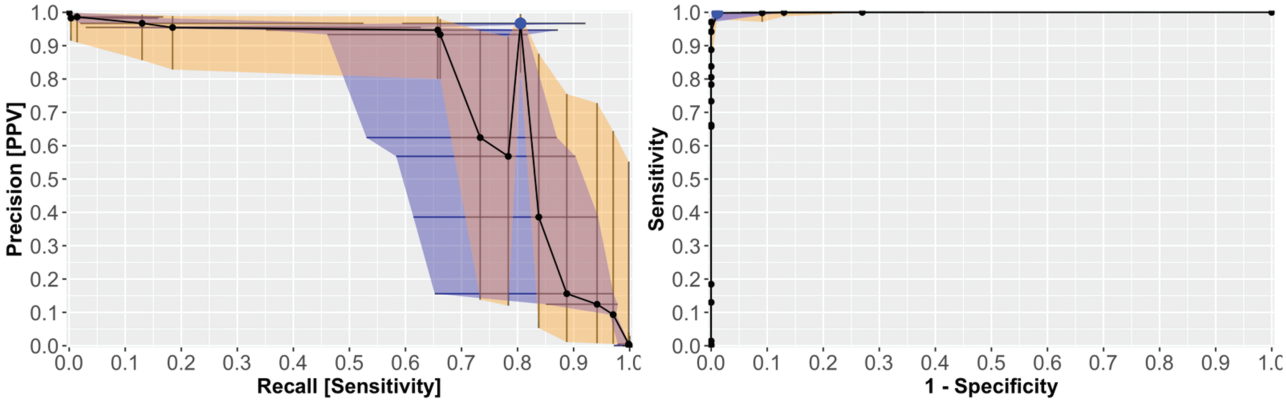

Figure 6. Performance of Voting Ensemble DL algorithm over various thresholds. (Left) PAR curve in which the blue ribbon represents the PI-95 for Recall and the orange ribbon represents the PI-95 for Precision. Area under the PAR curve was 0.81. (Right) receiver operating characteristic curve for comparison looks near perfect with an AUC of 0.99 demonstrating how unbalanced data can bias traditional measures of classifier performance. The larger blue point in each plot indicates the optimum threshold value. DL, deep learning; PAR, precision and recall.

Algorithm Scoring and Classifier Comparison.
Using PAR-AUC, we identified the most effective of the evaluated algorithms to be the Voting Ensemble, with an AUC of 0.99 and PAR-AUC of 0.812 (**Figure 6**). The Voting Ensemble was the only algorithm of the group to achieve both high sensitivity and PPV (**Table 3**). The next highest performing algorithm was the weighted ensemble followed by dFCN. For all algorithms, the specificity and NPV were estimated at 100%. This was a result of good algorithm performance combined with imbalanced data favoring GT-negative pixels. The PI-95 intervals were large for sensitivity and PPV since the residual standard deviation of the fitted GLMM was also large. These high standard deviations are expected from the low sample-size obtained in this study.

Effect of Time-Since-Burn.
Time-since-injury did impact the performance of the voting ensemble DL algorithm in both sensitivity and PPV. For sensitivity, the log-odds were predicted to increase by 1.54 (±1.17

Table 3. Performance comparison of DL algorithms used for nonhealing burn segmentation from MSI images

Algorithm	PAR-AUC	Sensitivity (%)	PPV (%)	Specificity (%)	NPV (%)
U-Net	0.167	6.6 (0.5, 51.2)	86.5 (70.9, 94.4)	100 (99.9, 100)	100 (99.9, 100)
SegNet	0.425	65.6 (30.4, 89.3)	34.7 (3.7, 88.1)	100 (99.9, 100)	100 (99.9, 100)
dFCN	0.561	43.5 (19.9, 70.5)	80.1 (42.3, 95.7)	100 (99.9, 100)	100 (99.9, 100)
Voting Ensemble	0.812	80.5 (60.6, 91.7)	96.7 (80.9, 99.5)	100 (99.9, 100)	100 (99.9, 100)

dFCN, Dilated fully connected neural network; *NPV*, negative predictive value; *PAR-AUC*, precision and recall curve; *PPV*, positive predictive value.

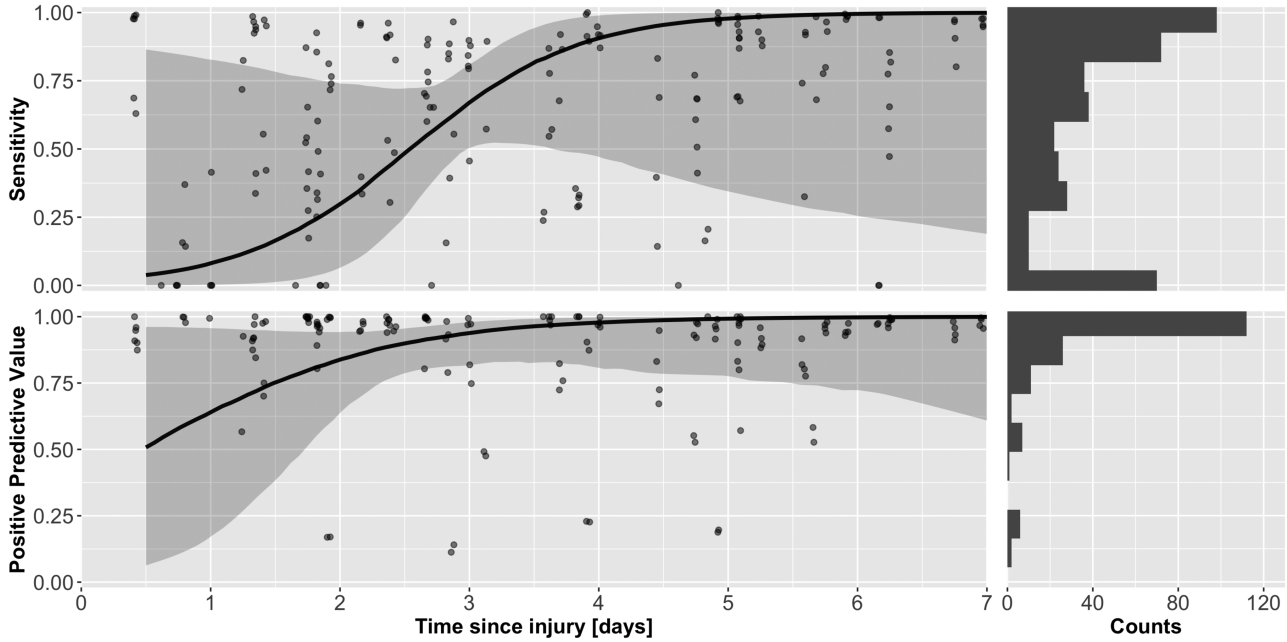

Figure 7. Performance of the Voting Ensemble DL algorithm over the first week of injury. (Left) This plot shows the change in PPV over time where the black line is the median PPV, and the gray ribbon is the PI-95. Points represent the sensitivity of individual images (Right) demonstrates the increase in Sensitivity over time from about 9 to 99%. Sensitivity shows a sharp increase between 1.5 and 3.5 days since injury whereas PPV increases only modestly throughout the 7-day timespan. DL, deep learning; PPV, positive predictive value.

CI-95; $P = .01$) each additional day postburn. This increase brought sensitivity of the Voting Ensemble from approximately 10% to over 99% within the first week since injury with a sharp increase from day-zero to day-four (**Figure 7**). The increase in PPV was less dramatic, with a log-odds increase of 1.02 (±1.35 CI-95; $P = .16$) and an increase from approximately 76% to 99% across the first week of burn injury. The variable time-since-injury did not impact the Specificity or PPV. Both metrics were estimated at 99.9% throughout the first week of injury.

A 57-year-old female study subject with DPT flame burns covering 30% TBSA provides an example of the change in Voting Ensemble algorithm performance over time (**Figure 8**). Their injury on the abdomen was imaged from 1 to 7 days post-injury. The truthing panel identified the burn areas in the image as DPT (ie, nonhealing) based on the five biopsies obtained on day-9 postburn - all showing greater than 75% adnexal structure necrosis. Visually, the edges of the burn began as hyperemic (pinkish-red) that progressed to develop an eschar that became completely yellow by day 4. When we reviewed the probability map outputs by the Voting Ensemble algorithm, there was an increase in the probability of deep burn across the region of hyperemia as the burn developed toward a more stable appearance on day 4. We found this behavior in 8 of the 27 nonhealing burn cases.

Burns of Indeterminate Depth

Clinicians at the study site classified 35 of the 56 study burns as indeterminate on their first day of clinical assessment. Of all 406 images collected in this study, 56% were collected while the clinicians believed the burn to be of indeterminate depth (**Table 4**). Day 4.7 after the burn injury was a key time for clinical diagnosis. At this time, clinicians were making determinations of burn severity more often than they were judging burns as indeterminate (ie, greater than 50% of the study burns had been diagnosed by clinicians).

Within the indeterminate group, the DL algorithm had 97.1% PPV and 70.3% sensitivity. The Specificity and NPV were unchanged after segmenting the images into these two groups. The algorithm's performance increase across the first week since injury within this subset of indeterminate burns, and the algorithm performance had a sensitivity above 50% by day 2 and a PPV above 50% by 12 hours (**Figure 9**).

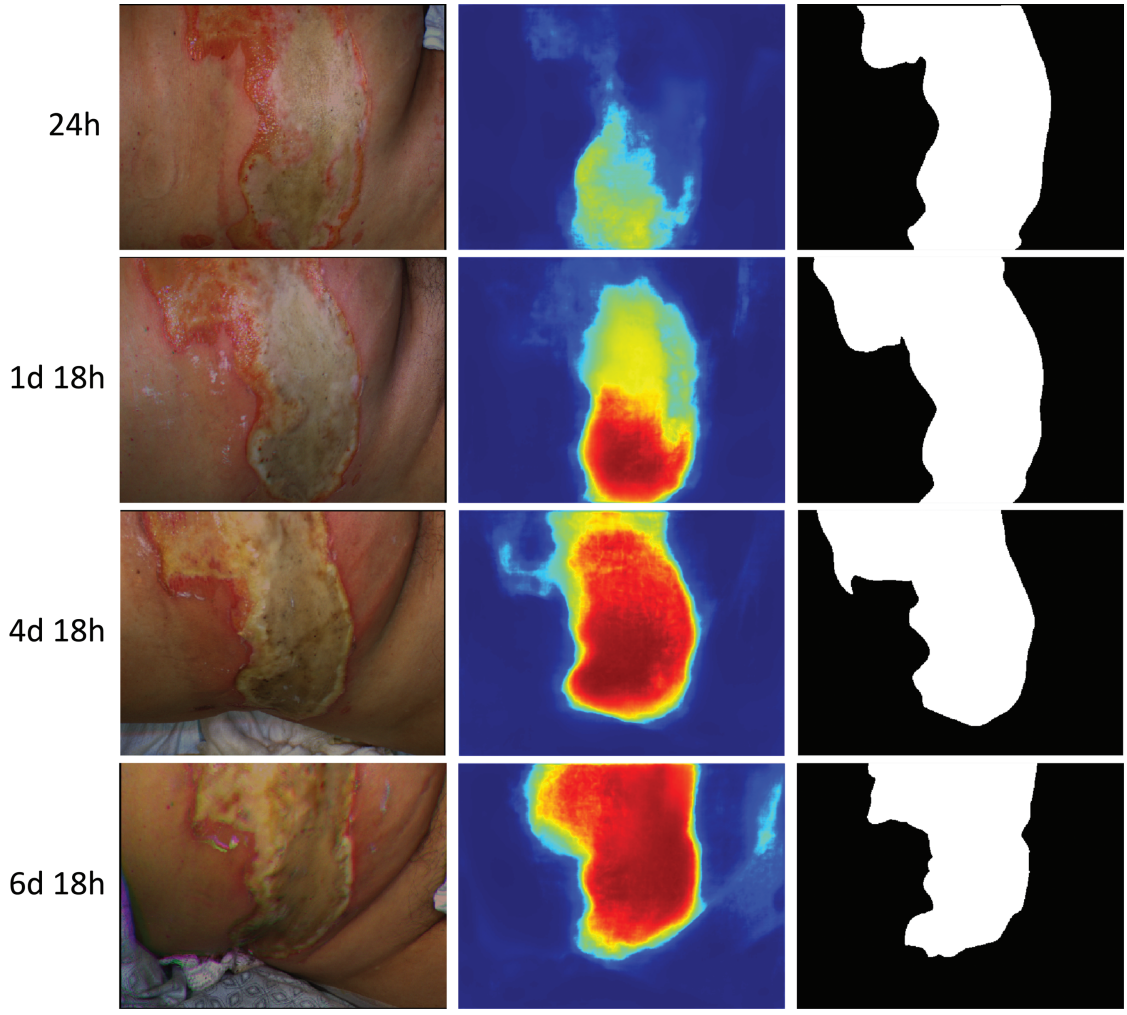

Figure 8. Repeated measurements from a nonhealing burn in the study: (Left column) color images showing the edges of the burn began as hyperemic at 24 h postburn (pinkish-red) and then went on to develop an eschar that became more yellow as time went on. (Middle column) predicted probability of nonhealing burn by the Voting Ensemble algorithm at each timepoint based on cross-validation. Dark red represents probability of 100% and dark blue probability of 0%. Probability of nonhealing burn increases as time-since-injury increases. (Right column) ground truth images indicating the true location of nonhealing burn in white.

Table 4. Estimated performance of DL algorithm on indeterminate depth burns

Indeterminate Diagnosis (Yes/No)	Number of Images Assessed as Indeterminate (N)	Sensitivity (%)	PPV (%)	Specificity (%)	NPV (%)
Yes	227	70.3 (46.0, 86.8)	97.1 (84.3, 99.5)	100 (99.9, 100)	100 (99.9, 100)
No	179	88.0 (70.3, 95.3)	96.4 (97.1, 99.4)	100 (99.9, 100)	100 (99.9, 100)

DL, deep learning; *NPV*, negative predictive value; *PPV*, positive predictive value.

DISCUSSION

This study was motivated by the need for burn assessment technologies and the importance of accurate clinical gold standards in their development. To the best of our knowledge, this was the first human study of a burn diagnostic device where *entire* burn images were labeled using a panel of surgeons with access to gold-standard methods for determining the true depth of the burn.

Previously, significant clinical investigation in technologies such as laser Doppler imaging, laser speckle, spatial frequency domain imaging, thermography, and polarized light imaging were conducted for the purpose of assisting clinicians during burn assessment.[3,36] Recent advancements include cases where AI, specifically DL algorithms, applied to burns have identified burn severity with high levels of prediction accuracy.[37–39] These investigations, while mainly focusing on pediatric scald injuries, represent important steps in noninvasive

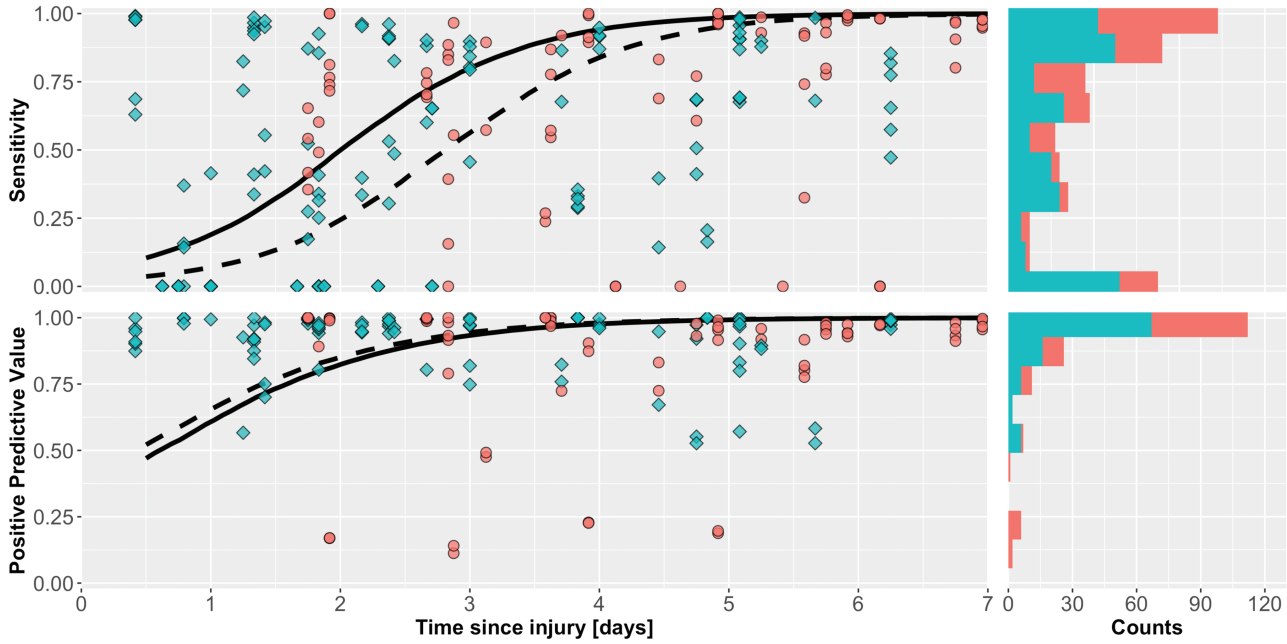

Figure 9. Performance of DL algorithm over time with burns assessed as indeterminate indicated by, and burns that were diagnosed by the clinician as. Dashed-line: performance of DL algorithm on the subgroup of indeterminate burns. Solid-line: performance of AI on diagnosed burns. Histograms represent the total number of images at each level of DL performance. DL, deep learning.

imaging for burn wound diagnosis, which we sought to advance by expanding into the more general case of thermal burns from both adult and pediatric patients.

In the current study, we employed rigorous methods for diagnosing burn severity, including biopsies and 21-day healing assessments performed by panel of burn experts, to translate diagnostic data into accurate labels for AI training. Using these diagnostic gold standards and evaluation by a clinician panel is best practice, because: (1) using clinician judgment alone for image labeling teaches the AI the same errors and biases of the clinician; and (2) biopsies and healing assessments inform on the true severity of the burn, but still require interpretation by clinicians, or panel of clinicians, which is preferred for reaching a consensus.

Moreover, handling strong correlations in burn image datasets is challenging for researchers. Correlation is high between regions within the same image, between images from the same burn, and between different burn areas on the same patient. Therefore, the separation of subject-level data is critical when training and reporting the performance of AI algorithms. These methods enable the generation of reliable pixel-level ground truth for algorithm development and performance assessment.

The DL algorithm developed using the pilot study data shows potential clinical benefit starting at 36 hours postburn. At this time the PPV was above 75% and NPV was 99%, indicating feasibility in using MSI combined with DL to identify nonhealing burns with lower risk of false positives, even in burns of indeterminate depth. We have already begun collection of additional data from more sites to train a more robust algorithm and to collect an independent set of burn images for algorithm validation.

Of clinical interest was the performance of the DL algorithm on indeterminate burns. We found that during periodic evaluation of IDBs, the imaging device's PPV was very high, 97.1%. The specificity for IDBs, 70%, suggests that 30% of the nonhealing burn areas will be missed, but the high PPV value tells us that the device is nearly always correct when it does find a nonhealing burn area. This is an especially important statistic when the decision may lead to surgical management of the burn. Although further evaluations are necessary, this suggest MSI combined with DL would be a reliable aid to clinical assessment of IDBs.

Spectral imaging techniques have progressed significantly since the studies of multi-spectral imaging of burns that occurred in 1970s. The device used in this study was able to acquire the entire multispectral image in less than seven seconds. With the high intensity illumination output, these images could be acquired indoors in the presence of ambient room lighting. Computer processing of MSI images into a final highlighted output using DL is accomplished in under one second using a consumer grade graphics processing unit. This makes the data acquisition and processing aspects of MSI practical in a variety of clinical environments.

In comparing the performance of the CNN architectures, we believe analysis of the PAR curve is a useful tool to investigate the performance of image segmentation algorithms in unbalanced datasets. The clinical utility of a diagnostic with high sensitivity and specificity is not guaranteed. For example, even when sensitivity is high, the majority of positive results from a diagnostic will be *false positives* when the disease prevalence is low. Therefore, we were careful to evaluate all four metrics of performance (sensitivity, specificity, PPV, and NPV) and found that classifiers could be compared better when PPV was considered.

The models used to evaluate the impact of time-since-injury on the performance of the ensemble algorithm showed

a substantial improvement in specificity and PPV occurring around day -2.5. This trend was expected. Other studies of optical techniques to measure burn depth also show peak effectiveness is reached at about the third day postburn.[14,16,40] It was previously attributed to there being little apparent difference between SPT and DPT burns prior to day 3. We also observed the optical characteristics of partial-thickness burns changing up to 72 hours postburn (**Figure 8**). There remains a question as to whether burn conversion is a factor contributing to the change in performance of optical measurements over time. As partial-thickness burns cause injury down to the adnexal structures of the skin, a key reservoir of dermal and epidermal progenitor cells,[41] there is a high degree of uncertainty as to which areas will retain their regenerative capacity.[27] We might hypothesize those areas with a low nonhealing in the early imaging timepoints were, in fact, healing burns that had not fully converted to nonhealing burns. However, it was not possible to ascertain the conversion status of the burn separately at each imaging timepoint in this study design.

Study Limitations

A key focus of this investigation was on the use of DL algorithms to process MSI image data. We view this study, comprising 406 images, as a first step in generating a large database of images required for the application of DL to burn depth assessment. The following compromises were made while working within this data set. First, we utilized LOOCV to estimate the real-world performance and selected the threshold from these results. Second, the current dataset was underpowered to perform subgroup analysis of demographics and geography. Lastly, the data was collected from subjects treated by one team of burn providers. This same team also constituted the truthing panels used to generate ground truth images for training the algorithm. This could have introduced bias, as the primary surgeon might influence the ground truth to reflect their treatment decision. Based on these identified limitations, a follow-on study has been conducted to expand the image dataset by including more subjects, extend the age range to include pediatric burns, and to perform subgroup analyses with variables important to burn variability and MSI imaging. In addition, we implemented multiple strategies to address truthing panel member bias by restricting practitioners from reviewing subjects treated at their institution.

Utility of Ground Truth (Healing Assessments and Biopsies) in Algorithm Development

Obtaining biopsies and/or 21-day healing assessments was an improvement over clinician judgment. For example, 23% of biopsies were taken from areas of healing burn despite these areas being excised during surgery. This demonstrates that relying on clinician diagnoses for ground truth could introduce unwanted errors in the algorithm development process.

One drawback to using biopsies for burn assessment is that a standard protocol for evaluating biopsies is an ongoing research topic.[42] Therefore, we relied on available published literature and consultation with an expert pathologist to derive the formula utilized in this study. Drawbacks to this method are: (1) the study design did not require biopsies be collected on the same day postburn from each study burn; (2) the panel of burn experts was required to extrapolate the depth of burns in the spaces between each biopsy to label the entire study burn area; and (3) burn depth was identified by H&E staining alone. The level of extrapolation was reduced by taking multiple biopsies and using visual indications of changes to burn depth when selecting biopsy location.

CONCLUSIONS

These results are an incremental step towards understanding the clinical performance and benefits of MSI paired with DL for burn assessment. A substantial database will be required to train the algorithm before it can be finalized and compared to the abilities of an experienced burn surgeon. Meanwhile, this pilot study demonstrated the value of continuing to build a larger database for algorithm training. Future work is aimed at increasing the size and variability of burns in the training dataset to improve on both DL performance metrics as well as generalizability to the broader population of burns.

ACKNOWLEDGEMENTS

This project has been supported in whole or in part with federal funds from the Department of Health and Human Services; Administration for Strategic Preparedness and Response; Biomedical Advanced Research and Development Authority, under Contract No. HHSO100201300022C.The findings and conclusions have not been formally disseminated by the Department of Health and Human Services and should not be construed to represent any agency determination or policy.

REFERENCES

1. White CE, Renz EM. Advances in surgical care: management of severe burn injury. Crit Care Med 2008;36:S318–24. doi:10.1097/CCM.0b013e31817e2d64.
2. Association AB. National Burn Repository, 2019 update, report of data from 2009-2018. 2019.
3. Jaskille AD, Shupp JW, Jordan MH, Jeng JC. Critical review of burn depth assessment techniques: Part I. Historical review. J Burn Care Res 2009;30:937–47. doi:10.1097/BCR.0b013e3181c07f21.
4. Monstrey S, Hoeksema H, Verbelen J, Pirayesh A, Blondeel P. Assessment of burn depth and burn wound healing potential. Burns 2008;34:761–9. doi:10.1016/j.burns.2008.01.009.
5. Jeschke MG, van Baar ME, Choudhry MA, Chung KK, Gibran NS, Logsetty S. Burn injury. Nat Rev Dis Primers 2020;6:11. doi:10.1038/s41572-020-0145-5.
6. Giretzlehner M, Ganitzer I, Haller H. Technical and medical aspects of burn size assessment and documentation. Medicina (Kaunas) 2021;57:242. doi:10.3390/medicina57030242.
7. Resch TR, Drake RM, Helmer SD, Jost GD, Osland JS. Estimation of burn depth at burn centers in the United States: a survey. J Burn Care Res 2014;35:491–7. doi:10.1097/BCR.0000000000000031.
8. Pape SA, Skouras CA, Byrne PO. An audit of the use of laser Doppler imaging (LDI) in the assessment of burns of intermediate depth. Burns 2001;27:233–9. doi:10.1016/s0305-4179(00)00118-2.
9. Karim AS, Shaum K, Gibson ALF. Indeterminate-depth burn injury-exploring the uncertainty. J Surg Res 2020;245:183–97. doi:10.1016/j.jss.2019.07.063.
10. Mayer Tenenhaus MD FACS H-ORM. Treatment of superficial burns requiring hospital admission. 04/27/2021.
11. Hoeksema H, Van de Sijpe K, Tondu T et al. Accuracy of early burn depth assessment by laser Doppler imaging on different days post burn. Burns 2009;35:36–45. doi:10.1016/j.burns.2008.08.011.
12. Xiao-Wu W, Herndon DN, Spies M, Sanford AP, Wolf SE. Effects of delayed wound excision and grafting in severely burned children. Arch Surg 2002;137:1049–54. doi:10.1001/archsurg.137.9.1049.

13. Anselmo VJ, Zawacki BE. Multispectral photographic analysis. A new quantitative tool to assist in the early diagnosis of thermal burn depth. Ann Biomed Eng 1977;5:179–93. doi:10.1007/BF02364018.
14. Afromowitz MA, Van Liew GS, Heimbach DM. Clinical evaluation of burn injuries using an optical reflectance technique. IEEE Trans Biomed Eng 1987;34:114–27. doi:10.1109/tbme.1987.326036.
15. Eisenbeiss W, Marotz J, Schrade JP. Reflection-optical multispectral imaging method for objective determination of burn depth. Burns 1999;25:697–704. doi:10.1016/s0305-4179(99)00078-9.
16. Afromowitz MA, Callis JB, Heimbach DM, DeSoto LA, Norton MK. Multispectral imaging of burn wounds: a new clinical instrument for evaluating burn depth. IEEE Trans Biomed Eng 1988;35:842–50. doi:10.1109/10.7291.
17. King DR, Li W, Squiers JJ et al. Surgical wound debridement sequentially characterized in a porcine burn model with multispectral imaging. Burns 2015;41:1478–87. doi:10.1016/j.burns.2015.05.009.
18. Heredia-Juesas J, Graham K, Thatcher JE, Fan W, DiMaio JM, Martinez-Lorenzo JA. Merging of classifiers for enhancing viable vs non-viable tissue discrimination on human injuries. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018;2018:726–9. doi:10.1109/EMBC.2018.8512378.
19. Heredia-Juesas J, Grahaml K, Thatcher JE, Fan W, Michael DiMaio J, Martinez-Lorenzo JA. Mahalanobis outier removal for improving the non-viable detection on human injuries. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018;2018:698–701. doi:10.1109/EMBC.2018.8512321.
20. Serrano C, Boloix-Tortosa R, Gomez-Cia T, Acha B. Features identification for automatic burn classification. Burns 2015;41:1883–90. doi:10.1016/j.burns.2015.05.011.
21. Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. ArXiv. 2018;abs/1811.12231.
22. Heimbach D, Engrav L, Grube B, Marvin J. Burn depth: a review. World J Surg 1992;16:10–5. doi:10.1007/BF02067108.
23. Watts AM, Tyler MP, Perry ME, Roberts AH, McGrouther DA. Burn depth and its histological measurement. Burns 2001;27:154–60. doi:10.1016/s0305-4179(00)00079-6.
24. Papp A, Kiraly K, Harma M, Lahtinen T, Uusaro A, Alhava E. The progression of burn depth in experimental burns: a histological and methodological study. Burns 2004;30:684–90. doi:10.1016/j.burns.2004.03.021.
25. Singer AJ, Berruti L, Thode HC, Jr, McClain SA. Standardized burn model using a multiparametric histologic analysis of burn depth. Acad Emerg Med 2000;7:1–6. doi:10.1111/j.1553-2712.2000.tb01881.x.
26. Meyerholz DK, Piester TL, Sokolich JC, Zamba GK, Light TD. Morphological parameters for assessment of burn severity in an acute burn injury rat model. Int J Exp Pathol 2009;90:26–33. doi:10.1111/j.1365-2613.2008.00617.x.
27. Shupp JW, Nasabzadeh TJ, Rosenthal DS, Jordan MH, Fidler P, Jeng JC. A review of the local pathophysiologic bases of burn wound progression. J Burn Care Res 2010;31:849–73. doi:10.1097/BCR.0b013e3181f93571.
28. Evers LH, Bhavsar D, Mailander P. The biology of burn injury. Exp Dermatol 2010;19:777–83. doi:10.1111/j.1600-0625.2010.01105.x.
29. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2015:234–41.
30. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95. doi:10.1109/TPAMI.2016.2644615.
31. Anthimopoulos M, Christodoulidis S, Ebner L, Geiser T, Christe A, Mougiakakou S. Semantic segmentation of pathological lung tissue with dilated fully convolutional networks. IEEE J Biomed Health Inform 2019;23:714–22. doi:10.1109/JBHI.2018.2818620.
32. Renton G, Soullard Y, Chatelain C, Adam S, Kermorvant C, Paquet T. Fully convolutional network with dilated convolutions for handwritten text line segmentation. Int J Doc Anal Recognit 2018;21:177–86. doi:10.1007/s10032-018-0304-3.
33. Dietterich TG. Ensemble methods in machine learning. Multiple Classifier Systems. Lect Notes Comput Sci 2000;1857:1–15. doi:10.1007/3-540-45014-9_1.
34. Zhang C, Zhang C, Ma Y, SpringerLink. Ensemble machine learning: methods and applications. 1st ed. New York, NY: Springer New York: Imprint: Springer; 2012.
35. Gelman A, Hill J, collection EBe. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2007.
36. Thatcher JE, Squiers JJ, Kanick SC et al. Imaging techniques for clinical burn assessment with a focus on multispectral imaging. Adv Wound Care (New Rochelle) 2016;5:360–78. doi:10.1089/wound.2015.0684.
37. Suha SA, Sanam TF. A deep convolutional neural network-based approach for detecting burn severity from skin burn images. Mach Learn Appl 2022;9:100371. doi: 10.1016/j.mlwa.2022.100371.
38. Cirillo MD, Mirdell R, Sjoberg F, Pham TD. Improving burn depth assessment for pediatric scalds by AI based on semantic segmentation of polarized light photography images. Burns 2021;47:1586–93. doi:10.1016/j.burns.2021.01.011.
39. Cirillo MD, Mirdell R, Sjoberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. J Burn Care Res 2019;40:857–63. doi:10.1093/jbcr/irz103.
40. Gill P. The critical evaluation of laser Doppler imaging in determining burn depth. Int J Burns Trauma 2013;3:72–7. (https://www.ncbi.nlm.nih.gov/pubmed/23638324)
41. Herndon DN. Total burn care. 3rd ed. Edinburgh: Saunders Elsevier; 2007.
42. Phelan HA, Holmes Iv JH, Hickerson WL, Cockerell CJ, Shupp JW, Carter JE. Use of 816 consecutive burn wound biopsies to inform a histologic algorithm for burn depth categorization. J Burn Care Res 2021;42:1162–7. doi:10.1093/jbcr/irab158.