Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging

Weizhi Li
Weirong Mo
Xu Zhang
John J. Squiers
Yang Lu
Eric W. Sellke
Wensheng Fan
J. Michael DiMaio
Jeffrey E. Thatcher

SPIE.

Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging

Weizhi Li,[a,*] Weirong Mo,[a] Xu Zhang,[a] John J. Squiers,[a,b] Yang Lu,[a] Eric W. Sellke,[a] Wensheng Fan,[a] J. Michael DiMaio,[a,b] and Jeffrey E. Thatcher[a]
[a]Spectral MD, Inc., 2515 McKinney Avenue, Suite 1000, Dallas, Texas 75201, United States
[b]Baylor Research Institute, 3310 Live Oak, Suite 501, Dallas, Texas 75204, United States

Abstract. Multispectral imaging (MSI) was implemented to develop a burn tissue classification device to assist burn surgeons in planning and performing debridement surgery. To build a classification model via machine learning, training data accurately representing the burn tissue was needed, but assigning raw MSI data to appropriate tissue classes is prone to error. We hypothesized that removing outliers from the training dataset would improve classification accuracy. A swine burn model was developed to build an MSI training database and study an algorithm's burn tissue classification abilities. After the ground-truth database was generated, we developed a multistage method based on *Z*-test and univariate analysis to detect and remove outliers from the training dataset. Using 10-fold cross validation, we compared the algorithm's accuracy when trained with and without the presence of outliers. The outlier detection and removal method reduced the variance of the training data. Test accuracy was improved from 63% to 76%, matching the accuracy of clinical judgment of expert burn surgeons, the current gold standard in burn injury assessment. Given that there are few surgeons and facilities specializing in burn care, this technology may improve the standard of burn care for patients without access to specialized facilities. [DOI: 10.1117/1.JBO.20.12.121305]

Keywords: multispectral imaging; burn; medical imaging; machine learning; model accuracy; outlier detection.

Paper 150329SSPR received May 14, 2015; accepted for publication Jul. 20, 2015; published online Aug. 25, 2015.

1 Introduction

1.1 *Multispectral Imaging Application*

Multispectral imaging (MSI) techniques originate from remote sensing technology that gathers information across the electromagnetic spectrum.[1] Recently, MSI has been used widely in various applications such as astronomy,[2] agriculture,[3] geology,[4] and medical imaging.[2] In the field of medical imaging, several studies have taken advantage of tissue–light interaction properties at key wavelengths to obtain spectral signatures that can differentiate epithelial tissue from connective tissue.[2,5] Furthermore, our group has employed this imaging method in ulcer assessment[6] and burn classification.[7]

In this paper, we introduce an application of MSI technology for burn wound analysis. For burn treatment, it is important to determine the depth of the initial injury. Shallower burns, known as superficial partial thickness burns, do not require surgical therapy and typically heal with supportive therapy. More severe burns, categorized as deep partial thickness or full thickness burns depending on their depth, require surgical excision of all necrotic tissue in order to expose a viable wound bed as a base for grafting surgery. Currently, the gold standard of burn wound classification is the clinical judgment of expert burn surgeons. However, the accuracy of such experts has been estimated to be only 60% to 80%, and the accuracy of nonexperts is no higher than 50%.[8] A technological solution to improve the accuracy of burn classification, particularly in medical centers where burn experts are not available, is needed to improve clinical decision making regarding burn treatment. MSI can classify burn tissue into different clinical categories with a potentially high degree of accuracy, allowing burn surgeons to more frequently and quickly select appropriate treatment solutions. During the debridement of necrotic tissue from severe burns, surgeons aim to minimize the removal of any excess healthy tissue. MSI has the further potential to aid surgical excision by categorizing burn tissue intraoperatively to differentiate burn injury from healthy wound bed, preventing unnecessary excision of healthy tissue.

Human skin is a multilayer tissue consisting of multiple chromophore components, of which there are four significant constituents: blood, water, melanin, and fat.[9] Blood, water, melanin, and fat in the various skin layers have well-established spectral responses to optical illumination with certain wavelengths of light, especially in the visible and near-infrared bands.[9,10] By capturing and analyzing different tissues' responses to multiple incident characteristic wavelengths with MSI, one can, e.g., identify the presence of blood among other tissues by its unique spectral response.[9] Tissue response to incident light is quantified by its absorbance. The collection of absorbance data over a range of wavelengths by MSI allows the classification of different tissue types based on the relative amounts of tissue constituents present within each tissue class.

*Address all correspondence to: Weizhi Li, E-mail: li@spectralmd.com

1083-3668/2015/$25.00 © 2015 SPIE

1.2 *Classification Model Development and Outlier Detection*

Although MSI is capable of capturing unique spectral data from various tissue types, a classification model must be developed to interpret new spectral images and correctly identify tissues. A difficulty arises when developing the model, because it must be built from the same type of data that it will later be used to classify, through a process called machine learning. Therefore, during initial model construction, a "training" dataset must first be collected and manually classified as the "ground truth." Establishing the ground truth is a key step in any machine learning application and is, therefore, one of the most scrutinized stages in the development of these applications. A highly accurate ground truth is necessary to build an accurate classification model. The manner by which the ground truth is established varies depending on what the classification model is being constructed to assess. In every instance, however, it must be established by clinical experts using the current gold standard to gather the necessary information. For burn wounds, the gold standard for tissue classification is histopathological assessment. We present the details of our technique for establishing the ground truth in Sec. 2.

The training set is then used to develop the classification model, which is subsequently tested on additional collected data to determine its accuracy against the ground truth. Various algorithms have been developed to build classification models from ground truth training datasets. For example, the support vector machine (SVM)[11] algorithm has been used previously in kernel-based machine learning[2,12,13] for hyperspectral imaging data analysis.

Ultimately, manual demarcation of training data establishes the ground truth, so there is a potential bias in the resulting model due to classification errors. For example, if healthy skin is inappropriately classified as blood in the training data, the resulting model would subsequently have difficulty in accurately classifying healthy skin versus blood. As the training data is the sample space used to build the classification model, reducing any such bias is the key to improving the model's accuracy.

The inevitable bias in any training set ultimately reduces the model accuracy when it is tested after development. To reduce variance and improve model accuracy, the identification and removal of "outliers" from the training dataset are helpful. An outlier is defined as an observed variable that is statistically different from other observed variables.[14] Outlier detection (also known as anomaly detection or novelty detection) is a key element of statistical pattern recognition research, with applications in fields such as credit card fraud, sensor events, medical diagnosis, and network security. There are several established methods of outlier detection.[15] One commonly implemented outlier detection technique is the model-based algorithm.

In model-based algorithms, a statistical test estimates the parameters of the sample distribution. For example, a Gaussian distribution is described by two parameters: mean and standard deviation. These parameters are determined by the maximum likelihood[16] or maximum a *posteriori* estimation.[17] In a univariate Gaussian distribution, outliers are the points that have significantly extreme probabilities (high or low) of being included within the model parameters as quantified by a Z–score (standard score). Traditionally, samples with probabilities greater than 0.95 or less than 0.05 are considered outliers in univariate analysis.

The model-based algorithm correctly identifies outliers in many cases. However, it is important to note that the parameters that define these models are sensitive to any potential outliers when they are initially calculated. That is, the parameters are generated using the entire sample set, before outliers can be identified and removed. Therefore, by identifying and removing outliers before these algorithms are used to generate classification models, the accuracy of these models can be increased. In this research, we present a machine learning algorithm in the medical space to which we apply the concept of outlier removal. MSI imaging data was first captured from an established porcine burn model. Then we assessed the multispectral images and provided a statistical solution to quantitatively improve the classification accuracy of a model designed to classify the different tissues present in the burn injury images.

2 Methodology

2.1 *Hardware*

The multispectral image data were acquired using a home-made bench top imaging setup. Figure 1 illustrates the schematics of this image acquisition system.[18] The lighting source and the image capture module were both placed in a reflective mode at a distance of 60 ± 1 cm away from the target surface. A tungsten light (ViP Pro-light, Lowel Inc.) provided a broad spectral projection on the target surface in DC-mode. One piece of frosted glass (iP-50, Lowel Inc.) was mounted in front of the tungsten light to diffuse the light and increased the uniformity of spatial illumination. Some incident light penetrated through the target surface, while any back-scattered optical signal was collected by the image capture module. The image capture module consisted of a high-performance IR-enhanced optical lens (model: Distagon T* F-2.8/25 mm, Zeiss), an eight-slot filter wheel, and a 12-bit monochromatic camera (BM-141GE, JAI Inc.). The optical bandpass filters were designed and selected to isolate a single wavelength of light for the camera. The

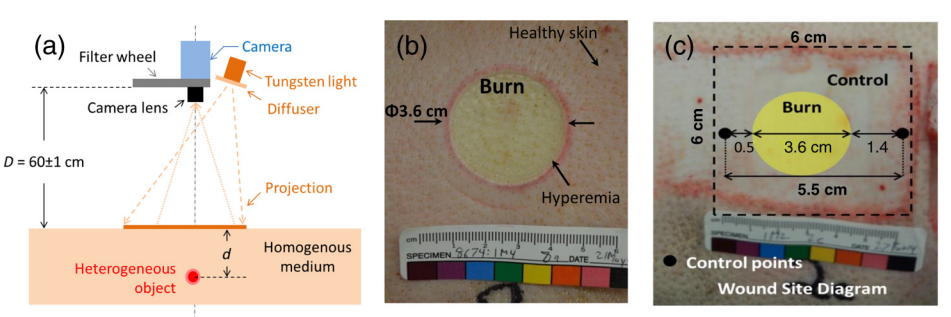

Fig. 1 (a) Hardware system set, (b) animal burn, and (c) first cut in burn tissue.[17]

following eight bandpass filters were installed in the filter-wheel. The center wavelength (CWL) and the full width at half maximum (FWHM) of the eight filters were (CWL-FWHM, both in nm): 420-20, 542-10, 581-20, 601-13, 726-41, 800-10, 860-20, and 972-10. Wavelength intensity was normalized by using a Reflectance Zenith Lite Panel (SphereOptics GmbH), and the maximum value of a pixel was 4098 (12 bits). The eight implemented wavelengths were selected based on known skin tissue absorption behavior at these wavelengths that would allow for accurate tissue differentiation for useful classification (see Ref. 7 for details). The camera sequentially captured single-wavelength images through each of the eight filters as the filter wheel rotated. Images were saved on the computer in an uncompressed format. All calculations and statistics were performed using MATLAB® software (version 2014 b).

2.2 *Animal Model and Image Acquisition*

We used the system above to collect imaging data by following an animal burn model protocol that was approved by and under the oversight of an Institutional Animal Care and Use Committer. In order to approximate human skin (epidermis thickness: 50 to 120 μm), male Hanford swine (epidermis thickness: 30 to 40 μm) were selected as the animal model.

Circular burns (diameter = 3.6 cm) were made on the backs of swine [Fig. 1(b)]. At this stage, three skin tissues were visualized: healthy, burned, and hyperemia (reddening of the skin due to increased blood perfusion following an injury). Debridement was carried out in serial 1-mm depth tangential excision layers, and the area of each debridement for each burn was 6 cm × 6 cm [Fig. 1(c)]. During debridement, six different skin tissues were appreciable: healthy, partial burn or full burn (depending on burn severity), blood, wound bed, and hyperemia. Each tangentially excised layer was stored in 10% neutral buffered formalin and sent for histopathological examination. Each specimen was sectioned and stained with hematoxylin and eosin (H&E). The purpose of the histological examination was to obtain the "gold-standard" identification of the tissue types previously mentioned, and their location in the multispectral images. The depth of burn damage and the precise excision layer at which viable tissue had been reached were determined by two pathologists.

Three pigs with six burn locations on each pig were used. For each burn location, we performed image acquisition using all eight wavelengths during at least five different time points—baseline images taken prior to injury, burn images taken directly after thermal injury, an image following the first 1-mm tangential excision with the dermatome, and two more images following the next two tangential excisions.

2.3 *Training Data Collection*

A supervised learning method was implemented to generate the classification model. To build a training database consisting of the six skin tissue classifications, we extracted the pixel intensity and the location of each of the six tissue types in every acquired image using the histology data as a reference. As shown in Fig. 2, each slice of tangentially excised skin was sectioned to show the burn depth as determined by board-certified pathologists according to well-established protocols.[19] We developed a drawing tool to mark the regions of healthy, partial burn injury, full burn injury, blood, wound bed, and hyperemia. The pathologists used the following parameters to determine

Fig. 2 Example of excised debridement layer used to establish the "ground truth" for tissue classifications. Tissue above the yellow line (more superficial) represents full burn injury. Tissue between the yellow and black lines represents partial burn injury. Tissue below the black line (deeper) represents the healthy wound bed. These regions were classified by board-certified pathologists.

these regions from the H&E-stained burned eschar: full burn injury is the zone of maximum damage. There is irreversible tissue loss due to coagulation of collagen and other tissue components. Histologically, this region is characterized by the loss of cellular detail. Partial burn injury has decreased tissue perfusion, with evidence of vascular occlusion. Collagen generally retains its structural integrity. However, there is some evidence of cellular necrosis with pyknotic nuclei. This tissue zone is considered to have the potential of being salvaged. Healthy wound bed was demarcated where essentially normal histological findings were present deep to burn tissue. These regions were then correlated with the previously acquired spectral imaging data, thereby establishing a ground truth by which our classification algorithms could be judged.

2.4 *Outlier Detection*

To reduce the influence of outliers on the model, an outlier detection algorithm utilizing two novel concepts was developed from the well-established foundation of maximum likelihood estimation as previously described.[16] First, a subset of samples located around the median of the sample space was taken as a subspace to calculate the mean and the standard deviation parameters for the model using the maximum likelihood estimation. We called this subspace the "first window," and its size was adjusted by novel coefficients α_1 and α_2 (from 0 to 0.5, unitless), defined as distances to the left and right, respectively, of the median of the sample space (thus, the width of the first window equals $\alpha_1 + \alpha_2$). As the width of the entire sample space was normalized to 1, setting $\alpha_1 = \alpha_2 = 0.5$ would result in the entire sample being selected as the "first window." By properly adjusting these coefficients, outliers may be excluded before calculating the distribution parameters [mean (μ) and standard deviation (σ) in Gaussian distribution] for the classification models. Second, the probabilities (from Z-score or other distribution function) were weighted (W_i) by a novel feature importance (w_i) to generate a threshold for detecting outliers within the first window. The technical details of these steps are as follows.

We began with a large sample space consisting of spectral data collected from the animal model. The foundation of the algorithm consisted of the well-established maximum likelihood estimation technique.[16] For an independent and identically distributed sample, the joint density function is

$$f(x_1, x_2, x_3, \cdots, x_n|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times f(x_3|\theta) \cdots \times f(x_n|\theta),$$

where $x_1, x_2, x_3, \ldots, x_n$ are the samples and θ denotes the parameters of the model. The likelihood of the function is

$$L(\theta, x_1, x_2, x_3, \cdots, x_n) = f(x_1, x_2, x_3, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

In practice, the logarithm of the likelihood function, known as log-likelihood, can be applied as follows:

$$\ln L(\theta; x_1, x_2, x_3, \cdots, x_n) = \sum_{i=0}^{n} \ln f(x_i|\theta).$$

To estimate θ_0, the value of θ that maximizes the following equation is calculated

$$\theta \subseteq \arg \max L(\theta; x_1, x_2, x_3, \cdots, x_n).$$

We can calculate the parameter θ_0 from the method of maximum likelihood.[16] If the sample distribution is Gaussian, the mathematical equations that describe the maximum likelihood parameters are as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma^2 = \frac{\sum_{i=1}^{n}(\mu - x_i)^2}{n},$$

where x_i is the value of the sample around the median. Our first novel outlier detection and removal method calls for these parameters to be controlled by the coefficients α_i as follows:

$$n = (\alpha_1 \times N) + (\alpha_2 \times N).$$

At this juncture, we apply the second of our novel outlier detection and removal methods. We designate weights to replace probabilities when detecting outliers. First, the probabilities (p_i) and feature importance (w_i) are determined. The probabilities, p_i, can be calculated with the distribution parameters of the sample distribution function. For example, for Gaussian distribution, p_i is generated from a standard Z-score, which is calculated as follows:

$$Z = \frac{x - \mu}{\sigma},$$

where μ is the mean of the samples and σ is the standard deviation of the samples. The Z-score determines p_i as follows:

$$\Phi(z) = p(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

For our outlier detection algorithm, we adjusted the probability p_i values according to the following:

$$\begin{aligned} p_i &= 2 \times p_i && \text{if } 0.05 \le p_i \le 0.5, \\ p_i &= 2 \times (p_i - 0.5) && \text{if } 0.5 < 2 \times p_i < 0.95, \\ p_i &= 0 && \text{if } 0.95 > p_i \quad \text{or} \quad p_i < 0.05. \end{aligned}$$

The feature importance, w_i, can vary depending on the desired application[20,21] and can be adjusted to improve the accuracy of any model. In our case, the feature importance was determined by the relative utility of each of the eight wavelengths implemented in the MSI machine toward distinguishing different tissue classes from one another. In the area of machine learning, the wavelength with more discriminant information was given higher weight values.

After calculating the probabilities, p_i, and feature importance, w_i, in the steps above, the sample weights (W_i) are calculated as follows:

$$W_i = p_1 \times w_1 + p_2 \times w_2 + \cdots + p_{n-1} \times w_{n-1} + p_n \times w_n = \sum_{i=1}^{n} p_i \times w_i.$$

Finally, a threshold weight ($W_{\text{threshold}}$) is assigned to generate a "second window" of data. If W_i is greater than $W_{\text{threshold}}$ for a given sample, this sample is assigned to the training set (the

Table 1 A summary of the classification and outlier detection algorithm.

Input: Dataset

1. Random selected N samples in whole dataset.
2. Sort the samples in each wavelength
3. Find the median index of these samples.
4. Set the "first window:"

   The left bound is:

   left = median index– ($\alpha_1 \times N$)

   if left < 0

   left = 0

   The right bound is:

   right = median index + ($\alpha_2 \times N$)

   If right > N:

   right = N

5. Calculate the means $[\mu_1, \mu_2, \mu_3 \ldots \mu_n]$ and standard derivations $[\sigma_1, \sigma_2, \sigma_3 \ldots \sigma_n]$ from "first window" data.
6. Assign feature importances $w_1, w_2, w_3 \ldots, w_n$ according to desired method
7. Calculate the weight W_i from the probability and feature importance of each wavelength in each sample

   For each sample in dataset:

   $W_i = p_1 \times w_1 + p_2 \times w_2 + \cdots + p_{n-1} \times w_{n-1} + p_n \times w_n = \sum_{i=1}^{n} p_i \times w_i$

8. Set threshold value to detect outlier.

   If $W_i > W_{\text{threshold}}$

   Will be considered to build the model

   Else:

   Will be considered as outlier and removed from the training set.

second window). Otherwise, this sample point is considered an outlier and is removed from the training set.

Empiric testing was repeated to find effective values for the algorithm coefficients (α_1, α_2, w_i, and $W_{\text{threshold}}$). A summary of the described algorithm is presented in Table 1.

3 Results

3.1 *Model Accuracy*

Prior to implementing the data classification and outlier removal algorithm, unfiltered spectral imaging data was analyzed by SVM and k-nearest neighbors (KNN) classification algorithms to train multiple burn classification models. When these models were given test data to classify after training, the average accuracy of classification was 63% overall as compared to the ground truth. After establishing this baseline accuracy for the burn model, the data classification and outlier removal algorithm was applied to the spectral imaging datasets before they were used to train these same classification algorithms.

Through empiric testing, effective values of the algorithm coefficients were found to be: $\alpha_1 = \alpha_2 = 0.2$, $w_1 = w_2 = \ldots = w_8 = 1$, and $W_{\text{threshold}} = 7$. With these parameters assigned, the mean and the standard deviation parameters of the "first window" were calculated for each of the eight wavelengths implemented by MSI (Table 2).

The results of the data classification algorithm after outlier detection and removal are presented in Fig. 3. For purposes of presentation, the sample space (red) is shown in two-dimensions with only two of the eight implemented wavelengths represented. After outlier detection and removal, the second window subspace (blue) used to train the burn classification model became more homogenous and tightly clustered, theoretically allowing for greater accuracy in the resulting model.

To visualize the results of the data classification and outlier detection algorithm across all eight MSI wavelengths, boxplots

Table 2 Mean and standard deviation (STD) of implemented wavelengths.

Wavelengths (nm)	420	542	581	601	726	800	860	972
Mean	2128.9	2164.7	2169.9	2202.9	2234.9	2241.0	2233.3	2172.2
3 × STD	21.3421	28.5486	22.8444	29.9799	23.2660	20.7308	22.1283	40.0605

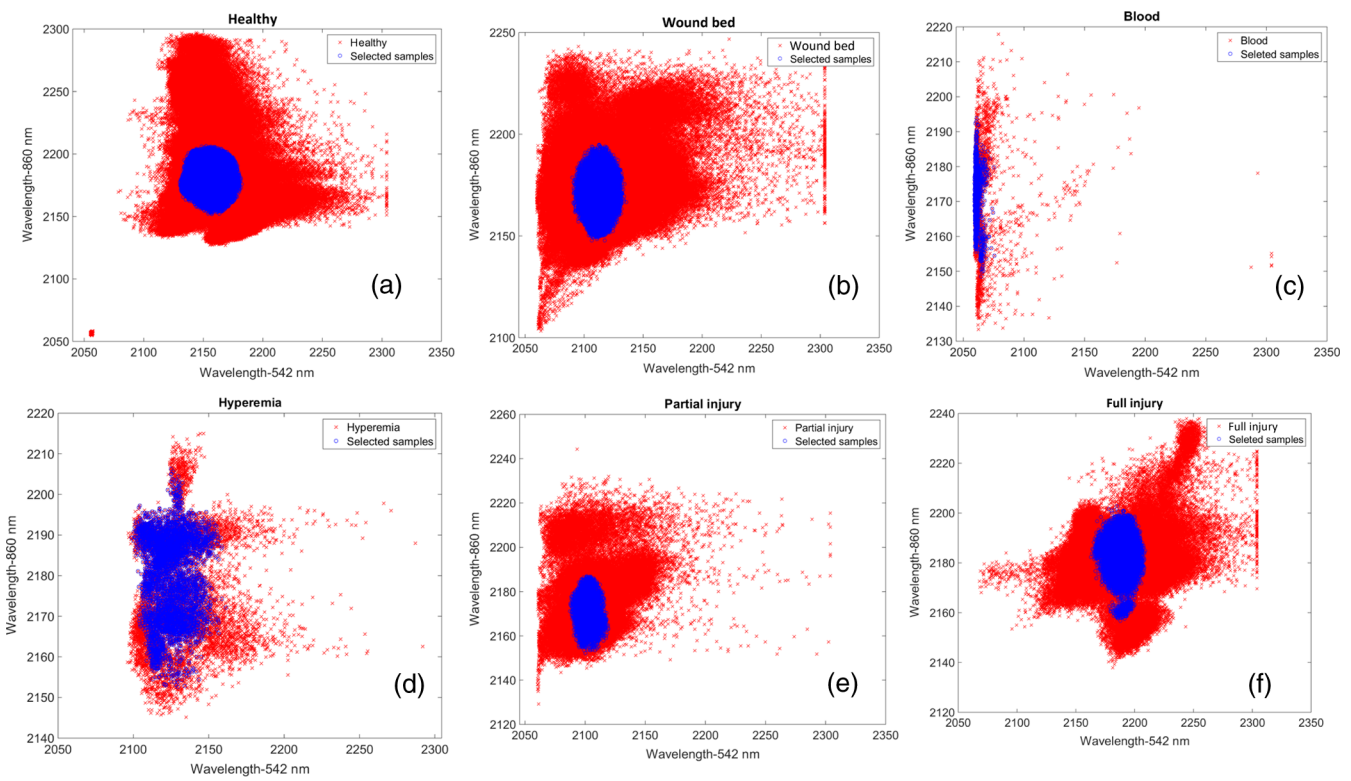

Fig. 3 Two-dimensional (2-D) representation of the experimental sample space before (red) and after (blue) outlier identification and removal for the six skin tissue classes: (a) healthy tissue, (b) wound bed, (c) blood, (d) hyperemia, (e) partial burn injury, and (f) full burn injury. Absorption at one of the eight total wavelengths were plotted on the *x* and *y* axes (542 and 860 nm, respectively) to generate 2-D plots representative of the eight-dimensional spaces used in the complete analysis. These wavelengths were selected arbitrarily, and any combination of two wavelengths would, in effect, generate similarly representative plots.

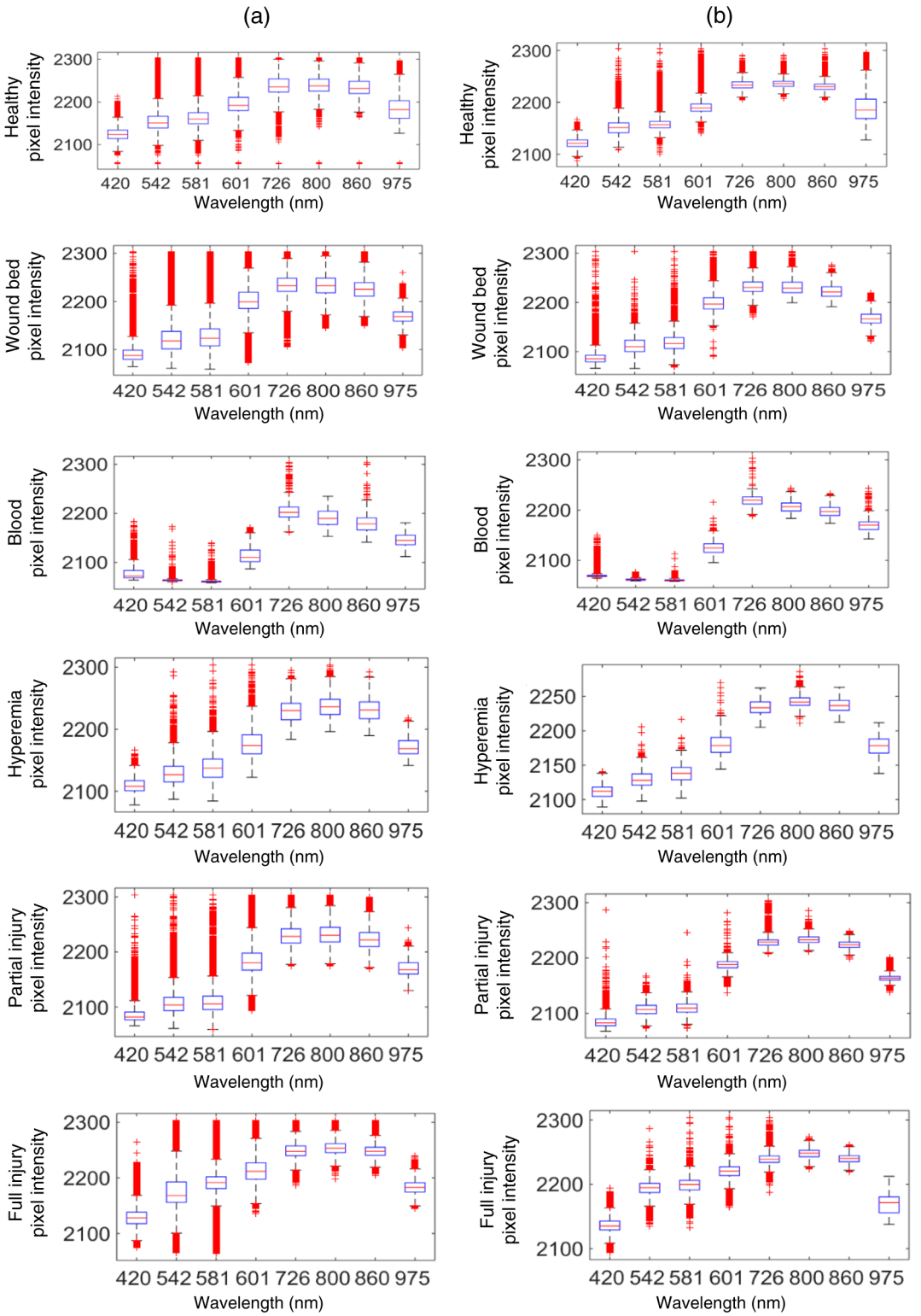

Fig. 4 Boxplots depicting sample spaces (a) before and (b) after outlier detection and removal for all eight wavelengths with each tissue classification. Boxes represent the interquartile range. Red plus signs demarcate data outlier. The number of outlier remaining in the sample space after outlier detection was significantly reduced in all tissue classes, most notably in the blood class.

representing the samples collected for all wavelengths in each tissue classification were plotted before and after outlier detection and removal. In the initial sample space [Fig. 4(a)], all tissue classifications, especially blood, included a significant number of outliers. After outlier and detection removal, the number of outliers remaining in the subspace was drastically reduced [Fig. 4(b)].

Representative two-dimensional sample spaces with spectral data for all six tissue classifications plotted together are represented in Fig. 5. Before outlier detection and removal, data from

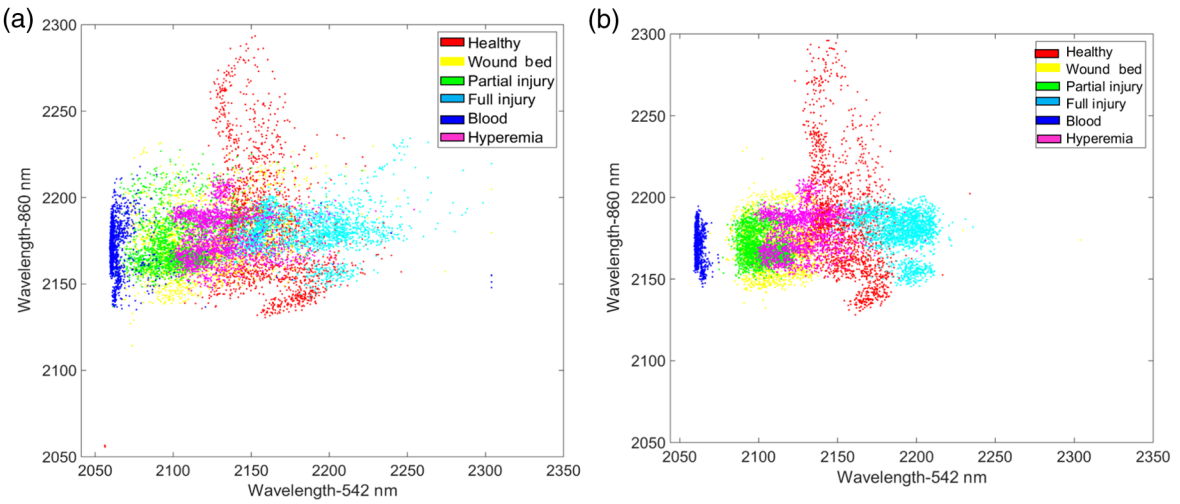

Fig. 5 Six classes in 2-D feature spaces (a) with outliers and (b) without outliers.

the various tissue classes were generally plotted in clusters, with the notable exception of blood, but a significant amount of overlap between the various clusters was appreciable. After applying the outlier detection and removal algorithm, a better separation between tissue classes was clear. After removal of outliers, new burn classification models were generated using the same classification algorithms (SVM, KNN, and so on). The overall average model accuracy improved from 63% to 76%.

The improvement in model classification accuracy is demonstrated in Fig. 6. Prior to outlier removal, the classification models could not accurately detect healthy skin or the hyperemic zone that physiologically surrounds a burn. The model also predicted several different classes of tissue where, in reality, healthy skin was present. In place of the hyperemic zone around the burn, the models predicted the presence of blood. Furthermore, healthy skin beyond the hyperemic zone was incorrectly classified as full burn injury. However, after outlier removal, the models accurately classified both healthy skin in the control image and burn image, as well as a hyperemic zone around a burn.

4 Discussion

4.1 *Multispectral Imaging Burn Model Classification Before and After Outlier Detection*

Several points from this experiment are worth highlighting. First, the assigned values for the algorithm coefficients (α_1,

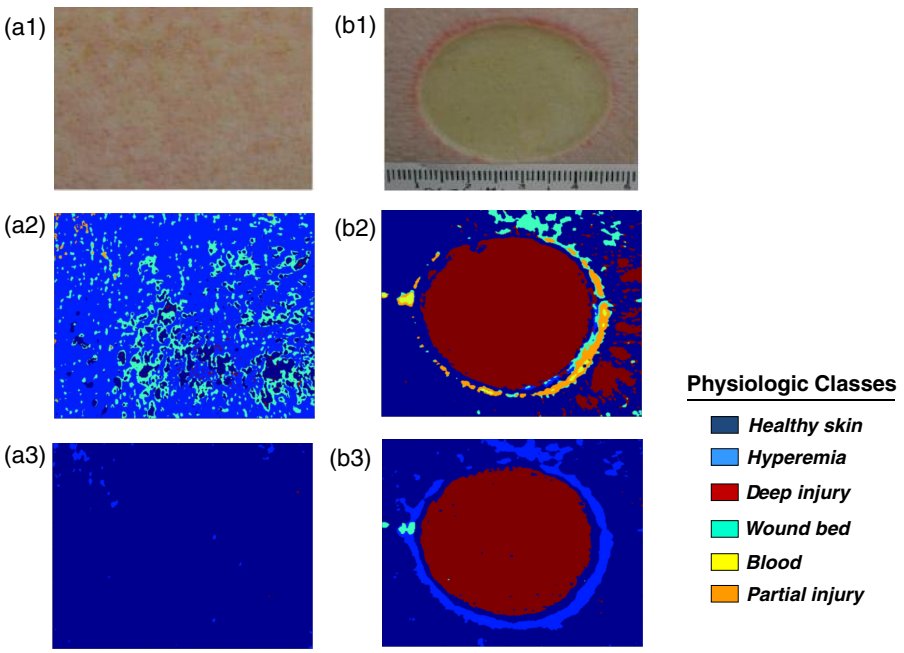

Fig. 6 Examples of burn classification model outputs. Series A represents: (a1) a healthy control image, (a2) model classification of the control prior to outlier removal, and (a3) model classification of the control after outlier removal. Series B represents: (b1) a full burn surrounded by a zone of hyperemia and healthy tissue, (b2) model classification of this image prior to outlier removal, and (b3) model classification of this image after outlier removal. Outlier detection and removal significantly increased the model accuracy in these cases.

α_2, w_i, and $W_{\text{threshold}}$) were determined through an empiric approach in a recursive process. The values were selected because they effectively increased the accuracy in the particular MSI application presented in this manuscript. However, with other applications, these values would likely need to be adjusted to achieve the desired result.

Interestingly, the optimal feature importance (w_i) for all wavelengths was set to a value of 1 after empiric testing to identify the best value for each wavelength. That all of the feature importances (w_i) were ultimately assigned a value of 1, reflects the fact that each of the eight wavelengths employed in our MSI device were selected to provide unique spectral information independently from one another. This result was not surprising given that the wavelengths were selected according to previously described[2,5] optical characteristics of skin tissue and burn tissue.

The most challenging tissue to accurately classify was blood. This was evident given the heterogeneous sample space collected for blood as represented in both Figs. 3 and 5. The bimodal distribution of spectral data characterizing blood is a result of blood's unique absorbance spectrum in the visible and near-infrared light bands, which is also bimodal. Each of the other tissue classes has a single absorbance peak, resulting in somewhat more homogenous distributions of spectral data in these other cases.

Ultimately, the outlier detection and removal algorithm significantly improved the accuracy of the MSI application for skin tissue classification. The algorithm successfully reduced the variance in the sample space for each of the tissue classes. By restricting the variance in this fashion, the overlap in spectral characteristics was reduced in a corresponding manner. With reduced overlap, the training of classification models was improved with a discernable increase in classification accuracy. By achieving a final accuracy of 76%, we improved our model to, at a minimum, meet the current clinical standard in burn tissue classification, clinical judgment by burn experts.[8] This model has the potential to aid decision-making for physicians treating burn victims in settings where burn experts may not be readily available.

4.2 *Application of Outlier Detection Algorithm to Multispectral Imaging and Beyond*

Although the outlier classification algorithm we have proposed can theoretically be applied to any number of machine learning applications,[22] we elected to apply it to MSI burn imaging classification. We made a key assumption regarding the spectral data acquired by MSI in order to simplify the outlier detection problem in this case. Namely, we have assumed the spectral data collected at each intensity-normalized wavelength represent a Gaussian distribution. This is a reasonable assumption because the sample spaces were sufficiently large to fit a Gaussian distribution according to the central limit theorem. The mean and the standard deviation of each sample space were, therefore, calculated using the appropriate equations for Gaussian distribution. In an application with data that does not conform to a Gaussian distribution, other methods such as maximum *a posteriori* estimation[17] can be used to extract these descriptive parameters.

Another key assumption in this case was that light–tissue absorption interactions at the eight wavelengths implemented in the MSI device were independent and identically distributed. To consider each wavelength as independent allowed for univariate analysis when determining the feature importance rank, which, as above, can be adjusted to achieve the desired data output for any given application. In other applications, in which the wavelengths (or any other input) cannot be considered independent and identically distributed, multivariate analysis would be required to assign feature importance ranks. Although we implemented only absorption data for each wavelength in our classification model, other descriptors of light–tissue interactions, such as scattering, could be additionally collected and incorporated into future models.

Ultimately, the extension of the proposed outlier identification and removal algorithm to other machine learning algorithms is possible, even if the simplifying assumptions used in this manuscript are not applicable in other cases. The algorithm successfully increased the accuracy of our burn model classification system.

5 Conclusions

Outlier detection and removal from training datasets is a key component of machine learning algorithms in order to improve the accuracy of the model. We have developed an outlier detection algorithm with two unique components. First, we identify a subsample from the initial sample space by selecting data points within a certain range of the median of the complete sample space. From this subsample, the distribution parameters (mean and standard deviation) can be determined. Second, instead of relying on traditional probability values to directly identify statistically significant outlier, we have developed a method to assign weights to each data feature according to the feature importance. Each of these mechanisms increases the accuracy of classification models generated from training datasets by facilitating the identification and removal of outliers. Finally, we demonstrate proof-of-concept with the application of MSI to classify skin and burn tissues. Burn model classification accuracy was significantly increased by the proposed algorithm.

Acknowledgments

This research is supported by the Biomedical Advanced Research and Development Authority (BARDA, Contract # HHSO1002013000022C). Animal model performance and support was received from Sinclare Research Center LLC. The authors would like to thank Leah Gaither for her editorial assistance.

References

1. C. Fischer and L. Kakoulli, "Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications," *Stud. Conserv.* **7**, 3–16 (2006).
2. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
3. R. Lu and Y. Chen, "Hyperspectral imaging for safety inspection of food and agricultural products," *Proc. SPIE* **3544**, 121–133 (1999).
4. F. Meer et al., "Multi- and hyperspectral geologic remote sensing: a review," *Int. J. Appl. Earth Obs. Geoinf.* **14**(1), 112–128 (2012).
5. V. T. Valery, "Light-tissue interactions," in *Biomedical Photonics Handbook*, pp. 1–27, CRC Press, Boca Raton, Florida (2003).
6. R. Moza, J. M. Dimaio, and J. Melendez, "Deep-tissue dynamic monitoring of decubitus ulcers: wound care assessment," *IEEE Eng. Med. Biol. Mag.* **29**, 71–77 (2010).
7. D. R. King et al., "Surgical wound debridement sequentially characterized in a porcine burn model with multispectral imaging," *Burns* (2015), Epub ahead of print.

8. A. Jaskille et al., "Critical review of burn depth assessment techniques: part II. Review of laser Doppler technology," *J. Burn. Care Res.* **31**(1), 151–157 (2010).
9. S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**(12), R37–R61 (2013).
10. W.-F. Cheong, S. A. Prahl, and A. J. Welch, "A review of the optical properties of biological tissues," *IEEE J. Quant. Electron.* **26**(12) (2002).
11. C. Cortes and V. Vapnik, "Support-vectors networks," *Mach. Learn.* **20**, 273–297 (1995).
12. H. Akbari et al., "Cancer detection using infrared hyperspectral imaging," *Cancer Sci.* **102**(4), 852–857 (2011).
13. D. Yodovsky et al., "Assessing diabetic food ulcer development risk with hyperspectral tissue oximetry," *J. Biomed. Opt.* **16**(2), 026009 (2011).
14. F. E. Grubbs, "Procedures for detection outlying observations in samples," *Technometrics* **11**(1), 1–21 (1969).
15. D. Coursineau, "Outliers detection and treatment: a review," *Int. J. Psychol. Res.* **3**(1), 58–67 (2010).
16. J. Aldrich, "R. A. Fisher and the making of maximum likelihood 1912–1922," *Stat. Sci.* **12**(3), 162–176 (1997).
17. J. Gauvain and C. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Process.* **2**(2), 291 (1994).
18. W. Mo et al., "The importance of illumination in a non-contact photoplethysmography imaging system for burn wound assessment," *Proc. SPIE* **9303**, 93030M (2015).
19. R. Gurfinkel et al., "Histological assessment of tangentially excised burn eschars," *Can. J. Plastic Surg.* **18**, e33–e36 (2010).
20. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
21. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1226–1238 (2005).
22. J. J. Squiers et al., "Quantifying regional left ventricular contractile function: leave it to the machines?" *J. Thorac. Cardiovasc. Surg.* **150**, 247–249 (2015).

Weizhi Li received his BSE in electrical engineering from Lanzhou University, China, in 2011 and received his Master of Science in electrical engineering, Southern Methodist University, 2013. He has been working as an algorithm engineer at Spectral MD, Inc., since January 2013. His research and work responsibility is to analyze tissue optical imaging data using machine learning, pattern recognition, computer version, and statistical analysis algorithms.

Weirong Mo received his bachelor's and master's degrees in biomedical engineering from Zhejiang University, China. He received his PhD from National University of Singapore in 2010. He joined Roswell Park Cancer Institute, Buffalo, New York, as a postdoc. Later, he joined Philips Research Asia as a senior scientist. Currently, he is a senior optical system engineer at Spectral MD, Dallas, Texas. His research and development experience is in biophotonics, optical instruments, imaging algorithms, and medical devices.

Xu Zhang is an application developer in the National Institutes of Health. She joined Spectral MD as a signal and image processing intern from January 2014 to November 2014 while seeking her master's degree in electrical engineering from Southern Methodist University. She received her MS degree in December 2014.

John J. Squiers received his BSE degree in chemical and biological engineering from Princeton University in 2012. He has worked with Spectral MD since 2010, when he completed a summer internship with the company. Currently, he is pursuing his medical degree at the University of Texas Southwestern Medical Center with plans to start a career in surgery upon his graduation in 2017.

Yang Lu is a consultant for Spectral MD, specializing in algorithm development. She received her Bachelor of Engineering from Tianjin University, majoring in telecommunications, before receiving her PhD in electrical and computer engineering from National University of Singapore. Prior to Spectral MD, she held a postdoctoral position at Oregon State University, where she researched vocalization patterns of marine mammals. Her research interests include signal processing, machine learning, motion artifact removal, and image processing.

Eric W. Sellke is a biomedical engineer at Spectral MD, where he is contributing to the development of a state-of-the-art burn care imaging device. He has publications in SPIE Proceedings, *Surgery*, and *Journal of Burn Care and Research*.

Wensheng Fan is an executive, entrepreneur, and innovator with over 20 years of experience in natural speech recognition and real-time imaging systems. Currently, he serves as an executive vice president and CTO for Spectral MD, Inc. He has experience in leading medical imaging system development from concept to market. He received his MSEE degree from Northeastern University and his BSEE from Tsignhua University in China.

J. Michael DiMaio is the founder and CEO of Spectral MD. He received his medical degree from the University of Miami and completed his thoracic surgery residency at Duke. He was faculty in cardiothoracic surgery and biomedical engineering at UT Southwestern Medical Center from 1998-2013, where he was the director of Cardiothoracic Surgery Research and held a distinguished chair in heart disease. This position allowed for Spectral MD's early development of its medical application directions.

Jeffrey E. Thatcher serves as a chief scientist at Spectral MD, where he oversees technology and applications research for medical imaging systems. He is a former HHMI undergraduate research fellow, served as PI on three NSF grants, and is currently the PI of an HHS/BARDA government contract to develop an imaging device for burn care. He received his PhD in biomedical engineering from UT Southwestern and his BS degree in molecular biology from Texas Tech University.