

Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging

Mark Witteveen[Ⓜ],^{a,b,*} Henricus J. C. M. Sterenberg[Ⓜ],^{a,c}
Ton G. van Leeuwen[Ⓜ],^c Maurice C. G. Aalders[Ⓜ],^{c,d}
Theo J. M. Ruers,^{a,b} and Anouk L. Post[Ⓜ]^{a,c}

^athe Netherlands Cancer Institute, Surgical Oncology, Amsterdam, The Netherlands

^bUniversity of Twente, Science and Technology, Nanobiophysics,
Enschede, The Netherlands

^cAmsterdam UMC, University of Amsterdam, Cancer Center Amsterdam,
Amsterdam Cardiovascular Sciences, Department of Biomedical Engineering and Physics,
Amsterdam, The Netherlands

^dUniversity of Amsterdam, Co van Ledden Hulsebosch Center,
Amsterdam, The Netherlands

Abstract

Significance: Hyperspectral reflectance imaging can be used in medicine to identify tissue types, such as tumor tissue. Tissue classification algorithms are developed based on, e.g., machine learning or principle component analysis. For the development of these algorithms, data are generally preprocessed to remove variability in data not related to the tissue itself since this will improve the performance of the classification algorithm. In hyperspectral imaging, the measured spectra are also influenced by reflections from the surface (glare) and height variations within and between tissue samples.

Aim: To compare the ability of different preprocessing algorithms to decrease variations in spectra induced by glare and height differences while maintaining contrast based on differences in optical properties between tissue types.

Approach: We compare eight preprocessing algorithms commonly used in medical hyperspectral imaging: standard normal variate, multiplicative scatter correction, min–max normalization, mean centering, area under the curve normalization, single wavelength normalization, first derivative, and second derivative. We investigate conservation of contrast stemming from differences in: blood volume fraction, presence of different absorbers, scatter amplitude, and scatter slope—while correcting for glare and height variations. We use a similarity metric, the overlap coefficient, to quantify contrast between spectra. We also investigate the algorithms for clinical datasets from the colon and breast.

Conclusions: Preprocessing reduces the overlap due to glare and distance variations. In general, the algorithms standard normal variate, min–max, area under the curve, and single wavelength normalization are the most suitable to preprocess data used to develop a classification algorithm for tissue classification. The type of contrast between tissue types determines which of these four algorithms is most suitable.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.27.10.106003](https://doi.org/10.1117/1.JBO.27.10.106003)]

Keywords: hyperspectral; preprocessing; normalization; classification; scatter correction; cancer; glare; machine learning.

Paper 220046GRR received Feb. 28, 2022; accepted for publication Aug. 23, 2022; published online Oct. 7, 2022.

*Address all correspondence to M. Witteveen, ma.witteveen@nki.nl

1 Introduction

Hyperspectral imaging can be used in medical applications to distinguish tissue types based on differences in their spectral signature.¹⁻⁵ Currently, many researchers use a variety of methods to develop classification algorithms for hyperspectral imaging, such as machine learning, statistical analysis, or fitting algorithms, which identify tissue types based on spectral differences between tissue types. The quality of the data used by the classification algorithm influences the accuracy and robustness of the developed algorithm. Ideally, the spectra used to distinguish tissue types are only influenced by the tissue composition, but unfortunately these spectra are also influenced by other factors, such as surface reflections and sample thickness variations. Removing influences not related to tissue composition is expected to improve the accuracy of the developed algorithm itself, and it would also make the developed algorithm more generalizable, because it does not depend on the amount of surface reflections and sample thickness variations present in the dataset used to develop the algorithm. Thus, reducing these influences is a valuable step in the development of classification algorithms for hyperspectral imaging.

The measured reflectance is the sum of diffuse [Fig. 1(a)] and surface reflections [Figs. 1(b) and 1(c)]. Surface reflections can be divided into mirror-like specular reflections from a smooth tissue surface [Fig. 1(b)] and specular reflections from a rough tissue surface, known as glare [Fig. 1(c)]. Glare is a form of specular reflection, albeit on a much smaller scale, that is influenced by the roughness of the tissue surface. To prevent mirror-like specular reflections from a smooth surface, the light source and camera are generally placed at an angle with respect to each other. However, this will not prevent glare. Due to the surface roughness of the tissue, light from the tissue surface will be reflected in many directions of which some will be detected by the camera. Assuming surface roughness is not homogeneous, the orientation of the tissue sample with respect to the camera will influence the detected amount of glare. Thus, the amount of glare in a measured spectrum depends not only on the tissue composition but also on the geometry of the hyperspectral setup.

In *ex-vivo* settings with benchtop systems, the influence of surface reflections can be reduced by polarization filters and the influence of variable sample heights by surface profilometry. However, for *in-vivo* applications, data analysis should be fast to provide clinicians with real-time information. Surface profilometry takes additional time during a measurement and has to be redone frequently to allow the clinician to move the hyperspectral imaging device.

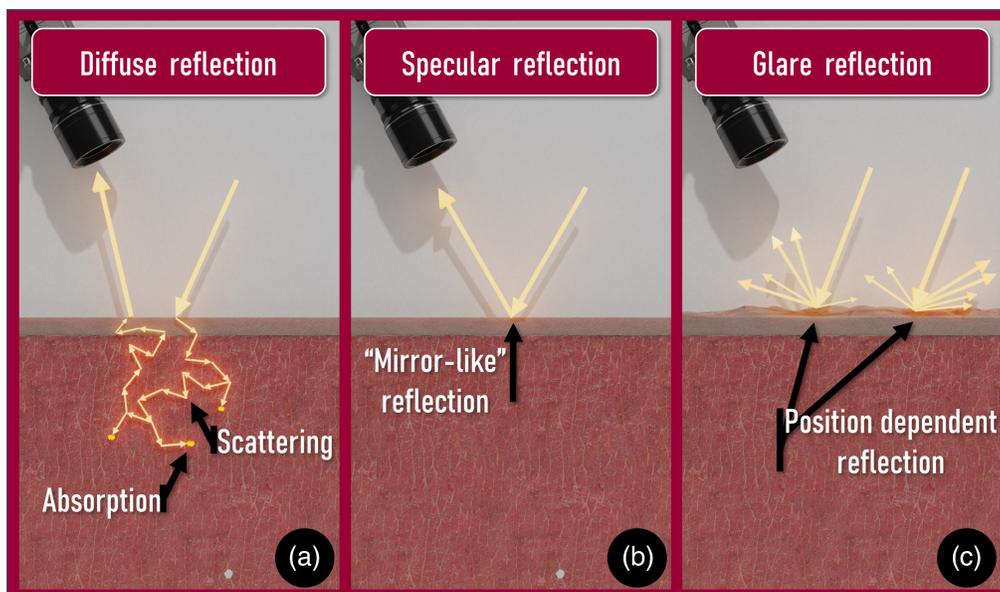


Fig. 1 Three different types of reflected light: (a) diffuse reflectance, where the light travels through the tissue, (b) specular reflectance from a smooth tissue surface, and (c) glare as a result of surface roughness, where light is reflected from the surface in many directions. Only the light that reaches the sensor should be considered for the measured signal.

Polarization filters not only remove glare but also reduce the intensity of the diffuse reflectance since the diffuse reflectance will be unpolarized. Thus, to obtain a good signal-to-noise ratio, this could increase measurement times, especially in low-light settings such as endoscopy. Instead of hardware solutions, the influence of surface reflections and sample height variations can also be reduced by preprocessing the measured spectra used by tissue classification algorithms.

In theory, when sophisticated algorithms such as convolutional neural networks are used on vast amounts of data, algorithms should be able to learn to identify tissue types in the presence of additional variations such as surface reflections. However, for clinical applications, algorithms are always developed with limited sample sizes, especially when hyperspectral imaging is used to identify tissue types since it requires correlation with histopathology, which is expensive and time-consuming. Algorithms developed on limited sample sizes with additional variations in the spectra that are not related to tissue composition could result in a poorer performance. Multiple studies have shown that, in general, preprocessing data before feeding it to a convolutional neural network improves the performance of the developed algorithm.⁶⁻⁸

An example of the influence of glare can be seen in the study by Kho et al.⁴ where they imaged breast tissue slices. In Fig. 2, spectra from two regions with different tissue types as indicated by the rectangles are shown. Within each rectangle, the tissue type is the same [Fig. 2(a)–2(c)], but large variations exist in the obtained spectra [Figs. 2(a) and 2(d)]. Because the size of these rectangles is small, no variation in tissue composition is expected. The large variations shown in Figs. 2(d) and 2(e) give the impression of a wavelength-independent offset. Specular reflection, and thus glare by extension, is related to the difference in refractive index of tissue and air. As the refractive index of tissue does not change dramatically with wavelength (for both pure water and pure fat it varies <2% over the entire 400 to 1600 nm range), the specular reflection and thus also the glare will only vary slightly with wavelength at lower wavelengths. Hence, glare would result in an offset, and thus the large variations in the measured spectra shown in Figs. 2(d) and 2(e) are likely to be attributed to glare.

The pipeline for classifying tissue using hyperspectral data can often be broken down into three steps: calibration, which is necessary to compensate for wavelength-dependent properties of the setup; preprocessing, which is needed to reduce unwanted variation in the data; and a classification algorithm, where the actual classification based on the data is made. For the development of a classification algorithm, there are several possibilities, such as fitting spectra with diffusion theory and using optical properties or principle component analysis. In recent years, machine learning is increasingly used in hyperspectral imaging to develop classification algorithms.

Hyperspectral imaging setups require calibration primarily to compensate for wavelength-dependent properties of the setup such as: sensor sensitivity, transmission of lenses used, variable properties of the light source, and the physical layout of the setup. Ideally, a calibration is done with a reference sample, performed at a location corresponding to the location of the imaged sample. However, in practice, this is rarely the case. Sample thickness variations introduce distance differences between the sample, light source, and camera. When the tissue is closer to the light source, the incident intensity will increase; if the tissue is closer to the camera, the surface area represented by a pixel on the camera will decrease. The character of these spectral variations is very different from the variations in glare: where glare adds an offset to the spectrum from within the sample, sample height variations influence the amount of light that is detected. Height differences will result in a multiplication of the entire spectrum with a wavelength-independent constant and introduce differences in the size of the tissue area that corresponds to a pixel in the camera.

Thus, glare and sample thickness variations introduce variations in the measured spectra, which are not related to tissue composition. This makes spectra from different tissue types harder to distinguish from each other. For the development of tissue classification algorithms in hyperspectral imaging, several preprocessing algorithms have been used to reduce the variations that are not related to tissue composition. An ideal preprocessing algorithm would reduce the differences in the spectra that are due to glare and height differences while retaining the spectral signatures that differ between tissue types. Currently, which preprocessing algorithm is more suitable for a specific clinical application is unknown. One approach of choosing a preprocessing algorithm would be to test combinations of preprocessing algorithms and classification

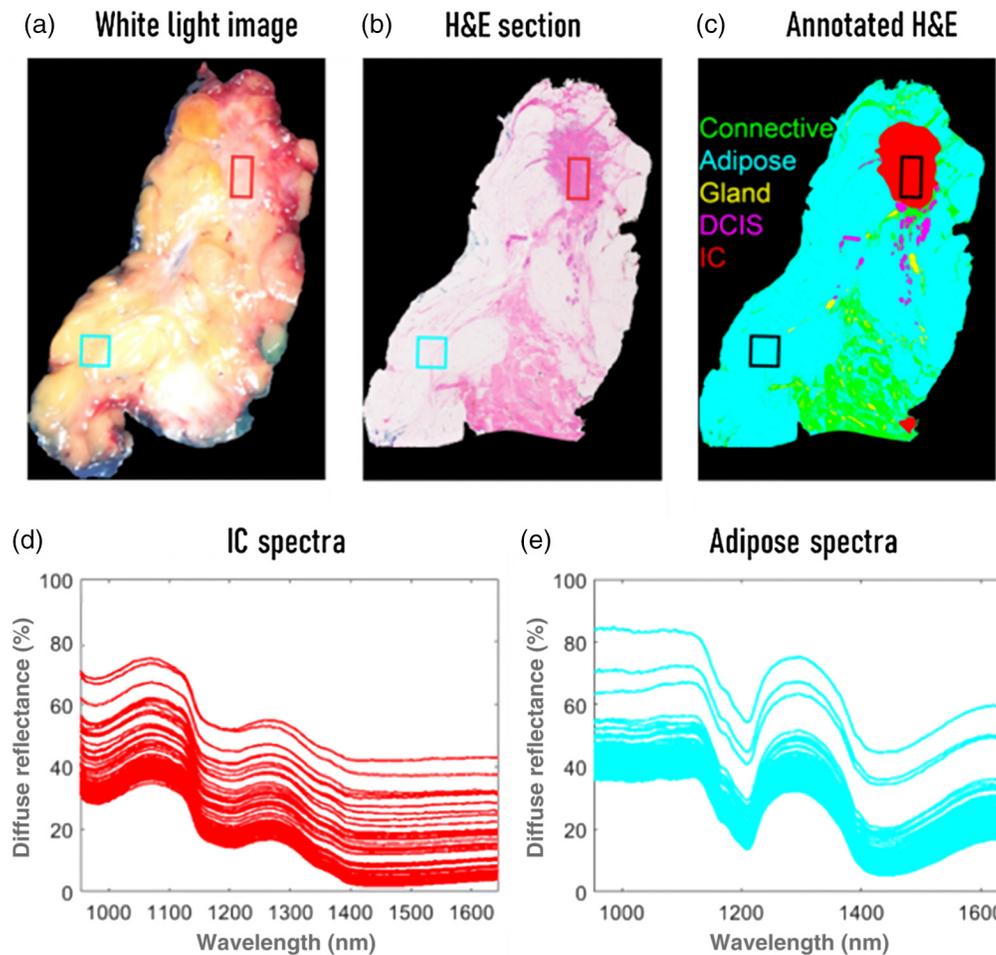
Adapted from Kho et al.⁴

Fig. 2 An illustration of variations in measured spectra, which are not directly related to tissue composition or sample thickness, obtained from resected breast tissue. (a) White light image of the tissue specimen, (b) stained H&E section, (c) image from the hyperspectral camera with annotations based on histopathology to indicate the tissue class of each pixel, (d) spectra from the red region of interest containing invasive carcinoma, (e) spectra from the blue region of interest containing adipose tissue. Because the size of these rectangles is small, no variation in tissue composition or height differences are expected. The variation between spectra within a rectangle are likely attributable to glare. Figure modified from Kho et al.⁴

algorithms and determine which combination results in the highest accuracy of the classifier. However, this approach (data-dredging—testing multiple hypotheses on a single dataset by extensively searching for the most optimal approach) could increase the probability of lucky shots and thereby nonreproducible results.⁹ Reducing the amount of considered preprocessing techniques and introducing selection criteria could reduce these effects. In this paper, we provide researchers with a solid basis to identify the most suitable preprocessing algorithm for their application.

By a literature search, we identified eight commonly used preprocessing algorithms applied to medical hyperspectral images: standard normal variate (SNV),^{10–17} multiplicative scatter correction (MSC),^{18–20} min-max normalization (MM),^{16,21–29} mean centering (MC),^{30,31} area under the curve normalization (AUC),^{32–37} single wavelength normalization (SW),^{1,38} first derivative (FD),^{16,18,39–41} and second derivative (SD).^{42,43} The details of each algorithm are discussed in Sec. 2.3.

In this study, synthetic reflection spectra (based on a simulated dataset resembling diffuse reflectance spectra from human tissue from 400 to 1600 nm) were created with known

differences in optical properties from absorption [blood volume fraction (BVF) and presence of absorbers] and scattering (scatter amplitude and slope). For each set of tissue properties, 100 spectra with variable amounts of glare and variable height differences and noise were generated. Ideally, a preprocessing algorithm would reduce the differences in spectra within one tissue type while maintaining the differences in spectra between tissue types. To quantify how well an algorithm does this, we calculated the “overlap coefficient” of the spectra, a measure of similarity (ranging from 0 to 1), for any of combination of two sets of tissue properties. For two sets of spectra with different tissue properties, a reduced overlap coefficient after preprocessing would imply that variations due to glare and height differences are reduced, while the contrast related to the difference in optical properties is retained. Thus, a lower overlap coefficient would likely improve the discrimination ability of tissue classification algorithms. Finally, we investigate the effect of the preprocessing algorithms on clinical data, measured on colon¹⁰ and breast¹¹ tissues to determine whether the trends we identify based on the synthetic data translate to clinical applications.

2 Methods

2.1 Synthetic Spectra Theoretical Background

The diffuse reflectance consists of light that has traveled through the tissue. For the illumination geometry commonly used in hyperspectral imaging, an infinite wide beam and infinite sample thickness can be assumed. Consequently, the diffuse reflectance ($R_{\text{diffuse}}(\lambda)$) of a homogeneous sample with $\mu_a \ll \mu'_s$ can be approximated using diffusion theory as:⁴⁴

$$R_{\text{diffuse}}(\lambda) = \frac{\alpha'(\lambda)}{1 + 2k(1 - \alpha'(\lambda)) + \left(1 + \frac{2k}{3}\right)\sqrt{3(1 - \alpha'(\lambda))}}, \quad (1)$$

where k is the internal reflection coefficient due to the tissue–air refractive index mismatch and $\alpha'(\lambda)$ is the transport albedo, which equals:

$$\alpha'(\lambda) = \frac{\mu'_s(\lambda)}{\mu'_s(\lambda) + \mu_a(\lambda)}. \quad (2)$$

Here, μ'_s is the reduced scattering coefficient and μ_a is the absorption coefficient, which both depend on the wavelength λ . The reduced scattering coefficient $\mu'_s(\lambda)$ can be described as $\mu'_s(\lambda) = a \cdot \left(\frac{\lambda}{\lambda_0}\right)^{-b}$, where a is the scatter amplitude, b is the scatter slope, and λ_0 is the reference wavelength to normalize the reduced scattering coefficient and make it dimensionless. In this paper, the reference wavelength was set to 500 nm.

For light incident perpendicular to a boundary where the refractive index changes from n_1 to n_2 , the specular reflection R_s is given by the Fresnel equation:

$$R_s(\lambda) = \frac{|n_1(\lambda) - n_2(\lambda)|^2}{|n_1(\lambda) + n_2(\lambda)|^2}. \quad (3)$$

We assume that glare will be proportional to but smaller than R_s —depending on the surface roughness of the tissue, the tissue refractive index, and the illumination/detection geometry. We will simulate this effect as a fraction of R_s reaching the detector, independent of the wavelength. For the synthetic spectra, glare is then simulated as

$$R_{\text{glare}}(\lambda) = R_s(\lambda) \cdot \text{rand}(0, 1), \quad (4)$$

where R_s is the maximum reflection given by Eq. (3), multiplied by a random number between 0 and 1 from a uniform distribution.

In most clinical applications, the imaged tissue specimen is not a flat surface. In addition to surface roughness, as discussed above, sample thickness variations introduce distance differences among the sample, light source, and camera. When the tissue is closer to the light source, the incident intensity per surface area will increase; if the tissue is closer to the camera,

the surface area represented by a pixel on the camera will decrease. Thus, sample thickness variations have a multiplicative effect on the measured spectra. We model the light source as an isotropic point source, in which case the intensity decreases with the square of the distance from the point source (d_{source}). Glare can be seen as a form of specular (mirror-like) reflection and does not interact with the tissue, therefore the light can be modeled as having traveled the combined distance between the source and the tissue (d_{source}) and from the tissue to the detector (d_{det}). Assuming the distance between the source and the tissue is equal to the distance between the tissue and the detector (d), the effect of distance on glare is then equal to the inverse square law over the combined total distance:

$$\beta_{\text{glare}} = \frac{1}{(d_{\text{source}} + d_{\text{det}})^2} = \frac{1}{4d^2}. \quad (5)$$

Diffuse light interacts with the tissue and can therefore not be modeled in the same manner. After the light has entered the tissue, under the diffusion approximation the light will exit the tissue with isotropic radiance.⁴⁵ For isotropic radiance, the diffusely reflected light can be modeled using the inverse square law:

$$\beta_{\text{diffuse}} = \frac{1}{d_{\text{source}}^2 \cdot d_{\text{det}}^2} = \frac{1}{d^4}. \quad (6)$$

Thus, glare and distance scale with different factors. The total measured reflectance $R_{\text{tot}}(\lambda)$ can be simulated as the sum of the diffusely reflected light ($R_{\text{diffuse}}(\lambda)$) multiplied by the scaling factor for the distance (β_{diffuse}), glare ($R_{\text{glare}}(\lambda)$) multiplied by the scaling factor for the distance (β_{glare}), and noise (γ):

$$R_{\text{tot}}(\lambda) = [R_{\text{diffuse}}(\lambda) \cdot \beta_{\text{diffuse}}] + [R_{\text{glare}}(\lambda) \cdot \beta_{\text{glare}}] + \gamma(\lambda). \quad (7)$$

2.2 Synthetic Spectra Generation

First, R_{diffuse} was created from 400 to 1600 nm with a step size of 1 nm, using diffusion theory [Eq. (1)] and the absorption coefficients of water,⁴⁶ fat,⁴⁷ hemoglobin,⁴⁸ and bilirubin⁴⁶ in combination with the optical tissue parameters from Table 1. Sets of spectra were made in which the contrast between tissue types was either a varying BVE, the presence of different absorbers, a varying scatter amplitude (a), or a varying scatter slope (b). Table 1 specifies the properties used for each set of simulations. The values were chosen to simulate general soft tissue based on the review by Jacques.⁴⁹

Next, for each R_{diffuse} , 100 spectra with variable amounts of glare, sample heights, and noise were generated [Eq. (7)]. The specular reflection was calculated using the Fresnel equations for the reflection of light at an interface between two media with different optical properties—in this case water and air [Eq. (3)]. Glare was added as a variable fraction (between 0 and 1) of the specular reflection, using the “rand” function in MathLab (MathWorks, Natick, Massachusetts) picking values from a uniform distribution. The added glare is then defined in Eq. (4) as R_{glare} .

The simulated height differences were based on the height differences observed in the study of Kho et al.¹¹ Within a single specimen, height differences of ~ 1 cm were observed, and between specimens, height differences are between 1 and 5 cm. The detector in the benchtop system was placed at 30 cm from the measurement plate, giving a maximum distance variation of 20% between all the measurements. Thus, the maximum value for d in Eqs. (5) and (6) is equal to 1.2.

We simulated realistic white noise and added this to all the spectra to make our results more comparable to real-life measurements. From the clinical datasets of the breast¹¹ and colon¹⁰ obtained previously within our group, a noise estimation was made. The noise estimation was done by taking the difference between a Savitzky–Golay smoothed reflectance spectrum and the unprocessed spectrum.⁵⁰ This was done for the breast and colon datasets and then the average of those two was taken, which was approximated as

Table 1 Tissue parameters used to create the synthetic data, based on the review by Jacques.⁴⁹ B, blood; F, fat; W, water; Bi, bilirubin; VF, volume fraction; a , scatter amplitude; b , scatter slope. For the oxygen saturation of blood, a value of 75% was used.

Contrast	Absorbers	Optical properties					
		BVF (%)	FVF (%)	WVF (%)	Bi (mg/dL)	a (cm ⁻¹)	b
BVF	B + F + W	0.5; 2.0; 3.5; 5.0; 6.5	35	35	0	15.35	1.25
Presence of different absorbers	B	4.0	0	0	0	15.35	1.25
	B + W;		0	70	0		
	B + W + F;		35	35	0		
	B + W + Bi;		0	70	13.5		
	B + W + F + Bi		35	35	13.5		
Scatter amplitude	B + F + W	4.0	35	35	0	8; 12; 16; 20; 24	1.25
Scatter slope	B + F + W	4.0	35	35	0	15.35	0.500; 0.875; 1.250; 1.625; 2.00

$$\text{Intensity}_\gamma(\lambda) = 1.0497 \cdot 10^{-15} \cdot |\lambda - \bar{\lambda}|^5 + 9.5469 \cdot 10^{-4}, \quad (8)$$

$\bar{\lambda}$ is the mean wavelength, such that for a spectra from 400 to 1600 nm, $\bar{\lambda} = 1000$ nm. The intensity of the noise is highest at the extremes of the spectra and reduces in the middle, as was seen in the clinical datasets. For the simulated spectra, we multiplied $\text{Intensity}_\gamma(\lambda)$ by a random number from a uniform distribution between -0.5 and 0.5 for each wavelength:

$$\gamma(\lambda) = \text{Intensity}_\gamma(\lambda) \cdot \text{rand}(-0.5, 0.5). \quad (9)$$

2.3 Preprocessing Algorithms

A literature search was done on hyperspectral imaging for medical applications, which resulted in almost 200 papers. In most of those papers, no preprocessing was mentioned. It was not always clear whether a preprocessing algorithm was used and not mentioned, or no preprocessing algorithm was applied. From the remaining papers, we identified 11 preprocessing algorithms, of which we excluded two since they are not widely applicable or require additional measurements. The algorithms that were excluded were: resonant Mie scattering extended multiplicative signal correction, due to its complexity and iterative nature being only selectively implementable;^{16,51} and retrieved background correction, due to the additional measurements that need to be made to correct for the spatial deviation of the light source.⁵²

The eight remaining algorithms that we discuss in this paper are: SNV, MSC, MM, MC, AUC normalization, SW normalization, FD, and SD.

2.3.1 Standard Normal Variate

SNV is commonly used in chemical analysis to remove the effects of scattering from a measured spectrum.⁵³ In medical hyperspectral imaging, this method was used by Baltussen et al.,¹⁰ Kho et al.,¹¹ Li et al.,¹² Collins et al.,¹³ Maktabi et al.,¹⁴ Malegori et al.,¹⁵ Peñaranda et al.,¹⁶ and Pardo et al.⁵⁴ In the most used version of SNV, the mean of each individual reflectance spectrum \bar{R}_{tot} is subtracted from that same individual spectrum $R_{\text{tot}}(\lambda)$, and the resulting values are divided by the standard deviation of the spectrum σ of that same individual reflectance spectrum:

$$R_{\text{SNV}}(\lambda) = \frac{R_{\text{tot}}(\lambda) - \bar{R}_{\text{tot}}}{\sigma}. \quad (10)$$

2.3.2 Multiplicative Scatter Correction

MSC is often used in the chemical field and food-sciences. As the name implies, this algorithm aims to minimize the effects of scattering on the obtained spectra of samples.⁵⁵ In medical hyperspectral imaging, MSC was used by Amigo et al.,¹⁸ Jian et al.,¹⁹ and Alrezj et al.²⁰ The goal of MSC is to correct for offsets and scaling in individual spectra, so they are as similar as possible to a “reference spectrum.” In general, the “reference spectrum” is taken as the mean spectrum of the entire dataset:

$$R_{\text{ref}}(\lambda) = \frac{\sum_{i=1}^m R_{\text{tot},i}(\lambda)}{m}, \quad (11)$$

where m is the number of spectra within the dataset. For any single spectrum within the dataset, $R_{\text{tot},i}$, it is assumed that it can be modeled as

$$R_{\text{tot},i}(\lambda) = c + d \cdot R_{\text{ref}}(\lambda). \quad (12)$$

The values of c and d are obtained for each individual reflectance spectrum by fitting Eq. (12) to each individual measured spectrum $R_{\text{tot},i}(\lambda)$ using mean square error minimization.⁵⁶ The final MSC corrected spectrum is then calculated as

$$R_{\text{MSC},i}(\lambda) = \frac{R_{\text{tot},i}(\lambda) - c}{d}. \quad (13)$$

Note that MSC and SNV both correct for scaling and an offset. However, in the implemented MSC approach, the reference spectrum R_{ref} will change when spectra are added or removed from the database, which will in turn change and affect the shape of all spectra $R_{\text{tot},i}(\lambda)$. SNV processing is done on each spectrum separately and thus is not influenced by other spectra.

2.3.3 Min–max normalization

MM, as used by Halicek et al.,^{21,57} Koprowski et al.,²² Wu et al.,²⁴ Fabelo et al.,²⁵ Martinez et al.,²⁶ Leon et al.,²⁷ Aboughaleb et al.,²⁸ Peñaranda et al.,¹⁶ and Luthman et al.,²⁹ is an algorithm that uses the minimum and maximum values of each individual measured reflectance spectrum $R_{\text{tot}}(\lambda)$ to scale and offset that same individual spectrum:

$$R_{\text{MM}}(\lambda) = \frac{R_{\text{tot}}(\lambda) - \min(R_{\text{tot}}(\lambda))}{\max(R_{\text{tot}}(\lambda)) - \min(R_{\text{tot}}(\lambda))}. \quad (14)$$

2.3.4 Mean centering

MC, as used by Lasch and Noda,³⁰ and Morais et al.,³¹ is aimed at reducing the effect of unwanted offsets to the signal. MC is defined as

$$R_{\text{MC}}(\lambda) = R_{\text{tot}}(\lambda) - \bar{R}_{\text{tot}}, \quad (15)$$

where the mean reflectance value of each individual spectrum \bar{R}_{tot} is subtracted from that same individual spectrum $R_{\text{tot}}(\lambda)$.

2.3.5 Area under the curve

AUC normalization, as used by Lu et al.,^{32,33} Kumashiro et al.,³⁴ Ma et al.,³⁵ Waterhouse et al.,³⁶ and Leitner et al.,³⁷ divides each measured spectrum $R_{\text{tot}}(\lambda)$ by the AUC of that same individual spectrum:

$$R_{\text{AUC}}(\lambda) = \frac{R_{\text{tot}}(\lambda)}{\sum_{\lambda=\lambda_{\text{begin}}}^{\lambda_{\text{end}}} R_{\text{tot}}(\lambda)}. \quad (16)$$

Note that since reflectance values in hyperspectral imaging are always positive, this algorithm is equal to l1-normalization as used by Wirkert et al.⁵⁸

2.3.6 Single wavelength

SW scaling, as used by Lu et al.³⁸ and Baltussen et al.,¹ is an algorithm in which each spectrum is divided by its value at a reference wavelength λ_0 :

$$R_{\text{SW}}(\lambda) = \frac{R_{\text{tot}}(\lambda)}{R_{\text{tot}}(\lambda_0)}. \quad (17)$$

A reference wavelength is chosen in a wavelength range where absorption is minimal. This low absorption wavelength can differ between tissue types, but in tissue applications a wavelength between 700 and 800 nm is often chosen, because in that wavelength range the absorption by blood, fat, and water is low. In our analysis, we use 730 nm as λ_0 , which was found to produce the best results.

2.3.7 First derivative

In FD processing, as used by Amigo et al.,¹⁸ Hu et al.,⁴⁰ Liu et al.,³⁹ Peñaranda et al.,¹⁶ and Mellors et al.,⁴¹ the FD of the spectrum, instead of the spectrum itself, is used for tissue classification:

$$R_{\text{FD}}(\lambda) = \frac{dR_{\text{tot}}(\lambda)}{d\lambda}. \quad (18)$$

FD processing results in a characterization of the localized slope of the spectrum. FD processing is very sensitive to noise. Therefore, in general, filtering or smoothing of the spectrum is performed before FD processing, such as Savitzky–Golay filtering.^{16,18,39–41} In our analysis, we used Savitzky–Golay filtering and optimized the window size so it would result in the lowest overlap coefficient, which was a window size of 199 nm.

2.3.8 Second derivative

SD processing, as used by Wang et al.⁴² and Zheng et al.,⁴³ uses the SD of the spectrum for tissue classification:

$$R_{\text{SD}}(\lambda) = \frac{d^2 R_{\text{tot}}(\lambda)}{d\lambda^2}. \quad (19)$$

SD processing results in a characterization of the rate with which the slope of the spectrum changes. Similar to FD processing, filtering or smoothing of a spectrum is performed before SD processing. In our analysis, we used again the Savitzky–Golay filtering and optimized the window size so it would result in the lowest overlap coefficient, which was a window size of 199 nm.

2.4 Data Analysis

2.4.1 Overlap coefficient

To quantify the effectiveness of each preprocessing algorithm, we use the Szymkiewicz–Simpson coefficient or overlap coefficient, O , which is defined as the number of reflectance values in the intersection of two sets of spectra, divided by the number of reflectance values in the smallest set:⁵⁹

$$O(\lambda) = \frac{|R_1(\lambda) \cap R_2(\lambda)|}{\min(|R_1(\lambda)|, |R_2(\lambda)|)}. \quad (20)$$

The overlap coefficient is a similarity measure for two sets of spectra, giving the overlap between the sets as a value between 0 and 1 for each wavelength.

Figure 3 shows how the overlap coefficient is calculated. Two types of tissues are simulated [Fig. 3(a)], where one type has a BVF of 0.5% (red lines) and the other has a BVF of 2% (blue lines). For each tissue type, 100 spectra were created with variable amounts of glare and height differences. In Fig. 3(b), the effect of preprocessing is shown on the spectra. In the unprocessed spectra [Fig. 3(a)], large nontissue-specific variations can be seen. In the processed spectra [Fig. 1(b)], these nontissue-specific variations are reduced and the spectra can be better distinguished from each other. To quantify how well the spectra can be distinguished from each other, the overlap coefficient is used. First, the reflectance values for a single wavelength are taken from two sets of spectra and a histogram is created with a total of 25 bins for both spectra combined [Figs. 3(c) and 3(d)]. The number of spectra in the intersection (purple bars) are then divided by the number of spectra in the smallest set (which is equal to 100 for the synthetic spectra). The overlap coefficient is calculated for each wavelength of the spectra for both the unprocessed [Fig. 3(e)] and the processed spectra [Fig. 1(f)]. The mean is then taken over all wavelengths, resulting in the mean overlap coefficient (dashed red line). As shown in Fig. 3, the mean overlap coefficient is a similarity measure between the types of tissue, where less similarity means that the two types of tissue are more distinguishable from each other.

To assess general trends, the improvement in the overlap coefficient for each preprocessing algorithm $O_{\text{improvement}}$, is calculated as

$$O_{\text{improvement}} = \frac{O_{\text{unprocessed}} - O_{\text{processed}}}{O_{\text{unprocessed}}} \cdot 100\%, \quad (21)$$

where $O_{\text{unprocessed}}$ is the average overlap coefficient over all wavelengths for the two sets before any preprocessing algorithm is applied and $O_{\text{processed}}$ is the average overlap coefficient over all wavelengths between the two sets after a specific preprocessing algorithm is applied.

2.5 Clinical Data

To test whether our results on synthetic data can be translated to clinical applications, we also investigated the overlap coefficient for different preprocessing algorithms on two clinical datasets from colorectal cancer patients¹⁰ and from breast cancer patients.¹¹ Both datasets were collected previously using a bench-top system and contain influence of glare and height differences. A detailed description of the materials and methods can be found in these papers, but a short description will be provided here. Both studies were performed on *ex-vivo* tissue samples with bench-top hyperspectral systems. Each pixel within a hyperspectral image was correlated to histopathology to determine the tissue type within the pixel. Data were obtained from Baltussen et al.,¹⁰ who imaged tissue slices from resected colorectal tissue from 32 patients with two hyperspectral line scanning camera (Spectral Imaging Ltd., Finland) in the visual (PFD-CL-65-V10E) and the near-infrared (VLNIR CL-350-N17E) range (400 to 1600 nm with an average resolution of 4 nm). 2170 spectra were present in the combined dataset, of which 857 from fat, 563 from muscle, and 750 from tumor tissue. Using the same setup, Kho et al.¹¹ imaged lumpectomy specimens and slices from resected breast tissue. Only the lumpectomy data were considered for this paper due to the reduced effect of height differences in the slices dataset. From the lumpectomy dataset, the spectra of eight patients were included, which gave a total of 1072 spectra containing 453 spectra of healthy breast tissue and 619 tissue of tumorous breast tissue, which included both ductal carcinoma *in situ* and invasive carcinoma. Please note that due to the use of two hyperspectral cameras in both setups, the preprocessing was done for each camera separately, meaning that the data from the visual spectra were processed separately from the infrared spectra.

A difference between the clinical and synthetic datasets lies in the variations that are included in them. First of all, in our simulations, we compared a dataset where all the spectra had a BVF of

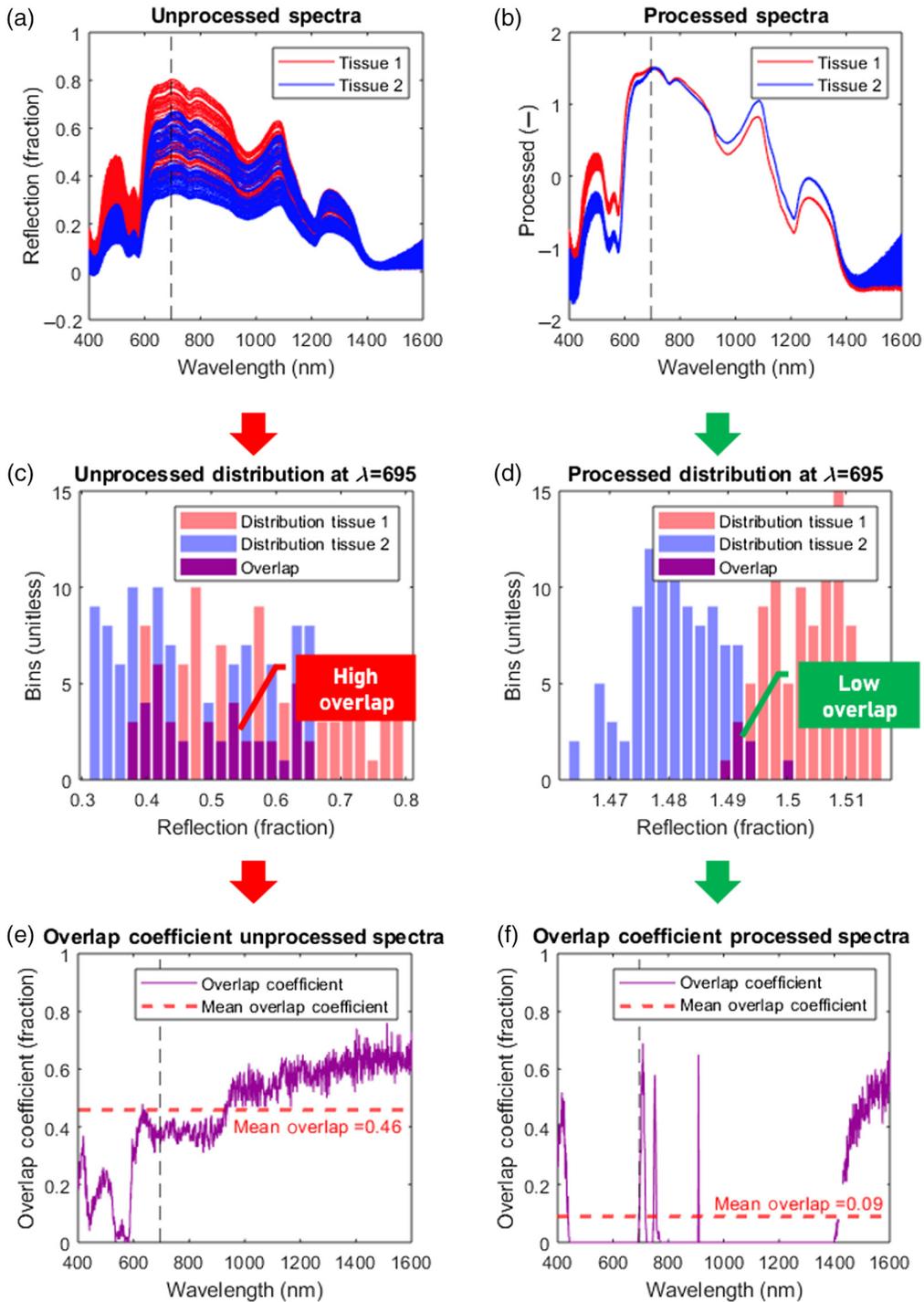


Fig. 3 Calculation of the overlap coefficient for (a) unprocessed synthetic spectra and (b) pre-processed synthetic spectra. In the unprocessed spectra, large nontissue-specific variations in the spectra are visible, which make it difficult to distinguish tissue 1 (red lines) from tissue 2 (blue lines). In the processed spectra, the nontissue-specific variations are reduced significantly. (c), (d) To calculate the overlap coefficient per wavelength histograms of the reflectance values at a single wavelength are created. The overlap coefficient is equal to the number of reflectance values in the intersection of the two sets (purple bars), divided by the number of reflectance values in the smallest set [Eq. (20)]. (e), (f) The mean overlap coefficient over all wavelengths is calculated (dashed red line). This shows that the unprocessed spectra have a high overlap, whereas the processed spectra have a low overlap.

Table 2 Tissue parameters used to create the synthetic data, based on the DRS data from de Boer et al.^{60,61} The numbers in the table indicate the ranges are given that were used in our simulated spectra. In total, 2000 spectra were simulated, of which 1000 for healthy and 1000 for tumor. The spectra were made by choosing optical properties within the range specified in the table for each spectra and tissue type.

	BVF (%)	Saturation (%)	WVF (%)	FVF (%)	Scatter amplitude	Scatter slope
Simulated healthy	0 to 6	30 to 80	2 to 4	30 to 68	10 to 15	0.8 to 1
Simulated tumor	2 to 10	5 to 60	35 to 37	33 to 35	12 to 35	0.9 to 1.3

2% to a dataset where all the spectra had a BVF of 5%. However, in reality, the difference in healthy and tumor tissues will not be such a clear dichotomy—both tissue types will have distributions of BVFs that most likely overlap. Second, while the synthetic dataset considers only one changing optical property, a clinical dataset will include variation of multiple optical properties at the same time. To illustrate both effects, we performed simulations with optical properties from a study with diffuse reflectance spectroscopy in the breast, which has shown that there are indeed large variations in optical properties between tissue types^{60,61} (note that DRS is a contact measurement and does not have any effect from glare or distance). Based on the optical properties they measured, we performed simulations to illustrate how these variations influence the effect of preprocessing algorithms.

The main contrast for healthy and tumor tissue in breast comes from the fat-water ratio, but as can be seen in Table 2, many parameters have a large variation within each tissue type. We made spectra by combining random values for each optical property that lie within the ranges given in Table 2. In total 2000 spectra were simulated, of which 1000 for healthy and 1000 for tumor. The spectra are created by choosing optical properties within the range specified in the table for each spectra and tissue type. The simulated spectra using the optical properties as stated above gives us two distributions of spectra belonging to simulated healthy tissue and simulated breast tumors. For the overlap coefficient a bin size of 50 was chosen.

3 Results

3.1 Synthetic Spectra

Figures 4–7 visualize the results of our analysis for the synthetic datasets, where we varied the BVF (Fig. 4), the presence of different absorbers (Fig. 5), the scatter amplitude (Fig. 6), and the scatter slope (Fig. 7). Each block within a figure represents the results for the unprocessed spectra or spectra processed with the algorithm. Each square represents the mean overlap coefficient between two sets of spectra with different tissue properties, where the properties of the sets are specified on the horizontal and vertical axes. A lighter square indicates a lower overlap coefficient and thus a better performance of the preprocessing algorithm. Overall, as expected, in the unprocessed spectra (upper left corner) the overlap between sets of spectra decreases when the differences in optical properties increase.

The results of the analysis on the synthetic data with different BVFs are visualized in Fig. 4. As shown in the top left panel, the mean overlap coefficients between the different BVF spectra of the unprocessed spectra are high (on average). Clearly, SNV and MSC preprocessing result in the lowest overlap coefficients (on average 15%), whereas FD and SD show the highest overlap coefficients (average 53% and 49%).

The results of the analysis on the synthetic data with adding absorbers water (W), fat (F), and bilirubin (Bi) to blood (B) are visualized in Fig. 5. Here, the type of absorbers is shown on the horizontal and vertical axes. The least overlap is seen after AUC, SNV, and MSC, which reduce the overlap between the spectra with different absorbers, resulting in average overlap coefficients of 13%, 13%, and 14%, respectively.

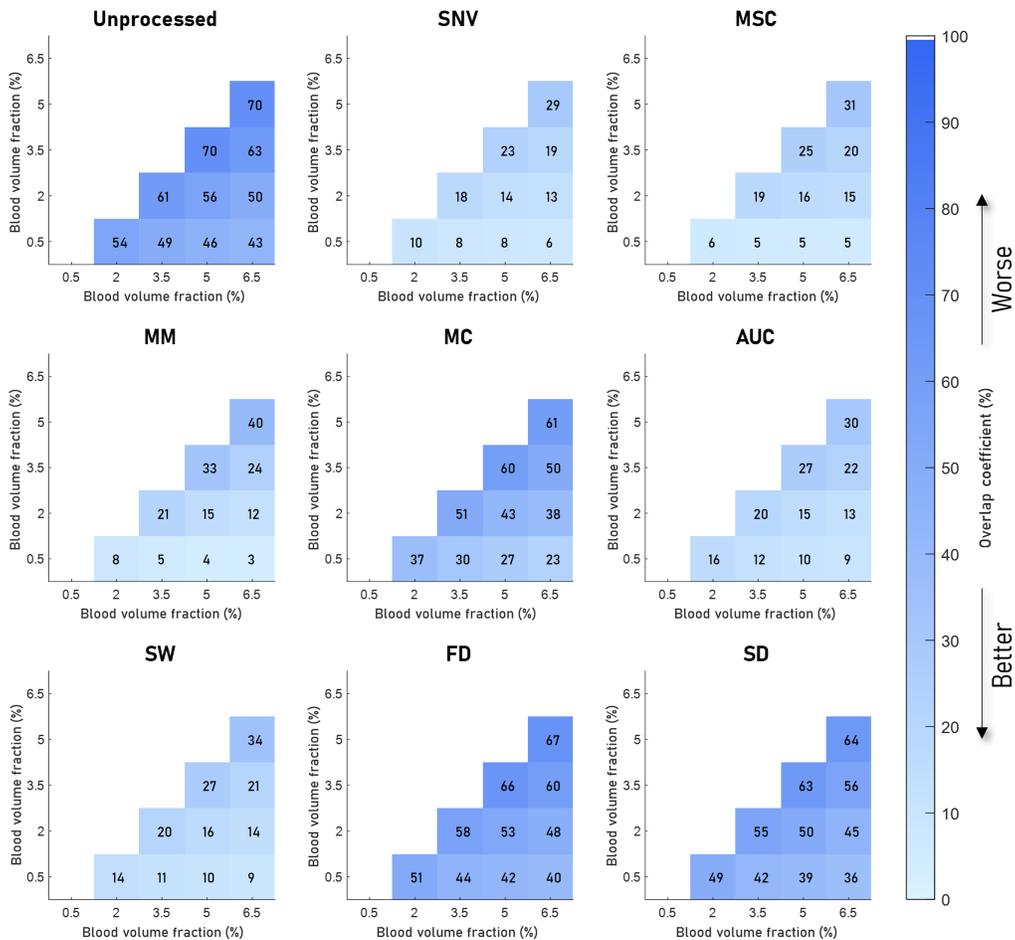


Fig. 4 Overlap between sets of spectra with different BVFs for eight preprocessing algorithms. The color and the number indicate the overlap coefficient, where a higher overlap coefficient indicates that more of the spectra overlap and would be worse for classifying. The mean overlap coefficient is shown for five different BVFs, where the unprocessed result is shown in the top left image, and the various preprocessing techniques in the rest of the image. Here, it can be seen that SNV and MSC result in a low overlap. AUC, SW, and MM have similar overlap but higher than SNV and MSC. MC, FD, and SD have significantly higher overlap after processing, with FD and SD having the highest.

The results of the analysis on the synthetic data with different scatter amplitudes are visualized in Fig. 6. For the scatter amplitude, AUC and SW reduce the overlap the most, resulting in an average overlap coefficient of 23% and 21%, respectively; MC, FD, and SD again reduce the overlap the least, with an overlap coefficient of 58%, 60%, and 58%, respectively, which is worse than no preprocessing

The results of the analysis on the synthetic data with different scatter slopes are visualized in Fig. 7. Here, AUC, SNV, and MSC reduce the overlap the most, resulting in an average overlap coefficient of 17%, 17%, and 18%, respectively. The algorithms that reduced the overlap the least were MC, FD, and SD, with average overlap coefficients of 52%, 53%, and 60%, respectively.

3.2 Clinical Data

The results are presented in the same manner as the synthetic data, but the horizontal and vertical axes indicate the tissue type to which the spectra belong instead of the tissue parameters shown in the synthetic data.

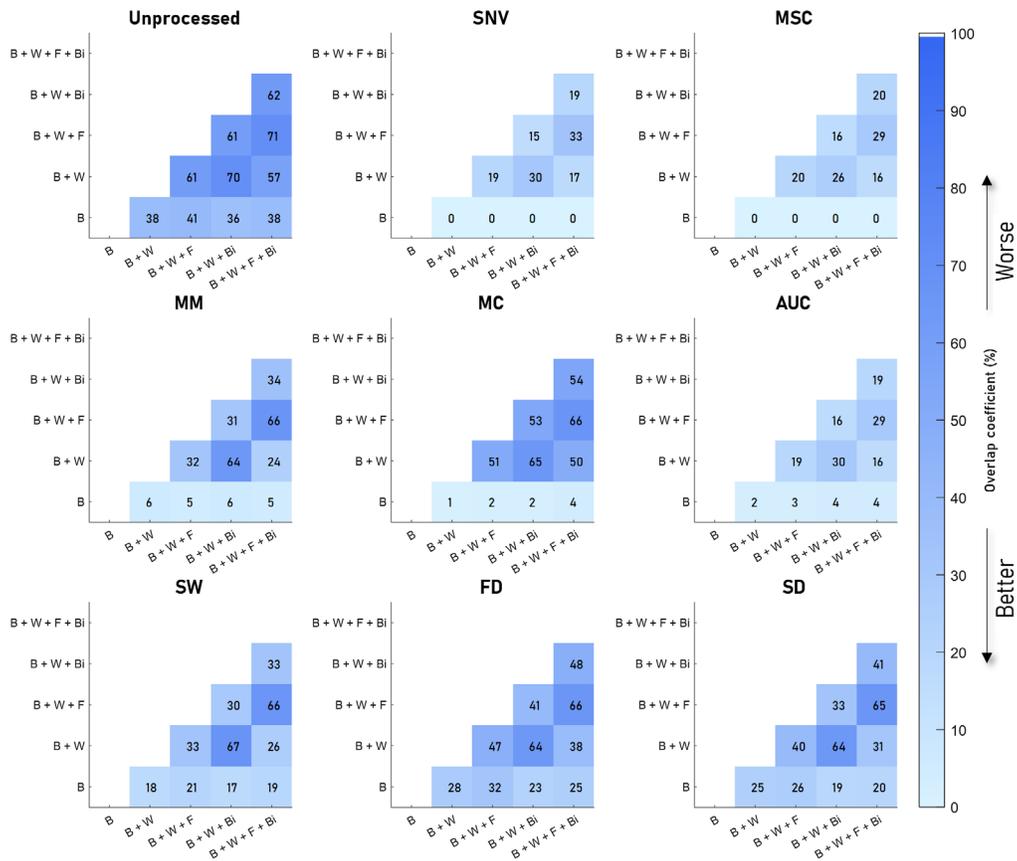


Fig. 5 Overlap between sets of spectra with different absorbers for eight preprocessing algorithms. The axes are labeled with the type of absorbers included, where B, blood; W, water; F, fat; and Bi, bilirubin. SNV, MSC, and AUC reduce the overlap the most, followed by MM, which reduces the overlap less than SNV, MSC, or AUC but still reduces the overlap significantly. MC reduces the overlap less. Finally, FD, SD, and SW reduce the overlap the least with some high values for a few combinations remaining.

3.2.1 Colon

The colon data were divided into three categories, muscle, fat, and tumor. In all cases, the overlap between spectra from tumor and muscle tissue was much higher than the overlap between tumor and muscle tissue compared with fat (Fig. 8). Most preprocessing algorithms reduce the overlap coefficient compared with the unprocessed spectra. However, AUC and MC increase the overlap coefficient for fat versus muscle. Overall, SNV and MSC resulted in the lowest overlap coefficients, whereas MC and AUC resulted in the highest overlap coefficients. SW produces the lowest overlap between tumor and muscle.

3.2.2 Breast

For the breast data, the tissue was divided into two categories: healthy and tumor tissues. Therefore, we only obtained a single value for the overlap for each algorithm (Fig. 9). All algorithms reduced the overlap coefficient compared with unprocessed data. SNV and MSC resulted in the lowest overlap coefficients.

3.2.3 Simulated mixed tissue

The results of the mixed tissue type that includes the variations discussed are shown in Fig. 10. In the simulated data, SNV, MSC, and SW reduce the overlap the most, whereas MC, FD, and SD perform the worst.

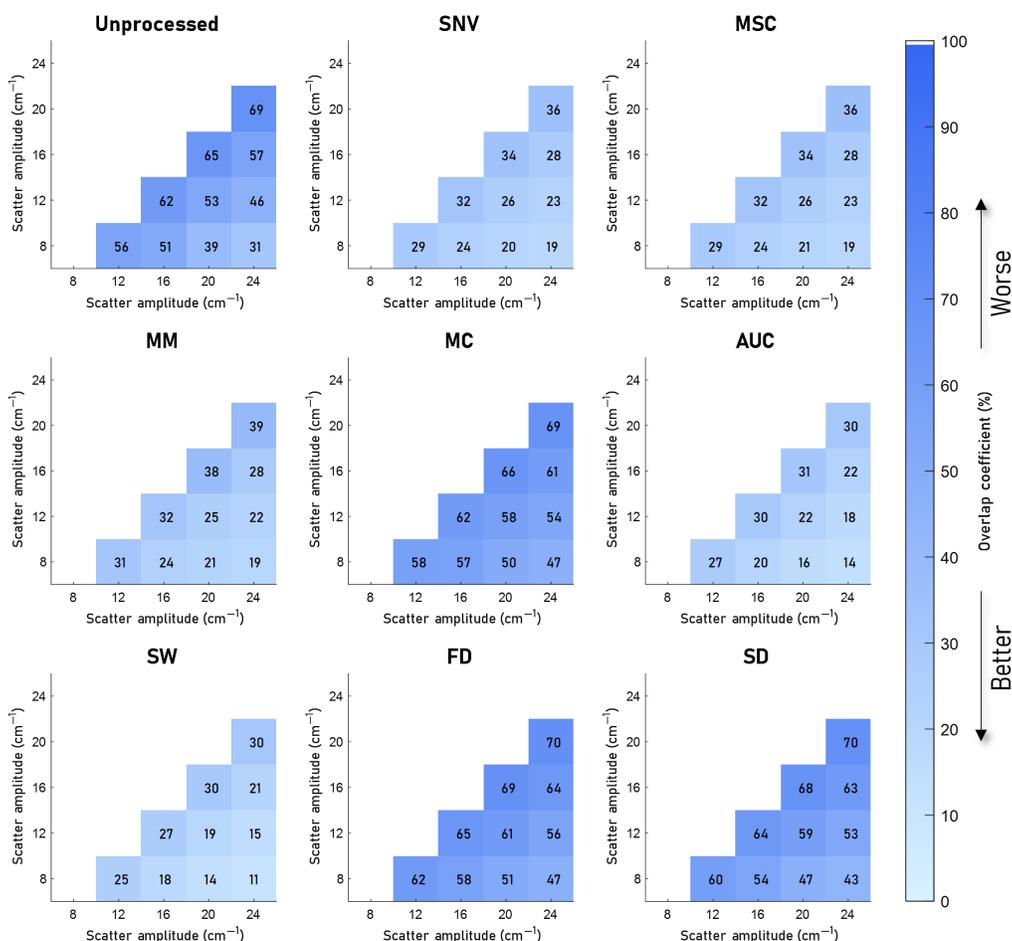


Fig. 6 Overlap between sets of spectra with different scatter amplitudes (a) for eight preprocessing algorithms. The mean overlap coefficient is shown for five different values of a . The unprocessed result is shown in the top left image, here a high overlap due to the nontissue-specific variations can be seen between the different values of a . AUC and SW reduce the overlap significantly. MM, SNV, and MSC remove the overlap less than AUC and SW. MC, FD, and SD reduce the overlap the least.

To facilitate easier comparison of the different algorithms for the tissue types and the clinical data, the mean improvement in overlap coefficient relative to the unprocessed data is shown in Fig. 11. The first four columns show the improvement for the synthetic spectra and the last two columns show the results for the clinical data. For the colon, the improvement in the overlap coefficient shown is the average improvement for all three classes of tissue (fat, muscle, and tumor). For the synthetic data, SNV and AUC have the highest overall improvement in the overlap coefficient. Algorithms that correct for scaling and an offset (blue bars; SNV, MSC, and MM) have the highest improvement in the overlap coefficient for the BVF, different absorbers, and scatter slope. MM generally performs less well compared to MSC and SNV. Algorithms that correct for scaling alone have the highest improvement for the scatter amplitude (red bars; AUC and SW). The two algorithms that use the derivative of the spectrum (green bars; FD and SD) and the algorithm that only subtracts an offset (MC) have a much smaller decrease in the overlap coefficient compared with all the other algorithms.

For the clinical data, the improvement in the overlap coefficient is less pronounced. For both the breast and colon tissues, similar trends can be observed as in the synthetic data, where the algorithms that correct for scaling and an offset (blue bars) and the algorithms that correct for scaling alone (red bars) perform better than the algorithms that only correct for an offset (yellow bar) and use differentiation (green bars). For the colon and the breast, MSC and SNV perform the best. For the colon, MC and AUC increase the overlap between the spectra from different tissue

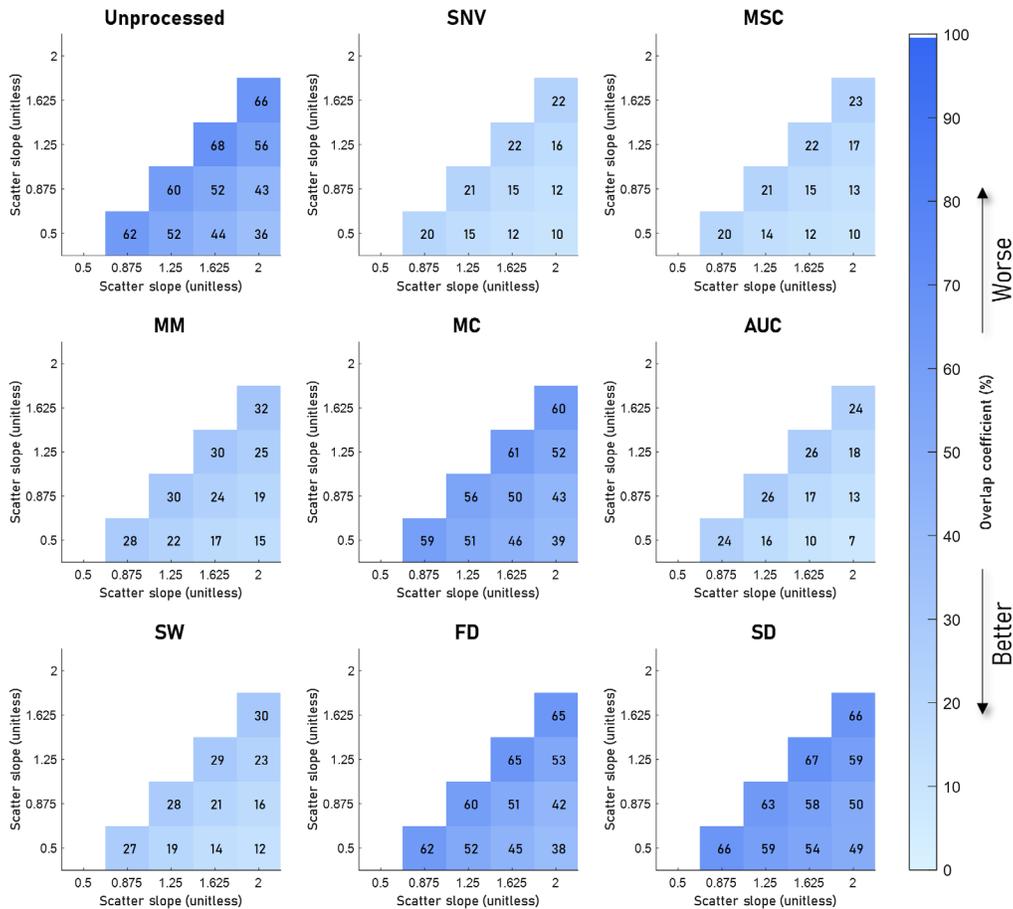


Fig. 7 Overlap between sets of spectra with different scatter slopes (b) for eight preprocessing algorithms. The mean overlap coefficient is shown for five different values of b . The unprocessed result is shown in the top left image, here a high overlap due to the nontissue-specific variations can be seen between the different values of b . Here, SNV, MSC, and AUC reduce the overlap the most. SW and MM perform comparably with SW reducing the overlap only slightly more. MC, FD, and SD reduce the overlap the least, with SD increasing the overlap.

types. For the simulated mixed breast data, the improvement in overlap is lower than for the synthetic data and closer to the clinical data. Again, similar trends hold for the simulated breast data, where both the scaling and subtraction techniques (blue bars), and the only scaling techniques (red bars) perform well in reducing the overlap.

4 Discussion

In this paper, we compared the suitability of preprocessing algorithms to reduce the nontissue-specific spectral variations caused by glare and sample thickness variations in a simulated dataset resembling diffuse reflection spectra from human tissue, and in a clinical dataset containing hyperspectral images of breast and colon tissues. Overall, our results indicate that preprocessing algorithms can reduce the unwanted spectral variations caused by variations in glare and sample thickness while retaining contrast due to differences in tissue optical properties.

Glare and sample thickness variations introduce both an offset and a multiplication. Therefore, it was expected that algorithms that subtract an offset and divide by a scaling factor would perform the best. SNV, MSC, and MM determine an offset and scaling factor for each individual spectrum, thereby reducing the influences of glare and sample thickness variations. In general, the algorithms that subtract an offset and divide by a scaling factor (SNV, MSC, and MM) perform the best, except for tissue contrast related to the scatter amplitude. Since glare and

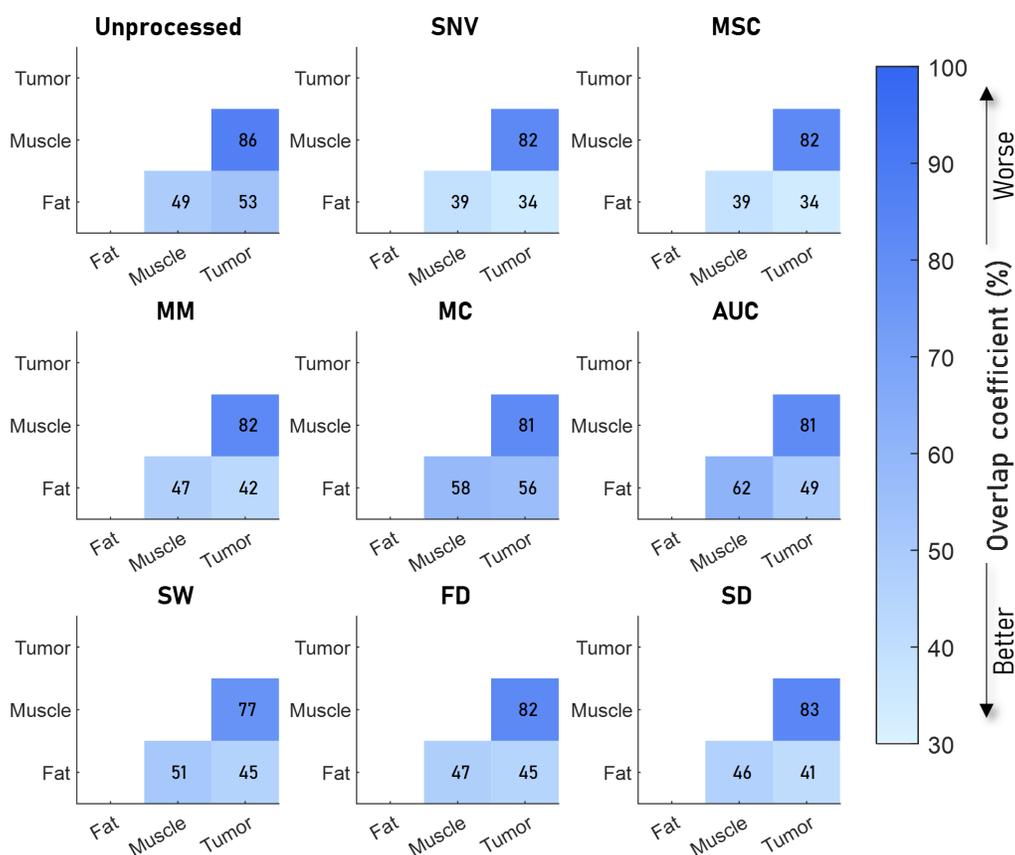


Fig. 8 Overlap between sets of spectra from different tissue types in colon samples measured by Baltussen et al.¹⁰ for eight preprocessing algorithms. The unprocessed result is shown in the top left image. SNV and MSC processing resulted in the lowest sum of overlap coefficients, whereas MC and AUC resulted in the highest overall overlap coefficients. SW produces the lowest overlap between tumor and muscle.

sample thickness variations introduce both an offset and a multiplication, it was expected that SNV, MSC, and MM reduce these variations the best. Since SNV and MSC were originally developed to reduce the effects of scattering, it was interesting that SNV and MSC do not only perform well for tissue contrast related to absorption but also related to the scatter slope. SNV and MSC produce very similar results, and MM generally performs slightly worse. Algorithms that only divide by a scaling factor but do not correct for an offset (AUC and SW) also perform well but slightly less than the above-discussed algorithms that also subtract an offset before dividing by a scaling factor. The algorithm that only removes the offset (MC) performs poorly on reducing the overlap coefficient, which can be explained by the fact that it does not reduce variations as a result of variations in the sample thickness, leaving overlap in the spectra as can be seen in the [Supplementary Material](#). The two algorithms that use the derivative of the spectrum (FD and SD) perform significantly worse for all types of tissue contrast compared with the other algorithms. We had expected that FD and SD would also perform well since the derivatives of the spectra are less influenced by offsets and scaling factors. In the [Supplementary Material](#), it can be seen that FD and SD are less able to remove the influence of glare and sample thickness variations compared with other preprocessing algorithms and also reduce the contrast between the spectra.

The general trends observed in the synthetic datasets translated well to the clinical datasets. The improvement in overlap coefficient is lower for the clinical dataset compared with the synthetic dataset, but this is to be expected, because in the clinical spectra the natural variation in tissue structure and composition in samples from the same tissue type is incorporated—not every tumor is the same. For example, in our synthetic dataset, we compare two “tissue types” where one has a BVF of 2% and the other of 5%. Even if on average, the BVF of healthy tissue is 2% and of tumor tissue 5%, most likely in reality the ranges of BVFs for both tissue types overlap.

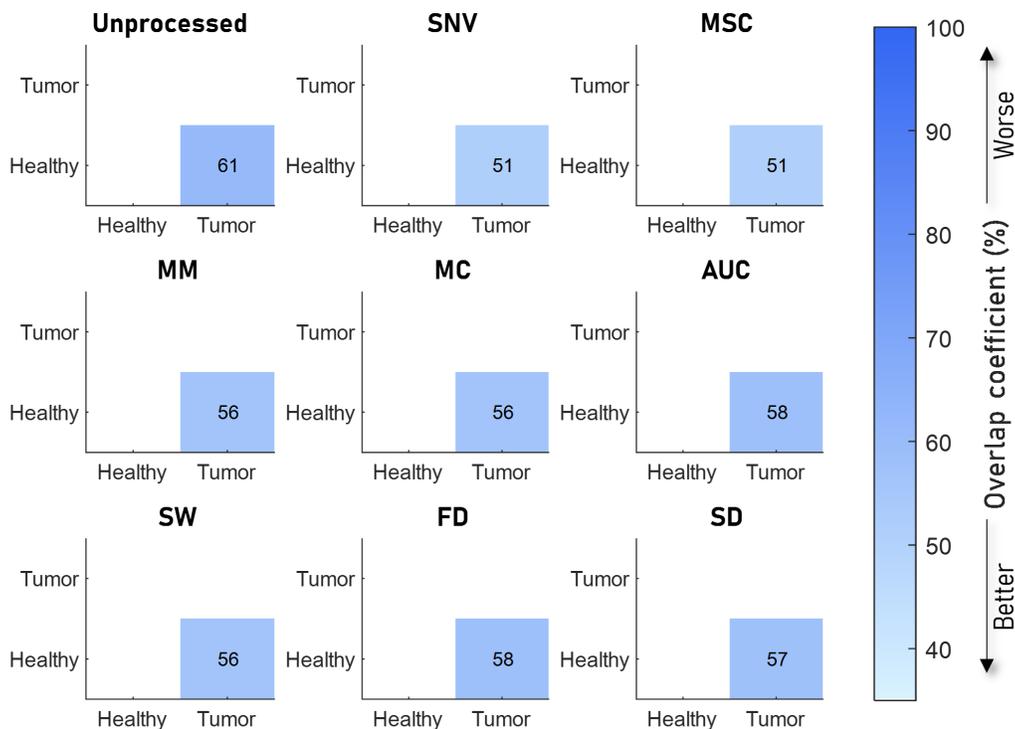


Fig. 9 Overlap between sets of spectra from different tissue types in breast samples measured by Kho et al.¹¹ for eight preprocessing algorithms. The unprocessed result is shown in the top left image. SNV and MSC processing resulted in the lowest overlap coefficients, whereas AUC and FD resulted in the highest overall overlap coefficients.

To verify that the heterogeneity between and within patients can explain the lower overlap coefficients in clinical datasets, we have simulated spectra with additional variations in optical properties within each tissue type. The results for the simulated breast spectra are comparable with the clinical breast data, most importantly, as was the goal for the simulated breast tissue data, it shows a clear reduction in the effectiveness of all preprocessing techniques due to the variations mentioned above. The general trends observed in the synthetic datasets translated well to the clinical dataset, except the performance of AUC in the colon dataset, where it resulted in worse overlap coefficients compared with the original spectra without preprocessing.

Our results can be useful for the development of classification algorithms in a dataset with variations in glare and sample thickness. For example, De Boer et al.^{60–62} found for breast cancer that the fat/water ratio could discriminate between healthy and tumor tissues with an accuracy of 100% and that the fat/water ratio in breast tumor tissue was close to zero due to the absence of fat in tumor tissue, indicating that the contrast is based on the presence of additional absorbers. For this type of contrast, MSC, SNV, and AUC perform the best (Fig. 11). For the clinical breast dataset, MSC and SNV indeed perform the best, whereas AUC performs worse.

For the colon, three tissue types are classified, which makes it more complex to analyze. When analyzing multiple classes of spectra, two approaches can be taken when encountering such a problem. First, one can identify which two classes are most important to distinguish (e.g., muscle versus tumor) and choose the algorithm based on the contrast between those two classes. A second approach would be to first calculate which tissue types have the highest overlap coefficient. The type of tissue contrast that gives the highest overlap, and thus the lowest distinguishability, should then be the leading choice for a preprocessing algorithm. For the colon data, we are most interested in muscle versus tumor, where we expect differences in BVF and scattering.^{63–66} Based on our synthetic data, SNV, MSC, and SW would be good candidates, which corresponds to our findings for the clinical dataset.

We did not directly test how the different preprocessing algorithms influence tissue classification algorithms since many different types of classification algorithms can be developed.

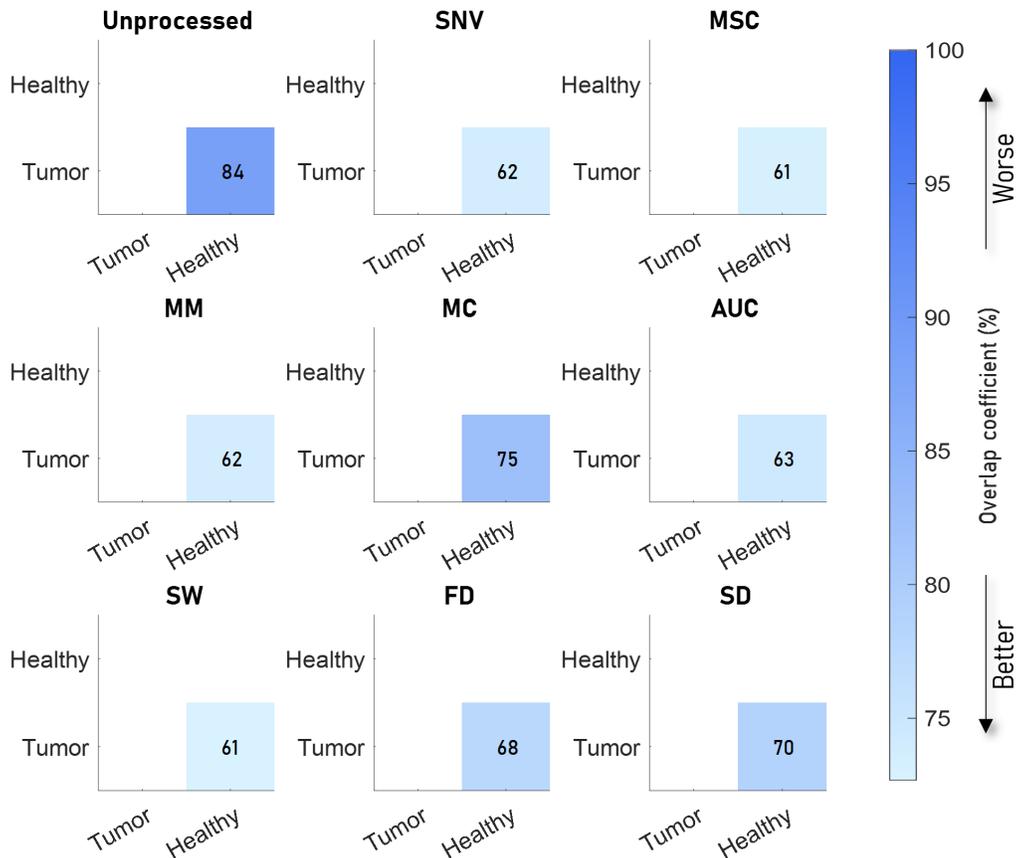


Fig. 10 Overlap between sets of spectra from mixed tissue types, which includes the variation between and within patients. The overlap is lowest for SNV, MSC and SW and highest for MC, FD and SD.

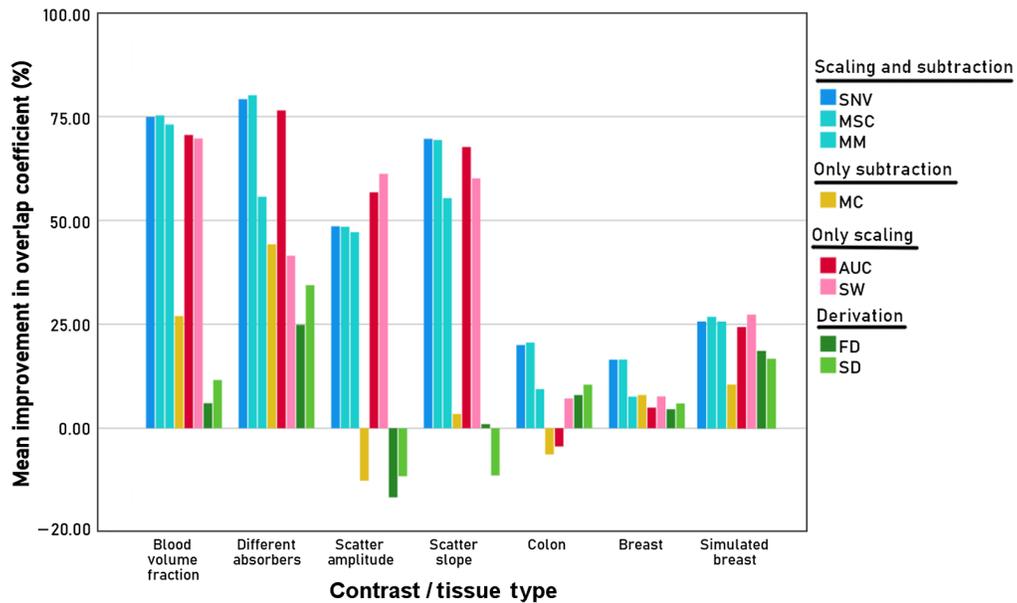


Fig. 11 Summary of the data presented in Figs. 4–10. For the four types of synthetic data with different tissue contrast (different absorbers, BVF, scatter amplitude, and scatter slope) as well as the colon and breast data; and the simulated breast data, the mean improvement in overlap coefficient relative to the unprocessed data is depicted per preprocessing algorithm. Preprocessing algorithms from the same category have a similar color. For the colon, the improvement in the overlap coefficient shown is the average of the improvement for each of the combinations of tissue classes.

Nevertheless, the quality of the data used to develop an algorithm is known to influence the quality of the developed algorithm. Therefore, preprocessing is often used for classification algorithms. For the development of a tissue classification algorithm, removing variations in the data not directly related to differences between tissue types is expected to improve the accuracy of the developed algorithm. Since glare and sample height differences are not discriminating features between tissue types, reducing unwanted spectral variations due to distance and glare while maintaining contrast between tissues with different optical properties should improve tissue classification algorithms.

Based on our results, for the development of tissue classification algorithms in a dataset with variations in glare and sample thickness, we do not recommend the use of MC, FD, and SD as preprocessing algorithms since they performed much worse compared to the other algorithms. Because SNV and MSC normalization perform equally well, it might not be useful to test both algorithms on the same dataset. Furthermore, since the MSC normalization of an individual spectrum depends on the other spectra within the dataset, individually processed spectra will change when new spectra are added to the dataset. Thus, for large datasets, the performance of four preprocessing algorithms could be investigated: SNV, MM, AUC, and SW. A power analysis for four algorithms could result in a larger dataset than practical for a study. If the contrast between tissue types is unknown, the best choice for a preprocessing algorithm would then be AUC or SNV since on average these outperform the other algorithms for the four types of contrast investigated in this paper. If the contrast between tissue types is known, we recommend the following preprocessing algorithms: SNV for contrast due to changes in BVF; AUC or SNV for contrast due to changes in the type of absorbers in the tissue; AUC or SW for contrast due to changes in the scatter amplitude; and AUC or SNV for contrast due to changes in the scatter slope.

In *ex-vivo* settings with benchtop systems, the influences of glare and distance differences can be reduced by polarization filters and surface profilometry. For *in vivo* applications, this would hamper real-time feedback. Since classification algorithms have to be developed on similar data as the data that will be acquired during its clinical application, it is essential that *ex-vivo* studies that are performed to develop classification algorithms for *in vivo* applications use the same preprocessing algorithms (and thus not polarization filters and profilometry).

5 Conclusion

This paper provides researchers with a solid basis to identify the most suitable preprocessing algorithm that decreases variation due to glare and sample thickness in spectra while maintaining as much contrast between tissue types as possible. We compared eight commonly used preprocessing algorithms and identified four algorithms that we found suitable to use before developing an algorithm for tissue classification: SNV, MM, AUC, and SW. For very large datasets, all four algorithms can be tested, but for smaller datasets, we recommend to choose one or two algorithms *a priori*, based on the expected contrast between tissue types.

Disclosures

The authors have no conflicts of interest to declare. All coauthors have seen and agree with the contents of the manuscript and there is no financial interest to report.

References

1. E. J. M. Baltussen et al., "Optimizing algorithm development for tissue classification in colorectal cancer based on diffuse reflectance spectra," *Biomed. Opt. Express* **10**(12), 6096 (2019).
2. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
3. M. Halicek et al., "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," *J. Biomed. Opt.* **24**(3), 036007 (2019).

4. E. Kho et al., "Hyperspectral imaging for resection margin assessment during cancer surgery," *Clin. Cancer Res.* **25**(12), 3572–3580 (2019).
5. S. Mascharak, B. J. Baird, and F. C. Holsinger, "Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning," *Laryngoscope* **128**(11), 2514–2520. (2018).
6. K. B. De Raad et al., "The effect of preprocessing on convolutional neural networks for medical image segmentation," in *Proc. – Int. Symp. Biomed. Imaging*, pp. 655–658 (2021).
7. N. M. Nawi, W. H. Atomi, and M. Z. Rehman, "The effect of data pre-processing on optimized training of artificial neural networks," *Proc. Technol.* **11**, 32–39 (2013).
8. C. C. Olisah and L. Smith, "Understanding unconventional preprocessors in deep convolutional neural networks for face identification," *SN Appl. Sci.* **1**(11), 1–12 (2019).
9. S. J. Lord, V. J. GebSKI, and A. C. Keech, "Multiple analyses in clinical trials: sound science or data dredging?" *Med. J. Aust.* **181**(8), 452–454 (2004).
10. E. E. J. M. Baltussen et al., "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," *J. Biomed. Opt.* **24**(1), 016002 (2019).
11. E. Kho et al., "Feasibility of *ex vivo* margin assessment with hyperspectral imaging during breast-conserving surgery: from imaging tissue slices to imaging lumpectomy specimen," *Appl. Sci.* **11**(19), 8881 (2021).
12. B. Li et al., "The age estimation of blood stains up to 30 days old using visible wavelength hyperspectral image analysis and linear discriminant analysis," *Sci. Justice* **53**(3), 270–277 (2013).
13. T. Collins et al., "Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging," *Diagnostics* **11**(10), 1810 (2021).
14. M. Maktabi et al., "Tissue classification of oncologic esophageal resectates based on hyperspectral data," *Int. J. Comput. Assist. Radiol. Surg.* **14**(10), 1651–1661 (2019).
15. C. Malegori et al., "Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics," *Talanta* **215**, 120911 (2020).
16. F. Peñaranda et al., "Discrimination of skin cancer cells using Fourier transform infrared spectroscopy," *Comput. Biol. Med.* **100**(June), 50–61 (2018).
17. A. Pardo et al., "Directional Kernel density estimation for classification of breast tissue spectra," *IEEE Trans. Med. Imaging* **36**(1), 64–73 (2017).
18. J. M. Amigo, H. Babamoradi, and S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial," *Anal. Chim. Acta* **896**, 34–51 (2015).
19. H. Jiang et al., "Integration of spectral and textural features of visible and near-infrared hyperspectral imaging for differentiating between normal and white striping broiler breast meat," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **213**, 118–126 (2019).
20. O. A. Alrezy et al., "Coupling scatter correction with bandpass filtering for preprocessing in the quantitative analysis of glucose from near infrared spectra," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1800–1803 (2017).
21. M. Halicek et al., "Cancer detection using hyperspectral imaging and evaluation of the superficial tumor margin variance with depth," *Proc. SPIE* **10951**, 109511A (2019).
22. R. KoproWSKI et al., "Automatic method for the dermatological diagnosis of selected hand skin features in hyperspectral imaging," *Biomed. Eng. Online* **13**(1), 47 (2014).
23. A. Banerjee et al., "Hypothesis testing, type I and type II errors," *Ind. Psychiatry J.* **18**(2), 127 (2009).
24. I. C. Wu et al., "Early identification of esophageal squamous neoplasm by hyperspectral endoscopic imaging," *Sci. Rep.* **8**, 13797 (2018).
25. H. Fabelo et al., "Deep learning-based framework for *in vivo* identification of glioblastoma tumor using hyperspectral images of human brain," *Sensors* **19**(4), 920 (2019).
26. B. Martinez et al., "Most relevant spectral bands identification for brain cancer detection using hyperspectral imaging," *Sensors* **19**(24), 5481 (2019).
27. R. Leon et al., "Non-invasive skin cancer diagnosis using hyperspectral imaging for *in-situ* clinical support," *J. Clin. Med.* **9**(6), 1662 (2020).
28. I. H. Aboughaleb, M. H. Aref, and Y. H. El-Sharkawy, "Hyperspectral imaging for diagnosis and detection of *ex-vivo* breast cancer," *Photodiagnosis Photodyn. Ther.* **31**(June), 101922 (2020).

29. A. S. Luthman et al., “Bimodal reflectance and fluorescence multispectral endoscopy based on spectrally resolving detector arrays,” *J. Biomed. Opt.* **24**(3), 031009 (2019).
30. P. Lasch and I. Noda, “Two-dimensional correlation spectroscopy (2D-COS) for analysis of spatially resolved vibrational spectra,” *Appl. Spectrosc.* **73**(4), 359–379 (2019).
31. C. L. M. Morais, P. L. Martin-Hirsch, and F. L. Martin, “A three-dimensional principal component analysis approach for exploratory analysis of hyperspectral data: identification of ovarian cancer samples based on Raman microspectroscopy imaging of blood plasma,” *Analyst* **144**(7), 2312–2319 (2019).
32. G. Lu et al., “Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging,” *J. Biomed. Opt.* **19**(10), 106004 (2014).
33. G. Lu et al., “Framework for hyperspectral image processing and quantification for cancer detection during animal tumor surgery,” *J. Biomed. Opt.* **20**(12), 126012 (2015).
34. R. Kumashiro et al., “An integrated endoscopic system based on optical imaging and hyperspectral data analysis for colorectal cancer detection,” *Anticancer Res.* **36**(8), 3925–3932 (2016).
35. L. Ma, M. Halicek, and B. Fei, “*In vivo* cancer detection in animal model using hyperspectral image classification with wavelet feature extraction,” *Proc. SPIE* **11317**, 113171C (2020).
36. D. J. Waterhouse et al., “Quantitative evaluation of comb-structure correction methods for multispectral fibroscopic imaging,” *Sci. Rep.* **8**(1), 1–14 (2018).
37. R. Leitner et al., “Multi-spectral video endoscopy system for the detection of cancerous tissue,” *Pattern Recognit. Lett.* **34**(1), 85–93 (2013).
38. G. Lu et al., “Quantitative wavelength analysis and image classification for intraoperative cancer diagnosis with hyperspectral imaging,” *Proc. SPIE* **9415**, 94151B (2015).
39. S. Liu et al., “Using hyperspectral imaging automatic classification of gastric cancer grading with a shallow residual network,” *Anal. Methods* **12**(30), 3844–3853 (2020).
40. B. Hu et al., “Tumor tissue classification based on micro-hyperspectral technology and deep learning,” *Biomed. Opt. Express* **10**(12), 6370 (2019).
41. B. O. L. Mellors et al., “Machine learning utilising spectral derivative data improves cellular health classification through hyperspectral infra-red spectroscopy,” *PLoS One* **15**(9), 1–21 (2020).
42. C. Wang et al., “Multi-scale hyperspectral imaging of cervical neoplasia,” *Arch. Gynecol. Obstet.* **293**(6), 1309–1317 (2016).
43. W. Zheng et al., “Hyperspectral wide gap second derivative analysis for *in vivo* detection of cervical intraepithelial neoplasia,” *J. Biomed. Opt.* **20**(12), 121303 (2015).
44. S. T. Flock et al., “Monte Carlo modeling of light propagation in highly scattering tissues—I: model predictions and comparison with diffusion theory,” *IEEE Trans. Biomed. Eng.* **36**(12), 1162–1168 (1989).
45. L. V. Wang and H. I. Wu, *Biomedical Optics*, John Wiley & Sons, Inc. (2009).
46. S. Prah, “Optical absorption spectrum of Bilirubin in chloroform,” <https://omlc.org/spectral/PhotochemCAD/data/119-abs.txt> (accessed 12 November 2019).
47. R. L. P. van Veen et al., “Determination of VIS- NIR absorption coefficients of mammalian fat, with time- and spatially resolved diffuse reflectance and transmission spectroscopy,” in *Biomedical Topical Meeting*, OSA, p. SF4 (2004).
48. S. Prah, “Optical absorption of hemoglobin,” (1999). <https://omlc.org/spectra/hemoglobin/> (accessed 12 November 2019).
49. S. L. Jacques, “Optical properties of biological tissues: a review,” *Phys. Med. Biol.* **58**(14), 5007–5008 (2013).
50. A. McAndrew, “An introduction to digital image processing with Matlab notes for SCM2511 image processing 1 semester 1,” (2004).
51. E. C. Mattson et al., “Toward optimal spatial and spectral quality in widefield infrared spectromicroscopy of IR labelled single cells,” *Analyst* **138**(19), 5610–5618 (2013).
52. J. Yoon, A. Grigoriou, and S. E. Bohndiek, “A background correction method to compensate illumination variation in hyperspectral imaging,” *PLoS One* **15**(3), 1–21 (2020).
53. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, “Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra,” *Appl. Spectrosc.* **43**(5), 772–777 (1989).

54. A. Pardo et al., "On the spectral signature of melanoma: a non-parametric classification framework for cancer detection in hyperspectral imaging of melanocytic lesions," *Biomed. Opt. Express* **9**(12), 6283 (2018).
55. I. Murray and A. Hall, "Animal feed evaluation by use of near infrared reflectance (NIR) spectrocomputer," *Anal. Proc.* **20**(2), 75–79 (1983).
56. K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark Version 20121115 (2012).
57. M. Halicek et al., "Hyperspectral imaging for head and neck cancer detection: specular glare and variance of the tumor margin in surgical specimens," *J. Med. Imaging.* **6**(03), 1 (2019).
58. S. J. Wirkert et al., "Robust near real-time estimation of physiological parameters from megapixel multispectral images with inverse Monte Carlo and random forest regression," *Int. J. Comput. Assist. Radiol. Surg.* **11**(6), 909–917 (2016).
59. M. K. Vijaymeena and K. Kavitha, "A survey on similarity measures in text mining," *Mach. Learn. Appl. Int. J.* **3**(1), 19–28 (2016).
60. L. L. de Boer et al., "Using DRS during breast conserving surgery: identifying robust optical parameters and influence of inter-patient variation," *Biomed. Opt. Express* **7**(12), 5188–5200 (2016).
61. L. L. de Boer et al., "Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries," *Breast Cancer Res. Treat.* **152**(3), 509–518 (2015).
62. L. L. De Boer et al., "Towards the use of diffuse reflectance spectroscopy for real-time *in vivo* detection of breast cancer during surgery," *J. Transl. Med.* **16**(1), 1–14 (2018).
63. A. N. Bashkatov, E. A. Genina, and V. V. Tuchin, "Optical properties of skin, subcutaneous, and muscle tissues: a review," *J. Innov. Opt. Health Sci.* **4**(1), 9–38 (2011).
64. A. N. Bashkatov et al., "Optical properties of human colon tissues in the 350–2500 nm spectral range," *Quantum Electron.* **44**(8), 779–784 (2014).
65. R. Marchesini et al., "Ex vivo optical properties of human colon tissue," *Lasers Surg. Med.* **15**(4), 351–357 (1994).
66. L. Oliveira et al., "Optical properties of colorectal muscle in visible/NIR range," *Proc. SPIE* **10685**, 106853D (2018).

Mark Witteveen graduated with honours in biomedical physics and technology from Vrije Universiteit Amsterdam in 2020. He is currently pursuing his PhD at the Netherlands Cancer Institute at the Antoni van Leeuwenhoek Oncological Hospital in Amsterdam. His current research is focused on imaging techniques for surgical navigation. Besides optical techniques and light tissue interactions, his interests are the involvement of machine learning and visualization for navigation during surgery.

Henricus J. C. M. Sterenberg (co)founded the Centre for Optical Diagnostics and Therapy at the Erasmus Medical Center in 1998 and was appointed as a professor of photodynamic therapy at the Erasmus University in 2008. Since 2013, he holds a joint position at the Department of Biomedical Engineering and Physics at the Amsterdam UMC and the Surgical Innovations Group at the Netherlands Cancer Institute, where his main focus is optical spectroscopy and hyperspectral imaging for evaluation of resection margins.

Ton G. van Leeuwen is the head of the BME and Physics Department at Amsterdam UMC. His research focuses on the physics of the interaction of light with tissue and to use that knowledge for the development, introduction, and clinical evaluation of (newly developed) optical imaging and analysis techniques. Key in this research is to use this knowledge and new devices to gather quantitative functional information of tissue or tissue sample. Techniques developed and used are optical coherence tomography (OCT), (single fiber) reflectance spectroscopy, fluorescence spectroscopy, hyperspectral imaging, Raman spectroscopy, flow cytometry, and three-dimensional imaging cryo-microtome. He is a fellow of the AIMB and SPIE.

Maurice C. G. Aalders graduated from VU University in Amsterdam in 1996 and received his PhD from the University of Amsterdam in 2001. Currently, he is a full professor at the Department of Biomedical Engineering and Physics of the Amsterdam UMC. His research is at the edge of applied physics, clinical- and forensic science; covering spectral imaging for age

determination of bloodstains, post-mortem time determination from body cooling, and spectral processing using AI and advanced light transport models.

Theo J. M. Ruers has been working as an oncologic surgeon since 1994. He did his training at the Hammersmith Hospital and King's College Hospital, both in London. Since 2007, he has been working at the NKI-AVL as a head of the surgical oncology division. His clinical expertise is focused on liver surgery and colorectal surgery. As head of the Surgical Innovations Group, his research focuses on the development and introduction of new image-guided treatment methods.

Anouk L. Post received her MSc degree in physics from the University of Amsterdam, in the Netherlands, in 2013. She obtained her PhD from the University of Amsterdam, for research into single fiber reflectance spectroscopy as a diagnostic tool for cancer. Since 2018, she has been a postdoc at the Surgical Innovations Group at the Netherlands Cancer Institute. The main focus of her research is the development of optical techniques to improve cancer detection.