Spectral areas and ratios classifier algorithm for pancreatic tissue classification using optical spectroscopy

Malavika Chandra,a James Scheiman,b,c Diane Simeone,c,d Barbara McKenna,c Julianne Purdy,e and Mary-Ann Mycek,c,f

Abstract. Pancreatic adenocarcinoma is one of the leading causes of cancer death, in part because of the inability of current diagnostic methods to reliably detect early-stage disease. We present the first assessment of the diagnostic accuracy of algorithms developed for pancreatic tissue classification using data from fiber optic probe-based bimodal optical spectroscopy, a real-time approach that would be compatible with minimally invasive diagnostic procedures for early cancer detection in the pancreas. A total of 96 fluorescence and 96 reflectance spectra are considered from 50 freshly excised tissue sites—including human pancreatic adenocarcinoma, chronic pancreatitis (inflammation), and normal tissues—on nine patients. Classification algorithms using linear discriminant analysis are developed to distinguish among tissues, and leave-one-out cross-validation is employed to assess the classifiers’ performance. The spectral areas and ratios classifier (SpARC) algorithm employs a combination of reflectance and fluorescence data and has the best performance, with sensitivity, specificity, negative predictive value, and positive predictive value for correctly identifying adenocarcinoma being 85, 89, 92, and 80%, respectively.

Keywords: spectroscopy; fluorescence; reflectance; tissues; pancreas; carcinomas.

Acquired fluorescence spectra in the 360 to 750 nm range were corrected for spectral instrument response after background correction. Reflectance spectra were background subtracted \( (R) \) and then scaled by the lamp reflectance spectrum \( (R_s) \) to obtain corrected reflectance spectra \( (R/R_s) \) in the 400 to 750 nm spectral range. All spectra were then normalized by scaling the peak intensity value to unity. Two sets of fluorescence and reflectance measurements were made on each site, except one. Three pairs of fluorescence and reflectance spectra were excluded based on the exclusion criteria: ratio of reflectance \( R_{515}/R_{600} < 0.1 \) or fluorescence signal-to-noise ratio \( (SNR) < 25 \), where \( SNR = \text{average intensity between 470 and 500 nm} / \text{standard deviation of intensity between 700 and 750 nm} \). This left 96 fluorescence and 96 reflectance measurements: 33 fluorescence and reflectance measurements were made on 17 adenocarcinoma sites, 41 were made on 22 pancreatitis sites, and 22 were made on 11 normal sites.

Figure 1 shows the mean of normalized measured reflectance and fluorescence spectra for adenocarcinoma (red)}
dashed line), pancreatitis (green dot-dashed line), and normal (blue solid line) pancreatic tissues, along with the standard error at select wavelengths. Tissue fluorescence spectra revealed cellular NAD(P)H (emission around 460 nm) and extracellular matrix collagen (emission around 400 nm), and characteristic hemoglobin absorption dips at around 420, 540, and 575 nm in the reflectance and fluorescence spectra. The adenocarcinoma sites showed markedly higher reflectance than pancreatitis and normal tissue sites between 450 to 540 nm, higher fluorescence at around 400 nm, and lower fluorescence between 450 to 700 nm. Due to the ex vivo nature of the study, it is difficult to draw conclusions on differing hemoglobin absorption between tissue types. The observed spectral differences are consistent with increased cellular nuclear size and collagen content in pancreatic adenocarcinoma tissues. To classify the tissue spectra based on these apparent differences, the ratio $R_{\text{ratio}} = R_{450}/R_{550}$ was calculated for each reflectance spectrum, the wavelength integrated fluorescence ($F_{\text{area}}$) was calculated as the area under each normalized fluorescence spectrum, and the ratio $F_{\text{ratio}} = F_{450}/F_{600}$ was calculated for each fluorescence spectrum. The gray dashed lines in Fig. 1 indicate the wavelength at which the ratios $R_{\text{ratio}}$ and $F_{\text{ratio}}$ were calculated.

Three different sets of these calculated spectral parameters (classification variables) were employed to develop tissue classification algorithms for distinguishing: 1. adenocarcinoma (A) from pancreatitis (P) and normal (N) tissue, 2. A from P tissue, 3. A from N tissue, and 4. P from N tissue. A leave-one-out cross-validation was undertaken to test the performance of the proposed tissue classification algorithms. Linear discriminant analysis (LDA) was employed using SPSS software (SPSS, Chicago, Illinois) to classify the test data using the three different sets of classification variables. LDA employs a linear combination of the classification variables to classify data. This process was repeated for each spectrum, and the sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) of the classification algorithms were calculated.

The first classification algorithm, the reflectance spectral areas and ratios (RSpARC) classifier employed $R_{\text{ratio}}$ as the sole classification variable for LDA, and Table 1 gives the algorithm’s performance. The second algorithm, the fluorescence spectral areas and ratios (FSpARC) classifier employed $F_{\text{area}}$ and $F_{\text{ratio}}$ as the classification variables for LDA, and Table 2 gives that algorithm’s performance. In the third algorithm, the spectral areas and ratios (SpARC) classifier step-wise LDA was performed using minimization of Wilks’ lambda (A) criterion (P-to-enter 0.085; P-to-remove 0.1) to assess the discriminating power of the variables, and to select the best set of variables from $R_{\text{ratio}}$, $F_{\text{area}}$, and $F_{\text{ratio}}$ for classification. The variables retained by the step-wise analysis were then employed to classify the data. Table 3 shows the performance of this algorithm, along with the variables retained for each classification. For example, classification of A versus P employed $R_{\text{ratio}}$, For A versus (P and N) and P versus N classification, the combination of classifiers extracted from both reflectance and fluorescence (Table 3) performed better than using information from either just reflectance (Table 1) or just fluorescence (Table 2), indicating that for the SpARC algorithm, both fluorescence and reflectance spectroscopies are required for accurate pancreatic tissue classification.

![Fig. 1](https://example.com/fig1.png)

**Fig. 1** (a) Mean fluorescence spectra obtained from human pancreatic normal (blue solid line), pancreatitis (green dot-dash line), and adenocarcinoma (red dashed line) tissues. The gray vertical dashed lines indicate the wavelengths at which $F_{\text{ratio}} = F_{450}/F_{600}$ was calculated for each spectrum. (b) Mean reflectance spectra obtained from normal (blue solid line), pancreatitis (green dot-dash line), and adenocarcinoma (red dashed line) tissues. The gray dashed lines indicate the wavelengths at which $R_{\text{ratio}} = R_{450}/R_{550}$ was calculated for each spectrum. N denotes the number of individual spectra. The standard error is shown at select wavelengths. (Color online only.)

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>NPV (%)</th>
<th>PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A versus P</td>
<td>85</td>
<td>86</td>
<td>92</td>
<td>76</td>
</tr>
<tr>
<td>A versus P</td>
<td>85</td>
<td>83</td>
<td>87</td>
<td>80</td>
</tr>
<tr>
<td>A versus N</td>
<td>88</td>
<td>95</td>
<td>84</td>
<td>97</td>
</tr>
<tr>
<td>P versus N</td>
<td>56</td>
<td>82</td>
<td>50</td>
<td>85</td>
</tr>
</tbody>
</table>

**Table 1.** RSpARC algorithm—reflectance only.

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>NPV (%)</th>
<th>PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A versus P</td>
<td>55</td>
<td>89</td>
<td>79</td>
<td>72</td>
</tr>
<tr>
<td>A versus P</td>
<td>52</td>
<td>88</td>
<td>69</td>
<td>77</td>
</tr>
<tr>
<td>A versus N</td>
<td>58</td>
<td>91</td>
<td>59</td>
<td>90</td>
</tr>
<tr>
<td>P versus N</td>
<td>56</td>
<td>64</td>
<td>44</td>
<td>74</td>
</tr>
</tbody>
</table>

**Table 2.** FSpARC algorithm—fluorescence only.
Hartwig et al. conducted a meta-analysis of 28 EUS-FNA studies, and reported median values of sensitivity at 83% (54 to 96%); specificity 100% (71 to 100%); NPV 72% (16 to 92%); and PPV 100% (92 to 100%) for adenocarcinoma distinction from normal tissue and chronic pancreatitis. The performance of the SpARC algorithm for adenocarcinoma distinction from normal tissue and chronic pancreatitis (Table 3, row 1) is comparable to that of a reported performance of EUS-FNA. Fritscher-Ravens et al. studied patients having both adenocarcinoma and pancreatitis, and reported that EUS-FNA had 54% sensitivity, 100% specificity, 91% NPV, and 100% PPV for distinguishing adenocarcinoma from pancreatitis in the setting of pancreatitis. All adenocarcinoma measurements in the study reported here were also made on pancreata that had concurrent histologic pancreatitis in addition to the carcinoma. Thus, the sensitivity of SpARC (85%) (Table 3, row 2) is well above that of EUS-FNA (54%) for distinguishing adenocarcinoma from pancreatitis in the setting of pancreatitis, which is an unmet clinical need in pancreatic cancer detection. The receiver operating curves for A versus (P and N) and A versus P classification using the SpARC algorithm were calculated using SPSS software and are shown in Fig. 2. The blue line suggests that in achieving a specificity of close to 100%, the A versus P sensitivity would be comparable to EUS-FNA. Thus, employing an optical technique to guide EUS-FNA could improve the sensitivity of pancreatic adenocarcinoma detection in the setting of pancreatitis.

While this study was undertaken in an ex vivo setting, the performance of the developed classification algorithm shows promise for clinical pancreatic tissue classification using bi-modal optical spectroscopy. Although the SpARC algorithm is rapid and simple to implement, its performance may be limited because it utilizes information at only select wavelengths. Future work will include the investigation of algorithms that consider information at all wavelengths for tissue classification. Separate studies are also underway to employ parameters extracted from a photon-tissue interaction model for tissue classification. The model provides quantitative links between spectroscopic measurements and biophysical tissue characteristics, including hemoglobin absorption. Future work will involve data acquisition in vivo, with the aim of deploying the fiber optic probe through a needle for optically guided EUS-FNA.

Acknowledgments

This work was supported in part by the National Institutes of Health (CA-114542), The National Pancreas Foundation, the Wallace H. Coulter Foundation, the University of Michigan Comprehensive Cancer Center, and the University of Michigan Medical School Translational Research Program. We thank C. Sonnenday and L. M. Colletti for allowing us to collect data during the pancreatic surgeries performed by them.

References