REVIEW

YJBM

Extracting Knowledge from Chemical Imaging Data Using Computational Algorithms for Digital Cancer Diagnosis

Saumya Tiwari[a,b] and Rohit Bhargava[a,b,c]*

[a]Beckman Institute for Advanced Science and Technology; [b]Department of Bioengineering; and [c]Departments of Chemical & Biomolecular Engineering, Electrical & Computer Engineering, Mechanical Science & Engineering and Chemistry, University of Illinois at Urbana Champaign, Urbana, Illinois

Fourier transform infrared (FTIR†) spectroscopic imaging is an emerging microscopy modality for clinical histopathologic diagnoses as well as for biomedical research. Spectral data recorded in this modality are indicative of the underlying, spatially resolved biochemical composition but need computerized algorithms to digitally recognize and transform this information to a diagnostic tool to identify cancer or other physiologic conditions. Statistical pattern recognition forms the backbone of these recognition protocols and can be used for highly accurate results. Aided by biochemical correlations with normal and diseased states and the power of modern computer-aided pattern recognition, this approach is capable of combating many standing questions of traditional histology-based diagnosis models. For example, a simple diagnostic test can be developed to determine cell types in tissue. As a more advanced application, IR spectral data can be integrated with patient information to predict risk of cancer, providing a potential road to precision medicine and personalized care in cancer treatment. The IR imaging approach can be implemented to complement conventional diagnoses, as the samples remain unperturbed and are not destroyed. Despite high potential and utility of this approach, clinical implementation has not yet been achieved due to practical hurdles like speed of data acquisition and lack of optimized computational procedures for extracting clinically actionable information rapidly. The latter problem has been addressed by developing highly efficient ways to process IR imaging data but remains one that has considerable scope for progress. Here, we summarize the major issues and provide practical considerations in implementing a modified Bayesian classification protocol for digital molecular pathology. We hope to familiarize readers with analysis methods in IR imaging data and enable researchers to develop methods that can lead to the use of this promising technique for digital diagnosis of cancer.

INTRODUCTION

Infrared (IR) spectroscopic imaging is a promising avenue for computerized disease diagnosis [1-7], especially for cancer [1,2,8-18] and a multitude of other diseases [19]. It is of particular relevance for recognizing features within solid tissues in which a variety of cell types and disease states may be present. Utilizing the tandem spatial and molecular information acquired using a combination of IR spectroscopy and optical microscopy, this technique relies on using the biochemical composition as a means to automate disease identification. In IR imaging, no stains are used. Instead, the chemical composition of the material is recorded via a local spectrum and computer algorithms are used to relate the data to underlying physiologic conditions. Since only light is used to record the necessary data, the technology is entirely non-perturbing to a prepared sample. The overall idea of using IR imaging for biological applications is shown in Figure 1. This approach is orthogonal to the current practice in histopathology, which

Figure 1. Overview of the use of IR imaging for biological analyses.

requires staining to visualize tissue morphology as well as intensive human involvement to recognize and categorize morphological features that are indicative of disease. The IR-based approach strongly relies on sophistication and utility of the numerical methods used. The focus of this article is to describe and highlight the salient features of numerical methods used in IR imaging.

IR Imaging to Address Current Cancer Pathology Needs

At present, the gold standard to identify many types of cancers is to perform a biopsy. The poorly quantitative procedures following the biopsy and staining are semi-automated at best and still suffer from user introduced variability [20,21]. This not only introduces subjectivity in examination [22] but also increases the load on the pathologist that could otherwise be devoted to more complicated cases. Misclassification of biopsies during screening and diagnosis may lead to overtreatment or undertreatment, posing significant concerns for patients. For example, a recently published report [23] evaluated an agreement among 115 pathologists who interpreted 240 cases of breast biopsy samples and compared it to the consensus-derived reference diagnoses from three expert pathologists. The researchers found out that the overall agreement between the participating pathologists' interpretations with the reference was 75.3 percent. Alarming under-interpretations were found in ductal carcinoma *in situ* (DCIS) cases (13 percent) and atypia cases (35 percent). Considering that DCIS accounts for 15 percent to 25 percent of

the newly diagnosed breast cancer cases currently in the United States [24] and identification of atypical cells often requires further rounds of biopsy to establish aggressiveness of possible tumor, large numbers of patients could be affected every year based on whether a second opinion is obtained. In another recent study [25], the researchers consulted 252 pathologists to assess the policy of obtaining a second opinion on a variety of specimens. Their response indicated that a second opinion was only required in 56 percent of the laboratories when DCIS was diagnosed and in 36 percent of laboratories when atypical ductal hyperplasia was observed. In many cases, a third opinion was required to resolve the differences between the first and second opinions. Studies like these and others [26-28] clearly show that there are a lot of breast cancer cases that are affected by confusions in classification of type and aggressiveness of tumor and current pathology practice is in need of better tools to aid diagnoses.

Multiple computer-aided detection systems have been used in the past to assist the pathologists and help them reduce occurrences of false positives and false negatives [29]. In current practice, the computer-aided detection systems that rely on pattern recognition software used by radiologists can be considered semi-automated in that some degree of human interaction is still needed before a final decision is given. In that sense, *detection* systems are different from *diagnosis* systems, which are capable of rendering a decision based on a consideration of a variety of factors such as mass of tumor, biochemical data from biopsy, and patient characteristics such as breast density and age. These sys-
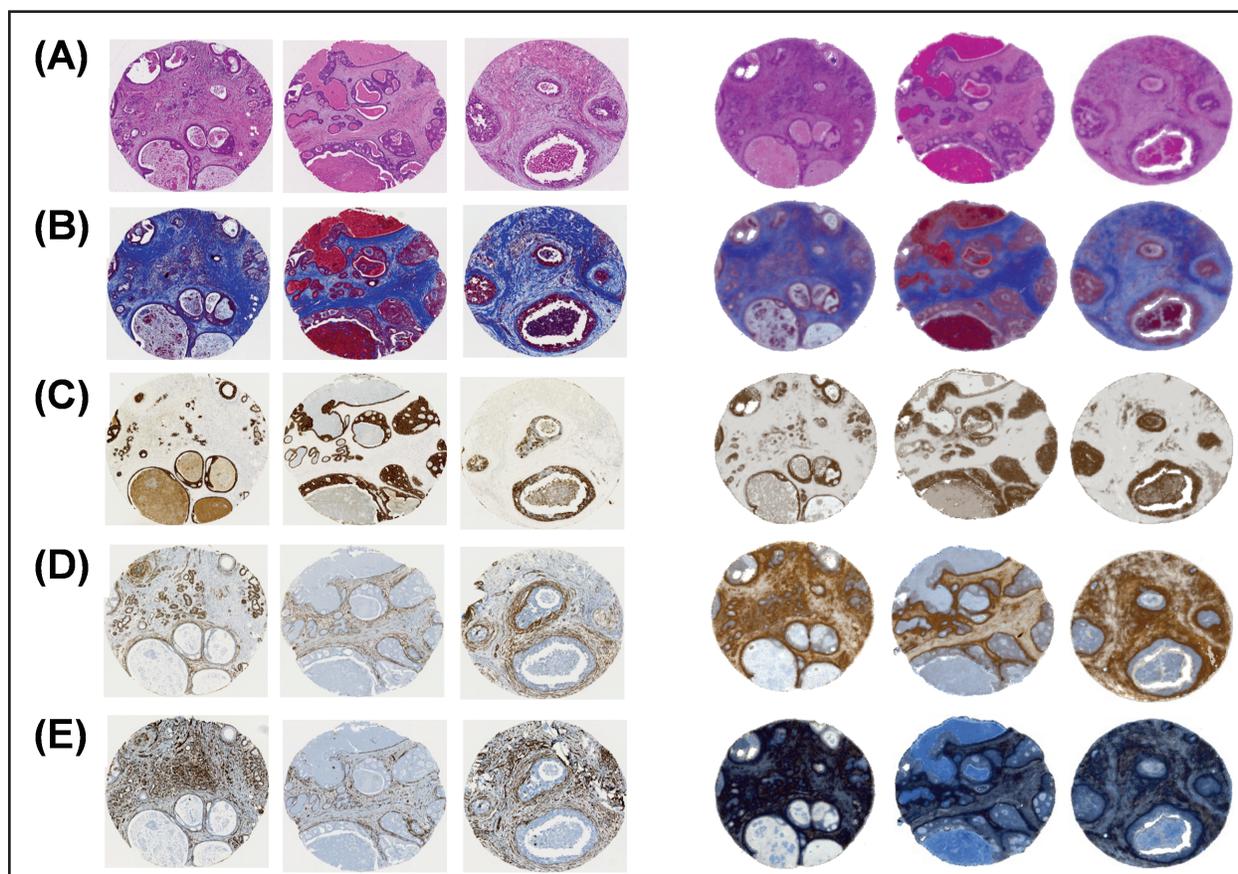
Figure 2. Molecular imaging (three sample panel on the left) can be reproduced by chemical imaging (right panel). In addition to H&E stained images, (A) we extend the concept of stainless staining to molecularly specific stains. (B) Masson's trichrome stain (collagen and keratin fibers). (C) High molecular weight (HMW) cytokeratin (epithelial type cell). (D) Smooth muscle alpha actin (myo-like cell). (E) Vimentin (fibroblast like cell). Each spot is 1.4 mm in diameter. Adapted with permission from World Scientific Publishing Co./Imperial College Press (Mayerich D, Walsh MJ, Kadjacsy-Balla A, Ray PS, Hewitt SM, Bhargava R. Stain-less staining for computed histopathology. Technology. 2015;3(1):27-31.)

tems thus require integration of two major fields: computation and imaging. In terms of imaging for diagnostic cancer pathology, the foremost requirement is the ability to generate contrast between diseased regions and healthy regions. Traditionally, chemical and immunohistochemical stains have been used to produce this contrast that is then referred to pathologist for evaluation. The second step now increasingly involves the use of computers to manage images and assist with decisions using numerical indices or other image analysis techniques. However, there are emerging alternatives to this long-standing instrumentation. For example, microscopic contrast also can be produced optically using Raman imaging or IR spectroscopy, two strongly emerging modalities that also place new requirements and provide new opportunities for the associated computational methods. IR spectroscopic imaging has some distinct advantages over other contrast-producing modalities. First, it requires minimal sample preparation. Freshly taken tissue can be snap frozen and imaged without further aids. This greatly reduces variations during experimental stages, making the procedure standardized and efficient. It can as easily be applied to archival samples. Second, IR imaging does not require contrast agents but utilizes the inherent biochemical contrast in the tissues for differentiation of diseased state. Third, the chemical changes recorded by infrared spectroscopy across the tissue are capable of giving the same information as achieved by histological stains [30]. In addition, since the information is computer generated, they provide greater contrast and statistical confidence, in turn enabling easier identification of problematic areas. A recently published report [31] showed that a single IR spectral image could reproduce staining patterns of multiple stains such as hematoxylin and eosin (H&E), Masson's trichrome stain, cytokeratin stain, smooth muscle alpha actin, and vimentin (Figure 2). This could allow the researchers to analyze the samples through multiple stains without putting in additional time, effort, or resources to develop the stains.

Along with reproducing classical stains with great accuracy, data generated by IR imaging is highly amenable to computational analysis, and pattern recognition algorithms are easily integrated for obtaining decisive reports. Currently, a major goal of the typical studies performed using IR imaging on tissue samples is to build classification sys-

tems that color code IR images to differentiate between different types of cellular and acellular components, much like H&E and IHC stains. Classes such as epithelium, endothelium, stroma, and muscle have been identified [32,33] and more cellular and acellular components are being added through current research. Although this approach provides high contrast images with minimal sample preparation for use by trained pathologists, in order to truly utilize the potential of IR imaging for cancer diagnosis, further computational prediction needs to be implemented. A recent report [34], for example, attempted to precisely predict recurrence of prostate cancer using IR imaging data and showed that this approach outperformed both Kattan nomogram and CAPRA-S scores for outcome predictions. Together, emerging studies are opening new avenues for utilization of IR-based models for cancer diagnosis and therapy by combining imaging, molecular detection, and computational cancer prediction to augment human decision-making. Owing to the practical requirements of speed of imaging and data acquisition and processing, no automated diagnosis systems have been clinically implemented until now; nevertheless, fast progress is being made to achieve this goal and will be discussed briefly in later sections. We first provide an overview of the methods, highlighting special considerations and challenges that use this data and lead to decision-making in cancer research and care.

Classification Models

A biological sample characteristically consists of many cell populations and extracellular matrix elements. All of these elements serve a function in the sample, and imbalance in the chemical composition and morphology of these can be a cause or an effect of a disease. Thus, these cellular and acellular components of tissue are carefully scrutinized by pathologists to obtain information about the ailment. We refer to all such functional elements as histological classes or, simply, classes. The idea underlying the use of IR spectroscopy for disease detection is that each such class will have a different biochemical composition and therefore unique spectral signature in IR absorbance spectra. Since digital spectral data is available for each pixel from the sample, we can employ pattern recognition algorithms to utilize these differences for recognition of classes. Various classification approaches have been used in the past to identify classes, termed as classification. Multiple studies have been performed for the analysis of data using various classification algorithms and are summarized here. For an in-depth theory on classification methods pertaining to biomedical imaging, the readers are directed to these references [33,35,36]. Typically, all methods can be classified into supervised or unsupervised methods, both of which are described briefly below. Subsequently, we focus here on describing the typical process of obtaining data, computational pipeline, and typical results obtained. We illustrate the entire process with representative examples to enable the reader to grasp the essential steps of extracting information from IR images.

Unsupervised Classification

The premise of unsupervised classification is that no prior information (i.e., spectral characteristics of the classes) is fed to the method for classification. Hence, distinction between classes is often a problem of finding clusters in which intra-cluster variation is smaller than inter-cluster variation. Unsupervised clustering approach has been applied previously to investigate tissue samples [37-39]. Since nothing is assumed known about the data classes, unsupervised processes can involve data reduction using the variance before applying a classification procedure. Such a methodology has been applied to classify IR imaging data from cervical cancer [40]. Principal component analysis (PCA) for data reduction followed by K means clustering was used elsewhere for classification of IR data [12]. Although unsupervised approaches work for exploratory analysis, they have been found to be computationally taxing and unable to differentiate between inter-class and intra-class variations, often necessitating the use of supervised classification algorithms [41,42]. In our opinion, the utility of these methods for IR imaging lies more in discovery rather than consistent knowledge extraction.

Supervised Classification

In supervised classification, prior information about the location and spectral properties of the classes is given to the classifier. Supervised algorithms such as discriminant analysis [2,43-45], neural network analysis [16,41,46-48], and Bayesian methods-based classification [32,49] have been used to classify tissue into various cellular and disease states. Underlying this method is the fundamental property of Bayes' theorem, indicating that known patterns provide a statistical probability for identification of each class. Methods based on this property and its application for biological specimens has been discussed elsewhere [32]. Here, we discuss the practical considerations for its implementation, in order to facilitate understanding and ease of use among spectroscopists and medical researchers alike.

IMAGE COLLECTION AND PRE-PROCESSING

Collecting a good quality image is the first step of any IR classification experiment. Often, this facet is overlooked. Good quality data reduces complexity of the methods and can provide faster as well as more accurate results. Multiple factors such as choice of substrate signal-to-noise ratio (SNR), spatial and spectral resolutions, and presence of contaminants such as paraffin residue can affect the overall classification accuracy. In this section, we will discuss methods employed to collect good quality data and prepare the data for classification.

Substrates

IR spectroscopic imaging data can be collected in both transmission and reflection mode. IR transparent sub-

strates, such as calcium fluoride (CaF_2) and barium fluoride (BaF_2) salt crystals are excellent substrates since they achieve greater than 95 percent transmission in mid IR region. An overview of properties and uses of CaF_2 and BaF_2 crystals can be found in reference [50]. Specifically for imaging biological samples, BaF_2 is preferable, since transmittance of CaF_2 cuts off at about $1000cm^{-1}$ and analysis at lower wavenumber (longer wavelengths) is not possible. BaF_2 is more prone to damage, however, due to higher water solubility compared to that of CaF_2 [51], making its handling and maintenance slightly more difficult but not substantially different compared to standard glass slides. Once a sample is placed on these crystals, imaging is rather simple. Using a microscope objective-condenser setup, light simply passes through the substrate and the sample in a "transmission" mode. The major problem with either substrate is cost, which can run to hundreds of dollars. Due to the high cost of these substrates and higher maintenance requirements compared with standard glass slides, many IR studies now utilize IR reflective substrates such as gold coated slides and Low-E slides (MirrIR, Kevley Technologies). Low-E slides, in particular, have been very useful for IR imaging, owing to their ability to transmit visible light and reflect infrared light. Thus, imaging is often conducted in the transflection mode with these substrates. In the transflection mode, light is incident upon the sample, passes through it, is reflected from the sample-substrate interface, and re-transmitted through the sample. Due to the sample typically being of a thickness that is the same order as the wavelength of light, passing through the sample twice results in distortions in the spectrum [52,53] compared to the transmission case. However, some pre-processing steps have been reported that can effectively encounter most of the side effects of transflection mode and are discussed below. With emerging methods and more flexibility in terms of cost and maintenance [54], Low-E slides are attractive options to carry forward IR based detection technologies to everyday use in clinics.

Signal-to-Noise Ratio

In IR imaging, the spectral signal-to-noise ratio (SNR) is the primary measure of the quality of data. It has been shown that high levels of noise in data negatively impact the classification accuracy [55]. Hence, SNR should be carefully considered in the design and use of any protocol. Modern infrared imaging instruments have combated the problem of low SNR quite well, and one can routinely obtain an SNR of greater than 200 on commercial instruments. There are multiple factors that can determine the SNR for data collected. For commonly used focal plane array (FPA) detectors in IR imaging instruments, each element in the detector records the spectrum from one pixel in the sample. As the number of co-additions is increased, the signal is recorded a multiple number of times and averaged. This improves the SNR by the square root of the number of co-additions. However, this also increases the time required for data acquisition almost linearly with the number of co-additions. Another option is to reduce spatial resolution (increase the size of the pixel at the sample plane), which can provide a higher SNR in smaller time due to a larger angle of light collected, but this may compromise identification of small cells in biological samples. An additional key factor with image collection is background spectrum. Every IR imaging experiment requires collection of background spectra used as a reference to obtain absorbance measurements. The number of co-additions for background spectrum should be much larger than the number of co-additions for the image in order to have minimal introduction of noise in signal from background [56]. Some limits on SNR also are imposed by the interferometer and other hardware, as well as multiple other factors such as spectral and spatial resolution, which is a result of complexities in the acquisition process. Some of the factors that affect SNR have been discussed in previous works [55-57]. Here, we want to emphasize that the data quality in IR imaging is a balance between optimum SNR, optical configuration needed, and the time required to achieve the desired SNR. One method we have not discussed thus far is the use of post-acquisition numerical processing techniques that can use statistical or other measures of noise reduction and lead to reduced noise in the images. The basic principle underlying these methods is to transform the data into a space that collapses all information into a minimum number of factors, for example, using principal components transform [58]. Fortunately, due to these computational noise reduction techniques (discussed later) SNR is not a limiting factor for classification accuracy for many of the common tasks in spectroscopic imaging [55].

Spatial and Spectral Resolution

The main constituents of biological samples are the different types of cells that comprise the tissue as well as the extracellular matrix (ECM) that holds tissue structures together. The size of eukaryotic cells can vary from about 5 to 30 microns. Sufficient spatial resolution is necessary to identify each cell type [59] and, thus, the instrumentation and experimental parameters must be carefully selected. Insufficient spatial resolution leads to the problem of mixed pixels, whereby, if the pixel is too large, it can have contribution from multiple cells, leading to greater confusion and low accuracy of classification [33]. Typically, pixel size used for IR microscopy has been approximately 5 µm x 5 µm. Pixel size in attenuated total reflectance (ATR) mode can be higher due to the use of a solid immersion lens [60-62]. A microscope equipped with transmission optics and ATR lenses can provide higher resolution depending on the solid immersion lens or ATR crystal material. For example, one commercial instrument provides a pixel size of 6.25 µm x 6.25 µm in the transmission mode and 1.56 µm x 1.56 µm sized pixels in ATR mode using a Germanium lens (refractive index ~ 4). High definition IR imaging instruments typically seek to provide 1 µm x 1 µm. It should be noted that the pixel size is not the same as resolution. Resolution
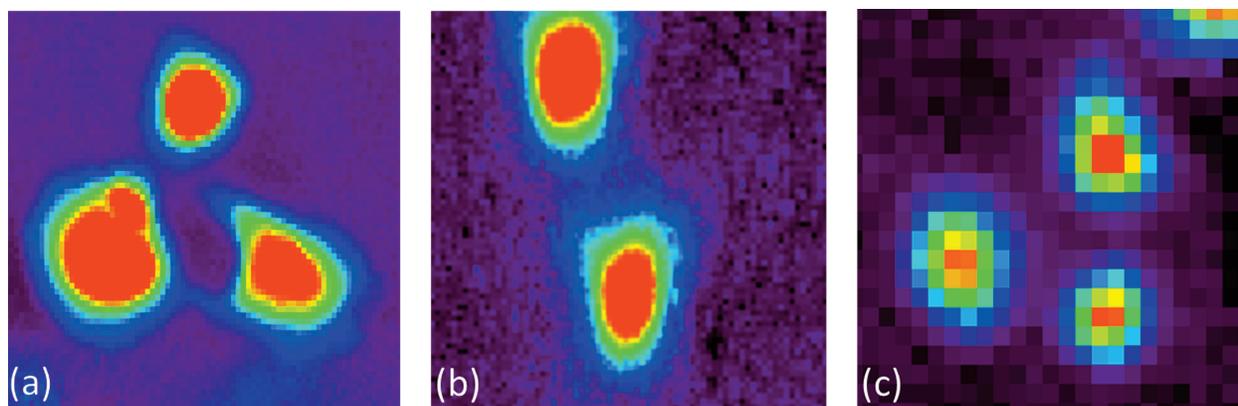
Figure 3. Images of eukaryotic cell at varying resolutions. (a) ATR mode, pixel size 1.56μm x 1.56μm; (b) transflection mode 74X, pixel size 1.1μm x 1.1μm; and (c) transmission mode 14X, pixel size 6.25μm x 6.25μm. All images are at amide I band (1652 cm^{-1})
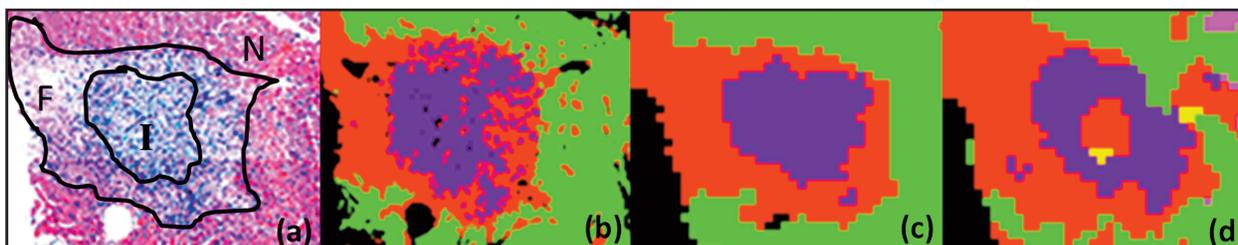
Figure 4. Factors affecting classification. (a) H&E image with marked regions (I: Infiltrate, F: Fibrosis, N: normal tissue); (b) classified image pixel size 6.25μm x 6.25μm (deparaffinized); (c) classified image pixel size 25μm x 25μm (deparaffinized); and (d) classified image pixel size 25μm x 25μm (paraffinized).

is still determined by the Rayleigh Criterion; for example, it is ~5 μm for transmission mode imaging and 1 μm for ATR imaging. A comparison of IR images taken at various pixel sizes for mammalian cells is shown in Figure 3. As can be seen, high amide absorbance region of nucleus is much better resolved with high spatial resolution as compared to the low resolution transmission image. Effect of varying pixel size on classification is shown in Figure 4b and c, where 6.25 μm x 6.25 μm pixel size data is compared to 25 μm x 25 μm pixel size data. H&E image with marked classes is shown for comparison (Figure 4a). Higher pixel density via smaller pixel sizes provides IR images that are closer to histologic stain image, whereas more averaging to increase SNR and larger scanning time to acquire more pixels is required. The large pixel sizes result in overlapping of signals from different cell types and the ECM, reducing confidence in classification. A large pixel can reduce the scanning time greatly and provide high SNR. For most biological problems involving complex tissues, however, a high spatial resolution is often needed. An equally important factor for good classification is spectral resolution. For very coarse spectral resolution, the peaks begin to overlap, causing significant reduction in classification accuracy [63]. Typical IR imaging experiments utilize a spectral resolution of 2 cm^{-1} to 16 cm^{-1}. For biological specimens, spectral resolution of 4 cm^{-1} to 8 cm^{-1} is able to differentiate most of the significant peaks and has been found to give good classification results in our experience.

Paraffin Removal

Since most samples are typically paraffin embedded and sectioned before IR imaging, the sections need to be deparaffinised in order to remove spectral contributions from paraffin that typically occur as a set of peaks from 2800 cm^{-1} to 3000 cm^{-1} due to C-H stretching vibrational modes and a strong peak at 1462 cm^{-1} due to C-H bending modes [64]. Deparaffinization carried out either in Xylene [44] or Hexane washes for 16 to 24 hours with mild stirring or with octane for 4 hours [65] have all shown to remove paraffin features from spectrum. Figure 4 compares classification results from a paraffinized sample (Figure 4d) and the same sample after paraffin removal (Figure 4c) for same spatial resolution. A spectrum from sample before and after de-paraffinization is shown in Figure 5. Even though we avoided using parts of the spectrum from paraffin-affected regions from 2800 cm^{-1} to 3000 cm^{-1} and around 1462 cm^{-1}, the classification was more accurate for the deparaffinized sample than the paraffinized sample for the same spatial and spectral resolution. Classification of samples imaged without deparaffinization also can work if appropriate corrections are performed [66]. Nevertheless, if paraffin retention is known or suspected, care should be taken to address any signals arising from paraffin while performing classification.

Preprocessing

Once IR images are acquired, minimal data processing is needed for performing classification. Based on the
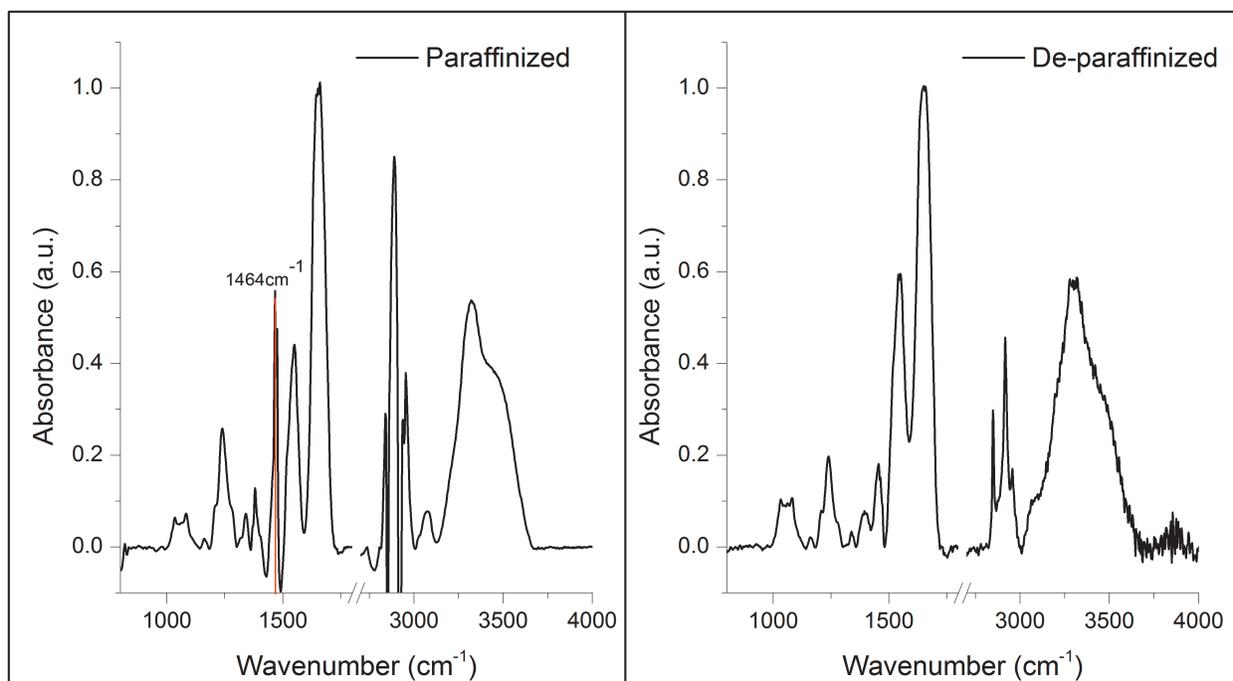
Figure 5. Difference in spectrum for tissue with and without paraffin.

SNR, computational noise reduction methods such as those based on the Minimum Noise Fraction (MNF) [55] may be needed before classification can begin. This is a modification of principal components analysis whereby the ordering of eigenimages is performed in decreasing order of SNR and high SNR eigneimages are chosen for analysis. Noise statistics are calculated to form the image data. MNF transform creates three files: covariance statistics of the noise, MNF statistics, and forward MNF transformed, which contains bands with descending eigenvalues. Based on the eigenvalues, the user can determine which bands contain data and which have predominant noise. Typically, the top 20 to 30 bands contain good quality data. Inverse MNF transformation is then applied on the forward MNF transformed file by taking high eigenvalue bands. Most commonly, this type of noise reduction is needed for ATR and high definition imaging data. Baseline correction of data is needed for comparing spectral features attributable to absorbance across classes and gives a good estimate of the differences before training can begin. A variety of baseline correction options are known, but all essentially approximate the known non-absorbing regions to zero. In one approach, all points where theoretically zero absorption is expected are first identified. Then, a linear two-point correction algorithm across peaks of interest is used. It should be noted that in imaging, the baseline points are often held the same for all spectra in the sample. To account for thickness variations in the sample or between samples, normalization with amide I peak (1650 cm^{-1}-1656 cm^{-1} based on location of peak) is often required. For biological tissues, this is the peak of highest absorbance and introduces the least decrease in SNR after normalization.

CLASSIFICATION PROTOCOL

The classification protocol followed here is based on modified Bayesian classification. The complex multistep method is explained through the flowchart shown in Figure 6. We describe the major steps in the workflow and discuss possible pitfalls while building and deploying a classification protocol.

Selection of Training and Validation Set

The goal of the classification protocol discussed here is to identify different types of cells and ECM elements called histological classes. The ultimate goal is to develop a computational algorithm that provides accurate recognition of classes in an unknown data set that can be encountered in practice. In the first step, the protocol needs to be developed and tested to perform optimally. To initiate this process, two separate data sets are selected. One is used for training and the other for independent validation. In some cases, the validation may come from the calibration data set itself. In such cases, one fraction of the data is selected for validation and the protocol is trained on the remainder. The fraction left off is changed, and a number of iterations of the process are averaged to train and validate. This "leave-one-out" procedure can be used when the numbers of samples or diversity of the data set is limited, but it is always ideal to have completely independent training and validation sets. It must be ensured that there is sufficient representation of all classes for getting satisfactory classification and to retain sufficient diversity for assessment of accuracy in validation dataset. In this approach, study design is critical as the method cannot predict conditions on which it has not been trained. The measures of success should also be carefully defined. We
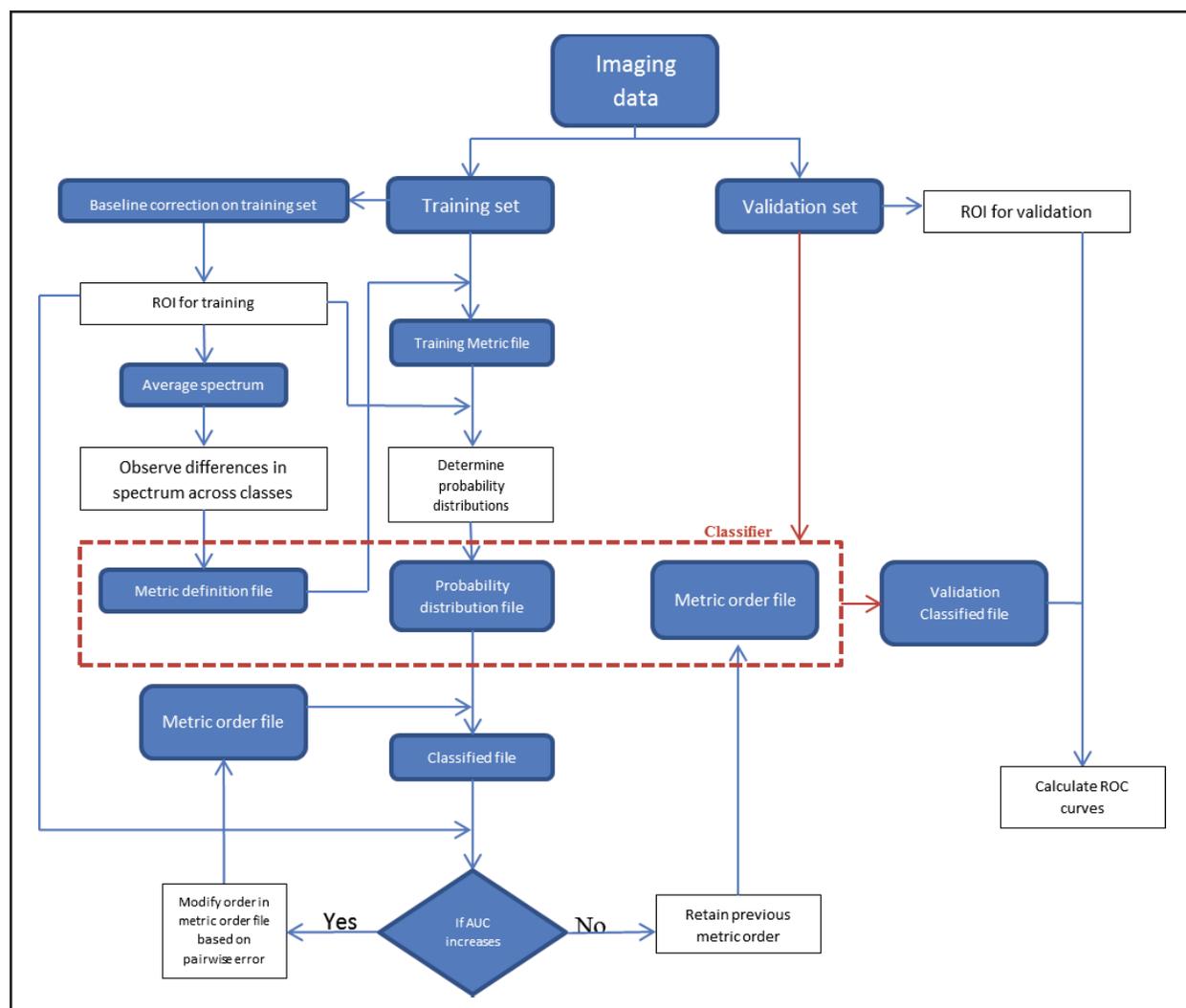
Figure 6. Flowchart of building supervised classifier.

favor the use of the receiver operating characteristic (ROC) curve that includes an assessment of both sensitivity and specificity of the method. Other approaches may be to maximize detection of any class or disease state (e.g., cancer) at a specified error rate or to evaluate errors in a holistic manner such as with confusion matrices. Finally, a statistically significant number of samples must be used to validate the protocol. While the numbers of samples needed for a diagnostic test is well understood, the sample size needed for satisfactory calibration in the IR is an ongoing subject of study [33,67]. In the absence of other guidance, the standard approach is to calibrate, validate, and calculate the ROC curves for the classifier in order to assess the accuracy. The errors in classification additionally must be carefully assessed. Based on these results, the investigator may need more data for accurate classification.

Metric Definition

Depending on the size of the sample, an FTIR imaging data set can be very massive, ranging from a few hundred megabytes to hundreds of gigabytes. Each pixel in

an IR image carries spectrum, which is usually recorded in the FT configuration across the entire bandwidth of the spectrometer but is usually truncated to reduce the size of the stored data to a smaller range, e.g., 750 cm^{-1} to 4000 cm^{-1}. However, not all spectral elements are useful for classification. For example, a region between 1900 cm^{-1} to 2500 cm^{-1} is biologically inactive and can be further removed if spectral corrections that depend on extensive refractive index measurements are not to be performed [68]. One approach to dealing with imaging data is the emergence of the so-called discrete frequency IR imaging in which, using filters [69,70] or a tunable laser [71-76], only a few frequencies of interest are collected. This approach likely will prove useful only after the calibration process. Hence, in general, the entire spectrum is acquired and needs to be handled for the calibration step. Data reduction discussed here simply suggests using data that gives qualitative and quantitative information about the sample and removing redundant data. While it is not necessary for single cell studies that do not require large computational power, biopsy sections and tumor micro arrays (of the order greater than 1mm X 1mm in size) would need much
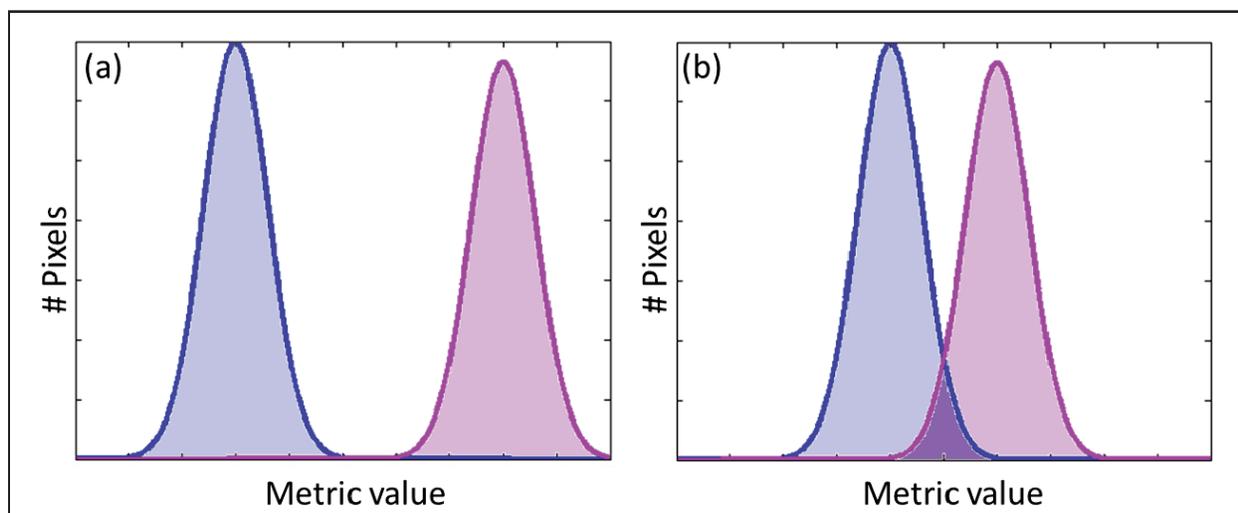
Figure 7. Histograms for metric evaluation. (a) An example of good metric and (b) an example of metric that will have errors in classification when the metric value lies in the overlapping area of the two curves.

computing time if raw spectrum is used without data reduction. Further, confounding information may become included unless a careful selection of informative spectral regions is used. We can utilize spectral features such as peak height ratios, peak area to height ratio, peak area to area ratio, and peak center of gravity to differentiate among classes. These parameters are known as spectral metrics. Metrics are defined by an expert spectroscopist by observing the spectrum in tissue to identify exact peak locations. Many metrics have biological relevance; for example, glycogen to phosphate ratio (1030 cm^{-1}/1080 cm^{-1}), but sometimes the physiological relevance is not intuitive. Even then, at this point, all possible metrics that show differentiation among classes based on class spectrum should be considered. For every new imaging experiment, it is necessary to define the metric definitions anew, in order to account for spectral differences among classes and small differences in peak locations.

Identification of Classes

Identification of classes is the major factor that can determine accuracy of classification. To feed class characteristics as prior information for supervised classification, one needs an accurate identification of pixels used for training. This is typically performed with the help of an expert pathologist, often guided by H&E stained images or immunohistochemical images of corresponding sections. Typically, the practitioner marks regions corresponding to different classes by microscopic examination of H&E stained section. Correspondingly, regions in IR image are marked as regions of interest (ROI). H&E stained section can be a serial section or neighboring section to the section utilized for IR spectroscopy. A much preferable approach is to first obtain infrared images from the sample and then perform H&E staining on the same sample so that an exact match can be obtained. Once the classes are marked, non-biological pixels from class layers should be removed by

setting an intensity threshold value for biologically active band such as amide I band (~1652cm^{-1}) to a high enough value to remove both tissue-less regions as well as those with excessive distortions due to edge effects [77-80]. Subjectivity is the biggest issue in identification of classes and there have been multiple studies in past that show that the interpretation of H&E stain suffers from inter-observer variability and can have a role in false positive and false negative results [20,21,81]. In the absence of any absolute identification criteria at present, we rely on the opinion of pathologist for identification of classes. This adds a human error to the classification, and care is taken to mark the regions on IR image exactly same as the regions identified by the pathologist on the H&E image (considered "gold standard" [33]). This prevents the addition of further error in prior information for classifier training that relies on manual identification and marking of classes in IR data. An alternative is to use immunohistochemical stains to identify cell types and overlay the IHC images with the IR images. However, IHC stains are not known to be reliable all the time, and staining intensity may be open to interpretation requiring the use of sophisticated methods [82].

Evaluating Metric Distributions

The distribution of values of metrics forms the basis on which the classifier identifies and learns the differences among classes. An example of histogram is shown in Figure 7 that is the type of data to evaluate the use of metrics. Here, the number of pixels versus the value of metric parameter for each class is plotted. For a large enough number of pixels for each class, the histograms are expected to follow a normal distribution unless there are subclasses within the data. Hence, the first check is to determine whether there may be more than one distribution in the pixels, which may cause an examination of the model used in turn. For N metric parameters, we obtain N histograms. This step is important in identifying metrics

that can potentially be useful in differentiating between classes. When comparing between two classes, the overlap between the distributions is the critical parameter to evaluate. A small overlap in distributions implies that the values of metric parameters are sufficiently different and can be used to differentiate between the classes. The actual efficacy of metric parameter depends on the fraction of overlap that the probability distribution functions of different classes have with each other (Figure 7). It must be noted that abundance also comes into consideration here. For a given pixel, the overlap in normalized histograms can be considered. However, the probability of the selected pixel belonging to any particular class also needs to be considered. This depends not only on the native distribution of classes in normal and diseased conditions but also on the sampling process (e.g., biopsy). For example, it is well known that there is significantly more epithelium in cancer and in the peripheral zone of the prostate. Hence, simple abundance probability can be augmented significantly by known characteristics of the disease, patients, and procedures performed. Caution must be exercised, however, in making models that are too specific. While such models may perform at high accuracy, their robustness is likely to be compromised. The metric distribution, hence, must be evaluated in light of the model, the classification methods, and the desired accuracy. A metric parameter that can differentiate between at least two classes is considered useful for the purpose of classification. Since multiple such parameters eventually can be used in conjugation to separate all classes from each other, the set of metrics to be used and the order of their usage will be evaluated next. Following this step, with appropriate user input in determining histogram limits, a probability distribution file is created that contains the prior information of the classes.

Determination of Metric Order

After determination of probability distribution for every class in our case, a Bayesian classifier, each metric can be considered as a rule that determines to which class the pixel belongs. Therefore, the classifier goes through a series of rules to come to a decision about the class, assigning a class value to the pixel after each step is executed. In this perspective, it becomes important to determine the most optimum order of metrics so that the end result is closest to the true histology. For this, average errors for metrics are calculated and pairwise errors are arranged in increasing order. This order is optimized by classifying with reordered metric, calculating the area under the receiver operating characteristic (ROC) curve and recalculating pairwise error if there is an increase in area under the ROC curve. It has been shown before that only a fraction of metrics actually is needed to achieve the highest accuracy in classification [32]. Thus, after the optimum metric order has been defined, metrics coming at the bottom of the order can be removed from the classifier. Typically, this set of 15 to 25 metrics is identified based on

the metric order and area under the ROC curve, but an additional step in optimization can be performed by manually removing one metric at a time and assessing whether any increase in accuracy is achieved [32].

Validation

Validation of the classifier is performed on an independent data set by comparing the classifier with pathologist annotated IR data in the manner similar to calibration. ROC curves and confusion matrices commonly are used to assess the accuracy of classification. The ROC curve leads to two measures: the area under the curve (AUC) of the ROC curve and the sensitivity and specificity operating point of the classifier. The AUC is a "global" measure of how accurate the designed protocol can be on an average. Comparisons of AUCs, statistical limits, and ordering of different models based on the AUC are all operations that can be used to refine the classifier and gain further insight [83-85]. The operating point (i.e., sensitivity and specificity) can be considered to be a local condition that determines a particular operation of a diagnostic test. For any selected protocol, there will always be an operating point, which trades off the specificity and sensitivity but is implemented for the test. This is often determined by the problem and the tolerable error in the test. Sensitivity of greater than 70 percent at high specificity (90 percent) is generally considered satisfactory for biomedical detection systems, although a much higher sensitivity and specificity is often desirable for tasks such as disease diagnoses or recognition of particular cells. For example, in one study, among multiple breast cancer surveillance methods such as MRI, mammography, ultrasound, and clinical breast examinations, the sensitivity ranged from 9.1 percent to 77 percent and specificity ranged from 95.4 percent to 99.8 percent [86]. When using IR-based staining for digital cancer diagnosis, it is desirable to have sensitivity and specificity reach close to 100 percent. This has been shown to be possible by various recent studies [87,88]. While ROC curves determine the specificity and sensitivity of classification, the confusion matrices give the investigator an idea of confusion between classes in classification, and both should be used to evaluate the performance of the classifier. In validation studies, these matrices often point to systematic errors in the development of the classification protocol and must be examined carefully.

CONCLUSIONS

Automated computational classification is a very powerful technique to utilize IR spectroscopic imaging data. We emphasize that due to multiple steps required in image acquisition and classification protocols, careful considerations throughout are needed to assure successful development of assays. Often, the process of development of a classifier is not linear, and careful analysis and examination at each step is needed to ensure that the proto-

col is both accurate and robust. The theory and practice of Bayesian classification is well developed for infrared imaging data [32,33]. The protocols for image acquisition also have been described in detail in the past [63,89-91]. However, practical considerations while performing classification that can greatly affect the classification accuracy have not been recorded in infrared imaging literature. Through this paper, with illustrative examples, we have attempted to provide an introduction and a practical guide to considerations in the development of a specific classification protocol. Many of these considerations can be adapted for similar classification procedures, and we do hope that this article would enable and encourage the readers to familiarize themselves with infrared spectroscopy and utilize the avenues it offers for cancer diagnosis.

REFERENCES

1. Petibois C, Déléris G. Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. Trends Biotechnol. 2006;24(10):455-62.
2. Gazi E, Baker M, Dwyer J, Lockyer NP, Gardner P, Shanks JH, et al. A Correlation of FTIR Spectra Derived from Prostate Cancer Biopsies with Gleason Grade and Tumour Stage. Eur Urol. 2006;50;750-61.
3. Mackanos MA, Contag CH. FTIR microspectroscopy for improved prostate cancer diagnosis. Trends Biotechnol. 2009;27:661-3.
4. Diem M, Romeo M, Boydston-White S, Miljkovic M, Matthaus C. A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004). Analyst. 2004;129(10):880-5.
5. Naumann D. Ft-Infrared and Ft-Raman Spectroscopy in Biomedical Research. Appl Spectrosc Rev. 2001;36(2-3):239-98.
6. Levin IW, Bhargava R. Fourier transform infrared vibrational spectroscopic imaging: integrating microscopy and molecular recognition. Annu Rev Phys Chem. 2005;56(1):429-74.
7. Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst. 2006;131(8):875-85.
8. Maziak DE, Do MT, Shamji FM, Sundaresan SR, Perkins DG, Wong PTT. Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: An exploratory study. Cancer Detect Prev. 2007;31:244-53.
9. Walsh MJ, German MJ, Singh M, Pollock HM, Hammiche A, Kyrgiou M, et al. IR microspectroscopy: potential applications in cervical cancer screening. Cancer Lett. 2007;246(1-2):1-11.
10. Park SC, Lee SJ, Namkung H, Chung H, Han SH, Yoon MY, et al. Feasibility study for diagnosis of stomach adenoma and cancer using IR spectroscopy. Vibrational Spectroscopy. 2007;44:279-85.
11. Petibois C, Drogat B, Bikfalvi A, Déléris G, Moenner M. Histological mapping of biochemical changes in solid tumors by FT-IR spectral imaging. FEBS Lett. 2007;581:5469-74.
12. Amharref N, Beljebbar A, Dukic S, Venteo L, Schneider L, Pluot M, et al. Brain tissue characterisation by infrared imaging in a rat glioma model. Biochim Biophys Acta. 2006;1758:892-9.
13. Paluszkiewicz C, Kwiatek WM, Banaś A, Kisiel A, Marcelli A, Piccinini M. SR-FTIR spectroscopic preliminary findings of non-cancerous, cancerous, and hyperplastic human prostate tissues. Vibrational Spectroscopy. 2007;43:237-42.
14. Gazi E, Dwyer J, Lockyer NP, Miyan J, Gardner P, Hart CA, et al. A study of cytokinetic and motile prostate cancer cells using synchrotron-based FTIR microspectroscopic imaging. Vibrational Spectroscopy. 2005;38:193-201.
15. Liu C, Zhang Y, Yan X, Zhang X, Li C, Yang W, et al. Infrared absorption of human breast tissues in vitro. Journal of Luminescence. 2006;119-120:132-6.
16. Fabian H, Thi NA, Eiden M, Lasch P, Schmitt J, Naumann D. Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy. Biochim Biophys Acta. 2006;1758:874-82.
17. Salman A, Sahu RK, Bernshtain E, Zelig U, Goldstein J, Walfisch S, et al. Probing cell proliferation in the human colon using vibrational spectroscopy: a novel use of FTIR-microspectroscopy. Vibrational Spectroscopy. 2004;34(2):301-8.
18. Kondepati VR, Keese M, Heise HM, Backhaus J. Detection of structural disorders in pancreatic tumour DNA with Fourier-transform infrared spectroscopy. Vibrational Spectroscopy. 2006;40:33-9.
19. Boskey A, Pleshko Camacho N. FT-IR imaging of native and tissue-engineered bone and cartilage. Biomaterials. 2007;28:2465-78.
20. Shanes JG, Ghali J, Billingham ME, Ferrans VJ, Fenoglio JJ, Edwards WD, et al. Interobserver variability in the pathologic interpretation of endomyocardial biopsy results. Circulation. 1987;75(2):401-5.
21. Allsbrook WC Jr., Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. Hum Pathol. 2001;32(1):81-8.
22. Edwards WD. Current problems in establishing quantitative histopathologic criteria for the diagnosis of lymphocytic myocarditis by endomyocardial biopsy. Heart Vessels Suppl. 1985;1:138-42.
23. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. JAMA. 2015;313(11):1122.
24. Hoffman AW, Ibarra-Drendall C, Espina V, Liotta L, Seewaldt V. Ductal carcinoma in situ: challenges, opportunities, and uncharted waters. Am Soc Clin Oncol Educ Book. 2012:40-4.
25. Geller BM, Nelson HD, Carney PA, Weaver DL, Onega T, Allison KH, et al. Second opinion in breast pathology: policy, practice and perception. J Clin Pathol. 2014;67(11):955-60.
26. Allison KH, Reisch LM, Carney PA, Weaver DL, Schnitt SJ, O'Malley FP, et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology. 2014;65(2):240-51.
27. Frable WJ. Surgical pathology--second reviews, institutional reviews, audits, and correlations: what's out there? Error or diagnostic variation? Arch Pathol Lab Med. 2006;130(5):620-5.
28. Staradub VL, Messenger KA, Hao N, Wiely EL, Morrow M. Changes in breast cancer therapy because of pathology second opinions. Ann Surg Oncol. 2002;9(10):982-7.
29. Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging. 2005;5:17-9.
30. Fernandez DC, Bhargava R, Hewitt SM, Levin IW. Infrared spectroscopic imaging for histopathologic recognition. Nat Biotechnol. 2005;23(4):469-74.
31. Mayerich D, Walsh MJ, Kadjacsy-Balla A, Ray PS, Hewitt SM, Bhargava R. Stain-less staining for computed histopathology. Technology. 2015;3(1):27-31.
32. Bhargava R, Fernandez DC, Hewitt SM, Levin IW. High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data. Biochim Biophys Acta. 2006;1758(7):830-45.
33. Bhargava R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. Anal Bioanal Chem. 2007;389(4):1155-69.

34. Kwak JT, Kajdacsy-Balla A, Macias V, Walsh M, Sinha S, Bhargava R. Improving Prediction of Prostate Cancer Recurrence using Chemical Imaging. Sci Rep. 2015;5:8758.

35. Reddy RK, Bhargava R. Chemometric Methods for Biomedical Raman Spectroscopy and Imaging. In: Matousek P, Morris M, editors. Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields. Berlin, Heidelberg: Springer; 2010. pp. 179-213.

36. Krafft C, Steiner G, Beleites C, Salzer R. Disease recognition by infrared and Raman spectroscopy. J Biophotonics. 2009;2(1-2):13-28.

37. Wood BR, Chiriboga L, Yee H, Quinn MA, McNaughton D, Diem M. Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium. Gynecol Oncol. 2004;93:59-68.

38. Tfayli A, Piot O, Durlach A, Bernard P, Manfait M. Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy. Biochim Biophys Acta. 2005;1724:262-9.

39. Mazur AI, Monahan JL, Miljković M, Laver N, Diem M, Bird B. Vibrational spectroscopic changes of B-lymphocytes upon activation. J Biophotonics. 2013;6(1):101-9.

40. Walsh MJ, Singh MN, Pollock HM, Cooper LJ, German MJ, Stringfellow HF, et al. ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. Biochem Biophys Res Commun. 2007;352(1):213-9.

41. Lasch P, Diem M, Hänsch W, Naumann D. Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. J Chemom. 2007;20(5):209-20.

42. Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. Biochim Biophys Acta. 2004;1688(2):176-86.

43. Fujioka N, Morimoto Y, Arai T, Kikuchi M. Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy. Cancer Detect Prev. 2004;28:32-6.

44. Patel II, Trevisan J, Singh PB, Nicholson CM, Krishnan RK, Matanhelia SS, et al. Segregation of human prostate tissues classified high-risk (UK) versus low-risk (India) for adenocarcinoma using Fourier-transform infrared or Raman microspectroscopy coupled with discriminant analysis. Anal Bioanal Chem. 2011;401(3):969-82.

45. Krafft C, Shapoval L, Sobottka SB, Schackert G, Salzer R. Identification of primary tumors of brain metastases by infrared spectroscopic imaging and linear discriminant analysis. Technol Cancer Res Treat. 2006;5(3):291-8.

46. Argov S, Ramesh J, Salman A, Sinelnikov I, Goldstein J, Guterman H, et al. Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. J Biomed Opt. 2002;7(2):248-54.

47. Lascha P, Schmitt J, Naumann D. Colorectal Adenocarcinoma Diagnosis by FT-IR Microspectrometry. BiOS 2000 The International Symposium on Biomedical Optics. International Society for Optics and Photonics. 2000:45-56.

48. Lasch P, Haensch W, Lewis EN, Kidder LH, Naumann D. Characterization of Colorectal Adenocarcinoma Sections by Spatially Resolved FT-IR Microspectroscopy. Appl Spectrosc. 2002;56(1):1-9.

49. Li X, Li Q-B, Zhang G-J, Xu Y-Z, Sun X-J, Shi J-S, et al. Identification of colitis and cancer in colon biopsies by Fourier Transform Infrared spectroscopy and chemometrics. ScientificWorldJournal. 2012;2012:936149.

50. Hahn D. Calcium Fluoride and Barium Fluoride Crystals in Optics. Opt Photonik. 2014;9(4):45-8.

51. Weber MJ. Handbook of optical materials. Volume 19. CRC Press; 2002.

52. Bassan P, Lee J, Sachdeva A, Pissardini J, Dorling KM, Fletcher JS, et al. The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications. Analyst. 2013:144-57.

53. Filik J, Frogley MD, Pijanka JK, Wehbe K, Cinque G. Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. Analyst. 2012;137:853.

54. Bird B, Miljkovic M, Romeo MJ, Smith J, Stone N, George MW, et al. Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology. BMC Clin Pathol. 2008;8(1):8.

55. Reddy RK, Bhargava R. Accurate histopathology from low signal-to-noise ratio spectroscopic imaging data. Analyst. 2010;135(11):2818-25.

56. Griffiths PR, De Haseth JA. Fourier transform infrared spectrometry. Volume 171. John Wiley & Sons; 2007.

57. Bhargava R, Levin IW. Fourier transform infrared imaging: theory and practice. Anal Chem. 2001;73(21):5157-67.

58. Bhargava R, Wang S-Q, Koenig JL. Processing FT-IR Imaging Data for Morphology Visualization. Appl Spectrosc. 2000;54(11):1690-1706.

59. Lasch P, Naumann D. Spatial resolution in infrared microspectroscopic imaging of tissues. Biochim Biophys Acta. 2006;1758(7):814-29.

60. Sommer AJ, Tisinger LG, Marcott C, Story GM. Attenuated Total Internal Reflection Infrared Mapping Microspectroscopy Using an Imaging Microscope. Appl Spectrosc. 2001;55(3):252-6.

61. Patterson BM, Havrilla GJ, Marcott C, Story GM. Infrared microspectroscopic imaging using a large radius germanium internal reflection element and a focal plane array detector. Appl Spectrosc. 2007;61(11):1147-52.

62. Ippolito SB, Goldberg BB, Ünlü MS. High spatial resolution subsurface microscopy. Appl Phys Lett. 2001;78(26):4071.

63. Keith FN, Reddy RK, Bhargava R. Practical protocols for fast histopathology by Fourier transform infrared spectroscopic imaging. Biomedical Optical Spectroscopy. 2008;6853:1-10.

64. Snyder RG, Schachtschneider JH. Vibrational analysis of the n-paraffins—I. Spectrochim Acta. 1963;19(1):85-116.

65. Pallua JD, Pezzei C, Zelger B, Schaefer G, Bittner LK, Huck-Pezzei VA, et al. Fourier transform infrared imaging analysis in discrimination studies of squamous cell carcinoma. Analyst. 2012;137:3965-74.

66. Wolthuis R, Travo A, Nicolet C, Neuville A, Gaub MP, Guenot D, et al. IR spectral imaging for histopathological characterization of xenografted human colon carcinomas. Anal Chem. 2008;80(22):8461-9.

67. Beleites C, Neugebauer U, Bocklitz T, Krafft C, Popp J. Sample size planning for classification models. Anal Chim Acta. 2013;760:25-33.

68. van Dijk T, Mayerich D, Carney PS, Bhargava R. Recovery of Absorption Spectra from Fourier Transform Infrared (FT-IR) Microspectroscopic Measurements of Intact Spheres. Appl Spectrosc. 2013;67(5):546-52.

69. Kodali AK, Schulmerich M, Ip J, Yen G, Cunningham BT, Bhargava R. Narrowband Midinfrared Reflectance Filters Using Guided Mode Resonance. Anal Chem. 2010;82(13):5697-706.

70. Liu J-N, Schulmerich MV, Bhargava R, Cunningham BT. Optimally designed narrowband guided-mode resonance reflectance filters for mid-infrared spectroscopy. Opt Express. 2011;19(24):24182-97.

71. Yeh K, Kenkel S, Liu J, Bhargava R. Fast Infrared Chemical Imaging with a Quantum Cascade Laser. Anal Chem. 2015;87(1):485-93.

72 Kole MR, Reddy RK, Schulmerich MV, Gelber MK, Bhargava R. Discrete frequency infrared microspectroscopy and imaging with a tunable quantum cascade laser. Anal Chem. 2012;84(23):10366-72.

73. Bassan P, Weida MJ, Rowlette J, Gardner P. Large scale infrared imaging of tissue micro arrays (TMAs) using a tunable Quantum Cascade Laser (QCL) based microscope. Analyst. 2014;139(16):3856-9.

74. Kölhed M, Haberkorn M, Pustogov V, Mizaikoff B, Frank J, Karlberg B, et al. Assessment of quantum cascade lasers as

mid infrared light sources for measurements of aqueous solutions. Vibrational Spectroscopy. 2001;29(1-2):283-9.

75. Brandstetter M, Koch C, Genner A, Lendl B. Measures for optimizing pulsed EC-QC laser spectroscopy of liquids and application to multi-analyte blood analysis. SPIE OPTO. 2013;8993.

76. Kröger N, Egl A, Engel M, Gretz N, Haase K, Herpich I, et al. Quantum cascade laser-based hyperspectral imaging of biological tissue. J Biomed Opt. 2014;19(11):111607.

77. Diem M, Miljković M, Bird B, Chernenko T, Schubert J, Marcsisin E, et al. Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises. Spectroscopy. 2012;27(5-6):463-96.

78. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, et al. Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst. 2010;135:268-77.

79. Mohlenhoff B, Romeo M, Diem M, Wood BR. Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy. Biophys J. 2005;88(5):3635-40.

80. Bird B, Miljković M, Diem M. Two step resonant Mie scattering correction of infrared micro-spectral data: Human lymph node tissue. J Biophotonics. 2010;3(8):597-608.

81. Cserni G, Bianchi S, Boecker W, Decker T, Lacerda M, Rank F, et al. Improving the reproducibility of diagnosing micrometastases and isolated tumor cells. Cancer. 2005;103(2):358-67.

82. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat Med. 2002;8(11):1323-7.

83. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997;30(7):1145-59.

84. Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. Arch Pathol Lab Med. 1986;110(1):13-20.

85. Zweig M, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1992;39(4):561-77.

86. Warner E, Plewes DB, Hill KA, Causer PA, Zubovits JT, Jong RA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. JAMA. 2004;292(11):1317-25.

87. Bassan P, Sachdeva A, Shanks JH, Brown MD, Clarke NW, Gardner P. Automated high-throughput assessment of prostate biopsy tissue using infrared spectroscopic chemical imaging. SPIE Medical Imaging. 2014;9041.

88. Khanmohammadi M, Bagheri Garmarudi A, Samani S, Ghasemi K, Ashuri A. Application of linear discriminant analysis and Attenuated Total Reflectance Fourier Transform Infrared microspectroscopy for diagnosis of colon cancer. Pathol Oncol Res. 2011;17(2):435-41.

89. Petter CH, Heigl N, Rainer M, Bakry R, Pallua J, Bonn GK, et al. Development and application of Fourier-transform infrared chemical imaging of tumour in human tissue. Curr Med Chem. 2009;16:318-26.

90. Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, et al. Using Fourier transform IR spectroscopy to analyze biological materials. Nat Protoc. 2014;9(8):1771-91.

91. Martin FL, Kelly JG, Llabjani V, Martin-Hirsch PL, Patel II, Trevisan J, et al. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat Protoc. 2010;5:1748-60.