# Statistical Considerations and Tools to Improve Histopathologic Protocols with Spectroscopic Imaging

**Shachi Mittal**[1,2],

**Jonathan Kim**[3],

**Rohit Bhargava**[1,4,5]

[1]Department of Bioengineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, IL, USA

[2]Department of Chemical Engineering, University of Washington, Seattle, WA, USA

[3]Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Champaign, IL, USA

[4]Departments of Mechanical Science and Engineering, Electrical and Computer Engineering, Chemical and Biomolecular Engineering, and Chemistry, University of Illinois at Urbana–Champaign, Urbana, IL, USA

[5]Cancer Center at Illinois, University of Illinois at Urbana–Champaign, Urbana, IL, USA

## Abstract

Advances in infrared (IR) spectroscopic imaging instrumentation and data science now present unique opportunities for large validation studies of the concept of histopathology using spectral data. In this study, we examine the discrimination potential of IR metrics for different histologic classes to estimate the sample size needed for designing validation studies to achieve a given statistical power and statistical significance. Next, we present an automated annotation transfer tool that can allow large-scale training/validation, overcoming the limitations of sparse ground truth data with current manual approaches by providing a tool to transfer pathologist annotations from stained images to IR images across diagnostic categories. Finally, the results of a combination of supervised and unsupervised analysis provide a scheme to identify diagnostic groups/patterns and isolating pure chemical pixels for each class to better train complex histopathological models. Together, these methods provide essential tools to take advantage of the emerging capabilities to record and utilize large spectroscopic imaging datasets.

## Introduction

The histopathology of cells and tissue remains the gold standard for clinical diagnostics and a significant part of biomedical research. It currently involves a visual examination of stained tissue slides to recognize the structural changes in specific cell types that are indicative of disease. This process is time-consuming and often prone to interobserver variability. During the last 15 years, infrared (IR) spectroscopic imaging has been studied as a potential route to obtaining diagnostic markers of disease in a variety of conditions: for example, for detection of breast micrometastases in axillary lymph nodes using microspectral IR imaging in combination with hierarchical clustering and artificial neural networks,[1] characterization of breast tumor tissue by developing supervised and unsupervised classifiers,[2] correlating breast microcalcifications with malignancy,[3] detection of microscopic differences in xenografted tumors,[4] prostate cancer recurrence prediction model,[5] and tumor microenvironment characterization of the breast.[6] Digital histopathological characterization has also been proposed for a variety of tissue types, including liver,[7,8] lymph node analysis,[9,10] kidney,[11,12] lung adenocarcinoma,[13] brain tumors,[14–16] head and neck tumors,[17] oral cavity,[18] prostate,[19,20] bone,[21–23] colon,[24–28] and breast.[6,29–35] In general, for these varieties of tissue types and diseases, spectroscopic imaging is coupled to a variety of data mining approaches. Recent developments in the capabilities of analytical instruments designed for vibrational IR spectroscopy to resolve microscopic structures have drastically advanced the feasibility of IR spectroscopic imaging for routine histopathology.[6,28,35–39] These instruments also allow an unprecedented volume of data to be acquired and allow us to revisit analytical considerations in using spectroscopic imaging for histopathology. New advances in feature extraction and computational analysis have enabled the differentiation of intricate disease categories. These biological classes have subtle variations in the given feature space like molecular imaging data, genetics, proteomics, or other systems biology experiments. In addition to measurement and computational factors, statistical considerations in the potential implementation of spectroscopy-based approaches are sample size estimation, generation of an annotated cohort for training/validation, and data analysis; individually and together, these factors can have a large influence on the development and use of spectral histopathology methods. Underestimation of the sample size can lead to statistically deficient diagnostic tests, for example, while an overestimation can significantly increase experimental costs and time to develop protocols. Finally, the sample size for the desired classification task is critical to achieving statistical significance in the estimates of performance.

Another concern in the development of analytical protocols is the issue of gold standards. Typically, annotations within images are used to train machine learning algorithms and validate their analytical performance. The transfer of the annotated regions from standard clinical images to images of interest (IR images in this case) can further add to data

variability and time input. To accurately train systems for automated histopathology, accurate mapping of the desired regions of interest from current pathology gold standard (stained) images to the spectral image is necessary. Multimodal image registration techniques can be applied to map data from the histologic images to the IR spectral image data. They have been utilized to correlate multiple biomarkers in conventional staining data.[40] Previous studies have also reported the use of image registration approaches for aligning mass spectroscopic images with Raman imaging data[41] and optical images.[42] Image registration by applying affine transformations to correlate hematoxylin and eosin (H&E) stained images with infrared imaging data has also been demonstrated.[43] However, an accurate and user-friendly approach for transferring clinical annotations to chemical imaging data is still a challenge. Multimodal image registration has gained popularity and advanced in the area of medical imaging.[44–46] While some aspects between these methods and our work here are related, there are also significant differences that do not allow a simple translation of these methods to vibrational spectroscopic imaging data. One of the chief ones is simply the size of the data set. While in vivo medical imaging sets that require registration are less than a few megapixels, emerging instruments in IR imaging now promise gigapixels per sample. A second major difference arises from the subtle structural changes that rapidly occur over a few pixel distances in microscopy. When the extant methods, which focus primarily on determining similarity between two images for developing a registration cost function, are applied to our images, major drawbacks emerge. The fully automated approaches require iterating through all possible affine transformations to find the optimal transformation. Performing the search to find the optimal registration would require sweeping over every possible combination of $x$-translation, $y$-translation, and rotation and then calculating the value of the registration cost function at every possible transformation. Due to the significant size disparity between the H&E image and the Fourier transform IR (FT-IR) region of interest (ROI) that we seek to register, this is a very large search space. It would be computationally expensive for most personal/work computers, potentially requiring hours of computing time for a single registration. To optimize computation time and ease of use, we made the conscious choice to allow manual user control point selection over an automated iterative process. Also, our control point-based registration approach provides higher accuracy and validity.

Finally, even though unsupervised methods have been previously implemented on IR data,[31,47–49] a combined approach of supervised and unsupervised analysis for identifying disease subgroups and isolating pure chemical pixels has not been widely explored. Here, we present a comprehensive toolbox containing sample size estimation for IR imaging-based histopathology utilizing multivariate analysis of variance, a user interface for automated annotations transfer, and the use of unsupervised clustering for identifying diagnostic patterns in breast tissue along with selecting pure IR pixels for building robust supervised models. Although the annotations transfer toolbox has been implemented and tested here on breast cancer tissue, the methods can easily be extended to other tissue types and diseases. It will enable the user to design and execute extensive studies utilizing IR spectroscopic imaging by estimating the diagnostic effect size to be captured by digital models, providing labeled data for model development, and understanding inherent pathologic and chemical distributions in the dataset.

## Methods

### Sample Preparation and Data Collection

Formalin-fixed, paraffin-embedded serial breast tissue microarray sections (BR1003 and BRC961) were obtained from US Biomax, Inc. The array (BR1003) consisted of a total of 101 cores of 1 mm diameter from 47 patients, and BRC961 consisted of 96 cores of 1.5 mm diameter from 48 patients. A few surgical specimens were imaged by selecting areas of interest representing diverse histological categories to closely simulate clinical settings as illustrated in Fig. 1. Unstained sections at 5 μm thick were placed on $BaF_2$ salt plates for transmission IR imaging. Corresponding adjacent sections were stained with H&E and imaged with a light microscope for ground truth generation. The sections were deparaffinized using an overnight hexane bath, and high-definition IR imaging was conducted using the Agilent Stingray imaging system as previously described.[50] Data collected at the spectral resolution of 4 cm$^{-1}$ was truncated to 3800–900 cm$^{-1}$ wavenumbers and ratioed to a background single beam scan acquired on an empty space of $BaF_2$ slide. Each sample was imaged and preprocessed by raster scanning a ~140 × 140 μm tile (2 min scanning, 30 s processing) and stitching individual tiles in ENVI+IDL 4.8 as previously described.[50]

### Variance Analysis and Sample Size Estimation

The sample size is typically calculated by a priori power analysis[51–54] for a given statistical power, significance level, and effect size. Statistical power[55] is defined as the probability of detecting an effect (difference in the class groups) under a given confidence interval and sample size constraints. It highlights the achievable sensitivity and specificity of a diagnostic test. Effect size, also referred to as minimum expected difference, is the difference between group scores based on the feature set being used; this is typically calculated based on previously reported studies or pilot studies. It is a function of within-group and between-group feature variance. For general clinical tasks, previously reported power analysis and lookup tables could also be utilized.[52] It can also be estimated by carrying out an analysis of variance (ANOVA). Previously, ANOVA models for identifying primary sources of variance in spectroscopic data of different diagnostic categories have been demonstrated.[56] Typically, in ANOVA, the effect of the disease states on infrared absorption at a particular frequency or a single feature vector is captured. In contrast, in multivariate analysis of variance (MANOVA), the effect of diagnostic groupings on multiple IR absorbance/metrics is estimated as illustrated in Fig. 2, closely simulating the multiclass classification downstream.

We performed MANOVA in Matlab (The Mathworks, Inc.) on IR images using eight spectral features (listed in Table S1, Supplemental Material) handcrafted based on known biochemical significance for the desired classification tasks and multiple histological classes. We have used these previously identified features for our classification task, but any set of features (even identified by automated schemes) can be used in this pipeline. The set of eight features illustrate our approach. Since the goal in our case is to determine the number of samples needed for a given statistical power and confidence, this is referred to as a priori power analysis. However, we also need to compute the effect size as the exact

detectable differences in the IR signature for our classification task are not available in the literature. To account for multivariate associations, the effect size is calculated by estimating the Pillai Trace in MANOVA analysis.[57] It is a test statistic in MANOVA to compute the within-group versus between-group variance (defined in the equation 1). The total number of groups (or classes in the classification task) and the types of features in a classification task change the Pillai trace due to a different within-group versus between-group variance, thereby changing the corresponding effect size. Since we are trying to estimate the effect of different diagnostic groups on the IR absorbances and see whether they are significantly different or not, these IR absorbance features are referred to as response/outcome variables. Pillai trace is calculated in Matlab using the in-built library "manova1"

$$
\begin{aligned}
\text{Piilai's trace} &= \sum \frac{1}{1 + \lambda_k} \\
\lambda &= \text{eigenvalues of A} \\
\lambda &= SS_W^{-1} SS_B
\end{aligned}
\tag{1}
$$

Next, the sample size needed for a given statistical power for the estimated effect size is calculated using F-tests (MANOVA global effects in our case) using the G*Power software.[58] Since in a clinical setting, the statistical power of at least 0.8 is recommended, all the sample size calculations in this manuscript are reported at 0.8 power and 0.05 statistical significance. With decreasing effect size to be captured for a given classification task, the sample size required to achieve the same statistical power will increase. To allow the user to change the desired statistical power, we have also generated sample size versus statistical power curves for a given classification task (see the Results and Discussion section).

## Multimodal Image Registration and Automated Ground Truth Transfer

Our image registration approach involved utilizing control point registration. The user can precisely annotate coordinates (three or more points) in the FT-IR image and the H&E image, corresponding to the same spatial architecture. Three or more points are needed to calculate an affine transform, including the rotational change accurately. The user needs to know the approximate location covered by the FT-IR image region within the larger H&E image to zoom in on the H&E image to select the points.

The user can alternate between the H&E and the IR image to precisely select these points, starting from the H&E image. First, mark the point 1 by double-clicking on the desired location on the H&E image, then mark the corresponding point 1 on the IR image, followed by point 2 on H&E, and so on. The first step is to calculate the needed dilations in both x- and y-axes to transform the H&E image to the same scale as the FT-IR image. Let (x1f, y1f) be the coordinates for a given control point 1 in the FT-IR image, and (x1h, y1h) be the coordinates for the corresponding control point in the H&E image. Similarly, let (x2f, y2f) be the coordinates for a different control point 2 in the FT-IR image, and (x2h, y2h) be the coordinates for the corresponding control point in the H&E image. A dilation value for the x-axis can be calculated thusly: dx = (x2f − x1f)/(x2h − x1h). This value represents the scaling factor needed to transform the x-dimension of the H&E image to the same scale

as that of the FT-IR image. Using the same formula, a dilation value for the *y*-dimension can also be obtained. Since there are three possible pairs of *x*- and *y*-axis dilations, we average the dilations to reduce any possible bias or random error. After applying the *x*- and *y*-axis dilation, we obtain the *x*- and *y*-axis translations needed to translate the scaled H&E image to precisely overlap the FT-IR image. Using the same notation as above, we can see that the *x*-axis translation, for example, would be equal to $x_{1f} - x_{1h}*d_x$, where dx is the calculated dilation in the *x*-dimension. We apply the same logic to obtain the *y*-axis translation. Finally, to obtain the rotation, we calculate the angle created by any two control points and a horizontal line, both for the FT-IR image and the H&E image. Then, we calculate the difference between the two calculated angles and apply the inverse rotation to correct any rotational misalignments. These transformations are then used to transfer the pathologist annotations to the FT-IR image automatically. Sometimes, accurate transfer of these annotations can be difficult due to histologic processing variations as the H&E and IR image coming from different tissue sections. While it is reasonable to assume a large degree of correspondence between the structures in two images, the sizes and proportions of these structures can vary.

For example, a duct that spanned 50 μm in the FT-IR image may have narrowed in the intervening difference in depth between the FT-IR image and H&E image and thus would have spanned 30 μm in the H&E image. As a result of this, transferred annotations can sometimes land on an incorrect structure. To overcome this issue, we utilized a pretrained random forest (RF) classifier that separates FT-IR pixels into the stroma, epithelium, and others[50] using 36 features (listed in Table S2). The model was trained on 20 000 pixels per class and validated on 5000 pixels per class (coming from different patients) for external validation. A confusion matrix was generated to validate the model, with the true positive rate being 99% rate for epithelium, 93% for stroma, and 99% for the "others" category. The confusion matrix is provided in the Supplemental Material, Table S3.

The transferred annotations are compared with the corresponding pixel-level classifications to ensure that each annotation lands on a correct cell type. For example, if a part of the pathologist annotation of "benign epithelium" happened to land on a stromal section (as defined by the classifier), in that case, the algorithm automatically discards that part of the annotation. Next, by using both the classified image and transferred annotations, we applied a recursive method known as the breadth-first search to expand the transferred annotations to neighboring pixels for which there is an uninterrupted path of pixels of the same cell type. This approach allowed us to expand the number of pixels with confident labels. We applied this flooding only to the epithelial classes since it is difficult to obtain a large number of pixels for different diagnostic categories in the epithelium except for the invasive tumor class. The choice of the distance criteria for flooding is detailed in Fig. S4. The overall workflow described above is illustrated in Fig. 3. We built a graphical user interface (GUI) that allows the user to perform these operations without coding knowledge. Step-by-step instructions for using the GUI are provided in Section 2, Supplemental Material. We have also made the GUIs available as executable files both for Mac and Windows users.

### Unsupervised Analysis

In this section, we wanted to further investigate the different disease signatures in an unsupervised manner. Since the tumor originates in the epithelium, we cluster the IR image by masking only the epithelium using k-means clustering with both three and four cluster models. We have presented three- or four-class models to achieve basic diagnostic segmentation of in situ tumors, invasive tumors, and other epithelial/cellular components. Euclidean distance criteria are used, and all pixels are grouped to the nearest class until the number of pixels in each class changes by less than 5%. The three-class model (discussed in the previous section) was used to mask out epithelial regions selected for clustering. The clustered image is then compared against the ground truth H&E-stained image to identify diagnostic patterns and select pure spectral pixels in each diagnostic class for precise training of supervised models. Here, we have presented this approach for clustering the epithelium, but the same method can be used to cluster the stroma. In that case, the three-class model should be used to mask the stromal regions and the same clustering method can be applied. This is further discussed in the Results and Discussion section of the manuscript.

## Results and Discussion

### Multivariate Analysis of Variance and Sample Size Estimation

To capture interclass and intraclass variation while accounting for the effect of multiple IR metrics on disease classification, we utilized multivariate analysis of variance to project the data to a canonical variable space and calculate the between-group and within-group variance. Next, eigenvalue analysis is used to estimate the Pillai Trace that can be utilized to calculate an effect size for a given classification problem. In this section, we used an example set of problems to highlight the effect size; we selected four different breast tissue classification models using 8 IR spectral features (listed in Table S1). Figs. 4a–d illustrate the analysis curves as a function of desired statistical power for sample size requirements. Conventionally in clinical practice, the statistical power of 0.8 and a significance level of 0.05 are often used to estimate sample size. Hence, we list in Fig. 4e the sample size requirements for this statistical power and significance level. It is interesting to note that the sample size can significantly change with different histologic classes due to alterations in the within-group versus between-group variations. For instance, the smallest sample size is for the three-class (benign, malignant, and stroma) model as the IR spectral signature of the stroma is significantly different from the other two epithelial classes, thereby increasing the effective between-group variance. On the other hand, the "others'" class contains many cell types such as necrotic tissue, blood cells, and secretions in the three-class (benign, malignant, and others) model. This results in the within-group variance being higher than the benign, malignant, and stroma model, thereby decreasing the effect size to be captured. Similarly, benign and malignant cells may not have as significant a spectral difference as the other cell types. Hence, though a (benign and malignant) model is only two classes, the effect size is small, and the sample size required is larger. As expected, a complicated model (four-class) that contains multiple cell types and small effect sizes require a larger sample size as detailed in Fig. S1 (Supplemental Material).

**Multimodal Registration and Annotations Transfer**

While the sample size is often considered at the level of the number of specimens, a robust sampling of sufficient annotated pixels in each specimen is also required to serve as a ground truth. This section illustrates an automated tool for the transfer of pathological annotations directly on IR spectroscopic images. Typically, when the regions of interest are manually transferred from the stained images to IR images, they can be sparse due to a lack of one-to-one correspondence between the two images and often take a considerable amount of time.

Figure 5a shows the manually transferred images that a pathologist confirmed for different histological categories. Figure 5b illustrates the transferred regions by utilizing the presented automated ROI transfer tool for the same diagnostic categories. The tool correctly transfers the annotations and expands them by using tissue architecture constraints and the three-class spectral classifier and flooding as discussed in the methods section. This provides additional pixels, especially for pathological epithelial categories that are often limited by the number of pixels available for supervised classification. Digitally transferred annotations with no flooding are shown in Fig. 5c for additional comparisons.

While this automated tool can find great utility, we also caution that its use should be in the context of the disease. For example, we do not do additional expansion when invasive epithelium may be present in breast tumor samples as multiple disease states, like ductal carcinoma in situ (DCIS), can be interspersed with invasive tumors. We also apply a more stringent flooding distance constraint on the other epithelial classes in that case (detailed in Fig. S2). A strength of this tool is that it reduces the barriers to annotation and can be used multiple times to assure that correct annotation is maintained while maximizing coverage. This capability cannot be achieved rapidly using manual marking alone. Here, three representative images are shown, but the tool was tested on multiple tissue architectures and patient samples shown in Fig. S2. This tool has been developed and tested on regions of interest imaged from large surgical specimens but should be easily applicable to other types of biopsy (core or needle) samples. To apply this tool on samples from tissue microarrays or for heterogeneous tumors on surgical sections, it would be best to transfer the annotations on parts of the image separately before re-forming a large field of view as the registration transform from the H&E-stained image to the IR image would be different for different patients and may differ in different regions of a large sample. For best results, we choose the control points carefully (maximum correspondence) and add as detailed annotations on the H&E image as possible. The presented approach is semi-automated and provides more accurate registration than the possible fully automated schemes for IR images. Let us consider one possible approach to a fully automated image registration algorithm, a two-step process consisting of (i) automated detection of control points followed by (ii) the mathematical calculation of the resulting registration/transformation. In the fully automated approach, the second step of calculating the registration/transformation is the same as in the manual approach, so we will disregard that step when comparing accuracy and validity. Considering the first step of selecting control points, if we assume some measure of domain knowledge on the user's part, manual control point selection will result in greater or equal accuracy in selecting control points compared to automatic control point selection. It would

also be problematic when there are small morphological differences between the FT-IR and H&E images due to processing or when a series of images may be used for registration (e.g., histochemical images). Indeed, our experiments with automatic control point detection yielded low accuracy, especially given the differences in images yielded by H&E and FT-IR. As such, we can generally state that our approach will yield equivalent or better accuracy than a fully automated approach.

Let us now consider a second possible approach to a fully automated registration algorithm, which would be to perform a stepwise search over all possible transformations, calculating a similarity function for each possible transformation, and then selecting the transformation that results in the highest similarity. The registration precision may be constrained by the step size used during the exploration of all possible transformations. In other words, if the optimal registration happens to require parameters not explicitly covered by the discrete search space, the resulting registration will be suboptimal. In contrast, our approach would yield an equal or higher "resolution of registration", allowing for a more optimal registration.

Finally, our approach ensures the validity of the transferred annotations. In our approach, the registration calculated from manual control point selection is supported and "double-checked" by the supervised cell type classifier; if an annotation happens to fall on an inappropriate tissue type determined by the classifier, it is discarded. This helps ensure the validity of the transferred annotations. As a final consideration, manual control point selection and registration allow for the earlier identification of intractable cases. Some of the H&E–FT-IR image pairs we attempted to register in the course of this work were so visually different that an automated registration approach would have a high chance of failure. Our manual approach allows for early identification and removal of these cases. This tool will allow rapid scaling of digital models based on spectroscopy and decrease interobserver variability in generating ground truth data. Currently, we have incorporated benign epithelium, DCIS, invasive tumor, dense stroma, reactive stroma, and others with defined color codes (listed in Section 4, Supplemental Material), but any additional classes can be easily added to the tool. The supplementary data includes the first version of this toolbox, and we anticipate that straightforward further development by us and others can make this widely usable for different problems.

## Clustering for Selecting Pure Chemical Pixels and Identifying Distinct Histological Patterns

Illustrated here is how a combination of supervised and unsupervised learning allows identifying distinct diagnostic patterns and selecting pure spectral pixels for each class. It can be seen in Figs. 6c and 6d that invasive tumors get separated from early-stage disease by clustering the pixels based on IR absorbance at specific wavenumber locations (1079, 1238, 1307, 1396, 1450, 1538, 1650, 2865, 2942, 3062, and 3286 cm$^{-1}$) used as a bar code for recognition of disease. In this case, recorded absorbances at these wavenumbers are used as features for the unsupervised clustering. It is evident that there are other distinct clusters (blue cluster in Fig. 6c and in the blue and yellow clusters shown in Fig. 6d that are primarily lymphocytes or scattered nuclear entities classified as epithelium by the simple three-class model. These nuclear architectures have IR spectral signatures closer to

epithelium than stroma or "others" class. It can be observed that within the invasive and DCIS class, some pixels are overlapping with each other, which can be removed from the supervised analysis of complex histologic models downstream. The presented approach will isolate pure chemical pixels within each class and help tackle the subcellular variations for precise training of complex histopathological models. The approach applies to large and small samples and can be used to discover the structure and more detailed annotations. It can also be used to explain heterogeneity in classification or sources of error in the image that might have a correlated spectral profile. While image segmentation methods typically rely on supervised or unsupervised techniques, this mixed method can provide a powerful tool for exploration as well.

## Conclusion

This study provides a systematic approach for extending IR spectroscopy-based digital models to large-scale studies by overcoming three challenges that are frequently encountered by practitioners. Precise sample size calculation will enable planning clinical studies and achieving the desired confidence in the diagnostic assessment. Even though we demonstrate effect size calculation from MANOVA, it can also be used as a canonical space to transform the IR feature space for better class differentiation or dimensionality reduction like the principal component analysis. The automated annotations transfer tool will allow rapid and consistent training and validation data for large-scale spectroscopic imaging studies. Modern IR imaging instruments are now allowing for rapid data acquisition. In studies involving large numbers of images, however, hand drawing of annotations on IR images using H&E annotations as a reference can be time-consuming and limit the yield and quality of ground truth pixels. This work tackles the broad problem of generating curated and annotated datasets in spectroscopy, allowing large-scale training and evaluation of spectroscopy-based digital histopathological models.

However, other registration approaches and further optimizations can be carried out to achieve desired efficacy for specific tasks. The existing methods and the approach presented in this paper can be extended for more sophisticated image registration approaches for spectroscopy. For example, one direction could include a preliminary step of establishing a relatively coarse registration through the manual control point tool and then having a subsequent step of iterative perturbation and registration for fine-tuning, according to one of the registration cost functions developed in the existing approaches. This would significantly reduce the search space and enable a fast, automated tuning step. Another possible future direction would be to develop this tool as a web application with a higher computational power in the backend. In this scenario, a powerful computational backend could be used to tackle the computationally expensive search problem, allowing for an automatic solution in a reduced amount of time, accessible to and usable by all researchers.

Finally, we present a combination of supervised and unsupervised analysis to identify distinct spectral signatures and isolate pure chemical pixels for each class in diagnostic images. It can be used both for an exploratory study focused on histopathology or to further segment the annotations for precise ground truth generation to further train complex histological models.[58]

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding

## References

1. Bird B, Bedrossian K, Laver N, et al. "Detection of Breast Micro-Metastases in Axillary Lymph Nodes by Infrared Micro-Spectral Imaging". Analyst. 2009. 134(6): 1067–1076. doi:10.1039/B821166c. [PubMed: 19475131]

2. Benard A, Desmedt C, Smolina M, et al. "Infrared Imaging in Breast Cancer: Automated Tissue Component Recognition and Spectral Characterization of Breast Cancer Cells as Well as the Tumor Microenvironment". Analyst. 2014. 139(5): 1044–1056. doi:10.1039/C3AN0145A. [PubMed: 24418921]

3. Baker R, Rogers KD, Shepherd N, Stone N. "New Relationships Between Breast Microcalcifications and Cancer". Br. J. Cancer 2010. 103(7): 1034–1039. doi:10.1038/Sj.Bjc.6605873. [PubMed: 20842116]

4. Camp CH Jr, Lee YJ, Heddleston JM, et al. "High-Speed Coherent Raman Fingerprint Imaging of Biological Tissues". Nat. Photonics 2014. 8(8): 627–634. doi:10.1038/Nphoton.2014.145. [PubMed: 25621002]

5. Kwak JT, Kajdacsy-Balla A, Macias V, et al. "Improving Prediction of Prostate Cancer Recurrence Using Chemical Imaging". Sci. Rep 2015. 5(8758): 1–10. doi:10.1038/Srep08758.

6. Mittal S, Yeh K, Leslie LS, et al. "Simultaneous Cancer and Tumor Microenvironment Subtyping Using Confocal Infrared Microscopy for All-Digital Molecular Histopathology". Proc. Natl. Acad. Sci. USA 2018. 115(25): E5651–E5660. doi:10.1073/Pnas.1719551115. [PubMed: 29866827]

7. Stenlund H, Gorzsás A, Persson P, et al. "Orthogonal Projections to Latent Structures Discriminant Analysis Modeling on in Situ FT-IR Spectral Imaging of Liver Tissue for Identifying Sources of Variability". Anal. Chem 2008. 80(18): 6898–6906. doi:10.1031/Ac8005318. [PubMed: 18714965]

8. Coe JV, Chen Z, Li R, et al. "Imaging Infrared Spectroscopy for Fixation-Free Liver Tumor Detection". In: Farkas DL, Nicolau DV, Leif RC, Editors. Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XII. 2014. P. 89470B.

9. Tian P, Zhang W, Zhao H, et al. "Intraoperative Detection of Sentinel Lymph Node Metastases in Breast Carcinoma by Fourier Transform Infrared Spectroscopy". Br. J. Surg 2015. 102(11): 1372–1379. doi:10.1002/Bjs.9882. [PubMed: 26198697]

10. Leslie LS, Wrobel TP, Mayerich D, et al. "High Definition Infrared Spectroscopic Imaging for Lymph Node Histopathology". PLoS One. 2015. 10(6): E0127238. doi:10.1371/journal.pone.0127238. [PubMed: 26039216]

11. Varma VK, Kajdacsy-Balla A, Akkina SK, et al. "A Label-Free Approach by Infrared Spectroscopic Imaging for Interrogating the Biochemistry of Diabetic Nephropathy Progression". Kidney Int. 2016. 89(5): 1153–1159. doi:10.1016/J.Kint.2015.11.027. [PubMed: 26924056]

12. Sreedhar H, Varma VK, Nguyen PL, et al. "High-Definition Fourier Transform Infrared (FT-IR) Spectroscopic Imaging of Human Tissue Sections Towards Improving Pathology". J. Visualized. Exp 2015. 95: e52332. doi:10.3791/52332.

13. Petibois C, Drogat B, Bikfalvi A, et al. "Histological Mapping of Biochemical Changes in Solid Tumors by FT-IR Spectral Imaging". FEBS Lett. 2007. 581(28): 5469–5474. doi:10.1016/J.Febslet.2007.10.052. [PubMed: 17983600]

14. Heraud P, Caine S, Campanale N, et al. "Early Detection of the Chemical Changes Occurring During the Induction and Prevention of Autoimmune-Mediated Demyelination Detected by

FT-IR Imaging". NeuroImage. 2010. 49(2): 1180–1189. doi:10.1016/J.Neuroimage.2009.09.053. [PubMed: 19796690]

15. Wehbe K, Pinneau R, Moenner M, et al. "FT-IR Spectral Imaging of Blood Vessels Reveals Protein Secondary Structure Deviations Induced by Tumor Growth". Anal. Bioanal. Chem 2008. 392(1–2): 129–135. doi:10.1007/S00216-008-2230-6. [PubMed: 18636246]

16. Krafft C, Sobottka SB, Geiger KD, et al. "Classification of Malignant Gliomas by Infrared Spectroscopic Imaging and Linear Discriminant Analysis". Anal. Bioanal. Chem 2007. 387(5): 1669–1677. doi:10.1007/S00216-006-0892-5. [PubMed: 17103151]

17. Bruni P, Conti C, Giorgini E, et al. "Histological and Microscopy FT-IR Imaging Study on the Proliferative Activity and Angiogenesis in Head and Neck Tumours". Faraday Discuss. 2004. 126(1): 19–26. doi:10.1039/B306787b. [PubMed: 14992397]

18. Conti C, Giorgini E, Pieramici T, et al. "FT-IR Microscopy Imaging on Oral Cavity Tumours, II". J. Mol. Struct 2005. 744–747(1): 187–193. doi:10.1016/J.Molstruc.2004.10.042.

19. Wrobel TP, Kwak JT, Kadjacsy-Balla A, Bhargava R. "High-Definition Fourier Transform Infrared Spectroscopic Imaging of Prostate Tissue". In: Progress in Biomedical Optics and Imaging -Proceedings of SPIE 9791. SPIE Medical Imaging, 2016, San Diego, California, United States. Pp. 4–9. 97911D. doi:10.1117/12.2217341.

20. 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): 3856–3859. doi:10.1039/C4an00638k. [PubMed: 24965124]

21. Chaber R, Łach K, Depciuch J, et al. "Fourier Transform Infrared (FT-IR) Spectroscopy of Paraffin and Deparafinnized Bone Tissue Samples as a Diagnostic Tool for Ewing Sarcoma of Bones". Infrared Phys. Technol 2017. 85: 364–371. doi:10.1016/J.Infrared.2017.07.017.

22. Bozkurt O, Bilgin MD, Evis Z, et al. "Early Alterations in Bone Characteristics of Type I Diabetic Rat Femur: A Fourier Transform Infrared (FT-IR) Imaging Study". Appl. Spectrosc 2016. 70(12): 2005–2015. doi:10.1177/0003702816671059. [PubMed: 27680083]

23. Boskey A, Pleshko Camacho N. "FT-IR Imaging of Native and Tissue-Engineered Bone and Cartilage". Biomaterials. 2007. 28(15): 2465–2478. doi:10.1016/J.Biomaterials.2006.11.043. [PubMed: 17175021]

24. Wolthuis R, Travo A, Nicolet C, et al. "IR Spectral Imaging for Histopathological Characterization of Xenografted Human Colon Carcinomas". Anal. Chem 2008. 80(22): 8461–8469. doi:10.1021/Ac801191x. [PubMed: 18847281]

25. Nallala J, Lloyd GR, Shepherd N, Stone N. "High-Resolution FT-IR Imaging of Colon Tissues for Elucidation of Individual Cellular and Histopathological Features". Analyst. 2016. 141(2): 630–639. doi:10.1039/C5AN01871D. [PubMed: 26549223]

26. Travo A, Piot O, Wolthuis R, et al. "IR Spectral Imaging of Secreted Mucus: A Promising New Tool for the Histopathological Recognition of Human Colonic Adenocarcinomas". Histopathology. 2010. 56(7): 921–931. doi:10.1111/J.1365-2559.2010.03563.X. [PubMed: 20500531]

27. Bird B, Rowlette J. "High Definition Infrared Chemical Imaging of Colorectal Tissue Using a Spero QCL Microscope". Analyst. 2017. 142(8): 1381–1386. doi:10.1039/C6AN01916A. [PubMed: 28098273]

28. Kuepper C, Kallenbach-Thieltges A, Juette H, et al. "Quantum Cascade Laser-Based Infrared Microscopy for Label-Free and Automated Cancer Classification in Tissue Sections". Sci. Rep 2018. 8(1): 1–10. doi:10.1038/S41598-018-26098-W. [PubMed: 29311619]

29. Fabian H, Lasch P, Boese M, Haensch W. "Infrared Microspectroscopic Imaging of Benign Breast Tumor Tissue Sections". J. Mol. Struct 2003. 661–662(1–3): 411–417. doi:10.1016/J.Molstruc.2003.07.002.

30. Mayerich DM, Walsh M, Kadjacsy-Balla A, et al. "Breast Histopathology Using Random Decision Forests-Based Classification of Infrared Spectroscopic Imaging Data". In: Proc. SPIE 9041, Medical Imaging Digital Pathology. SPIE Medical Imaging, 2014, San Diego, California, United States 904107. doi:10.1117/12.2043783.

31. Verdonck M, Denayer A, Delvaux B, et al. "Characterization of Human Breast Cancer Tissues by Infrared Imaging". Analyst. 2016. 141(2): 606–619. doi:10.1039/C5an01512j. [PubMed: 26535413]

32. Pounder FN, Reddy RK, Bhargava R. "Development of a Practical Spatial-Spectral Analysis Protocol for Breast Histopathology Using Fourier Transform Infrared Spectroscopic Imaging". Faraday Discuss. 2016. 187: 43–68. doi:10.1039/C5FD00199D. [PubMed: 27095431]

33. Wu LA, Kuo WH, Chen CY, et al. "The Association of Infrared Imaging Findings of the Breast with Prognosis in Breast Cancer Patients: An Observational Cohort Study". BMC Cancer. 2016. 16(1): 1–7. doi:10.1186/S12885-016-2602-9.

34. Mittal S, Wrobel TP, Leslie LS, et al. "A Four Class Model for Digital Breast Histopathology Using High- Definition Fourier Transform Infrared (FT-IR) Spectroscopic Imaging". In: Proc. SPIE 9791. Medical Imaging 2016: Digital Pathology. SPIE Medical Imaging, 2016, San Diego, California, United States 979118. doi:10.1117/12.2217358.

35. Pilling MJ, Henderson A, Gardner P. "Quantum Cascade Laser Spectral Histopathology: Breast Cancer Diagnostics Using High Throughput Chemical Imaging". Anal. Chem 2017. 89(14): 7348–7355. doi:10.1021/Acs.Analchem.7b00426. [PubMed: 28628331]

36. Yeh K, Kenkel S, Liu J, Bhargava R. "Fast Infrared Chemical Imaging with a Quantum Cascade Laser". Anal. Chem 2015. 87(1): 485–493. doi:10.1021/Ac5027513. [PubMed: 25474546]

37. Kole MR, Reddy RK, Schulmerich MV, et al. "Discrete Frequency Infrared Microspectroscopy and Imaging with a Tunable Quantum Cascade Laser". Anal. Chem 2012. 84(23): 10366–10372. doi:10.1021/Ac302513f. [PubMed: 23113653]

38. Bird B, Rowlette J. "A Protocol for Rapid, Label-Free Histo-chemical Imaging of Fibrotic Liver". Analyst. 2017. 142(8): 1179–1184. doi:10.1039/C6AN02080A. [PubMed: 27858020]

39. Pahlow S, Weber K, Popp J, et al. "Application of Vibrational Spectroscopy and Imaging to Point-of-Care Medicine: A Review". Appl. Spectrosc 2018. 72(Suppl. 1): 52–84. doi:10.1177/0003702818791939. [PubMed: 30265133]

40. Lippolis G, Edsjö A, Helczynski L, et al. "Automatic Registration of Multi-Modal Microscopy Images for Integrative Analysis of Prostate Tissue Sections". BMC Cancer. 2013. 13(1): 408. doi:10.1186/1471-2407-13-408. [PubMed: 24010502]

41. Bocklitz T, Bräutigam K, Urbanek A, et al. "Novel Workflow for Combining Raman Spectroscopy and MALDI-MSI for Tissue Based Studies". Anal. Bioanal. Chem 2015. 407(26): 7865–7873. doi:10.1007/S00216-015-8987-5. [PubMed: 26374565]

42. Gregory Schaaff T, McMahon JM, Todd PJ. "Semiautomated Analytical Image Correlation". Anal. Chem 2002. 74(17): 4361–4369. doi:10.1021/Ac025693b. [PubMed: 12236343]

43. Kwak JT, Hewitt SM, Sinha S, Bhargava R. "Multimodal Microscopy for Automated Histologic Analysis of Prostate Cancer". BMC Cancer. 2011. 11(1): 62. doi:10.1186/1471-2407-11-62. [PubMed: 21303560]

44. Aganj I, Fischl B. "Multimodal Image Registration Through Simultaneous Segmentation". IEEE Signal. 2017. 24(11): 1661–1665. doi:10.1109/LSP.2017.2754263.

45. Yang C, Niedieker D, Grosserüschkamp F, et al. "Fully Automated Registration of Vibrational Microspectroscopic Images in Histologically Stained Tissue Sections". BMC Bioinf. 2015. 16(1): 1–14 doi:10.1186/S12859-015-0804-9.

46. Goubran M, Leuze C, Hsueh B, et al. "Multimodal Image Registration and Connectivity Analysis for Integration of Connectomic Data from Microscopy to MRI". Nat. Commun 2019. 10(1): 1–17. doi:10.1038/S41467-019-13374-0. [PubMed: 30602773]

47. Augustyniak K, Chrabaszcz K, Jasztal A, et al. "High and Ultra-High Definition of Infrared Spectral Histopathology Gives an Insight Into Chemical Environment of Lung Metastases in Breast Cancer". J. Biophotonics 2019. 12(4): 1–8. doi:10.1002/Jbio.201800345.

48. Baker MJ, Trevisan J, Bassan P, et al. "Using Fourier Transform IR Spectroscopy to Analyze Biological Materials". Nat. Protoc 2014. 9(8): 1771–1791. doi:10.1038/Nprot.2014.110. [PubMed: 24992094]

49. Bird B, Miljkovic M, Romeo MJ, et al. "Infrared Micro-Spectral Imaging: Distinction of Tissue Types in Axillary Lymph Node Histology". BMC Clin. Pathol 2008. 8: 8. doi:10.1186/1472-6890-8-8. [PubMed: 18759967]

50. Mittal S, Wrobel TP, Walsh M, et al. "Breast Cancer Histopathology Using Infrared Spectroscopic Imaging: the Impact of Instrumental Configurations". Clin. Spectrosc 2021. 3(1): 100006. doi:10.1016/J.Clispe.2021.100006.

51. Sakpal TV. "Sample Size Estimation in Clinical Trial". Perspect. Clin. Res 2010. 1(2): 67–69. [PubMed: 21829786]

52. Flahault A, Cadilhac M, Thomas G. "Sample Size Calculation Should Be Performed for Design Accuracy in Diagnostic Test Studies". J. Clin. Epidemiol 2005. 58(8): 859–862. doi:10.1016/J.Clinepi.2004.12.009. [PubMed: 16018921]

53. Eng J. "Sample Size Estimation: How Many Individuals Should Be Studied?" Radiology. 2003. 227(2): 309–313. doi: 10.1148/Radiol.2272012051. [PubMed: 12732691]

54. Dell RB, Holleran S, Ramakrishnan R. "Sample Size Determination". ILAR J. 2002. 43(4): 207–213. doi:10.1092/Ilar.43.4.207. [PubMed: 12391396]

55. Jones SR. "An Introduction to Power and Sample Size Estimation". Emergency Medicine J. 2003. 20(5): 453–458.

56. Kwak JT, Reddy R, Sinha S, Bhargava R. "Analysis of Variance in Spectroscopic Imaging Data from Human Tissues". Anal. Chem 2012. 84(2): 1063–1069. doi:10.1021/Ac2026496. [PubMed: 22148458]

57. Steyn HS Jr., Ellis SM. "Estimating an Effect Size in One-Way Multivariate Analysis of Variance (MANOVA)". Multi. Behavio. Res 2009. 44(1): 106–129. doi:10.1080/00273170802620238.

58. Faul F, Erdfelder E, Lang AG, Buchner A. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences". Behav. Res. Methods 2007. 39(2): 175–191. doi:10.3758/Bf03193146. [PubMed: 17695343]
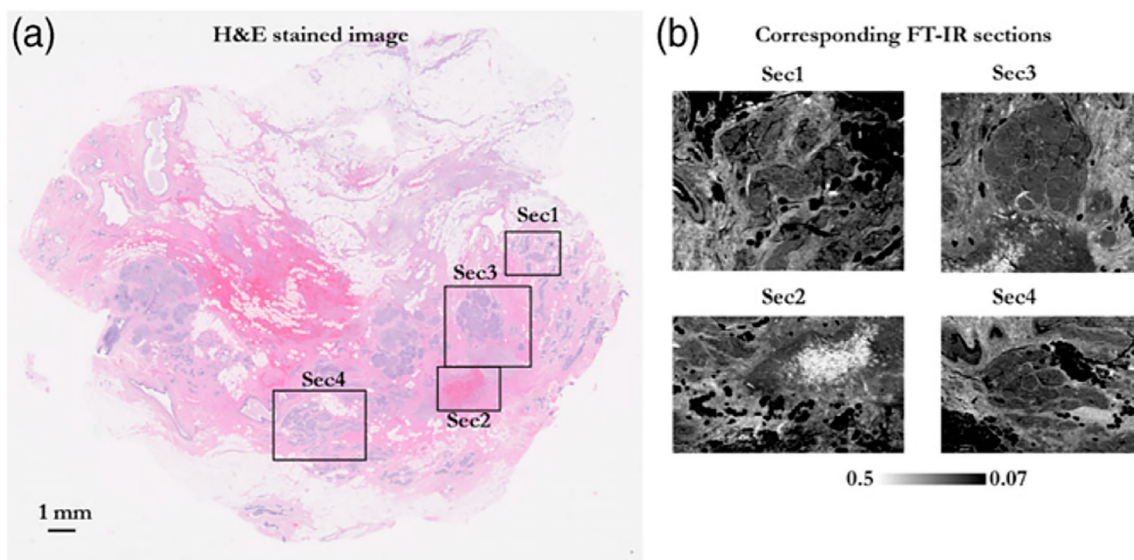
**Figure 1.**
Selected regions of interest from a surgical specimen imaged using high-definition IR imaging. (a) H&E-stained image of the surgical specimen with the marked areas imaged using IR imaging. (b) Corresponding IR images of the selected regions.
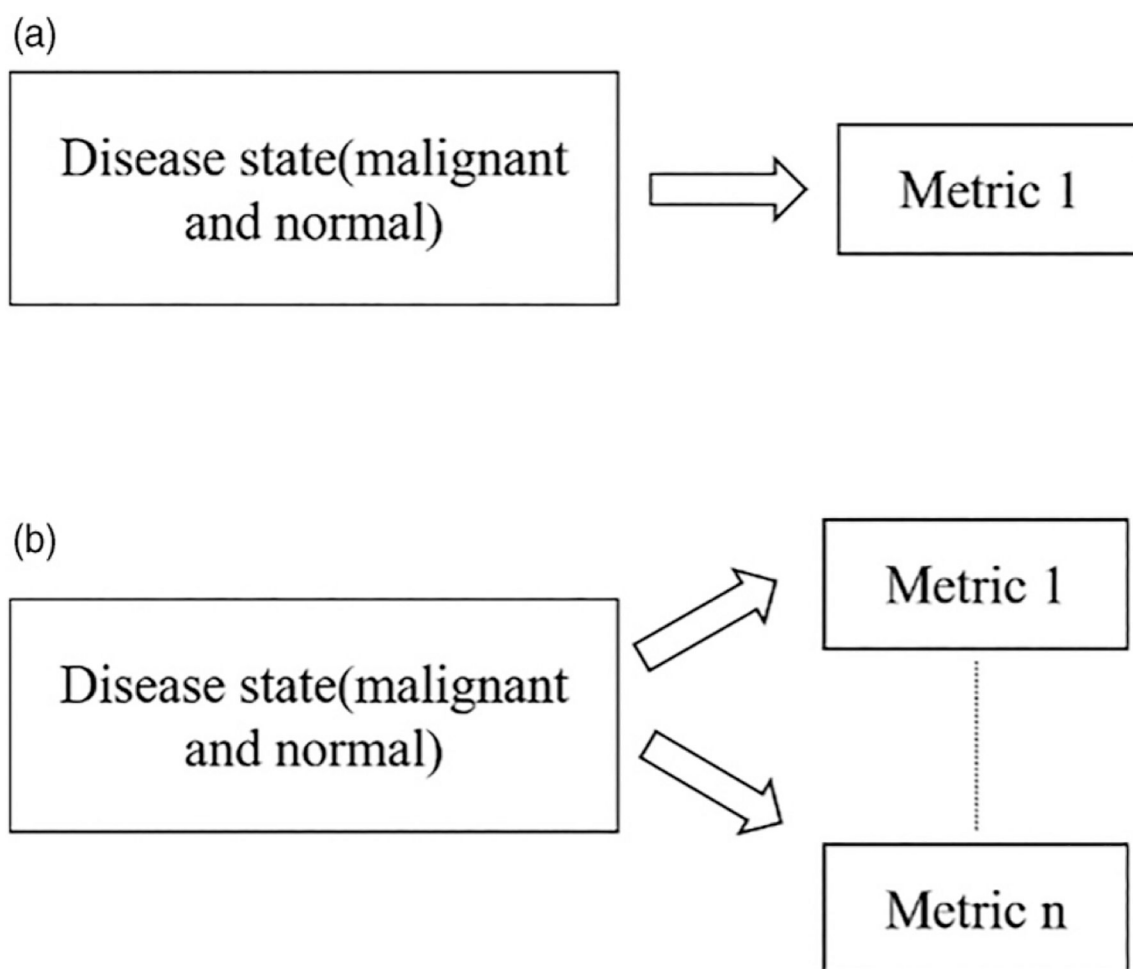
**Figure 2.**
(a) Estimating the effect of diagnostic classes using ANOVA (only one metric is considered).
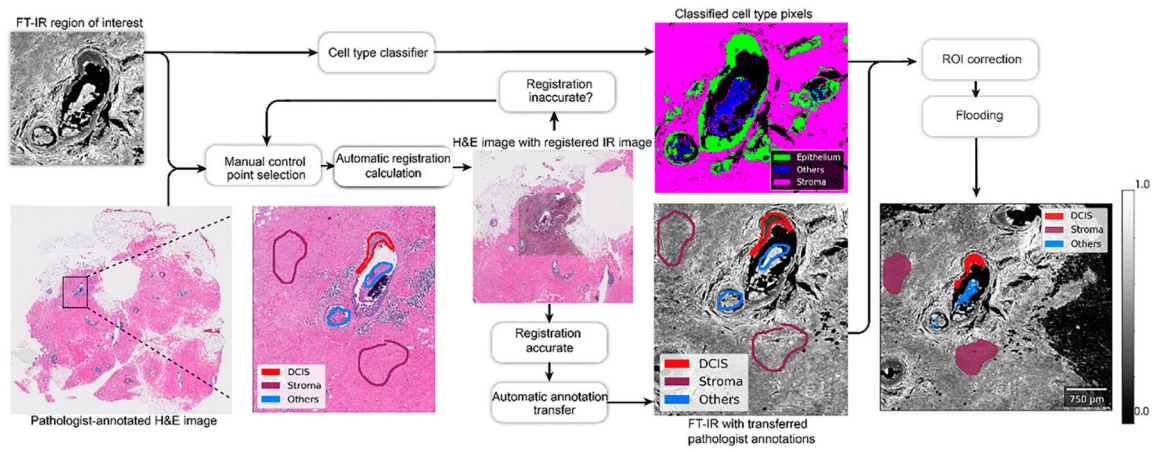(b) Effect size calculation using MANOVA where the interactions between multiple feature vectors can be incorporated.
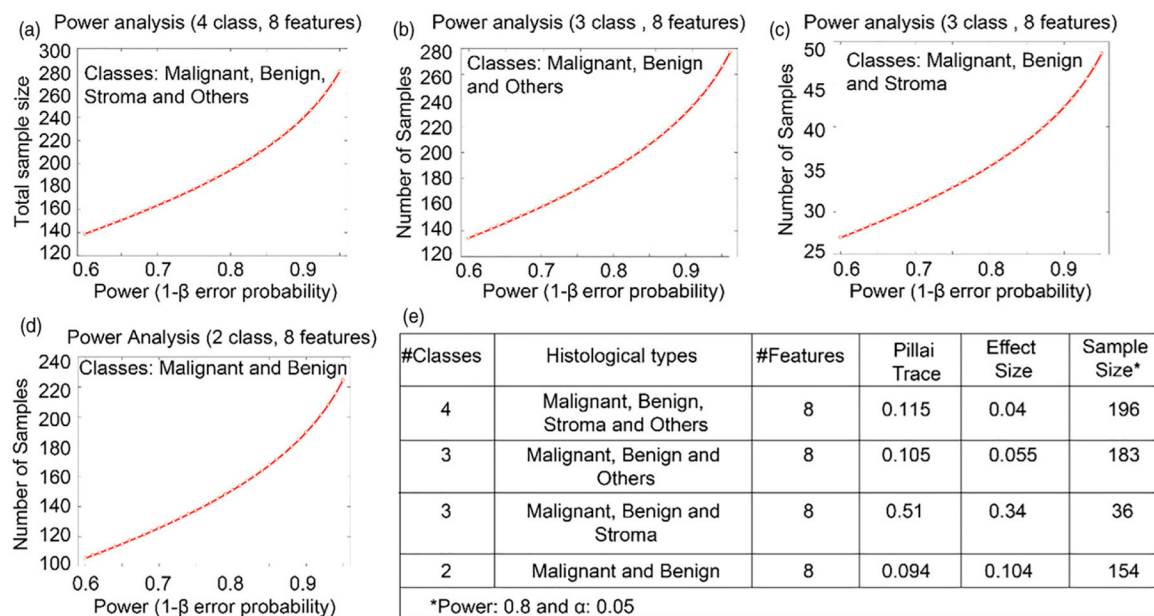
**Figure 3.**
Overall workflow for automated annotations transfers from H&E images to the corresponding IR images. The user needs to select points that spatially match the two images. Then, control point registration and tissue architecture feedback from the classifier is utilized for annotations transfer.

**Figure 4.**
Statistical power analysis using MANOVA global effects. (a) Number of samples required for the desired statistical power for a four-class model (malignant, benign, stroma, and others). (b) Sample size calculation as a function of power for a three-class model (malignant, benign, and others). (c) Sample size calculation for three-class stromal model (malignant, benign, and stroma). (d) Sample size calculation for two-class model (malignant and benign). (e) Effect size and sample size for 0.8 statistical power and significance level of 0.05.
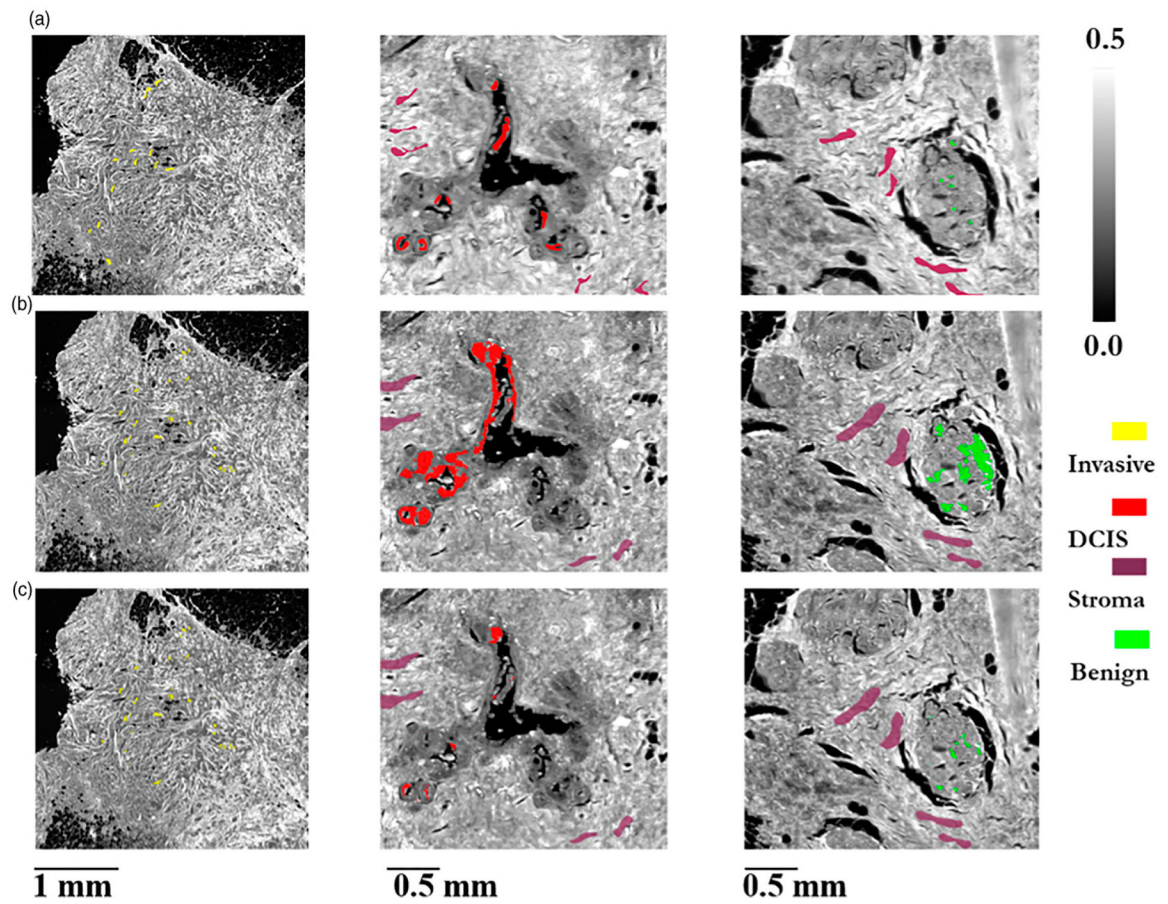
**Figure 5.**
Digital transfer of annotations from the H&E image to the corresponding IR image. We compare both the manually transferred annotations confirmed by a pathologist in (a) with the digitally annotated images in (b, with flooding) and in (c, without flooding/expansion).
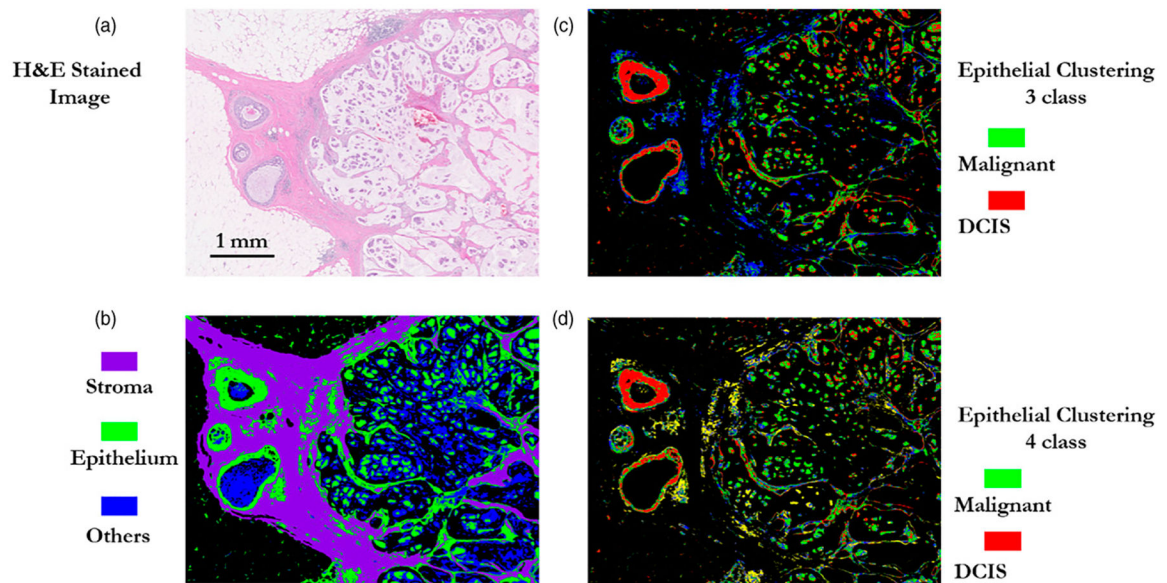
**Figure 6.**
K-means clustering of the epithelial component of a region of interest from a surgical specimen imaged using IR. (a) H&E-stained image of the selected area of a surgical specimen. (b) Digitally classified image using IR imaging and the three-class model random forest model described in the methods section separating epithelium, stroma, and others. (c) K-means clustered image of the epithelial component using three clusters. (d) K-means clustered image of the epithelial component using four clusters.