Chapter 8
Raman Spectroscopy and Advanced Statistics for Cancer Diagnostics

Nicole M. Ralbovsky and Igor K. Lednev

8.1 Introduction

Whether it be indirectly or directly, cancer has a vast impact on the majority of society. In 2018 alone, there was approximately 18.1 million newly diagnosed cases of cancer and an estimated 9.6 million deaths due to cancer, worldwide [1]. Cancer is the second-leading cause of death in the USA, where national expenses for care totaled $147.3 billion in 2017 [2].

Diagnosing the various forms of cancer often requires a myriad of methods. The main diagnostic approaches involve lab tests for blood, urine, and other body fluid samples; imaging tests, such as a CT scan, PET scan, X-ray, ultrasound, or MRI scan; and biopsy with either a needle, endoscope, or via surgery [2]. Detailed information regarding diagnosing cancers can be found elsewhere [3]—however, it is important to note that many cases of cancer can go undiagnosed or are misdiagnosed, and in some cases healthy individuals are even falsely diagnosed. Further, the time of diagnosis plays a crucial role in the survival rate of the afflicted. Early detection and diagnosis of cancer typically improves an individual's prognosis and increases the chances for successful treatment by allowing for care to be administered at the earliest opportunity. While early diagnostics and screening methods do exist, not all results are definitive or accurate. Even more, certain exams are invasive, expensive, and not accessible to all who require them. Financial burdens, as well as geographic and sociocultural barriers, prevent large groups of people from

This chapter is adapted from Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 219, N.M. Ralbovsky, I.K. Lednev, Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer, 463–487, Copyright (2019), with permission from Elsevier.

N. M. Ralbovsky · I. K. Lednev (✉)
Department of Chemistry, University at Albany, State University of New York, Albany, NY, USA
e-mail: ilednev@albany.edu

V. V. Tuchin et al. (eds.), *Multimodal Optical Diagnostics of Cancer*,
https://doi.org/10.1007/978-3-030-44594-2_8

seeking proper diagnostic opportunities. Another pressing issue, which is of further interest here, exists—there is no singular universal method that can accurately diagnose all forms of cancer early on. As such, there is a crucial unmet need for developing the first universal method for the non-invasive, inexpensive, and accurate diagnosis of all cancers which can be made accessible to all individuals who require testing. This chapter will discuss how the combination of Raman spectroscopy (RS) and advanced statistical analysis (or, chemometrics) has emerged with a strong potential to solve this imposing issue.

RS is advantageous over other techniques used for disease diagnostics due to its ability to produce a spectral "fingerprint" which specifically represents the total biochemical composition of a sample. As quoted by Mann and Vickers, RS "is unusually, if not uniquely, suited to be the process control star of the next century" [4]. This is because "the intrinsic selectivity of RS allows for accurate identification of organic, inorganic and biological species, an advantage that is lacking in many other analytical techniques, such as ultraviolet absorbance and fluorescence spectroscopies" [5]. RS provides considerable detail regarding the biochemical composition of a sample, and is thus able to detect changes that occur in biological samples during the onset and progression of a disease. RS has the ability to be non-invasive and has the potential for *in vivo* use, which makes it a much more appealing technique for diagnosing diseases over other methods, such as biopsies. It is much less expensive than imaging tests, and it is objective, making it a better choice than some diagnostic methods which require subjective analysis of the results. RS goes beyond simply ruling out other possible diagnoses, as it has the potential to definitively determine both the presence and the stage of disease progression. What's more, RS is a fast, easy-to-use, and reliable technique that can be easily incorporated into clinical settings, making it an exceptionally valuable diagnostic tool.

While RS has a high level of chemical specificity, the changes that occur between spectra of different classes of samples can oftentimes be minute and difficult to visually observe. Thus, advanced statistical analysis, or chemometrics, is utilized to better understand the information found within the obtained data. Chemometrics is essentially "the art of extracting chemically relevant information from data produced in chemical experiments." [6] Chemometrics is suitable to use with RS because the spectral data exists in the form of a data matrix consisting of wavenumbers (cm^{-1}) and corresponding intensities for each spectrum. Different algorithms can be applied to the spectral data matrix for the purpose of building statistical models. These models identify the most useful chemical data and separate it from less informative data and insignificant noise, all while learning how to recognize patterns and similarities within the data matrix. In this way, the models are able to learn to recognize similarities and differences between either labeled or unlabeled data, which it can then use to return either a qualitative or quantitative response.

Many chemometric algorithms exist to answer a wide array of questions a researcher may have. Notably, the two main categories of statistical models that can be built include supervised and unsupervised techniques. Unsupervised algorithms do not utilize sample labels or user-defined information when the model is being constructed [7]. Examples of unsupervised modeling include principal component analysis (PCA) and hierarchical clustering analysis (HCA). These models are

exploratory types of analysis and are not used for classification but rather can be used to display similarities or differences between groups of data. On the other hand, supervised models take into account user-defined labels, known as classes, for all samples in order to build prediction models. Supervised models can be further split into regression models or classification techniques. A regression model is often used for calibration purposes and will give a quantitative answer, while a classification model will give a qualitative response, such as the classification of a spectrum. A common regression model is multiple linear regression (MLR), while various classification techniques include partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), and artificial neural networks (ANN).

Because of the significant ability to pick out important information and recognize patterns and similarities within sets of data, chemometric techniques are ideal for analyzing Raman spectral data. Specifically, these advanced statistical methods are used in this chapter for the purpose of understanding the spatial distribution of biochemical components within a sample, identifying potential biomarkers, differentiating healthy biological samples from diseased ones, and for determining the stage of a disease, all for the purpose of diagnosing cancer.

This chapter presents research published between 2014 and 2018. Specifically, articles that focus on the application of RS combined with chemometrics for diagnosing cancer are considered. Modifications of RS will be considered, such as incorporating the use of fiber-optic techniques, which have not been well reviewed in the past. Studies using all forms of biological materials (cells, tissue, bodily fluids) will be included, and studies will not be limited by the Raman spectral range examined (i.e., high wavenumber region versus the fingerprint region). Research that does not utilize any advanced statistical techniques will not be considered. It is anticipated that in order for the RS methodology to be implemented clinically, an automatic data analysis procedure will be required to interpret the Raman spectral data and to make a diagnosis. Statistical analysis can be performed using software which can lead to automatic and definitive diagnoses in real-time, making diagnosing cancer much more accurate, rapid, and inexpensive. Furthermore, the goal of this chapter is to show that spontaneous RS is sufficient enough for cancer diagnostics, and that more complicated or expensive technology is not required to achieve highly accurate diagnoses. In this chapter, we aim to support and buttress the claim that RS in combination with chemometrics has a strong potential to be implemented as a novel universal method for diagnosing all forms of cancer in the near future.

8.2 Discussion

8.2.1 Spontaneous Raman Spectroscopy

Spontaneous Raman spectroscopy utilizes a monochromatic laser beam to radiate the sample being studied. Inelastically scattered light which interacts with molecular vibrations of the sample will be detected by the instrument. The outcome is a

very specific spectral fingerprint of the sample. Spontaneous RS is uniquely suitable for characterizing microheterogeneous environments; specifically, the collection of multiple spectra from a single sample will allow for the detection and spatial distribution of biochemical components within a sample to be determined [8]. RS can be used to collect and process spectral information obtained from multiple positions on a sample with the purpose of providing statistically significant characterization of a sample's heterogeneity and multicomponent composition. By collecting multiple spectra, biomolecules present at high local concentrations can be detected, allowing for identification of potential biomarkers, including those present at low average concentrations [9]. A great advantage of RS resides in its ability to probe the entire biochemical composition of a sample, thus producing a spectroscopic signature for different disease states which are based on the simultaneous integration of multiple biomarkers. This capability significantly improves the sensitivity and selectivity of the diagnostic technique. The following studies have capitalized on the advantages of spontaneous Raman spectroscopy for the purpose of diagnosing various cancers through analysis of either tissue, cells, or bodily fluids.

8.2.1.1 Tissue

Tissue is frequently analyzed in disease diagnostic studies due to its ability to indicate the presence of cancer in the body. As such, tissue is frequently biopsied and thus readily available for *in vitro* Raman spectroscopic analysis and especially for the purpose of diagnosing cancer.

In a study conducted by Kalkanis et al., 95 regions from 40 tissue samples were analyzed to distinguish normal brain tissue from glioblastoma multiforme (GBM) and necrosis using discriminant function analysis (DFA), achieving 99.6% and 97.8% accuracy in the training and validation datasets, respectively (Fig. 8.1) [10]. On the other hand, an average 87.6% accuracy for diagnosing a tissue sample as

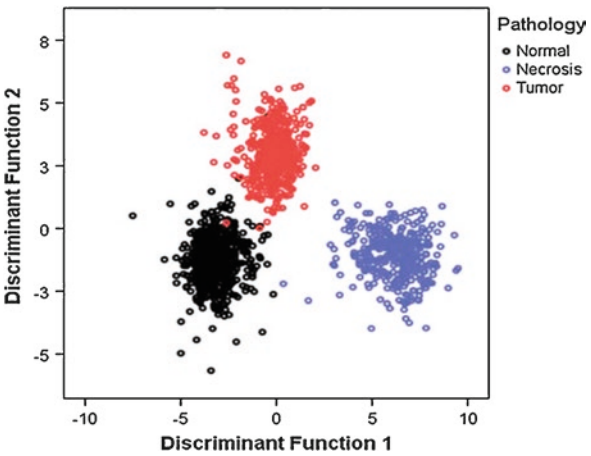

Fig. 8.1 Plot of discriminant function analysis scores for training data (Reprinted with permission from [10])

originating from a healthy donor or from an individual with a brain tumor was achieved through analysis of only 20 tissue samples by a learning vector quantization neural network (LVQNN) technique [11]. The apparent role of number of samples and chemometric technique used for obtaining successful results is already well demonstrated.

Several attempts for using RS to diagnose breast cancer have been made in the last 4 years. The carbonate intercalation signature in type II microcalcifications in tissue, a common diagnostic feature of breast cancer, was used to demonstrate the differences between benign and malignant breast lesions. Raman decision algorithms were developed to distinguish between benign and malignant lesions with type II microcalcifications. The differences in carbonate intercalation could differentiate benign and malignant lesions; specifically, empirical decision algorithms based on carbonate and cytoplasmic protein content achieved 77–83% accuracy for discrimination [12]. Raman spectroscopic analysis of 39 breast tissue samples was employed to understand the differences between normal, atypical ductal hyperplasia, ductal carcinoma *in situ*, and invasive ductal carcinoma lesions of the breast. A support vector machine (SVM) diagnostic model was built using the radial basis function (RBF) with leave-one-out cross-validation (LOO-CV) and achieved an overall accuracy of 74.39% for identifying a sample as belonging to one of the four classes [13]. Fallahzadeh et al. aimed to diagnose breast cancer by using ant colony optimization (ACO) to find the most useful Raman features for discrimination. With five spectral features selected by ACO, the algorithm could correctly classify the 11 tissue samples as normal, benign, or cancerous with 87.7% accuracy [14]. Based on the results of these small studies, RS analysis of tissue samples is not suggested as the most optimal method for diagnosing breast cancer. Results are greatly improved when cells are instead analyzed, as is later discussed.

The mortality rate of cervical cancer can be reduced if the disease is detected in the premalignant stage. As such, Rashid et al. utilized Raman spectral mapping to elucidate biochemical changes associated with premalignant stages of the cancer. When analyzed by K-means cluster analysis (KCA), cervical biopsies classified as negative for intraepithelial lesion and malignancy were divided into three different layers—stroma, basal/para-basal, and superficial—based on differences in collagen, DNA bases, and glycogen spectral features. For low-grade and high-grade squamous intraepithelial lesion (SIL) samples, KCA clustered regions of the basal layer together with the superficial layer. When morphological changes were not apparent, PCA could identify biochemical changes associated with the cancer, creating a useful method for detecting premalignant changes in cervical tissue [15]. Raman spectral mapping was further used to understand the differences between neoplasia and malignancy of cervical tissues. Gradual biochemical changes associated with cancer progression were identified using PCA and KCA, including changes in glycogen, collagen, lipids, protein, carotene, and the nucleus to cytoplasm ratio [16]. While both of the previous studies demonstrated the usefulness of RS to detect biochemical differences between different samples, Daniel et al. went on to improve the usefulness of RS by obtaining quantifiable results through analysis of spectral data by PCA in combination with ANN. The method could classify

tissue as normal, premalignant, or malignant with an overall accuracy of 99%. Following this, well differentiated, moderately differentiated, and poorly differentiated squamous cell carcinoma (SCC) samples were investigated using PCA combined with linear discriminant analysis (LDA), achieving 94% diagnostic accuracy [17].

Tissue obtained from healthy donors and from colorectal cancer patients was investigated by several groups. In one proof-of-concept experiment, different excitation wavelengths were used to study the disease. Near-infrared (NIR) Fourier transform-Raman (1064 nm), NIR-visible-Raman (785 nm) and visible-Raman (532 nm) excitation wavelengths were used to collect spectra from 14 samples. Each of the three sets of spectra was analyzed using PCA, and partial spectral differences in each dataset were observed between the normal and diseased samples. Interestingly, when the datasets were combined, the clearest separation between the two classes was seen [18]. While the previous study was useful for establishing biochemical differences between groups, Li et al. obtained quantitative results using ACO-SVM. ACO identified five diagnostically important Raman bands, which were then used to build the SVM diagnostic model. Results showed 93.2% accuracy for identifying colorectal cancer in 44 patients [19]. Two different chemometric systems were used in an additional study to evaluate which could best diagnose colorectal cancer based on tissue analysis of 81 patients. PCA-LDA and PLS-DA models were built and validated using leave-one-patient-out cross-validation (LOPO-CV). PLS-DA performed better, achieving a diagnostic accuracy of 84.3%, which was an improvement over the 79.2% accuracy achieved by PCA-LDA [20].

Tissue samples from patients with early-stage (stage 0 or I) esophageal cancer were examined *ex vivo*; Raman bands that showed a statistically significant difference in band intensity, determined using a *t*-test, were analyzed using LDA. The stage of tissue was correctly predicted with 81.0% sensitivity and 94.0% specificity [21]. However, it should be noted that the algorithm was not tested with comparison to healthy tissue samples.

Several studies were aimed at identifying gastric cancer. In the first study, which used a significant number of samples as well as a robust validation method, Jin et al. analyzed 105 tissue samples from cancerous and pre-cancerous lesions and normal gastric mucosa (NGM). Raman spectra showed differences between the samples related to protein, nucleic acid, and lipid content. Using PCA-LDA with LOO-CV, an average sensitivity of 88.9% and specificity of 94.6% were achieved for discriminating the three classes [22]. Yao et al. achieved 91.7% accuracy for distinguishing normal gastric tissue from cancerous tissue using Fisher discriminate analysis (FDA); however, only 12 samples were analyzed [23]. Two studies were conducted by Hsu and co-workers with a similar goal. In the first, PCA could differentiate the four main histological types of gastric adenocarcinoma (AC), including papillary adenocarcinoma (PAC), tubular adenocarcinoma (TAC), mucinous adenocarcinoma (MAC), and signet ring cell adenocarcinoma (SRC) by analyzing 79 tissue samples. PCA distinguished all gastric AC types from NGM in a binary system. SRC and MAC were able to be differentiated from TAC and PAC; however, TAC and PAC showed no significant differences between each other. Furthermore, LDA scatter

plots successfully differentiated all gastric AC types from NGM [24]. In the second study, the results improved to show gastrointestinal stromal tumors (GISTs) could be differentiated from AC and benign lesions from 119 patients. PCA-LDA was employed with CV, achieving an average sensitivity of 99.67%, specificity of 95.45%, and accuracy of 98.32% for distinguishing GISTs, gastric AC, and NGM; this information can be used to help clinicians determine an appropriate treatment path [25]. Based on the range of experiments performed, RS is clearly capable of detecting gastric cancer within tissue samples.

To diagnose nasopharyngeal carcinoma (NPC), tissue samples were collected from 15 individuals with the cancer and from 15 healthy donors. Investigators generated four models using PCA-LDA to discriminate spectra collected from NPC tissue and healthy tissue at various depths of the sample. Each model achieved greater than 95% sensitivity and specificity, with the exception of the deepest level of tissue [26]. Another group also utilized PCA-LDA with LOO-CV to distinguish normal tissue from NPC tissue, achieving a sensitivity of 81% and specificity of 87%. When the method was coupled with PLS, the sensitivity and specificity increased to 85% and 88%, respectively, showing how the chemometric technique selected plays a role in performance success [27]. Mian et al. constructed tissue engineered models of normal, dysplastic, and head and neck SCC using corresponding cell lines. PCA was used to analyze the Raman spectral data collected from the tissues to determine the maximum variance between the groups. LDA was then used to test the discriminatory capacity of the data and classify the tissue samples as normal, dysplastic, or cancerous. Predictions showed an average specificity of 70% and sensitivity of 100% in a binary model of normal versus cancerous tissue. For differentiating dysplastic versus cancerous tissue, 90% sensitivity and 98% sensitivity were achieved. To further validate the study, predictions were made for 12 blinded samples, obtaining 75% specificity for predicting normal tissue, 90% sensitivity for dysplastic tissue, and 98% sensitivity for cancer tissue [28].

Pence et al. utilized two different excitation wavelengths (785 and 1064 nm) to study a total of 15 healthy, AC, and hepatocellular carcinoma (HCC) liver tissue samples for the purpose of diagnosing liver cancer. Spectral data collected using the 1064 nm excitation wavelength was classified using sparse multinomial logistic regression (SMLR); data collected using the 785 nm laser suffered from an intense and highly variable fluorescence background that dominated the Raman spectra and was thus not analyzed further. Two different models were generated using the data collected with 1064 nm excitation. The binary model (healthy versus all tumor tissue) showed 100% sensitivity and 89% specificity and the tertiary model (AC versus healthy versus HCC tissue) achieved an average accuracy of 75.67% [29]. Notably, greater success for diagnosing liver cancer was achieved using Raman hyperspectral imaging, as is later discussed.

Wang et al. aimed to understand mutations of the epidermal growth factor receptor (EGFR), its relation to lung AC, and its potential to be used in future diagnostic studies through RS analysis of 156 lung AC tissues. Tissue samples of carriers without the mutation showed increased levels of amino acids and DNA, whereas samples from donors with a specific mutation group, L858R, exhibited increased

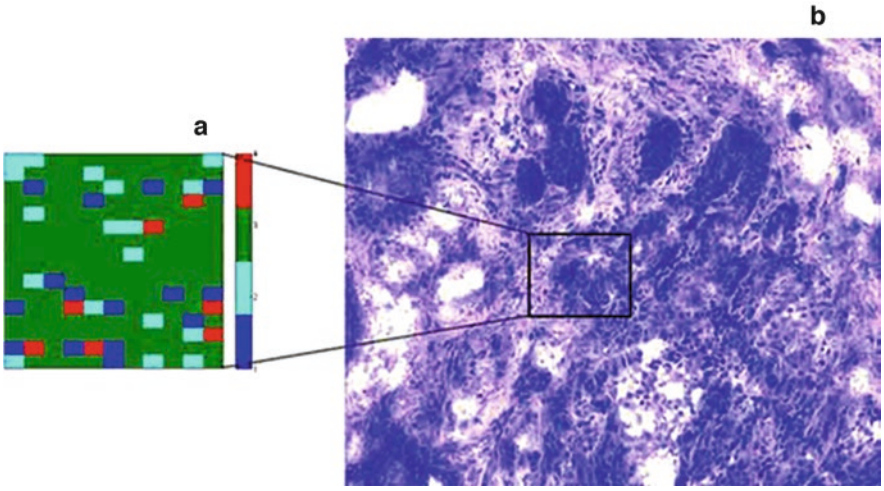

Fig. 8.2 Reconstructed Raman mapping (**a**) and their respective hematoxylin and eosin-stained image (**b**) of blinded sample; the box indicates the site of the Raman image (Reprinted with permission from [32])

arginine levels. To predict to which class a sample belonged, PCA-SVM with LOO-CV was used. The L858R and E19del mutation groups were differentiated from wild-type EGFR tissue with 87.8% accuracy; considering the sample-size and the level of accuracy, this study provides significant evidence for a novel lung cancer screening method based on RS analysis of EGFR mutations [30].

Oral tumor and healthy tissue (10 samples each) were qualitatively analyzed using KCA and PCA; KCA was used to generate Raman maps which correlated to the sample's histopathology. In healthy tissue sections, stratification of epithelial layers was observed. Each of the three layers detected within the normal epithelium tissue was successfully distinguished from the tumor section using PCA. In the unhealthy tissue samples, inflammatory regions of tumor cells and tumor-stroma regions were detected; while not providing quantifiable results, this study shows how Raman mapping can provide novel insight for understanding pathological states [31]. Continuing in the same manner, Raman maps of normal and cancerous oral tissue were obtained by Daniel et al. The maps showed an increase in glycogen, lipid, and protein content within the healthy tissue and an increase in nucleic acid content in the cancerous tissue. Similarly, PCA and KCA were again used to demonstrate the distribution of biochemical components within the samples. Dissimilarly, to improve the usefulness of the results, a discrimination line was computed, resulting in 98.9% accuracy for discerning the two groups. Raman spectral data was collected from a blind sample, which was then subjected to histological evaluation. A Raman image was generated, and the sample was determined to be dysplastic, which was confirmed by H&E staining (Fig. 8.2) [32]. In a third study, Raman spectral data was collected from 24 tissue samples of 14 donors with oral SCC and analyzed using multivariate curve resolution (MCR). The spectral maps of the tissue

were automatically and objectively compared through spectral matching of the MCR decomposed Raman spectra and the Raman spectrum of keratin, a biomarker of oral SCC. The oral SCC tissues were correctly identified with 77–92% sensitivity and 100% specificity, with the difference in sensitivity level depending on how positivity was defined [33].

RS was used to understand the pathological changes occurring in ovarian tissue for the purpose of distinguishing adenoma and early AC from benign tumors. Using PLS-DA and LOO-CV, the discrimination model provided an accuracy of 85.2% for diagnosing ovarian cancer [34].

Raman spectra from 25 malignant and benign pheochromocytoma and paraganglioma (PPGL) tissue samples were identified using PCA-LDA with a sensitivity of 80.0% and specificity of 100.0%. PPGLs are tumors that arise from adrenal or extra-adrenal chromaffin tissues. Notably, the obtained results were higher than those obtained using the pheochromocytoma of the adrenal gland scaled score, which is a current method for distinguishing between benign and malignant PPGLs [35].

In a thorough study performed by Liu et al., discriminate analysis (DA) with LOO-CV was applied to spectral data collected from tissue of 63 different patients to determine if RS could distinguish malignant and benign renal tumors using biopsy specimens. Results showed success not only in separating tumor and normal tissue samples (82.53% accuracy) but also in distinguishing malignant and benign tumors (91.79% sensitivity and 71.15% specificity) and low-grade and high-grade tumors (86.98% accuracy). Oncocytoma and angiomyolipoma, two different forms of benign tumors, were successfully differentiated from clear cell renal carcinoma with 100% and 89.25% accuracy, respectively, and subtypes of cell carcinoma were distinguished from each other with an accuracy of 93.48%. Notably, Raman spectroscopic analysis further resulted in successful diagnoses for 7 of 11 cases whose diagnoses were missed during biopsy, illustrating an improvement of the RS methodology over current diagnostic methods [36].

A selective-sampling method was used to collect Raman spectra of tissue samples from individuals with basal cell carcinoma (BCC) and healthy volunteers. A multinomial logistic regression classifier indicated 100% sensitivity and 92.9% specificity for correct classification of an independent set of skin tissue samples [37]. In a large study by Zhao et al., wavenumber selection based analysis was used to diagnose skin cancer. Multivariate techniques PCA-general discriminant analysis (GDA) and PLS with LOO-CV were employed; both were capable of classifying 645 lesions (including skin cancer, pre-cancer, and benign skin lesions) from 573 patients with skin cancer [38]. Interestingly, skin cancer is not well studied using biological fluids or cell samples, suggesting the greatest success for diagnosing skin cancer is through tissue analysis.

Raman spectral data collected from a total of 30 normal thyroid, goiter, and thyroid cancer tissue samples were analyzed by PCA and LDA in combination with CV and binary logistic regression (BLR). The results of LDA with CV showed normal versus cancerous tissues reached a discriminant value of 78.3%; goiter versus cancerous tissue reached a discriminant value of 75%; and normal versus goiter tissue reached a discriminant value of 68% when the spectral region was limited to

1200–1400 cm^{-1}. The results of the BLR model showed the same three groups each achieved greater than 80% concordance [39]. Senol et al. diagnosed papillary thyroid carcinoma (PTC) using an orthogonal PLS algorithm which discriminated 23 tumor and healthy tissue samples with 100% sensitivity and 81.8% specificity for the calibration dataset; the root mean square error of CV was about 47.8%, which is considered low [40]. Using 28 samples (18 for the calibration dataset and 10 for the test dataset) Palermo et al. was better able to differentiate healthy parathyroid tissue and parathyroid adenoma using PLS-DA, achieving 100% accuracy. Further, chief cell adenoma and oxyphil cell adenoma were distinguished from each other with 100% of oxyphil and 99.8% of the chief cell adenoma samples correctly predicted during external validation [41]. When observed together, these studies demonstrate the interesting effect that different chemometric techniques can have on developing successful prediction algorithms.

The Raman spectral data of biopsies from 27 women suspected to have vulval lichen sclerosus (LS), a condition associated with an increased risk of developing vulval carcinoma, were analyzed using PCA-LDA with LOPO-CV. LS tissue was separated from tissue of other inflammatory vulval conditions with 91.0% sensitivity and 80.0% specificity [42]. It is important to note that a comparison to healthy vulval tissue was not taken into consideration.

As these studies clearly show, tissue samples have an immense potential to diagnose various forms of cancer when studied by spontaneous RS and chemometrics. However, the collection of tissue samples can be invasive and uncomfortable for the afflicted patient. On the other hand, RS analysis of tissue samples can be used to confirm typical histopathological diagnosis, which can oftentimes be hindered by subjective and experience-based analysis, making RS advantageous for incorporating into diagnostic procedures. Further *in vivo* analysis is required, and has been conducted, to better understand the capacity of RS for diagnosing cancer, as is later discussed.

8.2.1.2 Cells

Cytology has been widely used for diagnosing cancer—this is most likely due to the fact that cytology specimens are usually easier to obtain while causing less discomfort, cost less money, and are less likely to result in complications when compared to biopsied tissue samples. In this regard, several recent studies have successfully applied spontaneous RS analysis of cells in combination with chemometrics for diagnosing cancers.

Kerr et al. conducted four different experiments to evaluate the potential of RS to diagnose bladder cancer. In each experiment, various standard clinical procedures were used in order to prepare the cell samples for analysis. Spectral data from each experiment was analyzed through PCA-LDA with LOO-CV, with each experiment achieving greater than 88% sensitivity and specificity. Importantly, it was determined that none of the standard procedures that was tested significantly impacted

the methodology's ability to diagnose bladder cancer, setting the foundation for RS analysis of cells to be used under a wide variety of clinical settings for diagnostic purposes [43].

Four brain cancer GBM cell lines were obtained from four different patients who each had grade IV astrocytoma. Raman spectra of single cells from each cell line were investigated using multivariate analysis. While this study did not focus on discriminating healthy and diseased states, the results do show that cell lines were similar among all four afflicted patients, thus confirming the reliability of RS analysis of cells for cancer diagnostics and staging for future studies [44].

RS was popularly used to examine cells for the purpose of detecting and diagnosing breast cancer. The results are generally an improvement over those achieved when tissue was analyzed and provide additional information regarding the effect of drugs, which is generally difficult or impossible to do through analysis of other biological samples. Marro et al. utilized RS to study cells undergoing an epithelial-to-mesenchymal transition, a process indicative of breast cancer metastasis. MCR was used to determine how the transition affected the lipid profiles of the cells; specifically, the transition resulted in increased levels of tryptophan and maintenance of a low fatty acid content as compared to highly metastatic cells. PLS-DA successfully discriminated cells within various stages of the transition process, achieving 94% sensitivity and 100% specificity, providing the ability to identify breast cancer in the earliest stages of malignancy [45]. Bi et al. studied the overexpression of human epidermal growth factor receptor 2 (HER2), which is associated with increased chances of developing breast cancer. Three different cell lines were studied—BT474 (HER2 positive breast cancer cell), MCF-10A (HER2 negative healthy control cell), and HER2+ MCF-10A (HER2 positive healthy control cell). The data was analyzed using lasso and elastic-net generalized linear models with CV, which achieved an average 99.8% sensitivity and 99.6% specificity for separating the three cell lines. Following this, Raman spectra of 104 Lapatinib-treated and 104 Lapatinib-resistant breast cancer cells were collected. Lapatinib, a tyrosine kinase inhibitor, is a common drug used to treat breast cancer patients. Significant differences between the spectral signatures of the two cell lines were observed, revealing vital biochemical information which could potentially identify cells resistant to important cancer-fighting drugs as well as demonstrating a novel method for studying the response of cancer cells to therapeutic interventions [46]. In a different study, the effect of pentoxifylline, a drug used to treat muscle pain, on human breast cancer cells was examined. Spectral changes suggested a linear relationship between alterations in DNA, protein, and lipid content with drug dosage. Further, PC-LDA with LOO-CV could separate the control group from cells treated with different levels of pentoxifylline, providing an opportunity to monitor changes occurring within cell lines as a result of medications [47]. Talari and co-workers published two studies on analyzing breast cancer cell lines. In one, a combination of PCA and LDA differentiated two different breast cancer cell lines and one normal breast cell line with 100% sensitivity and 91% specificity [48]. In a later study, normal proliferating, hypoxic, and necrotic regions of a T-47D human breast cancer spheroid model were analyzed by RS to identify chemical changes that occur as the regions

progress to necrosis. PCA showed lipid, amide I and III, and nucleic acid content differ significantly between the three regions, providing information for understanding the progression of cells to necrosis [49]. Winnard Jr. and researchers studied organ-specific isogenic metastatic breast cancer cell lines. PLS-DA with LOO-CV was used to classify the different cell lines with 96.8% accuracy; SVM was also used and provided similar results [50].

Ramos et al. evaluated the potential of RS to screen for cervical cancer using cell samples. Both the cervical intraepithelial neoplasia (CIN) and the SIL terminology systems for classifying cervical cancer cells were tested in the process. Biochemical fingerprints of normal and abnormal cell samples were used for discrimination by PCA. Subsequently, PCA-LDA models with LOO-CV were built for classification using either CIN or SIL terminology. The model built using SIL terminology, which characterizes lesions into low-grade and high-grade categories, achieved an average 93.45% sensitivity and 97.55% specificity. The model using CIN terminology gave better results—CIN divides classification of cells into three grades where CIN1 corresponds to mild dysplasia, CIN2 to moderate dysplasia, and CIN3 to both severe dysplasia and carcinoma *in situ*. The CIN model reached an average sensitivity of 96.3% and specificity of 98.27% [51]. Notably, the terminology system used plays a small but identifiable role in the classification efficiency of the chemometric technique employed in this study.

Hundreds of live colorectal cancer cells, derived from primary and secondary tumor cells of the same patient, were studied by Gala de Pablo et al. Using PCA-LDA with CV to analyze the RS data, an accuracy of 98.7% was achieved for classifying cells as either SW480 or SW620 (Fig. 8.3). These results were better than those achieved with SVM and discriminant trees, illustrating the usefulness of certain chemometric techniques over others. PCA-LDA was also used to classify HL60, HT29, HCT116, SW620, and SW480 cells with 92.4% accuracy [52].

Efforts to diagnose non-Hodgkin lymphoma was assessed through examination of normal B-cells and non-Hodgkin lymphoma B-cells using asymmetric least squares (ALS) baseline correction and PCA. K-nearest neighbor (KNN) was used

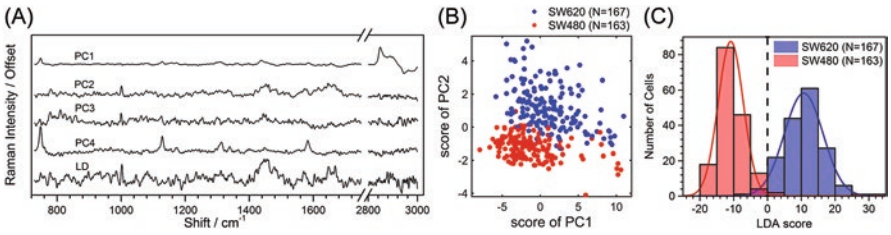

Fig. 8.3 PCA/LDA results. (**A**) Shape of the PCs 1 to 4 and of the LD (**B**) 2D plot of the scores for the first two PCs. (**C**) Histogram of the individual cell scores when projecting the cell data onto the LD from (**A**) with a vertical dashed line at the point of best separation. LD = linear discriminant; LDA = linear discriminant analysis; PC = principal component; PCA = principal component analysis (Reprinted with permission from [52])

to confirm the discriminatory powers of PCA, resulting in 100% accuracy, thus providing a potentially novel method for diagnosing the cancer [53]. It was further found that the Raman spectral data from peripheral blood mononuclear cells (PBMC) could be used to discriminate a significantly pure population of T-cell lymphocytes from other PBMC myeloid cells. Several classifiers, including PCA-LDA, SVM, and Random Forests (RF), were used for discrimination. SVM built using RBF performed the most optimally, achieving 98% sensitivity and 92% specificity. This study demonstrates fundamental differences between myeloid cells and lymphocytes which can be used to identify different PBMC subtypes for diagnostic functions, as well as the importance of testing different chemometric techniques for the purpose of optimizing diagnostic capabilities [54].

Carvalho and researchers collected Raman spectra of the nucleoli, nuclei and cytoplasm of oral epithelial cancer and pre-cancerous cell lines, as well as from normal oral epithelial primary cell cultures. PCA exhibited significant differences between the cell lines, and contributions from nucleic acids and proteins of nucleolar and nuclear sites and from lipids of the cytoplasmic area were primarily responsible for discrimination. This study shows the ability of RS analysis of cells to uncover incredibly useful information regarding cellular components which cannot be achieved through tissue or biological fluid analysis, and that can contribute significantly toward diagnosing cancer [55].

Cisplatin-resistant and cisplatin-sensitive ovarian carcinoma cells were discriminated using PCA-LDA with LOO-CV. Cisplatin, an anti-cancer chemotherapy drug, is often used to treat ovarian cancer. Using the classifier, a diagnostic accuracy of 82% was obtained [56]. These results are similar to those obtained from the previously described analysis of tissue samples; however, the analysis of cells provides the added benefit of monitoring the effect of anti-cancer drugs.

Corsetti et al. analyzed the Raman fingerprints of normal and metastatic hormone-resistant prostate cancer cells by PCA-LDA with CV, which reliably distinguished the two with 95% sensitivity and 88% specificity [57]. Alternatively, Olmos et al. aimed to understand the effect of the pesticide Aldrin on human prostate cancer cells. A portion of prostate cancer cells were exposed to Aldrin, which has been shown to increase the risk of developing prostate cancer in men exposed to it. To assess the differences between the normal and treated prostate cell populations, PLS-DA with CV identified biomarkers associated with pollutant stress, and the best classifier built achieved 91.3% specificity and 80.0% sensitivity for distinguishing the two cell classes [58]. Most recently, deep-ultraviolet Raman spectroscopy (excitation wavelength of 198 nm) also showed that normal human primary prostate epithelial cells and grade IV adenocarcinoma PC-3 prostate cancer cells could be successfully differentiated [59]. The results obtained here are comparable to those obtained using both high-throughput methods and fiber-optic probes, which are further discussed, suggesting the powerful diagnostic capability of spontaneous RS analysis of cells.

The combination of RS and chemometrics has the ability to detect chemical signatures of cells in order to quickly and accurately diagnose various types of cancer. Cells provide unique information regarding the mixtures of metabolites present at a

single point in the lifetime of the cell and can be used to probe cellular components which cannot be accessed in other biological samples. Furthermore, cells can be manipulated and exposed to different drugs as well as carcinogens in order to better understand the pathology of cancer as well as the effect of drugs on cancer, providing advantageous and unique information which cannot be easily accomplished using other biological samples. While the reported results are promising, it should be noted that in several of the previously reviewed studies, there is a slight problem of the number of samples analyzed—that is, Raman spectra were collected from a significant number of cells, but not a significant number of donors. This, however, is a straightforward criticism to address in future work; as such, because of the other advantages which outweigh this small issue, the potential for RS analysis of cells to diagnose cancer should not be disregarded.

8.2.1.3 Body Fluids

In an effort to create a more simple and less-invasive sample collection procedure, many studies have focused on studying various body fluids, including blood, urine, and saliva, for cancer diagnostics. These body fluids provide biochemical information which can be used not only for identifying cancer but also for determining the stage of the cancer. Body fluid analysis tends to be less costly and is a much more appealing option for reoccurring testing due to the ease of non-invasive collection; thus, many researchers have used RS and advanced statistical techniques to analyze body fluids for cancer diagnostic purposes.

Blood serum of 35 subjects with meningioma was investigated by RS and compared to blood serum collected from 35 control subjects. Through PCA and PC-LDA followed by LOO-CV, healthy and meningioma subjects were correctly classified with efficiency levels of 92% and 80%, respectively. Similar results were also obtained for identifying different grades of meningioma [60].

Blood serum and urine were both studied in an attempt to diagnose cervical cancer. In one study, González-Solís et al. utilized PCA to distinguish serum samples from 19 cervical cancer patients, 3 pre-cancer individuals, and 20 healthy controls. Differences in Raman spectra indicated a high amount of carotenoids and intense protein contribution in the control serum and higher concentrations of glutathione and tryptophan in the disease serum (Fig. 8.4). Using a LOPO-CV technique, 100% sensitivity and specificity were achieved [61]. Pappu et al. investigated 27 urine samples collected from healthy volunteers and patients with cancer. Using an LDA diagnostic algorithm with CV, 100% accuracy was achieved for discrimination [62]. Interestingly, these studies suggest that regardless of the body fluid type analyzed, markedly successful results are obtained for diagnosing cervical cancer.

The potential to diagnose colon cancer using blood serum was examined in a large study with 75 healthy volunteers, 65 colon cancer patients, and 60 post-operation colon cancer patients. Differences in Raman spectra were assigned to changes due to nucleic acids, amino acids, and chromophores. PCA and KNN

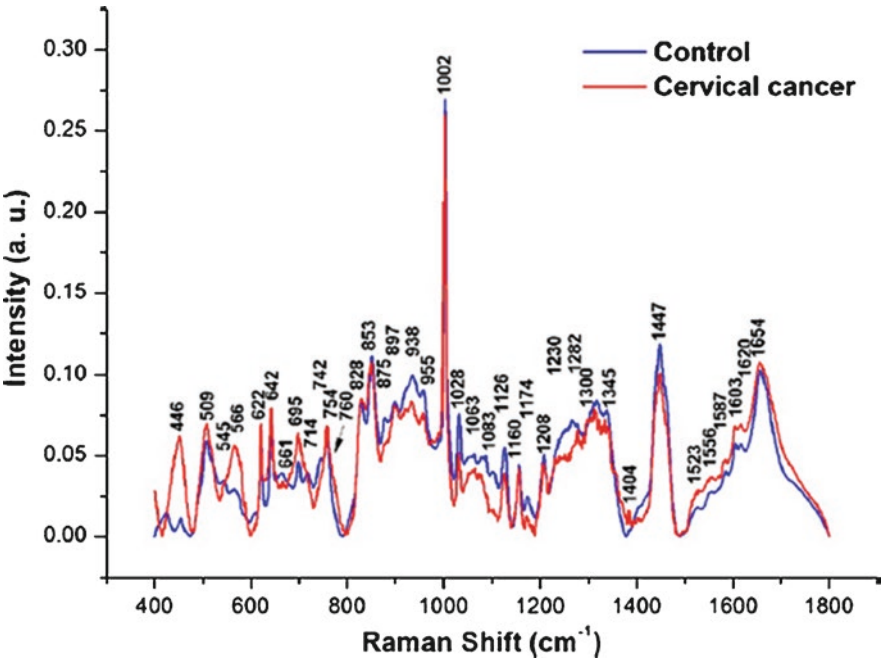

Fig. 8.4 Mean Raman spectra of the control and cervical cancer serum samples (Reprinted with permission from [61])

analyses were used to discriminate between the three classes, resulting in 91.0% accuracy [63].

Khan et al. used RS analysis of blood serum to diagnose NPC. PCA was used to highlight spectral differences and SVM with RBF and CV classified the serum as belonging to either the pathological class or the healthy class with 93% accuracy. Although a smaller dataset was used, these results are generally an improvement over those achieved through RS analysis of tissue [64].

Happillon et al. diagnosed chronic lymphocytic leukemia based on Raman spectral analysis of blood smears collected from 27 healthy volunteers and 49 individuals with the disease. Two SVM models were built with CV—the first could discriminate between the two main subpopulations of leukemia (lymphocytes and polymorphonuclears) with sensitivity and specificity levels both over 98.5%. The second SVM model discriminated neoplastic and healthy lymphocyte spectra with an average sensitivity of 88% and specificity of 91% [65]. These results are relatively comparable to those obtained through the analysis of cells, indicating this method should be considered further due to easier sample collection.

Interestingly, blood serum was used in two different studies to successfully diagnose lung cancer, suggesting the advantage of blood serum over other body fluids for this purpose. Li et al. tested several different modeling techniques to distinguish blood serum of 29 healthy donors and of 68 donors with lung cancer. Uncorrelated

linear discriminant analysis (ULDA) and LDA in combination with multiple scatter correction (MSC) pretreatment could each make the distinction with 100% sensitivity and specificity each. Interestingly, MSC combined with PLS-DA was unsuccessful in achieving the goal, further demonstrating the significance of chemometric technique selection [66]. Wang et al. analyzed 91 blood serum samples from healthy individuals and individuals with varying stages of non-small cell lung cancer (stages I–IV), and found LDA with CV could distinguish the five different groups with an overall accuracy of 92% [67].

Urine samples were obtained from patients with oral cancer and from healthy donors; the corresponding Raman spectra were analyzed using PCA-LDA with LOO-CV. The model achieved 98.6% sensitivity and 87.1% specificity, with an overall accuracy of 93.7% for identifying the cancer patients [68]. Pachaiappan et al. utilized both blood plasma and saliva to diagnose oral cancer. In one study, the blood plasma of 30 healthy individuals, 27 patients with oral sub mucous fibrosis, and 34 with oral SCC was analyzed by PCA-LDA. The algorithms could separate the normal group from the premalignant group with 96.3% sensitivity and 80.0% specificity and the normal group from the malignant group with 91.2% sensitivity and 96.7% specificity [69]. Saliva of 83 individuals from the same aforementioned groups was also subjected to analysis via PCA-LDA with LOO-CV. The algorithms separated normal from premalignant samples with 96.4% sensitivity and 70.2% specificity and normal from malignant samples with 93.8% sensitivity and 95.7% specificity [70]. These studies show that regardless of body fluid analyzed, high levels of performance can be achieved for diagnosing oral cancer.

Body fluid analysis is advantageous over analysis conducted using other biological materials for many significant reasons. Collection of body fluids is considerably less-invasive, and even non-invasive in certain cases; it is inexpensive and the process is quick, which allows for rapid results as well as repeat analyses as necessary, and can be conducted during routine exams. Biological fluids provide a great amount of biochemical information regarding the composition of the sample and have a great potential to diagnose all forms of cancer when analyzed by RS.

8.2.1.4 Spontaneous Raman Spectroscopy with Expanded Raman Spectral Range

The majority of the aforementioned studies using spontaneous RS in combination with chemometrics have focused on analyzing the “fingerprint” (FP) region of the Raman spectral data range. The FP region usually refers to the section of Raman spectral bands existing between 400 and 1800 cm^{-1}. It has been discovered that a wider Raman spectral range, which includes the high wavenumber (HWN) region, provides additional information that can be used for many analytical purposes, including disease diagnostics and biomarker detection. The HWN region of spectral data usually refers to the spectral range between 2800 and 3600 cm^{-1} which is found to contain important contributions from water, various C–H bond vibrational modes of lipids and proteins, as well as other N–H and O–H bond vibrations of

biomolecules. Notably, the HWN region does not usually suffer as much from autofluorescence signal as the FP region. Several studies have exhibited the usefulness of this region for diagnosing cancers.

Several studies have used the HWN region to successfully investigate oral cancer. Barroso et al. aimed to differentiate healthy tissue from oral SCC tumor tissue within 14 patients. Various bands attributed to water were used to quantify the water content in each sample. Specifically, the bands located between 3350 and 3550 cm^{-1}, for O–H-stretching vibrations, and 2910 and 2965 cm^{-1}, for C–H-stretching, were used. It was found that the water content values determined for the oral SCC samples were significantly higher than the healthy tissue values (Fig. 8.5). A receiver operating characteristic (ROC) curve determined that, using a water content cutoff value of 69%, tumor tissue could be identified with 99% sensitivity and 92% specificity [71]. In a different study, Pachaiappan et al. performed PCA-LDA with LOPO-CV of the HWN region (here, 2500–3500 cm^{-1}) of Raman spectra from blood plasma of 64 individuals. Results showed that oral malignancy could be identified with 92.2% accuracy for the training dataset and 84.4% accuracy for the CV dataset. Analysis of the HWN region allowed researchers to discover additional lipid and water spectral contributions useful for distinguishing the two classes [72]. Further, the HWN region of spectral data for 197 urine samples collected from healthy subjects, oral premalignant, and malignant patients was analyzed using PCA-LDA with LOO-CV. Three different models were built—normal and oral premalignant subjects were classified with 94.9% accuracy, normal and oral malignant groups with 92.1% accuracy, and all three groups with 91.2% accuracy for CV [73]. In a fourth study, Carvalho et al. showed that the HWN region of Raman spectra could differentiate the nucleolus, nucleus, and cytoplasmic areas of oral epithelial cancer, dysplastic, and normal epithelial primary cell lines. The combination of PCA and feature discriminate analysis showed that the cell type could be identified with 99.9% sensitivity and 97.4% specificity using the cytoplasm, 100% sensitivity and 99.1% specificity using the nucleus, and 100% sensitivity and 95.4% specificity using the nucleoli [74]. These studies clearly show the usefulness of the HWN region of Raman spectral data for diagnosing oral cancer; interestingly, these results are generally either comparable or an improvement over those obtained through analysis of other biological samples using only the FP region.

Melanoma and benign melanocytic lesions suspected of melanoma were investigated by Santos et al. Raman bands in the range of 2840–2930 cm^{-1} displayed significant spectral differences between the two groups; PCA-LDA with LOPO-CV of this region could distinguish samples which were considered difficult to distinguish by trained dermatologists. A ROC curved was used to set an optimal discrimination threshold; results showed that melanoma and benign melanocytic lesions often misdiagnosed as being melanoma could be discriminated based on the information found in the C–H-stretching region of HWN Raman data, thus suggesting the potential of the method for improving clinical diagnosis of skin malignancies [75].

While the HWN region provides novel useful and unique information, in several other recent studies, analysis of both the FP and HWN regions was considered for

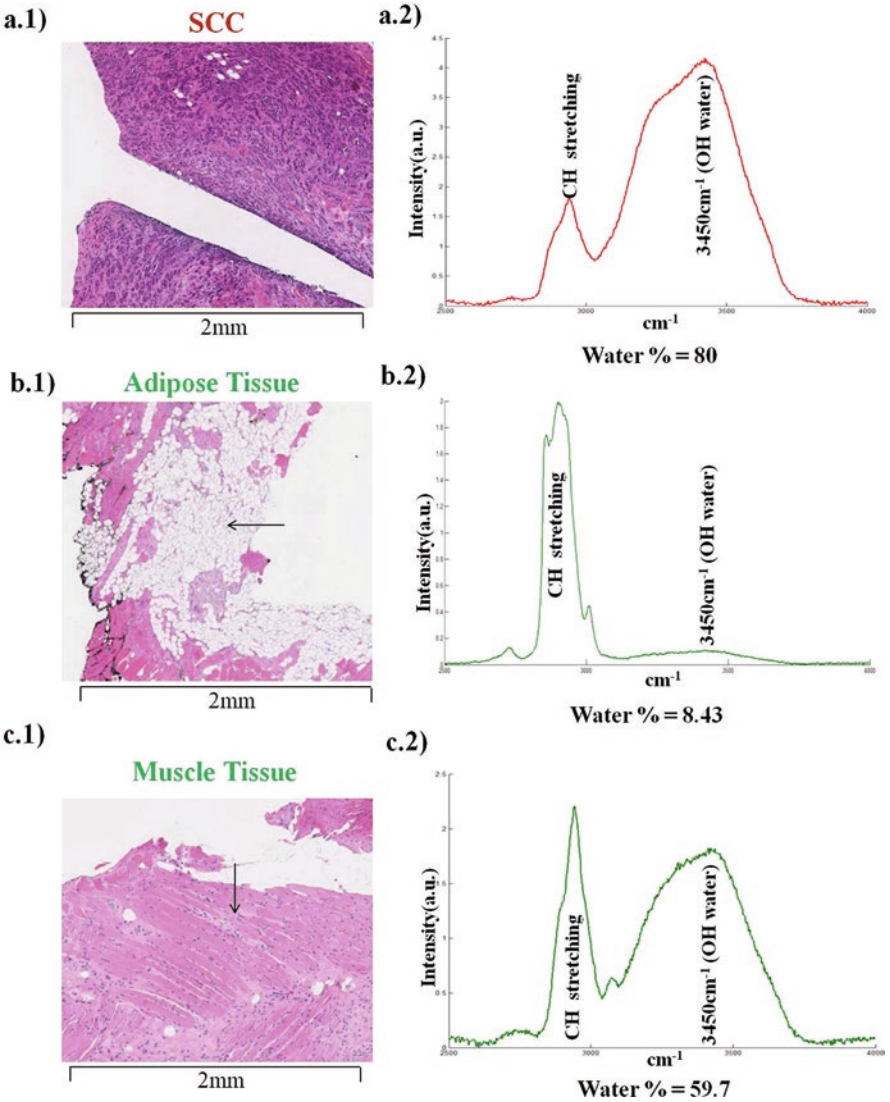

Fig. 8.5 Examples of HWNR spectra measured in (**a.1**) H&E stained thin tissue section of SCC, (**a.2**) typical Raman spectrum of SCC, (**b.1**) H&E stained thin tissue section showing adipose tissue (arrow), (**b.2**) Raman spectrum of adipose tissue, (**c.1**) H&E stained thin tissue section showing muscle tissue (arrow) and (**c.2**) representative Raman spectrum of muscle (Reprinted with permission from [71])

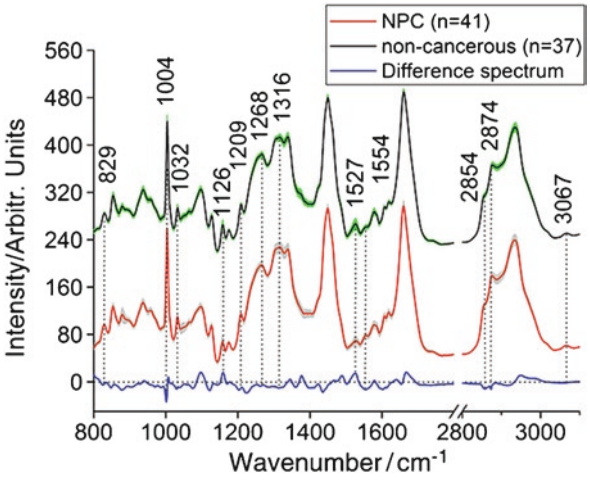

Fig. 8.6 Comparison of the mean intensities of FP/HW Raman spectra from NPC tissue (red line, $n = 41$) versus that of non-cancerous tissue (black line, $n = 37$) in the FP and HW spectral regions. For better visualization, the mean Raman spectra of nasopharyngeal non-cancerous tissue are shifted vertically. The shaded areas represent the respective standard deviations of the means. ((non-cancerous)-cancerous)∗2 was used to represent the corresponding mean difference spectrum (blue line), which is also shown at the bottom. The Raman spectral ranging from 1800 cm^{-1} to 2800 cm^{-1} was not shown by using the broken interval (—//—) to indicate which does not contain tissue biochemical information (Reprinted with permission from [77])

diagnostic purposes. The best results are most consistently seen when the two spectral regions are considered together.

The diagnostic potential of RS for gastric cancer was evaluated through a comparison of the FP and HWN regions. Raman spectra were collected from normal mucosa and gastric cancer tissue areas. Diagnostic algorithms were generated using PLS-DA with LOPO-CV, which yielded 94.59% sensitivity and 86.48% specificity for the FP region and 81.08% sensitivity and 71.05% specificity for the HWN region. Although both regions provide useful information, here, the FP region was better able to diagnose gastric cancer [76].

Huang et al. obtained FP and HWN region spectral data of nasopharyngeal tissue. Non-cancerous tissue was differentiated from cancerous tissue using only FP spectral data (800–1800 cm^{-1}), only HWN spectral data (2700–3100 cm^{-1}), and an integrated FP/HWN dataset (Fig. 8.6). The results, using PCA-LDA with LOPO-CV, showed the samples could be distinguished with 87.8% sensitivity and 86.5% specificity for the FP region, 85.4% sensitivity and 91.9% specificity for the HWN region, and 95.1% sensitivity and 89.2% specificity for the integrated dataset, thus demonstrating the potential of both FP and HWN regions to diagnosis NPC [77]. Sun et al. performed a study with a similar goal, with spectral data collected from biopsy tissue smear samples of 74 patients in the regions of 800–1800 cm^{-1} and 2800–3100 cm^{-1}. Using PCA-LDA of the combined FP/HWN regions dataset, 87.2% sensitivity and 85.7% specificity were achieved for classifying a sample as

belonging to the NPC group or to the non-cancerous group [78]. Clearly, the HWN region provides unique additional information useful for diagnosing cancer based on RS and chemometrics.

8.2.1.5 Raman Hyperspectral Imaging

Spontaneous RS in combination with chemometrics has shown great potential for generating a diagnosis based on the analysis of biological specimens; however, it is important to consider the advantages of Raman hyperspectral imaging (HSI). Raman HSI utilizes an imaging camera to collect additional information regarding the sample being analyzed; as such, the result is the combination of spectral information with spatial information. Raman spectral information is collected from each pixel of an image. The spectral signature from each pixel, or small volume of the sample, depends on the biochemical components present in that small volume—these components can vary within the sample itself and between different samples (i.e., tissue from a healthy donor or tissue from a donor with cancer). In Raman HSI, a three-dimensional (x, y, λ) hyperspectral data cube is formed where the x and y components are spatial dimensions and the λ component is the spectral signature. Together, this information can generate an image which provides information regarding the distribution of biochemical components within the sample. In fact, the hyperspectral images are useful for depicting the relative concentrations of various biomarkers in a biological sample, potentially indicating which areas of the sample, if any, are affected by the disease in question. Furthermore, Raman hyperspectral images of tissue in particular can be compared to the tissue after it has been stained using hematoxylin and eosin (H&E) staining. Thus, Raman HSI is useful for medical diagnostics as it can confirm the presence of a disease, distinguish between normal and diseased samples, and distinguish between disease stages, all through objective analysis. The following studies employ Raman HSI for the purpose of understanding the distribution of biochemical components within samples in order to identify cancer. Additionally, some studies go a step further and utilize advanced statistical analysis to build algorithms for quantitative diagnosis of various cancers.

In a study by Kast et al., the concentrations of Raman spectral bands corresponding primarily to lipid and protein content (1004, 1300:1344, and 1660 cm^{-1}) were imaged across forty brain tissue sections diagnosed as normal, GBM, necrosis, or infiltrating GBM. The goal was to understand the boundaries that exist between gray matter, white matter, and diseased tissue in an attempt to develop a novel method for rapid and non-destructive imaging of brain tissue for cancer diagnosis. The resulting Raman imaging maps corresponded with adjacent H&E-stained sections and could therefore successfully discriminate between the various regions of brain tissue [79].

In the last few years, several manuscripts were published concerning applying Raman HSI for breast cancer analysis. These studies were able to pinpoint valuable differences in biochemistry between diseased and healthy samples, which can be more easily detected with the advantages of HSI. In one study, tumor regions of

breast cancer tissue were discriminated from healthy tissue based on altered concentrations of nucleic acids, collagen, and fat as determined by Raman HSI and KCA. Furthermore, LDA could diagnose ductal carcinoma in breast tissue samples with 95.6% sensitivity and 96.2% specificity. Fresh samples were then subjected to Raman imaging using a selective-sampling strategy in order to decrease data acquisition time based on auto fluorescence imaging (AFI); results were in agreement with the diagnosis made by conventional histopathology [80]. In a different study, live non-malignant, mildly malignant, and malignant breast cancer cells as well as breast cancer tissue were analyzed. Results from Raman HSI suggested that lipid droplets in the various cell lines differ not only in concentration but also in biochemical composition, suggesting their potential role in breast cancer pathology. Differences were observed in the lipid composition within breast epithelial cells as well as in breast tissue. Further, PCA displayed identifiable differences in the Raman signatures of the cells, suggesting a method for predicting the state of the oncogenic pathway [81]. Brozek-Pluska et al. showed that RS and Raman HSI could detect relative amounts of acetylated and methylated lysine, which have been previously designated as biomarkers for breast cancer. The stretching vibration of the acetyl group observed near 2938–2942 cm^{-1} and of the methyl group around 2970 cm^{-1} allowed these molecular changes occurring in human breast tissue cancer cells to be monitored. Further, PLS-DA with CV provided 85.3% sensitivity and 91.3% specificity for detecting cancer [82].

Vanna et al. successfully distinguished the four subtypes of acute myeloid leukemia (AML), which include myeloblasts, promyelocytes, abnormal promyelocytes, and erythroblasts. Bone marrow samples of seven patients, each affected with one of the four AML subtypes, were collected. For each cell isolated from the bone marrow aspirate, 4096 spectra were collected in order to generate Raman images which could accurately demonstrate morphological features. When the Raman images were analyzed by HCA, automatic discrimination and localization of the nucleus, cytoplasm, myeloperoxidase-containing granules, and hemoglobin was achieved. The images provide additional biochemical information than what could be obtained using only spontaneous RS. Following this, the average Raman fingerprint of each cell was analyzed by PCA-LDA with LOO-CV. Myeloblasts, promyelocytes (both abnormal and normal), and erythroblasts were differentiated with 100% accuracy. Normal and abnormal promyelocytes were correctly classified with 95% accuracy, and all four subtypes could be classified with 98% accuracy [83].

Interestingly, when Raman HSI was used to study liver cancer specimens, the results were a great improvement and provided useful biochemical information as compared to those results obtained through spontaneous RS tissue analysis. Two liver cancer cell lines, HepG2—including HepG2 cells in different cellular growth phases—and SK-Hep1, were analyzed by Tolstik et al. The collected spectral data was used to generate color-coded images which were analyzed by HCA and PCA; this provided significant information regarding the biochemical composition of the samples. Spectral differences were mainly attributed to higher expression of unsaturated fatty acids in the HCC cells as well as during the proliferation phase of cellular growth. Through SVM analysis with CV, previously unknown cells were classified

as belonging to one of the two cell lines with 93% accuracy. Predictions of the unknown proliferation phase for HepG2 cells showed 100% sensitivity and 98% specificity. Raman HSI uniquely provides information regarding cell type and proliferation behavior, which are essential tools in identifying features of malignant tumors [84]. In a second study by the same group, Raman imaging of liver tissue was used to identify molecular information beneficial for diagnosing liver cancer. The most notable difference between HCC and fibrosis regions of tissue was found to be due to fatty acids, especially palmitic acid. A RF model with CV classified malignant and non-malignant tissue regions with 86% accuracy [85]. More recently, Ryabchykov et al. discriminated three different cell lines (HepG2, nondifferentiated hepatic stem cell line HepaRG, and differentiated hepatocyte-like HepaRG) using Raman HSI. KCA was used to visualize clusters of different cell components within the cells. Following this, a three-class LDA with LOO-CV model was constructed to achieve cell line classification, reaching 96% accuracy [86].

Raman HSI shows great capabilities for detecting oral cancer, in both paraffin-free and paraffin-embedded tissue. Oral SCC and healthy tissue samples were analyzed to assess the potential of RS to perform discrimination tasks at the histological level. 127 Raman images were generated from 25 unstained thin tissue sections; the images were comparable to corresponding histological evaluation obtained through H&E staining. After imaging, the spectra were labeled as cancerous or as a surrounding healthy tissue structure (squamous epithelium, connective tissue, adipose tissue, muscle, gland, or nerve) (Fig. 8.7). LDA models were built to analyze the labeled spectra for classification purposes. A total of six binary LDA models were built to distinguish oral SCC spectra from each of the surrounding healthy tissue structures, achieving an overall average accuracy of 93.17% [87]. In a unique study, Meksiarun et al. aimed to first understand if multivariate methods could extract the paraffin component of paraffin-embedded oral cancer tissue spectra. Typically, oral SCC tissue will be removed from a patient, fixed with formalin, and embedded in paraffin to prevent degradation. However, the Raman spectral features of paraffin overlap with main Raman spectral tissue bands, including the amide I and III bands. Three methods were tested for their ability to remove the paraffin spectral features while maintaining the integrity of the rest of the Raman spectrum, including PLS, independent component (IC) analysis, and IC-PLS. All methods were successful, however, PLS and IC-PLS were the most successful at removing the paraffin spectral component while still maintaining spectral integrity of the cancer tissue. The paraffin-removed spectra obtained via IC-PLS were analyzed by PCA to construct Raman images. Main Raman markers for discriminating healthy and malignant tissue were found to be collagen, phosphate, and DNA. The produced Raman images showed similarity to H&E stained tissue, thus demonstrating the ability of Raman HSI to diagnose oral cancer in paraffin-embedded tissue [88].

Human prostatic cells were collected and analyzed using Raman HSI. An emphasis was placed on the C–H vibration region (2800–3100 cm^{-1}) of the spectra due to its ability to pinpoint the main differences between normal and tumor cell lines. PCA was used for image processing and identified protein and lipid fractions which were important for differentiation. A self-modeling curve resolution (SMCR)

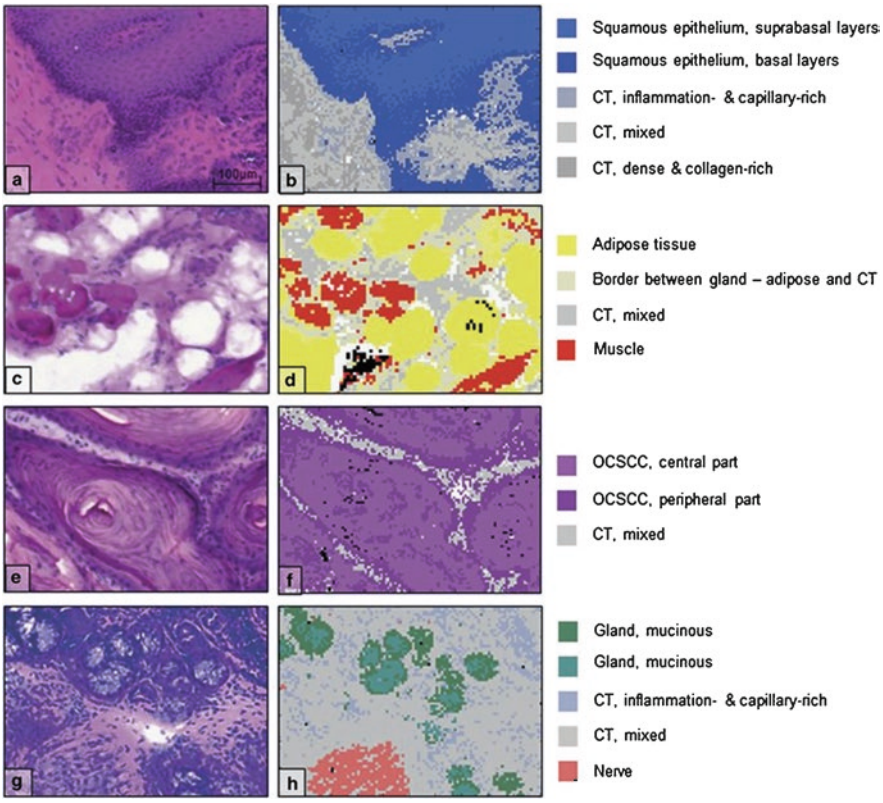

Fig. 8.7 H&E stained sections and corresponding pseudo-color Raman images. H&E-stained tissue sections (**a**, **c**, **e,** and **g**) and corresponding pseudo-color images (**b**, **d**, **f,** and **h**). The K-means cluster averages were annotated as one of the following tissue structures: OCSCC (central part, peripheral part, or n.o.s.), squamous epithelium (superficial layers, suprabasal layers, or basal layers), CT (dense and collagen-rich, mixed, or inflammation- and capillary-rich), gland (mucinous or serous), muscle, adipose tissue, or nerve. CT = connective tissue; n.o.s. = not otherwise specified; OCSCC = oral cavity squamous cell carcinoma (Reprinted with permission from [87])

algorithm was also employed and revealed tumor cells experience a 97% increase of the lipid fraction with respect to the control cells. Analysis by least squares curve fitting gave reproducible results for identifying differences at the molecular level between normal and tumor cells [89].

Raman imaging was performed on healthy and neoplastic thyroid tissue to improve the diagnosis of PTC (Fig. 8.8). Biochemical features of PTC were characterized by the significant presence of carotenoids in comparison to healthy tissue. LDA with LOO-CV was applied to estimate tissue classification. Healthy and PTC thyroid tissue were discriminated with 100% accuracy and classical and follicular variants of PTC were discriminated with 95% accuracy [90]. The performance of this study is generally an improvement over analysis of tissue by spontaneous RS alone.

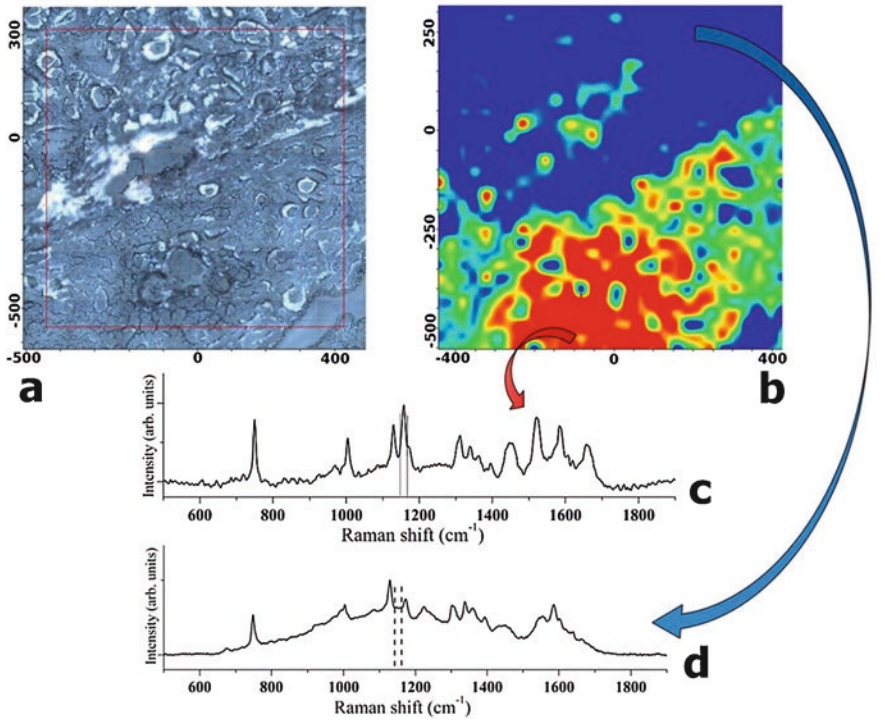

Fig. 8.8 Typical example of Raman chemigram map (1156 cm^{-1} band reference) of a mixed zone of thyroid tissue (blue-healthy; red-yellow-green–PTC): (**a**) dark field optical image, (**b**) Raman map, (**c**) average reference Raman spectrum corresponding to healthy tissue, (**d**) average reference Raman spectrum corresponding to PTC tissue. The red square on the right side (**a**) corresponds to the investigated tissue area shown on the left (**b**). The scale bars are expressed in μm (reprinted with permission from [90])

Raman HSI capitalizes on the advantages of spontaneous Raman spectroscopy while providing additional tools that can be used in diagnosing cancer. The images produced are comparable to those produced by H&E staining and the information provided is useful for understanding presence of a disease as well as the stage of a disease. Importantly, Raman HSI is able to identify incredibly valuable biochemical differences between healthy and diseased samples, further enabling the identification of biochemical changes that occur during pathogenesis as well as potential novel biomarkers that have not yet been considered. What's more, Raman HSI opens to the door for *in vivo* applications where the images can indicate tumor location which can be useful for surgical procedures.

8.2.2 *Spontaneous Raman Spectroscopy Combined with Other Analytical Techniques*

In an attempt to increase the amount of useful information obtained for cancer diagnostics, some research efforts have focused on combining RS with other analytical techniques. Ideally, these additional methods will provide complimentary information to that obtained by Raman spectroscopic analysis and will increase the confidence and statistical significance of the methodology for diagnosing cancer.

RS was used to study 12 healthy and 30 tumor bladder tissue samples. Using HCA and differences in peak ratios, the tissue type could be classified with 96.7% sensitivity and 66.7% specificity. Major differences between the two classes included higher tryptophan, cholesterol, and lipid content levels in healthy tissue, and increased levels of nucleic acids, collagen, and carotenoids in bladder tumor tissue. High-performance liquid chromatography (HPLC), an analytical technique useful for separating, identifying, and quantifying individual components within a mixture, was employed to analyze carotenoids extracted from the two tissue types. While the Raman spectra reflect contribution due to carotenoids, HPLC was able to further narrow down this contribution to a specific biomarker; it was found that β-carotene was the major carotenoid present in tumor tissue, marking the first time this biomarker has been identified for bladder cancer [91].

RS and Raman HSI were combined with atomic force microscopy (AFM) to discriminate brain tumor from normal brain tissue samples. AFM, a type of scanning probe microscopy, was used to obtain nanomechanical properties to form images of healthy and cancerous brain tissue, while RS was used to glean information regarding the biochemical composition of the tissues. High-grade medulloblastoma (grade IV) and non-tumor samples from tissue of the central nervous system were compared. After analyzing the Raman spectra and images, it was determined that proteins within medulloblastoma tumors exist in the β-sheet conformation at enhanced levels and in the α-helix conformation at decreased levels as compared to proteins within normal tissue. Upon comparison of Raman peak ratios, it was discovered that in normal brain tissue, the relative amount of lipids compared to proteins is considerably higher. Mechanical indentation by AFM discovered that medulloblastoma tissue mechanical properties are strongly heterogeneous. Lastly, RS data was analyzed using PLS-DA with CV, indicating 96.3% sensitivity and 92% specificity for separating the two tissue types. Through combination of Raman HSI and AFM, the biochemical and nanomechanical signatures obtained have the potential to identify biomarkers associated with the development of brain cancer [92]. Although these results are comparable to studies which use only spontaneous RS, the added information that is obtained improves the usefulness of the methodology for diagnosing brain cancer.

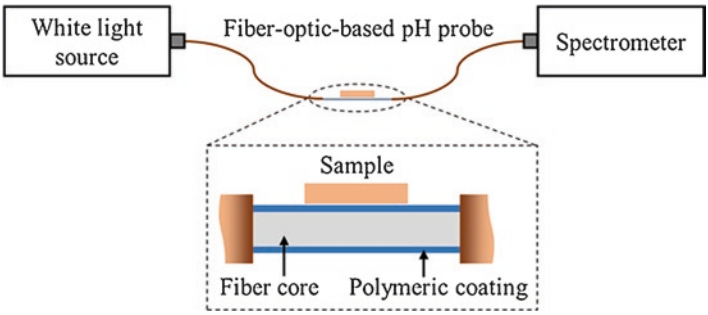

Fig. 8.9 Experimental setup using the fiber-optic-based pH probe for measuring the transmission spectra varying according to the pH level of the normal and cancerous breast tissue samples. The intensity of transmission spectra varies due to the absorbance change of the polymeric coating layer, which consists of neutral red/poly(acrylic acid) bilayers (Reprinted with permission from [93])

A novel approach combining spontaneous RS and optical pH sensing was used to differentiate healthy and cancerous breast tissue. To better prepare the Raman spectroscopic method for *in vivo* cancer detection, pH sensing can be first used to detect areas of tissue with lower pH levels, which is associated with cancer, thus ideally increasing the accuracy of the method as opposed to just using RS (Fig. 8.9). Fiber-optic-based Raman and pH probes were used to evaluate tissue samples; the pH sensing is based upon the pH level's dependence on the optical transmission spectrum. Raman spectra were collected first, followed immediately by collection of transmission spectra using the optical pH probe. The Raman spectra were combined with the transmission spectra from the same sample. PC-LDA with LOO-CV was employed for classification, first using only Raman spectra and then using the combined pH-Raman spectra. When Raman spectra were analyzed alone, the algorithm achieved 100% sensitivity and 91.5% specificity. When the algorithm analyzed the combined pH-Raman spectra, 100% sensitivity and 98% specificity were achieved, indicating the added advantage of pH sensing for diagnosing breast cancer using RS [93].

Both Raman and infrared (IR) spectroscopies were used in combination in several studies for the purpose of identifying various types of cancer. IR spectroscopy, another vibrational spectroscopic technique, is known to provide complimentary information to that obtained by RS. Specifically, IR spectroscopy analyzes the interaction of IR light with a molecule, generating an IR spectrum of energy that is absorbed or transmitted by the molecule as a function of either frequency or wavelength of light. The spectral information can then be used to identify and study the sample. The vibrational signatures of 164 invasive ductal carcinoma and invasive lobular carcinoma breast tissue samples were analyzed by both Raman and IR spectroscopies for the purpose of discriminating non-cancerous and cancerous tissue. Here, KCA followed by PCA and PLS-DA with CV were used to analyze the Raman spectral data. Raman imaging identified differences in spectral regions corresponding to vibrations of carotenoids, fatty acids, and proteins between normal

and cancerous tissue, while IR spectra depict differences in proteins and phospholipids. Results of statistical analysis showed 84.7% sensitivity and 71.9% specificity for determining if breast tissue displayed either normal biochemistry or cancer pathology [94]. Owens et al. aimed to determine whether attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy or RS could better characterize the biomolecular signatures of blood plasma or serum collected from patients with ovarian cancer as compared to healthy controls. FTIR is used to simultaneously collect data over a wide spectral range; the ATR attachment allows for surface properties of a sample to be measured rather than bulk properties, thus decreasing the potential for strong attenuation of the IR signal in samples that are highly absorbent. Here, 60 blood samples were analyzed using ATR-FTIR spectroscopy, while only 8 samples were studied using RS. All spectra were subjected to PCA-LDA, which showed statistically significant differences between healthy and cancerous samples using both spectroscopic methods. A SVM classifier successfully differentiated Raman spectral data of blood plasma with 74% accuracy; notably, the IR spectral data of blood plasma was successfully classified with 93.3% accuracy. It was further found that blood plasma was better suited for diagnostic discrimination than blood serum. Although ATR-FTIR spectroscopy is shown here to better diagnose ovarian cancer, one should consider the different sample sizes used in each part of the experiment [95]. In another study, Raman and ATR-FTIR spectroscopies were used to determine if either could identify the primary site of a metastatic tumor. Metastases were obtained from primary lung and colorectum AC as well as from metastatic melanoma. PCA-LDA determined points of dissimilarity between spectra; PCA in combination with a linear discriminate classifier (LDC) calculated classification accuracy. In a three-class algorithm built using Raman spectral data, 69% accuracy for predicting colorectal AC, 69% for lung AC, and 72% for melanoma were achieved. Using ATR-FTIR spectral data, 60% accuracy for predicting colorectal AC, 59% for lung AC, and 47% for melanoma were achieved. Interestingly, combination of the two AC groups improves results to 85% accuracy for predicting AC and 75.4% for melanoma using the Raman data and to 96% accuracy for AC and 72% for melanoma using the ATR-FTIR data [96]. In general, IR spectroscopy performs similarly to, if not better than, RS in these studies. It should be noted that differences in sample sizes may play a role, and that spontaneous RS has already been shown in other studies to successfully diagnose these same cancers.

In a unique study by Tatarkoič et al., blood plasma samples from 55 individuals were investigated using a combination of electronic circular dichroism (ECD), Raman optical activity (ROA), and conventional Raman and FTIR spectroscopies for the purpose of diagnosing colon cancer. ECD is a useful technique for analyzing stereochemistry; an ECD spectrum is the difference between absorption of left and right circularly polarized lights due to electronic transitions in the UV or visible regions of the spectrum [97]. Similarly, ROA measures the difference in intensity of Raman scattered left and right circularly polarized light which arises because of molecular chirality [98]. These techniques help to provide more specific information regarding the biochemical composition of a sample in order to better increase

the ability to identify cancer with chemometrics. The results of LDA showed that, for each of the individual methods, limited discrimination between control group subjects and patients with colon cancer was achieved. However, when spectra from all methods were combined and again evaluated using LDA with LOO-CV, sensitivity and specificity reached 93% and 81%, respectively, with an overall accuracy of 87% for discriminating the two classes of blood plasma samples [99]. Despite the combination of so many techniques, the performance of the model is not necessarily a significant improvement over those built using spontaneous RS data of various biological samples.

Lin et al. also used the combination of several different techniques, this time for diagnosing NPC. Here, a 4-modality endoscopy system comprised of white light imaging (WLI), AFI, diffuse reflectance spectroscopy (DRS), and RS was used for *in vivo* NPC detection. WLI can locate suspicious lesions, but has low diagnostic sensitivity and relies on subjective analysis. AFI, which has a higher diagnostic sensitivity, has the ability to monitor biochemical changes that occur in tissue based on the fluorescence profile of internal fluorophores which are associated with cancer progression. DRS can improve AFI by providing morphological and functional quantitative information regarding the tissue samples. RS and DRS data were collected from patients with NPC and from healthy subjects under the assistance of AFI and WLI. When the combined DRS/RS dataset was applied to PCA-LDA, the algorithm achieved 98.6% sensitivity and 95.1% specificity for separating the two groups of tissue samples, showing the usefulness of combining multiple methods to improve results [100].

The combination of RS and AFI was used in multiple studies to diagnose skin cancer. Zakharov et al. used fluorescence analysis first to quickly scan large areas of tissue samples for abnormality detection; when malignancy was suspected, Raman spectral analysis of the tissue was performed. Quadratic discriminant analysis (QDA) of the data provided a diagnosis of malignant melanoma with 89% sensitivity and 87% specificity [101]. In a proceeding study, Raman and auto fluorescence (AF) spectroscopies were used to identify skin neoplasms as melanoma, BCC, or benign tumors. Here, the Raman and AF signals were combined and analyzed via PLS-DA with LOO-CV. Results showed 98.3% accuracy for separating malignant and benign tumors [102]. Similarly, Bratchenko et al. differentiated skin melanoma and BCC tissue samples through the combination of Raman and AF spectra (Fig. 8.10). When considered separately, neither set of spectra was able to exceed 79% accuracy; however, PCA-DA analysis of a combined spectral dataset with six selected spectral features provided 97.3% accuracy for malignant skin detection [103]. Interestingly, these studies each show how AF can be used to increase the reliability of the RS method for diagnosing skin cancer.

While RS is oftentimes suitable for identifying cancer by itself, analysis can oftentimes be improved when additional methodologies are combined, as is displayed by the previously mentioned studies. However, it is important to note that the combination of multiple techniques increases the level of complexity of the methodology as well as potentially increasing the time, cost, and effort required to

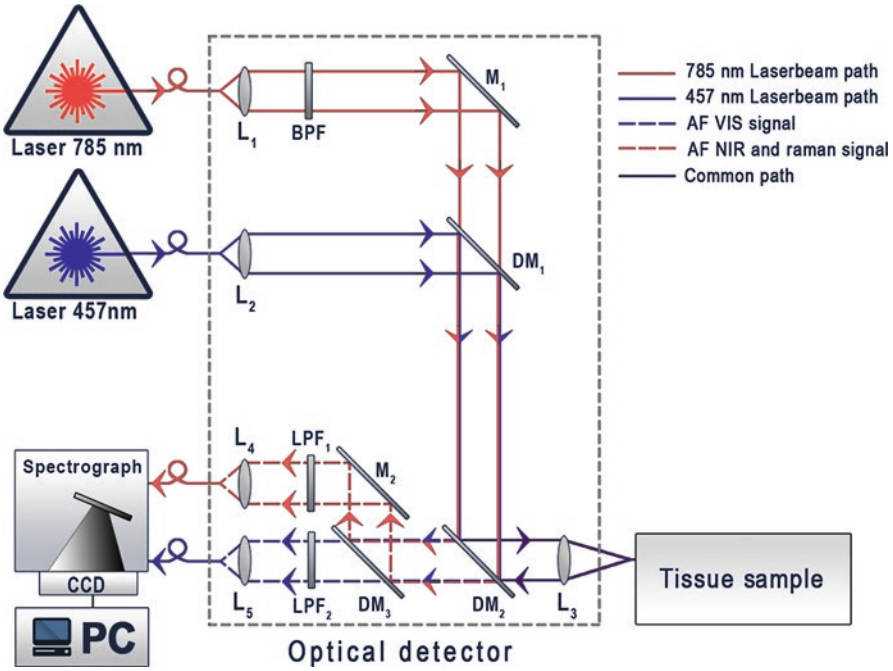

Fig. 8.10 Experimental setup: L_1, L_2, L_4, and L_5: matching lenses, L_3: focusing lens, BPF: band-pass filter, M_1 and M_2: mirrors, DM_1, DM_2, and DM_3: dichroic mirrors, and LPF_1 and LPF_2: long-pass filters (Reprinted with permission from [103])

achieve a diagnosis, indicating the importance of considering the costs and the benefits that accompany a more complex methodology system.

8.2.3 Modifications of Spontaneous Raman Spectroscopy

While conventional spontaneous RS has shown it is capable of diagnosing different cancers, some studies have advocated for the implementation of modifications of the technique. Variations of spontaneous RS have been proposed as effective methods to enhance diagnostic efforts. Those methods considered here are simple modifications of conventional RS; inclusion of techniques such as surface-enhanced and tip-enhanced RS are beyond the scope of this review chapter.

Fullwood et al. employed immersion Raman spectroscopy (IRS) to investigate brain cancer. Because IRS utilizes a specific immersion lens, the lens can have direct contact with a specific liquid; in this study, the liquid used was deionized water which covered the tissue sample being studied. Immersion of the sample in liquid protects the tissue from potential photo-damage and increases the spectral quality by reducing contribution of stray light to the spectral background. Both

spontaneous RS and IRS data were collected from 48 tissue samples. It was determined that a lower background contribution was observed in the IRS data as compared to the RS data. A PC-LDA diagnostic algorithm was therefore built using the IRS data which could successfully discriminate between normal, GBM, and metastatic brain tissue spectra. Following this, researchers effectively distinguished different primary sites of cancerous tissue and investigated the biochemical differences between primary and metastatic cancer using samples from the same patient [104].

The majority of previously mentioned manuscripts have used dispersive RS; in a paper published in 2017, Fourier transformation (FT)-NIR RS was used to diagnose oral epithelial dysplasia. FT-NIR RS excites samples using a laser, such as the Nd:YAG used in this study, at a wavelength of 1064 nm; excitation in the IR region of light helps eliminate fluorescence but provides a weaker Raman signal. To adjust for this, an interferometer is used to convert the Raman signal to an interferogram which allows the entire Raman spectrum to be collected simultaneously by the detector, improving the signal-to-noise ratio. The FT algorithm then converts the interferogram to a conventional Raman spectrum. In this study, the goal was to differentiate normal oral mucosa, oral SCC, and dysplastic tissue samples. After spectra were collected, a SVM classifier was built and results were verified using PCA-LDA. Through SVM, accuracies for distinguishing mild, moderate, and severe dysplasia from oral SCC were 100%, 44.44%, and 71.15%, respectively. PCA-LDA analysis did not allow for successful discrimination of the stages, either, suggesting the need for improvements to the classification system. However, PCA-LDA could still identify biochemical discrepancies between normal, oral SCC, and dysplastic tissue samples [105]. Interestingly, these results are not necessarily an improvement of those performed using spontaneous RS analysis of biological specimens.

Coherent anti-Stokes Raman scattering (CARS) imaging was used in an attempt to diagnose both bladder and lung cancer. Similar to spontaneous RS, CARS is sensitive to molecular vibrational modes. Dissimilarly, three laser beams will each emit photons of particular frequencies to produce a coherent optical signal, at the anti-Stokes frequency, with the goal of producing a much stronger signal as compared to normal RS. Weng et al. used CARS to collect cellular-level images of normal and cancerous lung tissue samples. A deep convolutional neural network (DCNN) learning algorithm automatically differentiated normal, small cell carcinoma, AC, and SCC lung images with 89.2% accuracy [106]. Yosef et al. collected both CARS and second harmonic generation (SHG) images. The CARS imaging of urine sediments was used to preselect urothelial cancer cells. Next, Raman HSI of the cells was performed (Fig. 8.11). Through HCA, it was found that the cancer cells displayed a decrease in glycogen and an increase in fatty acid levels as compared to healthy controls. A RF classifier was built which could identify cancerous urothelial cancer cells based on the analysis of full cells or cytoplasm with 100% accuracy and based on nuclei with 90% accuracy after LOPO-CV [107]. The results of using CARS for diagnosing cancer are inconsistent, seeming to depend on the type of sample analyzed and the cancer being targeted.

Shifted-excitation Raman difference spectroscopy (SERDS) was employed as a label-free and non-invasive method for diagnosing oral SCC. During SERDS

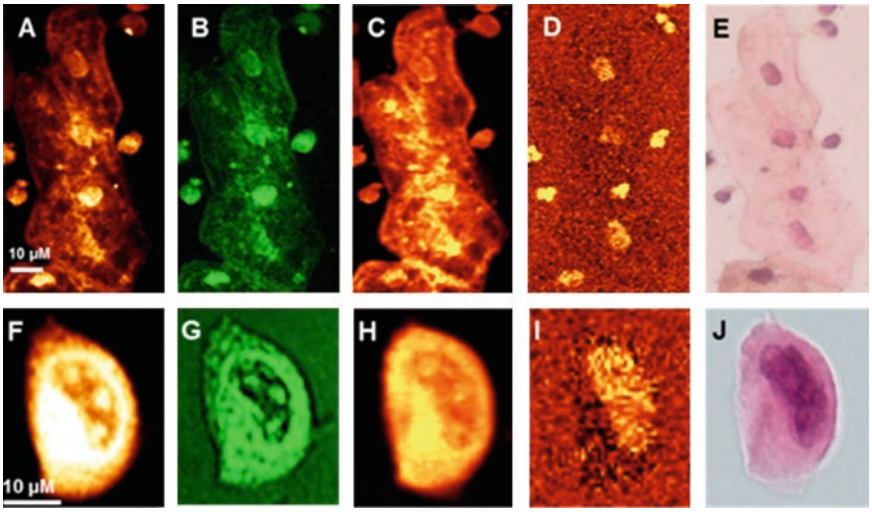

Fig. 8.11 Different imaging techniques applied to normal (**A**–**E**) and high-grade cancerous (**F**–**J**) urothelial cells in urine sediments: (**A**, **F**) SHG images, (**B**, **G**) CARS images, (**C**, **D**, **H**, **I**) integrated Raman intensity of cells in the (**C**, **H**) 2800–3050 cm^{-1} and (**D**, **I**) in 785–805 cm^{-1} regions, and (**E**, **J**) H&E-stained images (Reprinted with permission from [107])

measurements, spectra are first collected when the wavelength is set to a particular number (here, 783 nm). Then, the excitation wavelength undergoes a small shift to a second number (here, 785 nm), and a second spectral dataset is acquired. The dataset collected at each wavelength is averaged and the mean spectrum using the first excitation wavelength is subtracted from the mean spectrum collected using the second, ideally removing any contribution from fluorescence emission. In this manner, 72 SERDS spectra were collected, one from each of three different physiological tissue points and three different pathological lesions from 12 different patients. The SERDS spectra of malignant and benign tissues were discriminated using PCA-LDA, which achieved 86.1% sensitivity and 94.4% specificity for diagnosing oral SCC [108]. Although these results indicate success, it should be noted that many other research groups were able to accomplish similarly effective outcomes with much more simple RS technology.

Polarized Raman spectroscopy (PRS) was used by Daniel et al. for observing the biomolecular structural changes that occur in cervical cancer tissue samples. PRS is observed as the result of polarized light interacting with vibrating molecules, where the polarization is either parallel or perpendicular to the excitation laser's intrinsic polarization. Here, PRS provided information regarding the differences in orientation of biomolecules such as tyrosine, collagen, and DNA between normal and malignant cervical tissue samples. Depolarization ratios were analyzed by LDA with CV, yielding sensitivity, specificity and accuracy levels of 96%, 97.2%, and 96.7%, respectively. This is an improvement over the 92% sensitivity, 72.2% specificity, and 80.3% accuracy achieved using only spontaneous RS [109].

Stimulated Raman scattering (SRS) microscopy was used to reveal the infiltration of brain tumors in fresh surgical specimens collected from 22 neurosurgical patients. The basic mechanism of SRS is similar to spontaneous RS; however, SRS can enhance the signal of specific vibrational transitions due to the introduction of a second photon, a Stokes photon at a particular frequency, which can stimulate a specific molecular transition. By maintaining the pump laser beam at a constant frequency and scanning the Stokes laser beam, the spectral fingerprint of the sample can be improved. Here, results of SRS were in near-perfect agreement with results of H&E light microscopy. The data was interpreted using quasi-likelihood generalized additive models. Based on cellularity, axonal density, and protein/lipid ratios observed in SRS images, the classifier could successfully detect tumor infiltration with a sensitivity of 97.5% and specificity of 98.5%. The classifier was also able to distinguish between various categories of tumor infiltration including normal to minimal hypercellularity, infiltrating glioma, or dense glioma with high levels of accuracy [110]. Stimulated Raman histology (SRH) was used in a complimentary study for the intra-operative diagnosis of pediatric-type brain tumors. Based on RF analysis, 25 pediatric-type surgical specimens were correctly classified as normal versus lesional tissue and low-grade versus high-grade tumors all with 100% accuracy [111]. These results are generally better than those obtained through analysis of various biological samples by spontaneous RS for brain cancer diagnosis.

High-throughput (HT) RS was used for rapid screening of blood plasma samples collected from prostate cancer patients and healthy volunteers. In general, HT screening methods have the ability to automatically control and conduct millions of tests with a specific goal, saving time and effort for the user. Medipally et al. developed a HT-RS method which was optimized through testing a series of different instrumental and sample preparation parameters (Fig. 8.12). Once adjusted, the method was able to automatically record multiple Raman spectra from each of the well throughputs in a 94-well plate. To test the method, Raman spectra were obtained for blood plasma collected from 10 healthy volunteers and 10 prostate cancer patients using both 785 and 532 nm excitation. The best results were seen using the 785 nm excitation, with PCA-LDA yielding 96.5% sensitivity and 95% specificity after CV, demonstrating the ability of HT screening methods to be successfully incorporated with RS methodology [112].

A unique study performed by Stables et al. classified brain tumor spectra using spontaneous RS in combination with sound and listening tests. Metastatic brain cancer, glioblastoma, and non-cancer tissue samples were analyzed using RS. Three different chemometric techniques (SVM, KNN, and LDA) with CV were evaluated for their potential to identify brain cancer within the tissue samples using a feature extraction approach. Compared to using PCA for spectral dimensionality reduction, the feature extraction approach increased classification accuracy of the KNN classifier by 25% to 91.02% and of the SVM classifier by 26.25% to 97.01%. For LDA, the classification accuracy decreased from 96.54% to 95.38%. The results suggest feature extraction to be a more effective approach as opposed to dimensionality reduction for classification efficiency. Sonification was then used on the reduced Raman dataset of extracted features. Frequency modulation synthesis was used to

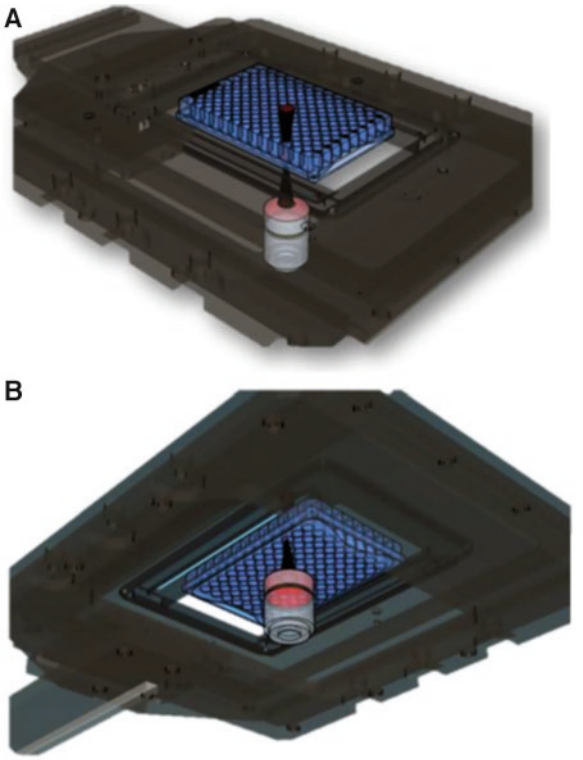

Fig. 8.12 Schematic representation of HT-Raman spectroscopy method. (**A**) Top view, (**B**) bottom view (these schematics are developed using Google Sketch up software) (Reproduced from [112] with permission from The Royal Society of Chemistry)

generate audio clips for each tissue sample based on the subset of extracted features, thus giving each one its own sound timbre, with similar tissue types having similar timbres. Listening tests were implemented with 25 participants, and based on the sound timbres, a mean classification accuracy of 71.1% was achieved, presenting a novel tool which can be used in addition to RS for clinicians to generate a diagnosis during endoscopic procedures [113].

Interestingly, not all modifications of spontaneous RS were completely effective in improving diagnostic accuracy. It should be observed that, on the other hand, there were some studies which did provide improvements. However, with the addition of more sophisticated methodology comes an increase in difficulty for bringing the technology into clinical settings. The more complicated the method, the less likely it is to be introduced as a new technology for universal cancer detection. While these aforementioned studies provide unique variations of RS, spontaneous RS alone has still shown great success in diagnosing cancers, suggesting the previously summarized alterations of the methodology may not ultimately be necessary for bringing the method to clinical settings.

8.2.4 *Fiber-Optic Studies*

To reach the ultimate goal of *in vivo* diagnoses, probes have been increasingly incorporated into RS studies. Special instrument setups have been created in which a fiber-optic probe can analyze tissue *in vivo* and collect Raman spectral data. Fiber-optic probes have the advantage of being less bulky and less expensive than typical Raman spectrometers; probes can be used intra-operatively, preventing the need for additional biopsy or *ex vivo* studies. They provide a shorter analysis time than typical histopathological examinations of biopsied tissue do, while still objectively capturing vital biochemical compositional changes that occur during disease progression. Probes can also allow for the identification of tumors, signaling where a surgeon should make excisions. When Raman spectral data collected through probes is analyzed using advanced statistical methods, research scientists are able to greatly reduce false positive biopsy results and increase the ease of and success of diagnosing cancers. Recent research which has incorporated probes into Raman spectroscopic systems, through either *in vivo* or *ex vivo* studies, for the purpose of diagnosing cancer are reviewed here. A schematic of a general fiber-optic probe setup is seen in Fig. 8.13.

Chen et al. used a low-resolution fiber-optic Raman sensing system to evaluate its diagnostic potential for *ex vivo* identification of different bladder pathologies. Spectra of 32 normal bladder tissue and low- and high-grade tumor bladder tissues were analyzed using a PCA fed ANN with CV. An overall accuracy of 93.1% was obtained for predicting to which class a sample belonged, introducing the possibility for further experiments to be successfully conducted *in vivo* [114].

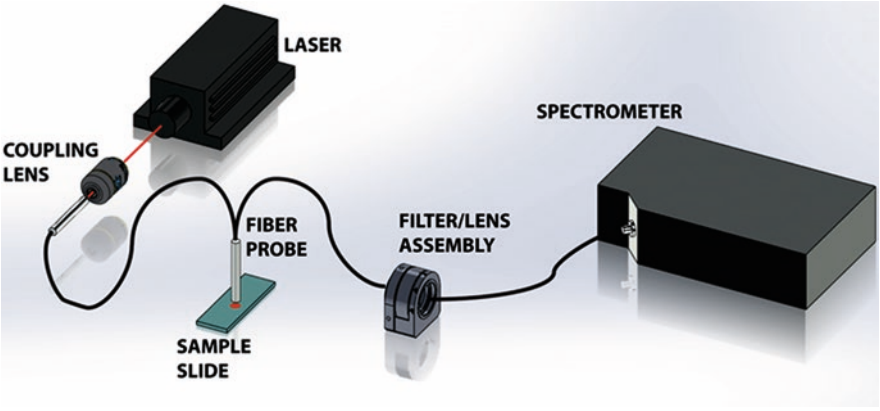

Fig. 8.13 Schematic of the experimental setup showing the 785 nm laser directed into the Raman probe via the 10× objective lens. The probe illuminates the tissue sample and collects the scattered light. The elastically scattered signal is removed via a long pass filter in the filter/lens assembly before the light is transmitted into the Maya Pro 2000 NIR spectrometer for dispersion and storage (Reprinted with permission from [153])

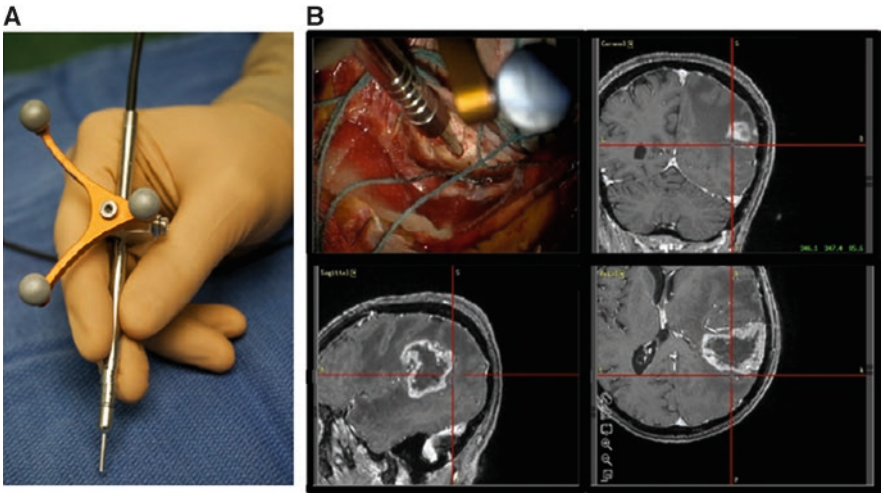

Fig. 8.14 Raman spectroscopy system for intra-operative detection. (**A**) Photograph of the handheld contact probe, with the attached neuronavigation tracking unit. (**B**) Illustration of the probe being used intra-operatively, with the neuronavigation system showing the location of the tip of the probe (cross hairs) on the preoperative magnetic resonance images (Reprinted with permission from [115])

Three different studies show the strength of Raman probe systems for diagnosing brain cancer *in vivo*. A handheld RS probe system was used to collect *in vivo* spectral data from normal, cancerous, and necrotic brain tissue of ten patients (Fig. 8.14). Using PCA and a boosted trees (BT) classification algorithm with LOO-CV, an accuracy of 87% for distinguishing necrosis from tumor and normal brain tissue was achieved [115]. Jermyn et al. used a handheld contact RS probe to differentiate normal brain, dense cancer, and normal brain invaded by cancer cells with 93% sensitivity and 91% specificity, using the BT machine learning method with CV. The RS probe system was also able to detect previously unidentifiable invasive brain cancer cells in patients with grade II through IV gliomas, showing the usefulness of fiber-optic probes for detecting cancerous cells which are oftentimes missed during normal surgery [116]. In a following study, RS data was collected intra-operatively from 17 patients with grades II through IV gliomas. Both BT and ANN were used for classifying the spectra. ANN performed better than BT when algorithms for distinguishing cancer from normal brain were built including light artifacts but performed the same when they were built excluding light artifacts due to operating room sources. Specifically, when light artifacts were excluded from the spectra, ANN achieved 92% classification accuracy, an improvement over 90% accuracy achieved when light artifacts were not excluded [117].

Li and co-researchers performed several studies using a miniature Raman spectrometer equipped with a fiber-optic probe for the purpose of diagnosing breast cancer. In the earliest study, 16 breast tissue samples were analyzed and an adaptive weight k-local hyperplane (AKWH) algorithm was used for differentiation. Three

different data processing schemes were generated based on varied splitting of the Raman spectral dataset; on average, the AWKH algorithm gave a 95.8% accuracy for classifying breast tissue as either cancerous or healthy [118]. The same samples were then analyzed using an adaptive net analyte signal AWKH pattern recognition method. Again, three different data processing mechanisms were generated based on different splitting of the Raman spectral dataset; the average accuracy of classification was 94.83% [119]. In their last study, new normal and malignant breast tissue samples were obtained, with the cancerous tissue existing at various stages of the disease. An adaptive local hyperplane K-nearest neighbor method was used for binary classification, achieving 93.2% accuracy [120]. While these successful results are generally comparable to those previously reviewed, they indicate the vital diagnostic potential of RS to be used intra-operatively.

A fiber-optic Raman system was used to obtain 68 spectra from benign and low- and high-grade SIL of 25 cervical tissue specimens. Multiclass PLS-DA with LOPO-CV showed an average sensitivity of 86.6% and specificity of 93.6% for classification [121]. Shaikh et al. performed two studies to explore *in vivo* classification of normal and cervix tumor tissue Raman spectra. First, 314 Raman spectra were collected from 63 subjects; the data was subjected to PC-LDA with LOO-CV, and classification efficiency reached 96.7% and 100% for the normal and cancerous conditions, respectively [122]. In the second study, PC-LDA was used to distinguish between normal and cancerous tissue as well as tissue collected from the vagina of both healthy controls and cancer patients, in an attempt to design an internal control. PC-LDA could classify normal and tumor spectra with 97% efficiency. When a PC-LDA algorithm was built to discriminate between all controls (normal cervix, and vagina of tumor and normal subjects) high misclassification levels were seen, suggesting similarities in biochemical composition among the control samples. Results of classification between tumor tissue and all controls support the idea of using the vagina as an internal control in cervical cancer diagnostics [123].

Wood et al. evaluated biopsy samples collected during colonoscopy using probe-based RS. The *in vitro* study examined 356 colon biopsies, including from normal colon mucosa, hyperplastic polyps (HP), adenomatous polyps, AC, and ulcerative colitis specimens. PC-LDA with LOO-CV was used to make two-group and three-group classification systems. For the binary models, accuracies ranged between 72.1% and 95.9% with ten-second acquisition times and between 61.5% and 95.1% with one-second acquisition times. For the tertiary model, normal tissue, adenomas, and AC tissue were identified with an overall accuracy of 74.1% for the ten-second acquisition time and 63.5% for the one-second acquisition time [124]. Raman fiber-optic measurements of colon biopsy samples, which were categorized as AC, tubular adenomas (TA), HP, and normal tissue, were analyzed from 151 patients. A SVM classifier was trained and validated using a LOPO-CV approach. For classifying AC versus normal tissue, 75% accuracy was achieved. To improve results, three different methods for outlier identification were applied: One Class Classification with SVM, Local Outlier Factor, and Refinement of Training Data (RoTD). The best improvement was seen with RoTD, which increased the accuracy of AC versus normal tissue classification to 81%. To classify high-risk (AC and TA) and low-risk

(HP and normal tissue) lesions, the SVM model without outlier identification reached 71% accuracy; with RoTD, accuracy increased to 77% [125]. Although other studies performed using typical spontaneous RS were more successful, it is important to note the beginning successes of an *in vivo* approach toward diagnosing colorectal cancer using RS and chemometrics.

A custom-built fiber-optic endoscopic Raman probe was used to analyze 673 *ex vivo* esophageal tissue samples from patients with Barrett's esophagus (BE). BE is known to increase the risk of developing esophageal cancer. The tissue was evaluated with PCA-fed LDA with LOPO-CV, which discriminated BE-associated high-grade dysplasia (HGD) and AC from low-grade dysplasia, nondysplastic BE, and normal squamous esophagus with 86% sensitivity and 88% specificity. AC was differentiated from normal squamous esophagus with 94% sensitivity and 91% specificity. Finally, BE and gastric mucosa were differentiated with 96% sensitivity and 92% specificity [126]. A beveled fiber-optic confocal Raman probe was evaluated for *in vivo* diagnosis of BE using epithelial tissue from 373 patients, obtained at endoscopy. Trichotomous probabilistic PLS-DA was used to discriminate columnar-lined epithelium, nondysplastic BE, and HGD BE. For *in vivo* detection of HGD BE, 87.0% sensitivity and 84.7% specificity were attained [127]. In another study, a Raman endoscopic probe measured 673 *ex vivo* benign and esophageal cancer specimens from 62 patients. The results of using a semi-supervised LDA technique, where some of the data is labeled and some is left unlabeled, was compared to standard (supervised) LDA results. Identification of intestinal metaplasia versus dysplasia improved from sensitivity and specificity levels of 73% and 78% with standard PCA-LDA to 78% and 84% for the semi-supervised method. Similarly, performance for differentiating intestinal metaplasia and low-grade dysplasia increased from 44% and 66% using standard PCA-LDA to 63% and 72% sensitivity and specificity levels, respectively, with semi-supervised LDA [128]. In a different study, Maeda et al. performed *ex vivo* experiments using a portable Raman system equipped with a micro-Raman probe. Spectra collected of normal and early-stage (stage 0) cancerous regions within six esophageal samples were analyzed by PC-LDA, which predicted the tissue type with 80% accuracy [129]. Interestingly, these studies all used either PCA-LDA or PLS-DA, showing that the number of samples and the method of sample probing can have a significant impact on the results of a study.

A fiber-optic depth-resolved NIR Raman endoscopic technique was integrated with diagnostic algorithms for *in vivo* epithelial diagnosis of gastric cancer with the assistance of wide-field imaging techniques. Generated diagnostic models using probabilistic PLS-DA with LOPO-CV identified gastric dysplasia with 81.3% sensitivity and 88.3% specificity [130]. Wang et al. compared the performance of two different endoscope-based fiber-optic Raman probe methods. Beveled and volume Raman probes were used for real-time *in vivo* detection of gastric dysplasia. The beveled probe consists of a central flat fiber used for laser light delivery, surrounded by 18 beveled collection fibers positioned in a ring formation; the volume probe also consists of a central flat fiber for excitation but is surrounded by 18 flat collection fibers positioned in a ring formation. A total of 1050 Raman spectra of normal and

dysplastic sites were collected from 66 gastric patients using the beveled Raman probe, while 1913 Raman spectra were collected from 98 gastric patients using the volume Raman probe. PLS-DA with LOPO-CV yielded diagnostic accuracies of 93.0% and 88.4% for the beveled and the volume fiber-optic probes, respectively, suggesting the beveled probe is better suited for further studies [131].

A miniature fiber-optic probe was used to investigate NPC in patients. Raman spectra were collected from nasopharynx tissue of patients with newly diagnosed NPC, post-irradiated nasopharynx (received radiotherapy greater than 6 months ago), or normal nasopharynx. A posterior probability model using PLS distinguished normal nasopharynx and NPC with 91% sensitivity and 95% specificity; the same method could distinguish post-irradiated nasopharynx versus NPC tissue with 77% sensitivity and 96% specificity [132]. A Raman spectrometer with a beam-steered fiber-optic probe was used to detect normal parotid gland and parotid gland tumors, including pleomorphic adenoma, Warthin's tumor, and mucoepidermoid carcinoma, for the purpose of identifying head and neck cancer. SVM with CV was used to distinguish each parotid gland tumor type against normal parotid glands, achieving an average accuracy of 99.43%. Three additional binary models were then built to distinguish the three tumor types from each other, achieving an average accuracy of 97.23% [133]. Here, it is observed that fiber-optic probes are successful for detection of head and neck cancers through both *in vivo* and *ex vivo* studies.

Lung cancer was studied using an endoscopic RS system. Spectra were collected *in vivo* from 280 tissue sites (including 72 HGD/malignant lesions and 208 normal/benign lesions) of 80 patients. Using stepwise multiple regression PLS with LOO-CV, HGD and malignant lesions were detected with 90% sensitivity and 65% specificity [134].

Oral cancer has been widely studied using Raman fiber-optic systems. In one study, Raman spectra were collected from the oral cavity of 18 human subjects *in vivo*, and premalignant/malignant lesions were correctly distinguished from normal and benign tissue with 100% sensitivity and 77% specificity using PCA-LDA

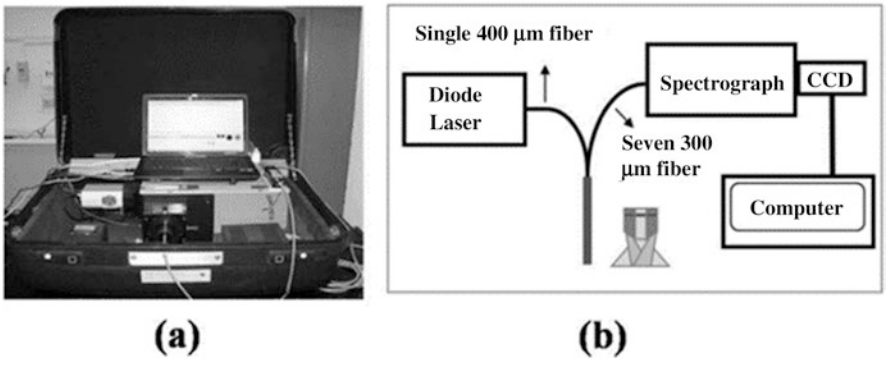

Fig. 8.15 (**a**) A photograph and (**b**) a schematic of the portable clinical Raman spectroscopy system for *in vivo* Raman measurements (Reprinted with permission from [136])

with LOO-CV [135]. Krishna et al. aimed to diagnose malignant and potentially malignant lesions of the oral cavity from 28 healthy volunteers and 171 patients. *In vivo* Raman spectra were collected from normal oral mucosa sites as well as histopathological sites including oral SCC, oral sub-mucous fibrosis, and leukoplakia using a portable clinical RS system (Fig. 8.15). A maximum representation and discrimination feature coupled with SMLR provided an average accuracy of 85.25% for classifying all four sites based on LOPO-CV; a binary model resulted in a sensitivity and specificity each of 94% for discriminating normal tissue spectra from all abnormal tissue spectra [136]. Research led by Sahu resulted in four studies published on diagnosing oral malignancies. In the earliest study, *in vivo* Raman spectra were collected from sera of buccal mucosa, tongue cancer, and healthy subjects using a fiber-optic Raman microprobe. Through PC-LDA with LOO-CV, binary models showed normal and cancer serum groups could be differentiated with about 70% classification efficiency and buccal mucosa and tongue cancer groups with about 68% efficiency [137]. Raman spectra were collected from oral exfoliated cells from healthy volunteers, healthy tobacco users, and subjects with oral cancer (from both tumor and healthy sites). PCA and PC-LDA showed distinct differences between the two healthy groups and the two cancer groups. Furthermore, PC-LDA with LOO-CV showed about 67% efficiency for predicting to which class the samples belonged, based on spectra-wise classification. Using a patient-wise approach, about 69% classification efficiency was achieved [138]. Following this, serum was collected from 22 oral cancer patients before and after surgery and analyzed using a Raman microprobe. PC-LDA followed by LOO-CV was again employed for discrimination, providing 78% classification efficiency for distinguishing recurrence and non-recurrence groups after surgery. The same method could distinguish recurrence and non-recurrence groups before surgery with only about 53% efficiency [139]. In the last study by Sahu et al., oral exfoliated samples were analyzed from healthy volunteers, healthy volunteers with tobacco habits, and patients with oral premalignant conditions (OPC) by the Raman microprobe. In the three-group model, OPC spectra were classified with 77% and 70% sensitivity for PC-LDA with spectra-wise and patient-wise CV methods, respectively. The sensitivity improved to 86% (spectra-wise) and 83% (patient-wise) using a binary model [140]. Interestingly, each of the studies performed by Sahu et al. show the ability of fiber-optic probes to be used for *ex vivo* studies performed on samples other than tissue. Yasser et al. analyzed the Raman spectra from parental oral cancer cell lines and from two different developed radio-resistant sublines using a fiber-optic microprobe system. Spectral differences were observed between the three different cell lines, and PCA showed distinct clustering, depicting the ability of RS to predict radioresistance in cells, which can be used for improved prognosis of oral cancer [141]. Notably, the best results for diagnosing oral cancer were achieved using the probe systems *in vivo* rather than *ex vivo*.

Spectra of fresh and non-processed post-prostatectomy specimens were collected using a macroscopic handheld RS probe. The areas of the tissue were labeled with tissue type (extra-prostatic or prostatic), malignant or benign, cancer grade (grade groups I–V), and tissue glandular level. Neural networks were used to

classify the spectra in binary models. Prostate and extra-prostatic tissue were distinguished with 82% sensitivity and 83% specificity, whereas benign and malignant tissue were correctly classified with 87% sensitivity and 86% specificity. Benign spectra were differentiated from each of the five cancer grade groups in multiple binary models, achieving an average sensitivity of 81.8% and specificity of 85.2% [142]. Silveira Jr. et al. collected 160 spectra from 16 benign tissue and 16 prostate cancer tissue samples. A discrimination model was built using Euclidean distance based on the relative concentrations of phosphatidylcholine and water in the tissue samples. The two kinds of tissues were discriminated with 74% accuracy [143]. While these *ex vivo* studies are less successful than others previously reviewed, the results are still accomplished enough to indicate the potential for the method to be used intra-operatively.

A Raman instrument equipped with a fiber-optic probe was used to collect spectra *in vivo* from 137 lesions in 76 skin cancer patients; biopsies of the lesions were classified as malignant melanoma (MM), non-melanoma pigmented lesion (PL), BCC, actinic keratosis (AK), and SCC. The collected data was analyzed by PCA, and LOO-logistic regression classifiers were built, the results of which were compared to the histopathology of the lesions. The sensitivity and specificity for binary classification of MM versus PL were 100% and 100%; of SCC and BCC versus AK was 95% and 71%; and of AK, SCC, and BCC versus normal tissue was 90% and 85%, respectively [144]. Zakharov et al. investigated the potential for diagnosing malignant tumors in both skin and lung tissue. 40 *ex vivo* lung tissue samples and 50 *in vivo* skin tumor samples were investigated through a combination of LDA, QDA, and SVM. It was discovered that MM could be diagnosed with 88.9% sensitivity and 87.8% specificity, lung AC with 100% sensitivity and 81.5% specificity, and lung SCC with 90.9% sensitivity and 77.8% specificity [145]. In a different study, lesions suspected of being MM, BCC, or SCC were subjected to *in vivo* Raman spectral analysis through a fiber-coupled probe. Non-melanoma skin cancers were discriminated from normal skin through PLS-DA with accuracies of 73% (BCC) and 85% (SCC). MM and pigmented nevi (moles) were discriminated with 91% accuracy [146]. A dispersive spectrometer connected to a Raman probe collected data from non-melanoma (BCC and SCC), pre-cancerous (AK), and benign lesions and from normal tissue. Using PCA-DA and PLS-DA algorithms, non-melanoma and pre-cancerous lesions were differentiated from benign and normal tissue with accuracies of 82.8% and 91.9%, respectively [147]. Zhao et al. used PC-GDA and PLS, built with selected wavenumber windows, to classify 645 cases of pre-cancerous, benign, and skin cancer lesions. Malignant and benign skin lesions could be identified *in vivo* with high levels of diagnostic accuracy [148]. The success for skin cancer detection *in vivo* through Raman probe systems is clear; these results are comparable with those from other studies, and indicate that the method should be strongly considered for real-time diagnosis of skin cancer in clinical settings.

8.2.4.1 Fiber-Optic Studies with Expanded Raman Spectral Range

In several additional papers, the HWN region of spectral data collected using Raman probe systems was considered in addition to the FP region for real-time diagnoses of cancers. Specifically, a fiber-optic Raman endoscope was used to collect *in vivo* Raman spectra in the FP (800–1800 cm^{-1}) and HWN (2800–3600 cm^{-1}) regions from colorectal tissue. Raman measurements were made at five different anatomical locations of normal colorectal tissue and PLS-DA with LOPO-CV was used to identify the different tissue sites. An average sensitivity of 29.27% and specificity of 83.51% were achieved, indicating low levels of inter-anatomical molecular variability between normal colorectal tissue areas. For discriminating between normal tissue areas and tissue affected by colorectal cancer, PLS-DA with LOPO-CV of the FP/HWN dataset was again performed, attaining a diagnostic accuracy of 88.8% [149]. In a follow-up study, researchers again simultaneously acquired *in vivo* FP and HWN region Raman spectra from colorectal tissue. Adenoma and HP were differentiated with 90.9% sensitivity and 83.3% specificity using PLS-DA with LOPO-CV, which is superior to results achieved using only FP or only HWN region spectral data [150]. Both studies show the advantages of collecting FP/HWN spectral data via Raman probes for intra-operative diagnoses of colorectal cancer. *In vivo* diagnosis of esophageal SCC at the time of clinical endoscopy was investigated by Wang et al. FP and HWN region Raman spectra were collected from 48 patients using the developed fiber-optic RS technique. Through PLS-DA with LOPO-CV, a sensitivity of 92.7% and specificity of 93.6% for esophageal SCC identification were achieved. Again, these results were found to be superior to those obtained using only FP or HWN region spectral data [151]. Wang et al. applied their rapid fiber-optic RS technique for diagnosing gastric pre-cancer during endoscopic examination. FP/HWN region Raman spectra from normal, dysplasia, and AC tissue sites were collected. PLS-DA with LOPO-CV reached an average sensitivity of 88.67% and specificity of 92.53% for detecting each of the three groups. A binary model could detect gastric dysplasia with 90.9% sensitivity and 95.9% specificity [152]. Further, the discrimination of diseased tissue and adjacent healthy tissue from patients who have head and neck cancer was accomplished using a wide Raman spectral range of 100–4300 cm^{-1}. PCA showed effective separation between healthy controls and malignant tissue samples, which included SCC and tonsil SCC; the separation was better observed through analysis of the full spectrum then it was of only the FP region [153].

Lin et al. utilized fiber-optic RS to develop a method for *in vivo* diagnosis of NPC at the time of endoscopy using the FP and HWN regions. Spectral data was collected from 204 different tissue sites of 95 subjects; PCA-LDA with LOO-CV provided a diagnostic accuracy of 93.1% [154]. In another study, Lin et al. acquired FP and HWN region spectral data of 101 healthy and diseased tissue sites from 60 patients with laryngeal cancer undergoing endoscopic examination. Here, PLS-DA with LOPO-CV could discriminate the two classes with an accuracy of 91.1% [155]. In both of those studies, and in general, the results were improvements over those achieved using just FP or HWN region spectral data.

The combination of Raman spectroscopic analysis with optical probing systems provides the first crucial step toward bringing the methodology to the clinical setting. By allowing for collection of spectral data *in vivo*, the need for additional sample collection which can be time-consuming, expensive, and invasive is eliminated. The spectral data can still be analyzed using a wide variety of chemometric techniques, as was shown, in order to develop an automatic diagnostic system which can be incorporated into clinical settings for quick and accurate diagnoses. Although not every study was an improvement upon those performed without using probes, each of the aforementioned projects serves to indicate that the technology is capable of use within clinical settings. This is an exciting first step toward introducing RS as a universal method for cancer detection that can be used *in vivo* and provide accurate results in real-time.

8.3 Critical Evaluation

There has been a vast amount of research published on utilizing Raman spectroscopy and advanced statistical analysis for the purpose of diagnosing cancer. Obviously, the methodology has great potential. Regardless of the biological sample analyzed, the exact variation of RS used, or the statistical technique applied—it is impossible to disagree that based on the incredible amount of research conducted and published, RS and advanced statistical analysis have a great potential for creating the first universal method for cancer detection.

While the potential of the method is obvious, it is important to note that some results published in the aforementioned studies may overstress their significance, and as such there is a need to address the risk associated with overestimating the capabilities of the methodology based on the reported results. A small number of studies do not report quantitative results, and focus more on the qualitative success of the methodology; while these findings are still important for indicating the ability of RS to detect biochemical differences between different sample types, it is necessary to remember that quantitative results are necessary for supporting efforts to bring the methodology into clinical settings. Several studies suffered from too small of a dataset to be considered significant. Some research achieved sensitivity, specificity, or accuracy levels which are not necessarily an improvement of those achieved using current methods for diagnosing cancer. The balance between number of samples used and performance results is something that needs to be kept in mind when evaluating the significance of different experiments with comparison to each other. Ideally, the most reliable studies are those which use a statistically significant number of samples and achieve impressive performance levels. Further, the methods of validation for several experiments are considered “internal” validation—this means that the model was tested with the same spectral data that was also used to build it. This can lead to the potential for the model to “over-fit” itself to the data it sees, preventing it from being able to accurately predict spectral data from new unknown samples.

On the other hand, many more studies can be considered reliable—those experiments that utilized external validation are more trustworthy than those that did not. External validation utilizes an independent spectral dataset, which the algorithm has not yet seen, in order to test the performance of the model. In this regard, there are much lower chances of the model becoming over-fit, and the results are abundantly more dependable. Furthermore, while a small number of samples is useful to accomplish a proof-of-concept study, many more published manuscripts used a statistically significant number of samples which also obtained impressive levels of accuracy. Most importantly, the number of strong research studies heavily outweigh those that are less ideal.

Based on the extensive amount of recent evidence which suggests the potential for RS to diagnose cancer, not to mention the plethora of significant research published greater than 4 years ago, one important question remains: why has this methodology not yet been introduced into clinical settings as a solitary method for diagnosing cancer? Although the answer is complicated and multifaceted, the truth remains that research has already shown RS is being used intra-operatively. In fact, many of the previously reviewed experiments performed using fiber-optic probes were conducted *in vivo*. In 2013, Kallaway et al. summarized the clinical usage of Raman spectroscopy for diagnosing colorectal, esophagus, breast, and bladder diseases and cancers [156]. Pence et al. detail the necessary considerations required for clinical implementation of RS as well as review large (considered as greater than 50 samples) *in vivo* applications of the method [157]. More recently, Santos et al. greatly covered the translation of RS into clinical settings for detecting cancer, reviewing many studies which successfully detected cancer with high levels of accuracy under conditions that resemble the intended clinical environment [158].

Although research repeatedly advocates for the potential of the method, as Santos importantly points out, there needs to be an increase in communication as well as trust between spectroscopists and clinicians in order to bring this methodology closer to the goal of true clinical applications. Understanding the methods behind the technology as well as understanding the needs of clinicians are important areas that are still being developed. However, the stepping stones have been laid, and there remains a clear path forward for introducing Raman-based technology to clinical oncology departments for cancer detection purposes.

Lastly, there is an essential need to consolidate the classification techniques used in order to unite the automatic analysis and disease detection portion of the methodology. A multiclass discrimination technique will need to be developed and fine-tuned which considers all cancers, or at least considers many groups of similar cancers. Based on typical expectations of the medical field, a major clinical trial needs to be conducted with a wide cohort of participants ranging in age, gender, race, ethnicity, location of residence, prescribed medications, and comorbidities. While research has repeatedly shown success on a smaller scale, a successful large-scale clinical trial is needed in order to convince all involved parties of the method's capabilities. Although there is work to still be done, RS has shown every indication that it has the potential to be used in clinical settings in the future for universal cancer detection.

8.4 Conclusion

Cancer affects millions of individuals every year and is a leading cause of death worldwide. Because of its association with mortality, it is vitally important to diagnose cancer as early on in its progression as possible. Early diagnoses provide the best chance for the afflicted individual to seek effective treatment options. While many methods exist to diagnose cancers individually, there is a strong unmet need to accurately, definitively, and with minimal invasiveness diagnose all cancers using one universal method. This chapter proposes Raman spectroscopy as a potential solution for this task. In countless different ways, the combination of Raman spectroscopy with chemometric analysis has proven its usefulness for diagnosing cancer. Innumerous studies have been published on over twenty different forms of cancer in the last few years alone. Regardless of the sample studied or which variation of the technique is used, if a probe is used to collect data or a second analytical technique is used to provide complimentary information, the obvious capability of RS for diagnosing cancers cannot be denied. The abundance of information that is obtained through Raman spectral data provides not only helpful material for classification purposes, but also delivers insight into the biochemical composition of samples, revealing both useful information as well as potential biomarkers indicative of different cancers. The plethora of herein reported studies have repeatedly shown evidence that RS is sensitive, specific, and overall a reliable technique for differentiating healthy samples from diseased samples. The next step for this methodology is to unite researchers to work toward conducting a large-scale clinical trial where a combined algorithm can be developed which can diagnose all forms of cancer. The results of such a trial would need to be confirmed using already established methods for cancer diagnosis; however, if such a trial is successful, it is possible that Raman spectroscopy with advanced statistical analysis may become the first singular universal method for diagnosing cancer.

References

1. Bray, F., et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**(6), 394–424 (2018)
2. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2019. CA Cancer J. Clin. **69**(1), 7–34 (2019)
3. Smith, R.A., et al.: Cancer screening in the United States, 2018: a review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J. Clin. **68**(4), 297–316 (2018)
4. Mann, C.K., Vickers, T.J.: In: Lewis, I.R., Edwards, H.G.M. (eds.) Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, pp. 251–274. Marcel Dekker, Inc., New York (2001)
5. Williams, T.L., Collette, T.W.: In: Lewis, I.R., Edwards, H.G.M. (eds.) Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, pp. 683–731. Marcel Dekker, Inc., New York (2001)

6. Wold, S.: Chemometrics; what do we mean with it, and what do we want from it? Chemom. Intel. Lab. Syst. **30**(1), 109–115 (1995)
7. Geladi, P.: Chemometrics in spectroscopy. Part 1. Classical chemometrics. Spectrochim. Acta B: Atomic Spectrosc. **58**(5), 767–782 (2003)
8. Sikirzhytski, V., Sikirzhytskaya, A., Lednev, I.K.: Multidimensional Raman spectroscopic signatures as a tool for forensic identification of body fluid traces: a review. Appl. Spectrosc. **65**(11), 1223–1232 (2011)
9. Muro, C.K., et al.: Forensic body fluid identification and differentiation by Raman spectroscopy. Forensic Chem. **1**, 31–38 (2016)
10. Kalkanis, S.N., et al.: Raman spectroscopy to distinguish grey matter, necrosis, and glioblastoma multiforme in frozen tissue sections. J. Neurooncol. **116**(3), 477–485 (2014)
11. Liu, T., et al.: Evaluation of Raman spectra of human brain tumor tissue using the learning vector quantization neural network. Laser Phys. **26**(5), 055606 (2016)
12. Sathyavathi, R., et al.: Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy. Sci. Rep. **5**, 9907 (2015)
13. Han, B., et al.: Differences and relationships between normal and atypical ductal hyperplasia, ductal carcinoma *in situ*, and invasive ductal carcinoma tissues in the breast based on Raman spectroscopy. Appl. Spectrosc. **71**(2), 300–307 (2017)
14. Fallahzadeh, O., Dehghani-Bidgoli, Z., Assarian, M.: Raman spectral feature selection using ant colony optimization for breast cancer diagnosis. Lasers Med. Sci. **33**(8), 1799–1806 (2018)
15. Rashid, N., et al.: Raman microspectroscopy for the early detection of pre-malignant changes in cervical tissue. Exp. Mol. Pathol. **97**(3), 554–564 (2014)
16. Daniel, A., et al.: Biochemical assessment of human uterine cervix by micro-Raman mapping. Photodiagnosis Photodyn. Ther. **17**, 65–74 (2017)
17. Daniel, A., Aruna, P., Ganesan, S.: Near-infrared Raman spectroscopy for estimating biochemical changes associated with different pathological conditions of cervix. Spectrochim. Acta A Mol. Biomol. Spectrosc. **190**, 409–416 (2018)
18. Synytsya, A., et al.: Raman spectroscopy at different excitation wavelengths (1064, 785 and 532 nm) as a tool for diagnosis of colon cancer. J. Raman Spectrosc. **45**(10), 903–911 (2014)
19. Li, S., et al.: Identification and characterization of colorectal cancer using Raman spectroscopy and feature selection techniques. Opt. Express. **22**(21), 25895–25908 (2014)
20. Liu, W., et al.: Raman spectroscopy in colorectal cancer diagnostics: comparison of PCA-LDA and PLS-DA models. J. Spectrosc. **2016**, 1603609 (2016)
21. Ishigaki, M., et al.: Diagnosis of early-stage esophageal cancer by Raman spectroscopy and chemometric techniques. Analyst. **141**(3), 1027–1033 (2016)
22. Jin, S., Mao, H.: Near-infrared Raman spectroscopy for diagnosis of gastric cancer. J. South. Med. Univ. **34**(3), 391–395 (2014)
23. Yao, C., et al.: A preliminary study of protein alterations accompanied with malignant transformation of gastric mucosa using Raman spectroscopy. Chongqing Med. J. **29**, 3875–3878 (2014)
24. Hsu, C.-W., et al.: Novel method for differentiating histological types of gastric adenocarcinoma by using confocal Raman microspectroscopy. PLoS One. **11**(7), e0159829 (2016)
25. Hsu, C.-W., et al.: Differentiating gastrointestinal stromal tumors from gastric adenocarcinomas and normal mucosae using confocal Raman microspectroscopy. J. Biomed. Opt. **21**(7), 075006 (2016)
26. Li, J.H., Li, W.T., Zhang, G.H.: Detection of nasopharyngeal carcinoma using deep NIR Raman spectroscopy. Laser Phys. **24**(12), 125601 (2014)
27. Li, Y., et al.: Rapid detection of nasopharyngeal cancer using Raman spectroscopy and multivariate statistical analysis. Mol. Clin. Oncol. **3**(2), 375–380 (2015)
28. Mian, S.A., et al.: Raman spectroscopy can discriminate between normal, dysplastic and cancerous oral mucosa: a tissue-engineering approach. J. Tissue Eng. Regen. Med. **11**(11), 3253–3262 (2017)

29. Pence, I.J., et al.: Discrimination of liver malignancies with 1064 nm dispersive Raman spectroscopy. Biomed. Opt. Express. **6**(8), 2724–2737 (2015)
30. Wang, L., et al.: Evaluation of Raman spectroscopy for diagnosing EGFR mutation status in lung adenocarcinoma. Analyst. **139**(2), 455–463 (2014)
31. Behl, I., et al.: Raman microspectroscopic study of oral buccal mucosa. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2014)
32. Daniel, A., et al.: Raman mapping of oral tissues for cancer diagnosis. J. Raman Spectrosc. **45**(7), 541–549 (2014)
33. Chen, P.-H., et al.: Automatic and objective oral cancer diagnosis by Raman spectroscopic detection of keratin with multivariate curve resolution analysis. Sci. Rep. **6**, 20097 (2016)
34. Lu, S.-y., et al.: Raman spectroscopy in ovarian cancer diagnostics. Spectrosc. Spectr. Anal. **37**(6), 1784–1788 (2017)
35. Du, Y., et al.: A novel diagnostic method of Raman spectroscopy for malignant pheochromocytoma/paraganglioma. RSC Adv. **6**(103), 101178–101184 (2016)
36. Liu, Y., et al.: Renal mass biopsy using Raman spectroscopy identifies malignant and benign renal tumors: potential for pre-operative diagnosis. Oncotarget. **8**(22), 36012–36019 (2017)
37. Kong, K., et al.: Increasing the speed of tumour diagnosis during surgery with selective scanning Raman microscopy. J. Mol. Struct. **1073**, 58–65 (2014)
38. Zhao, J., et al.: Wavenumber selection based analysis in Raman spectroscopy improves skin cancer diagnostic specificity. Analyst. **141**(3), 1034–1043 (2016)
39. Medeiros Neto, L.P., et al.: Micro-Raman spectroscopic study of thyroid tissues. Photodiagnosis Photodyn. Ther. **17**, 164–172 (2017)
40. Senol, O., et al.: Application of photonics in diagnosis of papillary thyroid carcinoma tissues through Raman spectroscopy-assisted with chemometrics. Anal. Lett. **51**(1–2), 229–235 (2018)
41. Palermo, A., et al.: Raman spectroscopy applied to parathyroid tissues: a new diagnostic tool to discriminate normal tissue from adenoma. Anal. Chem. **90**(1), 847–854 (2018)
42. Frost, J., et al.: Raman spectroscopy and multivariate analysis for the non invasive diagnosis of clinically inconclusive vulval lichen sclerosus. Analyst. **142**(8), 1200–1206 (2017)
43. Kerr, L.T., et al.: Methodologies for bladder cancer detection with Raman based urine cytology. Anal. Methods. **8**(25), 4991–5000 (2016)
44. Banerjee, H.N., et al.: Deciphering the finger prints of brain cancer glioblastoma multiforme from four different patients by using near infrared Raman spectroscopy. J. Cancer Sci. Ther. **7**(2), 44–47 (2015)
45. Marro, M., et al.: Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy. Biochim. Biophys. Acta (BBA)-Mol. Cell Res. **1843**(9), 1785–1795 (2014)
46. Bi, X., et al.: Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy. J. Biomed. Opt. **19**(2), 025001 (2014)
47. Goel, P.N., et al.: Investigating the effects of Pentoxifylline on human breast cancer cells using Raman spectroscopy. J. Innov. Opt. Health Sci. **8**(02), 1550004 (2015)
48. Talari, A.C.S., et al.: Raman spectroscopic analysis differentiates between breast cancer cell lines. J. Raman Spectrosc. **46**(5), 421–427 (2015)
49. Talari, A.C.S., et al.: Analyzing normal proliferating, hypoxic and necrotic regions of T-47D human breast cancer spheroids using Raman spectroscopy. Appl. Spectrosc. Rev. **52**(10), 909–924 (2017)
50. Winnard Jr., P.T., et al.: Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes. Oncotarget. **8**(12), 20266–20287 (2017)
51. Ramos, I., et al.: Raman spectroscopy for cytopathology of exfoliated cervical cells. Faraday Discuss. **187**, 187–198 (2016)
52. Gala de Pablo, J., et al.: Biochemical fingerprint of colorectal cancer cell lines using label-free live single-cell Raman spectroscopy. J. Raman Spectrosc. **49**(8), 1323–1332 (2018)

53. Shiramizu, B., et al.: Unique Raman spectroscopic fingerprints of B-cell non-Hodgkin lymphoma: implications for diagnosis, prognosis and new therapies. J. Biol. Med. Sci. **2**(1), 105 (2018)
54. Maguire, A., et al.: Competitive evaluation of data mining algorithms for use in classification of leukocyte subtypes with Raman microspectroscopy. Analyst. **140**(7), 2473–2481 (2015)
55. Carvalho, L.F.C.S., et al.: Raman micro-spectroscopy for rapid screening of oral squamous cell carcinoma. Exp. Mol. Pathol. **98**(3), 502–509 (2015)
56. Moradi, H., et al.: Raman micro-spectroscopy applied to treatment resistant and sensitive human ovarian cancer cells. J. Biophotonics. **10**(10), 1327–1334 (2017)
57. Corsetti, S., et al.: Raman spectroscopy for accurately characterizing biomolecular changes in androgen-independent prostate cancer cells. J. Biophotonics. **11**(3), e201700166 (2018)
58. Olmos, V., et al.: Preprocessing tools applied to improve the assessment of Aldrin effects on prostate cancer cells using Raman spectroscopy. Appl. Spectrosc. **72**(3), 489–500 (2018)
59. Ralbovsky, N.M., et al.: Deep-Ultraviolet Raman spectroscopy for cancer diagnostics: a feasibility study with cell lines and tissues. Cancer Stud. Mol. Med. - Open J. **5**(1), 1–10 (2019)
60. Mehta, K., et al.: An early investigative serum Raman spectroscopy study of meningioma. Analyst. **143**(8), 1916–1923 (2018)
61. González-Solís, J.L., et al.: Cervical cancer detection based on serum sample Raman spectroscopy. Lasers Med. Sci. **29**(3), 979–985 (2014)
62. Pappu, R., et al.: Raman spectroscopic characterization of urine of normal and cervical cancer subjects. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2017)
63. Li, X., et al.: Raman spectroscopy combined with principal component analysis and k nearest neighbour analysis for non-invasive detection of colon cancer. Laser Phys. **26**(3), 035702 (2016)
64. Khan, S., et al.: Optical screening of nasopharyngeal cancer using Raman spectroscopy and support vector machine. Optik-Int. J. Light Electron Opt. **157**, 565–570 (2018)
65. Happillon, T., et al.: Diagnosis approach of chronic lymphocytic leukemia on unstained blood smears using Raman microspectroscopy and supervised classification. Analyst. **140**(13), 4465–4472 (2015)
66. Li, Y., et al.: Pattern recognition methods combined with Raman spectra applied to distinguish serums from lung cancer patients and healthy people. J. Biosci. Med. **5**(9), 95–105 (2017)
67. Wang, H., et al.: Screening and staging for non-small cell lung cancer by serum laser Raman spectroscopy. Spectrochim. Acta A Mol. Biomol. Spectrosc. **201**, 34–38 (2018)
68. Brindha, E., et al.: Raman spectroscopic characterization of urine of normal and oral cancer subjects. J. Raman Spectrosc. **46**(1), 84–93 (2015)
69. Pachaiappan, R., et al.: Near infrared Raman spectroscopic characterization of blood plasma of normal, oral premalignant and malignant conditions—a pilot study. J. Raman Spectrosc. **46**(9), 735–743 (2015)
70. Pachaiappan, R., et al.: Near-infrared Raman spectroscopic characterization of salivary metabolites in the discrimination of normal from oral premalignant and malignant conditions. J. Raman Spectrosc. **47**(7), 763–772 (2016)
71. Barroso, E.M., et al.: Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy. Anal. Chem. **87**(4), 2419–2426 (2015)
72. Pachaiappan, R., et al.: High wavenumber Raman spectroscopic characterization of normal and oral cancer using blood plasma. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2017)
73. Brindha, E., et al.: High wavenumber Raman spectroscopy in the characterization of urinary metabolites of normal subjects, oral premalignant and malignant patients. Spectrochim. Acta A Mol. Biomol. Spectrosc. **171**, 52–59 (2017)
74. Carvalho, L.F.C.S., et al.: Raman spectroscopic analysis of oral cells in the high wavenumber region. Exp. Mol. Pathol. **103**(3), 255–262 (2017)

75. Santos, I.P., et al.: Raman spectroscopic characterization of melanoma and benign melanocytic lesions suspected of melanoma using high-wavenumber Raman spectroscopy. Anal. Chem. **88**(15), 7683–7688 (2016)
76. Zhou, X., et al.: Evaluation of the diagnostic potential of *ex vivo* Raman spectroscopy in gastric cancers: fingerprint versus high wavenumber. J. Biomed. Opt. **21**(10), 105002 (2016)
77. Huang, W., et al.: Study of both fingerprint and high wavenumber Raman spectroscopy of pathological nasopharyngeal tissues. J. Raman Spectrosc. **46**(6), 537–544 (2015)
78. Sun, L., et al.: Preliminary study of differentiating smears from cancerous and non-cancerous nasopharyngeal tissue using confocal Raman spectroscopy. J. Cancer Res. Clin. Oncol. **142**(4), 823–831 (2016)
79. Kast, R., et al.: Identification of regions of normal grey matter and white matter from pathologic glioblastoma and necrosis in frozen sections using Raman imaging. J. Neurooncol. **125**(2), 287–295 (2015)
80. Kong, K., et al.: Towards intra-operative diagnosis of tumours during breast conserving surgery by selective-sampling Raman micro-spectroscopy. Phys. Med. Biol. **59**(20), 6141–6152 (2014)
81. Abramczyk, H., et al.: The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue. Analyst. **140**(7), 2224–2235 (2015)
82. Brozek-Pluska, B., Kopeć, M., Abramczyk, H.: Development of a new diagnostic Raman method for monitoring epigenetic modifications in the cancer cells of human breast tissue. Anal. Methods. **8**(48), 8542–8553 (2016)
83. Vanna, R., et al.: Label-free imaging and identification of typical cells of acute myeloid leukaemia and myelodysplastic syndrome by Raman microspectroscopy. Analyst. **140**(4), 1054–1064 (2015)
84. Tolstik, T., et al.: Discrimination and classification of liver cancer cells and proliferation states by Raman spectroscopic imaging. Analyst. **139**(22), 6036–6043 (2014)
85. Tolstik, T., et al.: Classification and prediction of HCC tissues by Raman imaging with identification of fatty acids as potential lipid biomarkers. J. Cancer Res. Clin. Oncol. **141**(3), 407–418 (2015)
86. Ryabchykov, O., et al.: Raman spectroscopic investigation of the human liver stem cell line HepaRG. J. Raman Spectrosc. **49**(6), 935–942 (2018)
87. Cals, F.L.J., et al.: Investigation of the potential of Raman spectroscopy for oral cancer detection in surgical margins. Lab. Invest. **95**, 1186–1196 (2015)
88. Meksiarun, P., et al.: Comparison of multivariate analysis methods for extracting the paraffin component from the paraffin-embedded cancer tissue spectra for Raman imaging. Sci. Rep. **7**, 44890 (2017)
89. Musto, P., et al.: Hyperspectral Raman imaging of human prostatic cells: an attempt to differentiate normal and malignant cell lines by univariate and multivariate data analysis. Spectrochim. Acta A Mol. Biomol. Spectrosc. **173**, 476–488 (2017)
90. Rau, J.V., et al.: Raman spectroscopy imaging improves the diagnosis of papillary thyroid carcinoma. Sci. Rep. **6**, 35117 (2016)
91. Zha, W.L., et al.: HPLC assisted Raman spectroscopic studies on bladder cancer. Laser Phys. Lett. **12**(4), 045701 (2015)
92. Abramczyk, H., Imiela, A.: The biochemical, nanomechanical and chemometric signatures of brain cancer. Spectrochim. Acta A Mol. Biomol. Spectrosc. **188**, 8–19 (2018)
93. Kim, S., et al.: Dual-modal cancer detection based on optical pH sensing and Raman spectroscopy. J. Biomed. Opt. **22**(10), 105002 (2017)
94. Surmacki, J., et al.: The lipid-reactive oxygen species phenotype of breast cancer. Raman spectroscopy and mapping, PCA and PLSDA for invasive ductal carcinoma and invasive lobular carcinoma. Molecular tumorigenic mechanisms beyond Warburg effect. Analyst. **140**(7), 2121–2133 (2015)

95. Owens, G.L., et al.: Vibrational biospectroscopy coupled with multivariate analysis extracts potentially diagnostic features in blood plasma/serum of ovarian cancer patients. J. Biophotonics. **7**(3–4), 200–209 (2014)
96. Bury, D., et al.: Phenotyping metastatic brain tumors applying spectrochemical analyses: segregation of different cancer types. Anal. Lett. **52**(4), 575–587 (2018)
97. Polavarapu, P.L.: Determination of molecular stereochemistry using optical rotatory dispersion, vibrational circular dichroism and vibrational Raman optical activity. In: Busch, K.W., Busch, M.A. (eds.) Chiral Analysis, pp. 461–504. Elsevier (2006)
98. Ostovar Pour, S., et al.: Raman optical activity. In: Polavarapu, P.L. (ed.) Chiral Analysis (Second Edition): Advances in Spectroscopy, Chromatography and Emerging Methods, pp. 249–291. Elsevier (2018)
99. Tatarkovič, M., et al.: The potential of chiroptical and vibrational spectroscopy of blood plasma for the discrimination between colon cancer patients and the control group. Analyst. **140**, 2287–2293 (2015)
100. Lin, D., et al.: Autofluorescence and white light imaging-guided endoscopic Raman and diffuse reflectance spectroscopy for *in vivo* nasopharyngeal cancer detection. J. Biophotonics. **11**(4), e201700251 (2018)
101. Zakharov, V.P., et al.: Combined Raman spectroscopy and autofluorescence imaging method for *in vivo* skin tumor diagnosis. In: SPIE Optical Engineering + Applications. International Society for Optics and Photonics, San Diego (2014)
102. Khristoforova, Y.A., et al.: *In vivo* diagnostics of malignant and benign tumors with low-cost Raman spectrometer. In: Frontiers in Optics 2017. Optical Society of America, Washington, D.C. (2017)
103. Bratchenko, I.A., et al.: Combined Raman and autofluorescence *ex vivo* diagnostics of skin cancer in near-infrared and visible regions. J. Biomed. Opt. **22**(2), 027005 (2017)
104. Fullwood, L.M., et al.: Investigating the use of Raman and immersion Raman spectroscopy for spectral histopathology of metastatic brain cancer and primary sites of origin. Anal. Methods. **6**(12), 3948–3961 (2014)
105. Li, B., et al.: Evaluating oral epithelial dysplasia classification system by near-infrared Raman spectroscopy. Oncotarget. **8**(44), 76257–76265 (2017)
106. Weng, S., et al.: Combining deep learning and coherent anti-Stokes Raman scattering imaging for automated differential diagnosis of lung cancer. J. Biomed. Opt. **22**(10), 106017 (2017)
107. Yosef, H.K., et al.: Noninvasive diagnosis of high-grade urothelial carcinoma in urine by Raman spectral imaging. Anal. Chem. **89**(12), 6893–6899 (2017)
108. Knipfer, C., et al.: Raman difference spectroscopy: a non-invasive method for identification of oral squamous cell carcinoma. Biomed. Opt. Express. **5**(9), 3252–3265 (2014)
109. Daniel, A., et al.: Polarized Raman spectroscopy unravels the biomolecular structural changes in cervical cancer. Spectrochim. Acta A Mol. Biomol. Spectrosc. **152**, 58–63 (2016)
110. Ji, M., et al.: Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. Sci. Transl. Med. **7**(309), 309ra163 (2015)
111. Hollon, T.C., et al.: Rapid intraoperative diagnosis of pediatric brain tumors using stimulated Raman histology. Cancer Res. **78**(1), 278–289 (2018)
112. Medipally, D.K.R., et al.: Development of a high throughput (HT) Raman spectroscopy method for rapid screening of liquid blood plasma from prostate cancer patients. Analyst. **142**(8), 1216–1226 (2017)
113. Stables, R., et al.: Feature driven classification of Raman spectra for real-time spectral brain tumour diagnosis using sound. Analyst. **142**(1), 98–109 (2017)
114. Chen, H., et al.: Identification and characterization of bladder cancer by low-resolution fiber-optic Raman spectroscopy. J. Biophotonics. **11**(9), e201800016 (2018)
115. Desroches, J., et al.: Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification. Biomed. Opt. Express. **6**(7), 2380–2397 (2015)
116. Jermyn, M., et al.: Intraoperative brain cancer detection with Raman spectroscopy in humans. Sci. Transl. Med. **7**(274), 274ra19 (2015)

117. Jermyn, M., et al.: Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts. J. Biomed. Opt. **21**(9), 094002 (2016)
118. Li, Q., Gao, Q., Zhang, G.: Classification for breast cancer diagnosis with Raman spectroscopy. Biomed. Opt. Express. **5**(7), 2435–2445 (2014)
119. Li, Q., Hao, C., Xu, Z.: Diagnosis of breast cancer tissues using 785 nm miniature Raman spectrometer and pattern regression. Sensors. **17**(3), 627 (2017)
120. Li, Q.-B., et al.: Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics. J. Appl. Spectrosc. **82**(3), 450–455 (2015)
121. Duraipandian, S., et al.: Near-infrared Raman spectroscopy for assessing biochemical changes of cervical tissue associated with precarcinogenic transformation. Analyst. **139**(21), 5379–5386 (2014)
122. Shaikh, R.S., et al.: *In vivo* Raman spectroscopy of cervix cancers. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2014)
123. Shaikh, R.S., et al.: *In vivo* Raman spectroscopy of human uterine cervix: exploring the utility of vagina as an internal control. J. Biomed. Opt. **19**(8), 087001 (2014)
124. Wood, J.J., et al.: Evaluation of a confocal Raman probe for pathological diagnosis during colonoscopy. Colorectal Dis. **16**(9), 732–738 (2014)
125. Petersen, D., et al.: Raman fiber-optical method for colon cancer detection: cross-validation and outlier identification approach. Spectrochim. Acta A Mol. Biomol. Spectrosc. **181**, 270–275 (2017)
126. Almond, L.M., et al.: Endoscopic Raman spectroscopy enables objective diagnosis of dysplasia in Barrett's esophagus. Gastrointest. Endosc. **79**(1), 37–45 (2014)
127. Bergholt, M.S., et al.: Fiberoptic confocal Raman spectroscopy for real-time *in vivo* diagnosis of dysplasia in Barrett's esophagus. Gastroenterology. **146**(1), 27–32 (2014)
128. Lloyd, G.R., et al.: Utilising non-consensus pathology measurements to improve the diagnosis of oesophageal cancer using a Raman spectroscopic probe. Analyst. **139**(2), 381–388 (2014)
129. Maeda, Y., et al.: Measurement of the human esophageal cancer in an early stage with Raman spectroscopy. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2014)
130. Bergholt, M.S., et al.: Real-time depth-resolved fiber optic Raman endoscopy for *in vivo* diagnosis of gastric precancer. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2014)
131. Wang, J., et al.: Comparative study of the endoscope-based bevelled and volume fiber-optic Raman probes for optical diagnosis of gastric dysplasia *in vivo* at endoscopy. Anal. Bioanal. Chem. **407**(27), 8303–8310 (2015)
132. Ming, L.C., et al.: Real time near-infrared Raman spectroscopy for the diagnosis of nasopharyngeal cancer. Oncotarget. **8**(30), 49443–49450 (2017)
133. Yan, B., et al.: An intraoperative diagnosis of parotid gland tumors using Raman spectroscopy and support vector machine. Laser Phys. **24**(11), 115601 (2014)
134. McGregor, H.C., et al.: Real-time endoscopic Raman spectroscopy for *in vivo* early lung cancer detection. J. Biophotonics. **10**(1), 98–110 (2017)
135. Guze, K., et al.: Pilot study: Raman spectroscopy in differentiating premalignant and malignant oral lesions from normal mucosa and benign lesions in humans. Head Neck. **37**(4), 511–517 (2015)
136. Krishna, H., et al.: *In vivo* Raman spectroscopy for detection of oral neoplasia: a pilot clinical study. J. Biophotonics. **7**(9), 690–702 (2014)
137. Sahu, A., et al.: Classification of oral cancers using Raman spectroscopy of serum. In: SPIE BiOS. International Society for Optics and Photonics, San Francisco (2014)
138. Sahu, A., et al.: Raman spectroscopy and cytopathology of oral exfoliated cells for oral cancer diagnosis. Anal. Methods. **7**(18), 7548–7559 (2015)
139. Sahu, A., et al.: Recurrence prediction in oral cancers: a serum Raman spectroscopy study. Analyst. **140**(7), 2294–2301 (2015)

140. Sahu, A., et al.: Raman exfoliative cytology for oral precancer diagnosis. J. Biomed. Opt. **22**(11), 115003 (2017)
141. Yasser, M., et al.: Raman spectroscopic study of radioresistant oral cancer sublines established by fractionated ionizing radiation. PLoS One. **9**(5), e97777 (2014)
142. Aubertin, K., et al.: Mesoscopic characterization of prostate cancer using Raman spectroscopy: potential for diagnostics and therapeutics. BJU Int. **122**(2), 326–336 (2018)
143. Silveira Jr., L., et al.: Discrimination of prostate carcinoma from benign prostate tissue fragments *in vitro* by estimating the gross biochemical alterations through Raman spectroscopy. Lasers Med. Sci. **29**(4), 1469–1477 (2014)
144. Lim, L., et al.: Clinical study of noninvasive *in vivo* melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis. J. Biomed. Opt. **19**(11), 117003 (2014)
145. Zakharov, V.P., et al.: Two-step Raman spectroscopy method for tumor diagnosis. In: SPIE Photonics Europe. International Society for Optics and Photonics, Brussels (2014)
146. Schleusener, J., et al.: *In vivo* study for the discrimination of cancerous and normal skin using fibre probe-based Raman spectroscopy. Exp. Dermatol. **24**(10), 767–772 (2015)
147. Silveira, F.L., et al.: Discrimination of non-melanoma skin lesions from non-tumor human skin tissues *in vivo* using Raman spectroscopy and multivariate statistics. Lasers Surg. Med. **47**(1), 6–16 (2015)
148. Zhao, J., et al.: Real-time Raman spectroscopy for automatic *in vivo* skin cancer detection: an independent validation. Anal. Bioanal. Chem. **407**(27), 8373–8379 (2015)
149. Bergholt, M.S., et al.: Characterizing variability of *in vivo* Raman spectroscopic properties of different anatomical sites of normal colorectal tissue towards cancer diagnosis at colonoscopy. Anal. Chem. **87**(2), 960–966 (2015)
150. Bergholt, M.S., et al.: Simultaneous fingerprint and high-wavenumber fiber-optic Raman spectroscopy enhances real-time *in vivo* diagnosis of adenomatous polyps during colonoscopy. J. Biophotonics. **9**(4), 333–342 (2016)
151. Wang, J., et al.: Simultaneous fingerprint and high-wavenumber fiber-optic Raman spectroscopy improves *in vivo* diagnosis of esophageal squamous cell carcinoma at endoscopy. Sci. Rep. **5**, 12957 (2015)
152. Wang, J., et al.: Fiber-optic Raman spectroscopy for *in vivo* diagnosis of gastric dysplasia. Faraday Discuss. **187**, 377–392 (2016)
153. Holler, S., et al.: Raman spectroscopy of head and neck cancer: separation of malignant and healthy tissue using signatures outside the "fingerprint" region. Biosensors. **7**(2), 20 (2017)
154. Lin, K., et al.: Real-time *in vivo* diagnosis of nasopharyngeal carcinoma using rapid fiber-optic Raman spectroscopy. Theranostics. **7**(14), 3517–3526 (2017)
155. Lin, K., et al.: Real-time *in vivo* diagnosis of laryngeal carcinoma with rapid fiber-optic Raman spectroscopy. Biomed. Opt. Express. **7**(9), 3705–3715 (2016)
156. Kallaway, C., et al.: Advances in the clinical application of Raman spectroscopy for cancer diagnostics. Photodiagnosis Photodyn. Ther. **10**(3), 207–219 (2013)
157. Pence, I., Mahadevan-Jansen, A.: Clinical instrumentation and applications of Raman spectroscopy. Chem. Soc. Rev. **45**(7), 1958–1979 (2016)
158. Santos, I.P., et al.: Raman spectroscopy for cancer detection and cancer surgery guidance: translation to the clinics. Analyst. **142**(17), 3025–3047 (2017)