1-1-2012

# Raman spectroscopy and diffuse reflectance spectroscopy for diagnosis of human cancer and acanthosis nigricans

Suneetha Devpura
*Wayne State University,*

## Recommended Citation

**RAMAN SPECTROSCOPY AND DIFFUSE REFLECTANCE SPECTROSCOPY FOR DIAGNOSIS OF HUMAN CANCER AND ACANTHOSIS NIGRICANS**

by

**SUNEETHA DEVPURA**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2012

MAJOR: PHYSICS

Approved by:

_____

(Advisor)                              Date

_____

(Co-Advisor)                           Date

_____


_____


_____

**DEDICATION**

**"**_To my loving husband, my parents, and my sisters_**"**

pediatric tumors and giving useful tips to train my eyes to identify these tumor or normal cells under a microscope.  I also thank Dr. Olena Palyvoda for her kind support.

I would also like to thank all the professors who strengthened my Physics knowledge, Dr. Gavin Lawes, Dr. Paul Keyes, Dr. Sergei Voloshin, Dr. Jhy-Jiun Chang, Dr. Paul Karchin, Dr. William Rolnick, Dr. Robert Harr, Dr. Zhixian Zhou and all the other faculty members, especially Dr. Jogindra Wadehra for advising me and directing me to complete all the necessary documents to complete my thesis work.  Also I thank Dr. Chandran Sudakar and Dr. Ambesh Dixit for helping me with the Raman instrument at Physics and for the useful discussions.

There are not enough words to express my gratitude towards my parents, so I'm going to simply say "thank you" to my loving parents, Dharmasena Devpura and Lakshmi Wijeweera.  I also thank all my sisters for helping me in their own ways, especially my third sister, Neluka Devpura for her kind guidance, support, and encouragements to be the person I am today.

There's a saying that behind every successful man there is a woman, in my case it's quite the opposite.  My loving husband, Akila Kumarasiri, was with me every step of the way.  Thank you very much for everything.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

**CHAPTER 1**

**INTRODUCTION AND BACKGROUND**

**1.1 Cancer/Acanthosis Nigricans**

Cancer and diabetes are common chronic diseases in the world causing numerous deaths in adults as well as children.  Cancer can occur in various organs of adults and children.  Some of the common types of cancers in adults are prostate, breast, colon, head and neck squamous cell carcinoma, while among children; neuroblastoma and ganglioneuroma are quite common.  In 2008, 903,500 (14%) new prostate cases were recorded and about 258,400 (6%) men died.  The highest incidence rates are observed in the developed countries such as Oceania, Europe, and North America.[1]  There are 25.8 million (8.3% of US population) children and adults in America who suffer from diabetes and nearly 7 million people are not aware of their diabetic condition.[2]  If untreated, many long term health problems such as kidney failures, blindness, neuropathies, risk of heart diseases, stroke and complications in pregnancy can happen.

What is common to both these illnesses is the necessity to detect these conditions in their early stages so that appropriate medical treatments can be provided to avoid complications in their later stages and possibly prevent deaths.  Most common early cancer detection techniques include screening tests such as mammography, cytology, prostate specific antigen (PSA), trans-rectal ultrasound tests, Pap exams, biopsies, and blood tests. In addition, imaging tests such as CT, MRIs, and nuclear medical scans are also used widely.

Most of the above diagnostic procedures are lengthy and expensive.  With the advancement in optical technologies, it is becoming desirable to have non-invasive and reliable optical detection methods whose results are interpreted objectively.  Raman spectroscopy and diffuse reflectance spectroscopy are two of the widely used analytical optical spectroscopic methods where the spectral output contains information about the

molecular composition of the target tissue.  For example, Raman spectroscopy can measure changes in the bio-molecular composition of a tissue.  At the onset of the disease in an organ, the biochemical composition begins to change from its normal values.

Diffuse reflectance spectroscopy provides direct measurement of light absorption and scattering from a medium.  For example, diffuse reflectance spectrum can provide information about the main chromophores (light absorbing chemical groups) present in skin (hemoglobin and melanin) and other optical properties such as scattering coefficients and scatterer size.[3]

The main goal of our research work is to investigate the applications of (1) Raman spectroscopy to diagnose head and neck squamous cell carcinoma, prostate cancer, and pediatric tumors, and (2) diffuse reflectance spectroscopy to detect acanthosis nigiricans skin disorder.  The optical spectroscopic methods can possibly detect the onset of a disease which otherwise can be elusive to conventional methods of detection or to the person suffering from it.

### 1.1.1 Head and Neck Squamous Cell Carcinoma

The most common malignancy in the head and neck region is the head and neck squamous cell carcinoma.  Worldwide, 644,000 new cases are diagnosed every year, of which about 49,260 cases are reported in the United States, resulting in 16,320 deaths every year.[4]  The main tumor sites are oral cavity (tongue, lip or floor of mouth), nasal cavity, paranasal sinuses, salivary glands, pharynx (nasopharynx, oropharynx, and hypopharynx), larynx, trachea, and esophagus (see Fig. 1.1).  The use of tobacco including smokeless tobacco products is the major cause for all of the head and neck cancers.  The risk of developing squamous cell carcinoma in oral cavity is higher for the people who are habitual smokers, alcohol consumers, and betel quid chewers.[5]
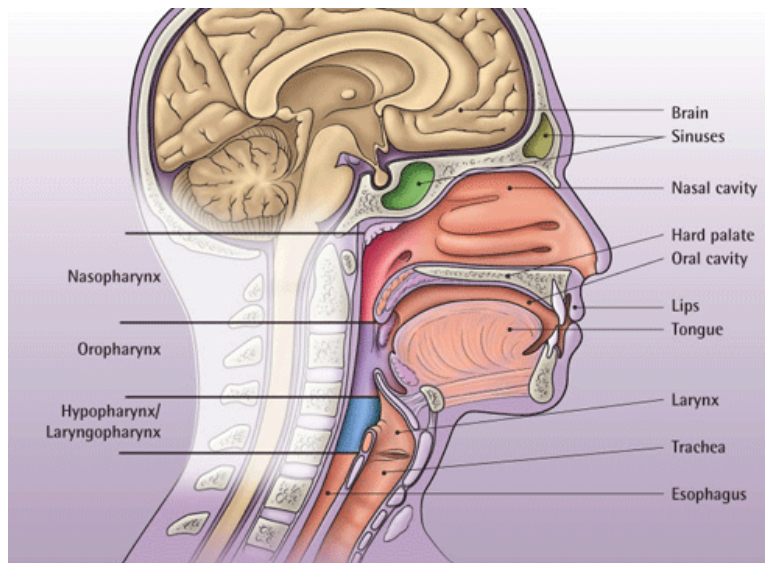
Figure 1.1 Tumor sites for head and neck cancers[6]

Carcinoma *in situ* is the precursor lesion of the head and neck squamous cell carcinoma. Squamous dysplasia, an abnormality of the squamous epithelium is also considered as the precursor to invasive squamous cell carcinoma. However, it is difficult to separate squamous dysplasia from carcinoma *in situ* as the morphological changes show a continuum. The key pathological findings to distinguish between normal and carcinoma *in situ* are architectural disorder, nuclear size variability, nuclear irregularity, and the nuclear chromatin changes.[7] During the progression of carcinoma, downward growth, complex and irregular branching, paradoxical keratinization, excessive volume of atypical epithelial proliferation in the stroma, and inflammatory stromal response to the invading tumor can happen. An example of epithelial cells invading the stroma can be seen from Fig. 1.2.
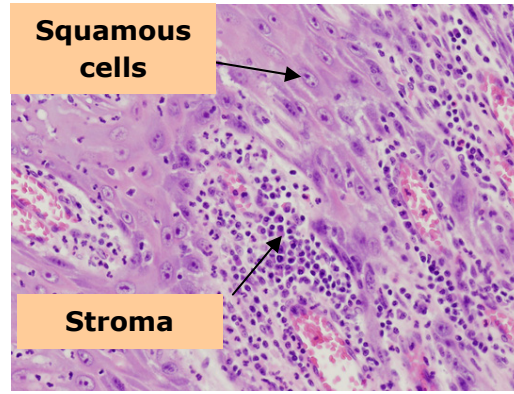
Figure 1.2 Tongue squamous cell carcinoma: Invasion of epithelial cells into the stroma

Currently, the tumor-node-metastasis (TNM) staging is one of the most important factors clinically used for estimation of survival. The clinical stages are listed in the Table 1.1. Reliable estimates of survival are important not only for patient counseling but it may also impact the physicians' choice of therapy. Physicians take into consideration the TNM stage when recommending treatment options. Early stage patients are offered single modality treatment such as radiation or surgery, while multimodality treatment is offered for advanced stage cancer. Head and neck squamous cell carcinoma is a heterogeneous disease and TNM staging provides only a rough estimation.[8] Not all patients in the same tumor stage respond in a uniform manner to the available treatment options. With the availability of novel technologies, there is an urgent need to improve diagnostic accuracy, prognosis, and treatment of head and neck squamous cell carcinoma patients.

| Clinical stage | Description |
|---|---|
| 0 | Tis, N0, M0 |
| I | T1, N0, M0 |
| II | T2, N0, M0 |
| III | T3, N0, M0/ T1-3, N1,M0 |
| IV | T4, N0-1-M0/Any T, N2-3, M0/Any T, and N, M1 |

Table 1.1 A description of TNM clinical staging[8]

The nomenclatures of the Table 1.1 are; Tis: carcinoma *in situ,* T1: Tumor less than 2 cm in dimension, T2: Tumor size is between 2 cm and 4 cm, T4: Tumor invades adjacent structures, N0: No regional lymph node metastases (M), N1: M to a single ipsilateral lymph node (≤3 cm), N2: N1 but less than or equal 6 cm, M0: No evidence of distance M, and M1: Distance M can be seen.

### 1.1.2. Prostate Cancer

Prostate cancer is one of the most common type of cancers among North American men. The mortality rate for this disease can be dramatically reduced if it can be diagnosed in its early stages with some reliable technique. Cancer occurs in the prostate gland which is a part of the male reproductive system as shown in Fig. 1.3. It produces and stores an alkaline fluid which provides an extra medium for mobility and life of sperm. The prostate consists of three zones: peripheral zone, transition zone and central zone.
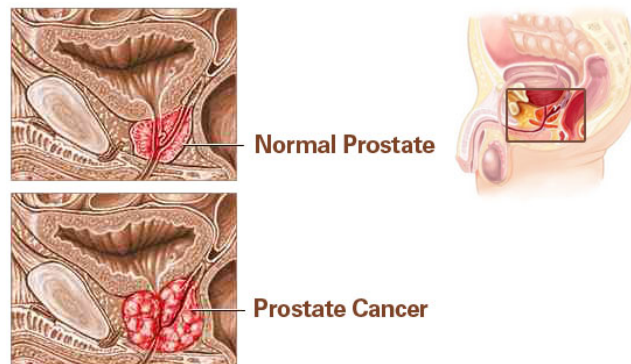


Figure 1.3 Prostate Gland[9]

The peripheral zone is nearly 70% of the glandular mass of the prostate. Most of the prostate cancers originate predominantly in this zone. The central zone is nearly 25% of the prostate and rarely contributes to prostatic carcinoma. If the cancer develops from the epithelial cells, it is referred to as "carcinoma". However, it contributes to the spread of the

cancer since it is surrounded and accompanied by the ejaculatory ducts from the seminal vessel.[10]   The transition zone which is 5% of prostate volume is the origin of basal cell hyperplasia, clear cell cribriform hyperplasia and clear cell carcinoma.

### a) Prostatic Intraepithelial Neoplasia

Prostatic Intraepithelial Neoplasia (PIN) is a precursor lesion in prostate cancer which can be of high grade or low grade category.  Usually, PIN is classified as high grade capable of developing into cancer within the next 10 years or so.[11]   One study has suggested that anti-tumor activity of certain dietary flavonoids prevents the progression of high grade PIN to cancer.[12]   However, this hypothesis could not be established in a randomized double-blind study performed with 303 men in twelve Canadian centers.  These men were given soy, vitamin E, and selenium on daily basis for three years.  The results were not statistically significant to show their effects on decreasing the progression of cancer or eliminating it.[13]   When the cancer progresses, the epithelium thickens due to the development of atypical cells.  In addition, intra-luminal papillae are formed as seen in Fig. 1.4 (b) and ultimately the glandular lumen will be linked creating a papillary arrangement.



(a)                              (b)
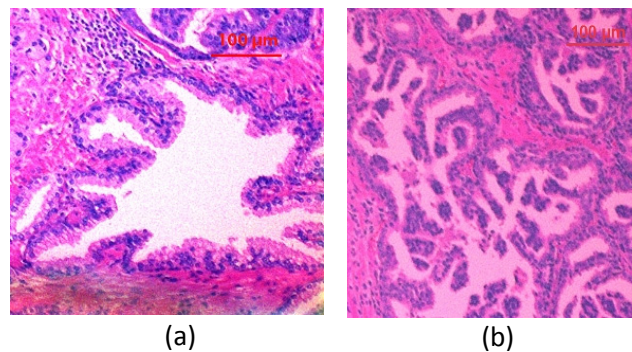
Figure 1.4 (a) Low-grade PIN which has a star shaped glandular form with a prominent basal layer (b) High-grade PIN shows increased number of atypical cells in the lining of the duct

PIN is mainly seen in needle biopsies of benign prostatic hyperplasia (an enlargement of prostate in men over 50) and in cancer patients as well. The patients who have high grade PIN histology have high probability of developing prostate cancer.

### b) Categorization of Cancer: The Gleason Grading System

Prostate cancer is categorized using the Gleason system which is used to evaluate the prognosis of men with prostate cancer. In this system, a tissue is assigned five grades ranging from 1 to 5 to indicate the degree to which the cancerous tissue differs, histologically, compared to normal prostate tissue. Grade 1 is more towards the normal prostate morphology which contains proliferated microacinar structures contoured by prostatic luminal cells without a basal cell layer.[14] Grade 5 contains irregular, poorly differentiated infiltrating individual cells, and grade 2, 3, and 4 fall in between as shown in Fig. 1.5.
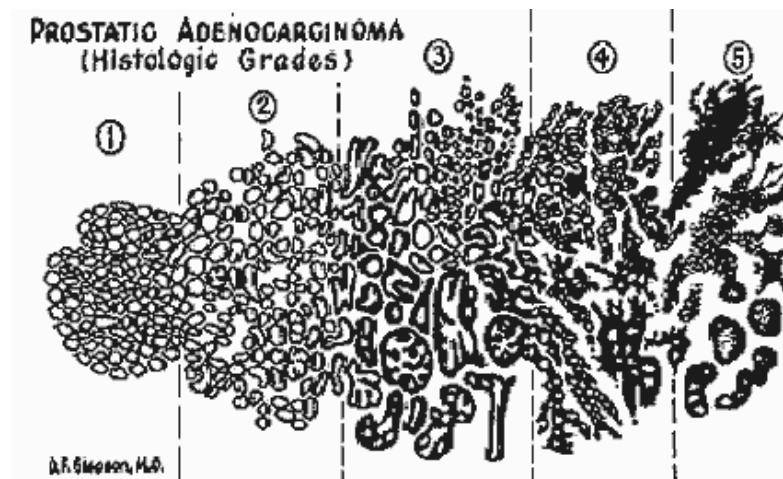


Figure 1.5 Gleason system of cancer cell architecture[15]

Since the different parts of cancer may have different grades, pathologists grade two areas and add the two grades. An example of grading is as follows: consider Gleason grade 7, this can be described as a combination of 3 and 4 as 3+4=7. Here, the first score

represents the most prevailing grade of the tissue and the second score represents the next most prevailing grade.  Therefore, the Gleason grade varies between 2 to 10.  A Gleason grade of 6 is the most commonly seen Gleason grade.  It has been reported that the DNA content of PIN is in between the DNA content of benign glands and cancer.[10]  In addition, as the number of aneuploid nuclei or abnormal number of chromosomes increases, so does the level of PIN.  However, the proliferation activity and ploidy values (the amount of DNA content in a cell) of PIN and carcinoma are similar to each other.  The progression of PIN stage to carcinoma stage is evident in the basal layer with an increase of mean nuclear volume in PIN, which was discovered by Teba *et al*.[16]  However, the nuclear features of the luminal layer in PIN remain the same as in carcinoma.

### 1.1.3 Pediatric Tumors

Neuroblastoma and ganglioneuroma are peripheral neuroblastic tumors.  These are derived from immature sympathetic nervous system and adrenal gland.  Neuroblastoma, the third most common type of cancer, occurs in embryonic neuroblasts of the sympathetic peripheral nervous system which can spontaneously be matured into ganglioneuroma.[17]  It is a solid malignant tumor which can be observed in abdomen, around the spinal cord near the chest region, or pelvis.  Neuroblastoma has been called the "great masquerader" since its symptoms mimic many other diseases.[18]

Prognostic Evaluation of Neuroblastic Tumors According to the International Neuroblastoma Pathology Classification (Shimada System)

| International Neuroblastoma Pathology classification | | Original Shimada classification | Prognostic group |
|---|---|---|---|
| Neuroblastoma<br>Favorable | (Schwannian stroma-poor)[a] | Stroma-poor<br>Favorable | Favorable |
| < 1.5 yrs | Poorly differentiated or differentiating<br>& low or intermediate MKI tumor | | |
| 1.5–5 yrs | Differentiating & low MKI tumor | | |
| Unfavorable | | Unfavorable | Unfavorable |
| < 1.5 yrs | a) undifferentiated tumor[b]<br>b) high MKI tumor | | |
| 1.5–5 yrs | a) undifferentiated or poorly<br>differentiated tumor<br>b) intermediate or high MKI tumor | | |
| ≥5 yrs | All tumors | | |
| Ganglioneuroblastoma, intermixed | (Schwannian stroma-rich) | Stroma-rich Intermixed<br>(favorable) | Favorable[c] |
| Ganglioneuroma<br>Maturing<br>Mature | (Schwannian stroma-dominant) | Well differentiated (favorable)<br>Ganglioneuroma | Favorable[c] |
| Ganglioneuroblastoma, nodular | (composite Schwannian stroma-rich/<br>stroma-dominate and stroma-<br>poor) | Stroma-rich nodular (unfavorable) | Unfavorable[c] |

MKI: mitosis-karyorrhexis index.

[a] Subtypes of neuroblastoma were described in detail elsewhere.[10]

[b] Rare subtype, especially diagnosed in this age group. Further investigation and analysis required.

[c] Prognostic grouping for these tumor categories is not related to patient age.

Table 1.2 International Shimada classification[19, 20]

According to International Neuroblastoma Pathology Classification, there are four basic morphological categories: neuroblastoma (Schwannian stroma-poor), ganglioneuroblastoma-intermixed (Schwannian stroma-rich), ganglioneuroma (Schwannian stroma-dominant), ganglioneuroblastoma-nodular (Composite, Schwannian stroma-rich/Schwannian stroma-dominant and Schwannian stroma-poor) as described in Table 1.2.

### 1.1.3.1. Neuroblastoma (Schwannian Stroma-poor)

Neuroblastoma frequently occurs in infants and children. This is subdivided into 3 groups: differentiating, undifferentiated, and poorly differentiated. The poorly differentiated and differentiating neuroblastoma are further categorized as favorable and unfavorable cases. For favorable case, there is a probability of appropriate age maturation, and while for biologically unfavorable peripheral neuroblastic tumors, there is no such a probability. The survival rate of favorable neuroblastoma is much higher than the unfavorable

histopathology.  Figure 1.6 shows a pathology image of favorable neuroblastoma with poor surrounding stroma.
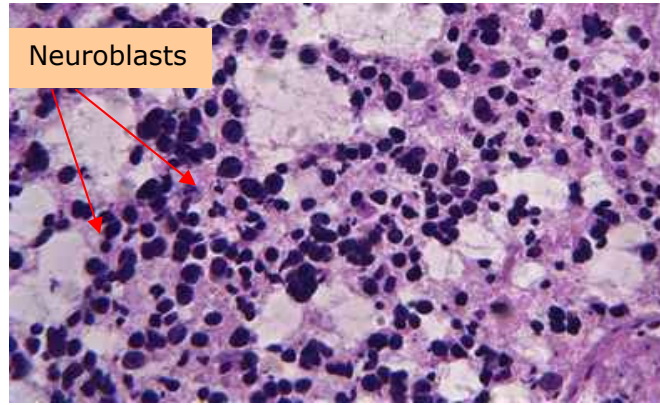


Figure 1.6 Neuroblastoma with stroma poor in the background

1. ***Undifferentiated Neuroblastoma:*** When a tumor tissue is composed of undifferentiated neuroblastic cells without identifiable neuropil (region between neuronal cell bodies in the gray matter of the brain and spinal cord), it falls under this subcategory.  These neuroblasts have indistinct nuclei cytoplasm borders where size can vary from small to medium.

2. ***Poorly Differentiated Neuroblastoma***: In this category most of the tumor cells are undifferentiated but still there is a presence of varying amount of neuropil corresponding to Schwannian differentiation.  Ganglion cells appear up to 5% of the tumor cells but still can be distinguished from neuroblasts.

3. ***Differentiating Neuroblastoma***: More than 5% of the tumor cells are differentiating neuroblasts and ganglion cells.  These neuroblasts have distinct borders of cytoplasm and they also show synchronous enlargement of nuclei and cytoplasm.

### 1.1.3.2. Ganglioneuroma (Schwannian Stroma-dominant)

Ganglioneuroma contains distributed maturing or mature ganglion cells with predominantly dominant Schwannian stroma in the background.  This type of tumor can be commonly seen in locations such as posterior mediastinum and retroperitoneum.  However, it is rarely seen in adrenal gland.  Ganglioneuroma can be subdivided as maturing and mature.  Mature ganglioneuroma contains fully matured ganglion cells and Schwannian stroma (see Fig. 1.7) while maturing ganglioneuroma has maturing ganglion cells and less contents of differentiating neuroblasts.
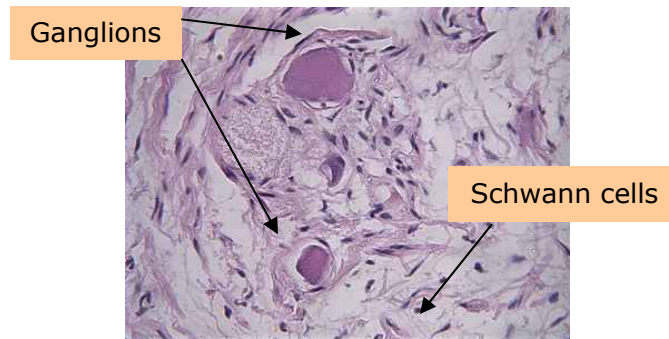


Figure 1.7 Ganglioneuroma

### 1.1.3.3 Wilms' Tumors

Wilms' tumor is the most common childhood kidney cancer.  Due to therapeutic advancement, the survival rate in the United States has increased to 90% in 1996.  However, the survivors are at a greater risk of dying after several years.[21]  Nephrogenic rest (persistent embryonal remnants in the kidney) is considered as the precursors of Wilms' tumor.  When multiple or diffuse nephrogenic rests are observed, they are called "nephroblastomatosis".  The common characteristics of nephrogenic rest are dormancy, maturation, involution, hyperplastic overgrowth, and neoplastic induction.[22]

### 1.1.4. Acanthosis Nigricans

Obesity is one of the major health problems in the USA. According to one overweight and obesity prevalence estimation, over two thirds of American adults are overweight and over one third of adults are obese.[23] Among insulin-resistant diabetic patients who also suffer from obesity, acanthosis nigricans is a very common dermatologic condition causing skin darkening and roughening mostly occurring in the posterior (see Fig. 1.8) and lateral folds of the neck, the axilla, inframammary, groin, and other areas.



Figure 1.8 Acanthosis nigicrican skin condition in the neck area

Classically, it was described as a herald for internal malignancy, but in 1976, acanthosis nigricans was shown to have a direct association with hyperinsulinemia.[24, 25, 26] It is considered a cutaneous marker for insulin resistant states including type 2 diabetes mellitus and polycystic ovarian syndrome.[27, 28, 29, 30]

In the year 2000, the American Diabetes Association established acanthosis nigricans as a criterion for identifying children at risk for developing diabetes mellitus.[31, 32] It has been demonstrated that hyperinsulinemia or insulin resistant states can lead to acanthosis nigricans. At higher concentrations, insulin interacts with insulin-like growth factor-1 receptors particularly on keratinocytes and fibroblasts in skin; this results in acanthosis nigricans, a state of hyperproliferation.[33] As hyperinsulinemia is treated, acanthosis

nigricans should improve and may entirely resolve.   However, the exact relationship between insulin, obesity, and insulin resistance of diabetic patients is not yet fully understood.[34]   Acanthosis nigricans can occur due to several other conditions such as a result of glandular disorder, Addison disease (deficiency of hormones from the adrenal gland), disorder from the pituitary gland, low level of thyroid hormones, and oral contraceptives.  However, obesity remains the major cause.[35]

Skin has a complex structure as shown in Fig. 1.9.  The human skin has many layers.  The three main layers are called the epidermis, dermis, and hypodermis (subcutis).
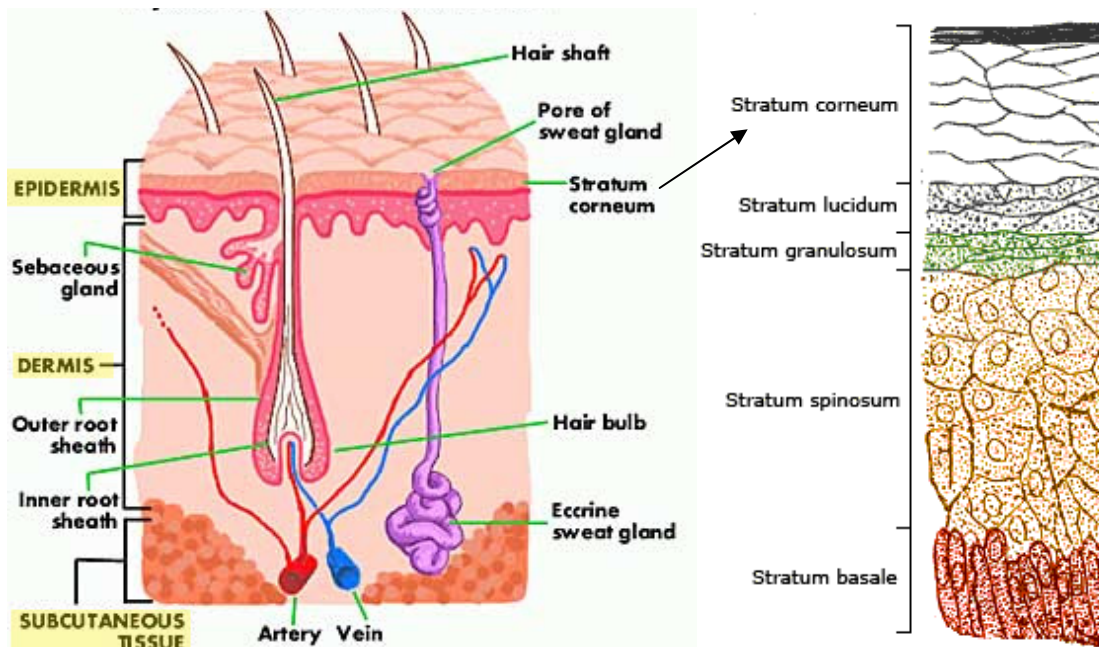


Figure 1.9 Skin Structures[36, 37]

The epidermis, stratified epithelium that forms the outer layer of the skin can further be subdivided into 5 layers, namely, stratum basale, stratum spinosum, stratum corneum, stratum lucidum, and stratum granulosum.  The bottom layer of epidermis is the stratum basale where cell division occurs, followed by keratinization in the layer of stratum spinosum.[38]   The keratinization is the skin renewal process where daughter cells are

transformed into pickle cells and then they die eventually. These dead cells are pushed to the top layer of stratum corneum. The next adjacent layer, stratum lucidum helps to reduce friction and shear forces between stratum corneum and stratum granulosum which can be seen in thick skin. It also protects the skin from ultra-violet rays. The dermis consists of two layers; papillary layer which contains thin collagen fibers that provides strength to the skin and reticular layer which contains thick collagen fibers that will add support to the surrounding cells.

The flexibility of skin is due to elastic fibers which are in the reticular layer. Hair follicles, lymph vessels, smooth muscles nerve cells, and sweat glands also grow in this layer. The blood flows to the epidermis through the capillaries in the dermis. The hypodermis contains loosely connective tissues and bases of hair follicles and sweat glands.

## 1.2. Optical Techniques Used in Diagnosis

Detection and confirmation of any disease/cancer is very crucial for its successful treatment in order to improve survival rate. Sometimes, the standard screening programs can provide misleading results leading to wrong or over-treatments, and occasionally, to fatal consequences. The interpretations of histological examination of biopsies, considered as the "gold standard" for diagnosis, are often subjective and can vary significantly from one pathologist to another.[39] Hence, it is imperative to detect the state of the disease with a method which is objective and capable of providing results within a very short period of time. Optical spectroscopy techniques, namely Raman spectroscopy, diffuse reflectance spectroscopy, etc. are very well suited for these types of goals as they are also capable of probing diseases at cellular level.

### 1.2.1 Raman Spectroscopy

Raman spectroscopy can provide information about the changes in the concentrations of the constituent biomolecules of tissues and detect the progression or state of the disease.  In Raman spectroscopy measurements, a monochromatic laser light incident on a sample and the scattered light is collected and analyzed to derive the information about the contents of the sample.  A majority of the light scatters elastically; however, a very small fraction of it scatters inelastically carrying information about the molecular nature of the sample, their mutual interaction, and their relative concentrations in the sample.[40]  The chemical nature of molecules can be uniquely determined from their vibrational energy levels and Raman spectroscopy has the capability to measure these vibrational energy levels.[41, 42]

Typical Raman spectra from some biopolymers and cells obtained by Notingher *et al*[43] are shown in Fig. 1.10.  This study demonstrates that Raman spectroscopy can become a powerful investigation method to determine the biochemical composition of a biological material.
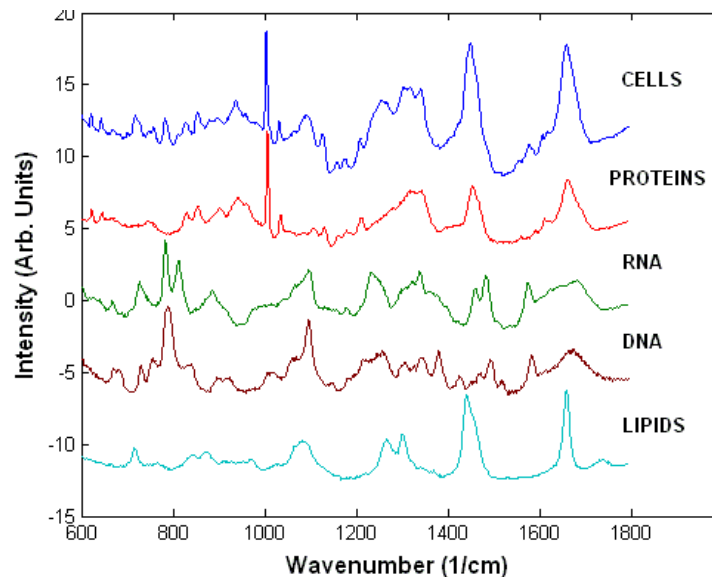


Figure 1.10 A typical Raman spectra of cells and biopolymers[43]

The peaks observed in a Raman spectrum of a tissue sample are often assigned to the vibrational energy states of the chemical groups and the backbones structures of the molecules. Table 1.3 lists the assignments found in the literature for the Raman bands (in the 600 - 1800 cm$^{-1}$ region) observed in tissue samples.

| Peak (cm$^{-1}$) | Assignment | | | |
| --- | --- | --- | --- | --- |
| | Nucleic Acids | Proteins | Lipids | Carbohydrates |
| 1736 | | | C=O ester | |
| 1680-1655 | | Amide I | C=C str. | |
| 1617 | | C=C Tyr, Trp | | |
| 1607 | | C=C Phe, Tyr | | |
| 1578 | G, A | | | |
| 1480-1420 | G, A, CH def | C-H | CH def | CH def |
| 1342 | A, G | C-H | | CH def |
| 1320 | G | C-H | | |
| 1301 | | | CH$_2$ twist | |
| 1284-1220 | T, A | Amide III | =CH bend | |
| 1209 | | C-C$_6$H$_5$. Phe, Trp | | |
| 1176 | | C-H bend Tyr | | |
| 1158 | | C-C/C-N str. | | |
| 1128 | | C-N str. | | C-O str |
| 1095-1060 | PO$_2^-$ str. | | Chain C-C str. | C-O, C-C str |
| 1033 | | C-H in-plane Phe | | |
| 1005 | | Sym. Ring br Phe | | |
| 980 | | C-C BK str. β-sheet | =CH bend | |
| 937 | | C-C BK str. α-helix | | C-O-C glycos. |
| 877 | | | C-C-N$^+$ sym str | C-O-C ring |
| 854 | | Ring br Tyr | | |
| 828 | O-P-O asym.str. | Ring br. Tyr | | |
| 811 | O-P-O str. RNA | | | |
| 788 | O-P-O str. DNA | | | |
| 782 | U,C,T ring br | | | |
| 760 | | Ring breath Trp | | |
| 729 | A | | | |
| 717 | | | CN$^+$(CH$_3$)$_3$ str | |
| 667 | T, G | | | |
| 645 | | C-C twist Tyr | | |
| 621 | | C-C twist Phe | | |

Table 1.3 Peak assignments of Raman spectra[43]

A Raman mapping study was done by Jong *et al* using tumor and healthy bladder tissues.[44] Figure 1.11 shows the mean spectra corresponding to those areas.
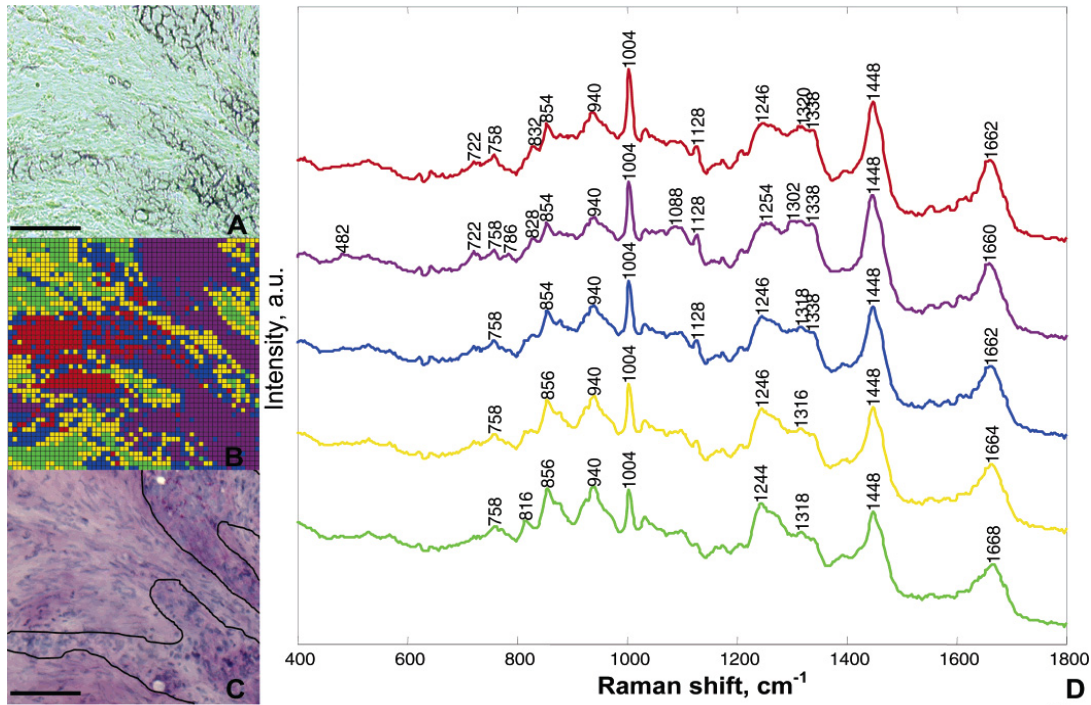


Figure 1.11 (A) and (C) are unstained and stained tissue sections which have non-tumor and tumor areas, (B) represents matching colors and (D) average spectra: red-muscle tissue, purple-tumor, green-collagen fiber and blue and yellow transition between the two types[44]

Using near-infrared Raman spectroscopy, a significant spectral differences between normal and dysplastic gastric mucosa tissues were found, particularly in the spectral ranges of 850–1150, 1200–1500, and 1600–1750 cm$^{-1}$, and were able to differentiate dysplasia from normal gastric mucosa.[45] Haka *et al* studied freshly excised surgical breast cancer tissues and successfully developed an algorithm to diagnose their pathological states.[46] Crowe *et al* reported a study of both the adenocarcinoma biopsies[47] and cell lines[48] of prostate cancer and successfully classified the benign and cancerous cells/tissues.

The studies on nasopharyngeal biopsy tissues showed that the spectral features of three Raman bands varied significantly when tissue changed from normal to malignant

state[49], and in another study on frozen tissues from larynx showed that the characteristic Raman spectral features for normal, carcinoma, and papilloma tissues are different.[50] Rabah *et al* reported Raman investigation of neuroblastoma and ganglioneuroma in fresh tissue specimens[51] and found distinct spectral differences between tumor and normal adrenal gland. Subsequently, an investigation was done to distinguish neuroblastoma and related tissues from healthy tissues in fresh vs. frozen specimens.[52] A high correlation between fresh and frozen tissues was found and a faster method of identifying pathological conditions was developed. In addition, normal kidney specimens were distinguished from Wilms' tumor.[53] In this study, both fresh and frozen tissue specimens from Wilms' tumor treated or untreated, normal kidney, and other related renal tumors which show various similar morphologies were used, but still Raman technique could distinguish between them successfully.

### 1.2.2. Diffuse Reflectance Spectroscopy

Diffuse reflectance spectroscopy provides information about the relevant physiological parameters and the biochemical composition. When light is incident on a sample, reflection, scattering, absorption and transmission through the sample can occur. The reflection of light obeys the law of reflection and the transmission of light obeys Snell's law. The Fresnel's equations describe reflection and transmission of light through multiple media having different indices.[54] Further, the reflection can be either specular or diffuse. Specular reflection is the reflection off of a smooth or a glossy surface in which an incoming ray is reflected into a single outgoing direction obeying the law of reflection. Conversely, when parallel rays of light are incident on a rough surface, the direction of the reflected light rays may differ due to different orientations of the surface normals for various incident rays. This type of reflection is called "diffuse reflection".

Many theoretical models have been developed to describe the diffuse reflectance phenomenon. The model proposed by Farrell *et al* in 1992[55] is quite successful in extracting tissue optical properties. This model allows determination of optical properties of tissues from diffuse reflectance spectra, using only the shape of the reflectance curve. This model was improved by Zonios *et al* in 1999[56] and was applied to biological tissues, human colon specimens of normal colonic mucosa, and adenomatous colon polyps (precursor lesion) by introducing four main parameters: hemoglobin concentration, hemoglobin saturation, effective scatterer density, and effective scatterer size of the tissue.

Hemoglobin and melanin are the main chromophore absorbers in the skin. Absorption (scattering) can be explained using the characteristic parameter called absorption (scattering) coefficient $\mu_a$ ($\mu_s$-cm$^{-1}$) which is defined as the probability of photon absorption (scattering) per unit path length in a medium. This coefficient depends on the cross section and the number density of the absorbers (scatterers).[3] Since hemoglobin can be either in the form of oxyhemoglobin or deoxyhemoglobin, their concentrations can be estimated using the known absorption coefficients of biological tissues at different wavelengths. Subsequently, quantitative assessment of human skin melanin, hemoglobin, and light scattering properties were determined using diffuse reflectance spectroscopy in the visible and near-infrared regions based on the same analytical model.[57] Since all these experiments were done using fiber optic probes, attempts to design a ball lens coupled fiber optic probe for depth resolved spectroscopy of epithelial tissues were also reported.[58, 59] Katika *et al* investigated optical properties of human skin using steady state directional diffuse reflectance through numerical simulations using a seven-layer skin model.[60] Nuclei size and density distribution of the mucosal tissues were also extracted through the backscattered light by analyzing the frequency and the amplitude of the fine structure.[61]

Benign and malignant breast tissues were also investigated through diffuse reflectance spectroscopy. Brestlin *et al* found adipose tissues do not contain β-carotene,

and thus, an indicator to differentiate malignant from benign.[62]  In another *in vivo* breast cancer study, hemoglobin saturation and total hemoglobin content was determined from the diffuse spectra and found that the hemoglobin saturation is significantly lower in benign tissues and also negatively correlated with the tumor size.[63]

### 1.2.3. Fluorescence Spectroscopy

Fluorescence spectroscopy has been widely tested to characterize skin.  The principal fluorophores of the skin are tryptophan, pepsin digestible collagen cross-links, collagenase digestible collagen cross-links, elastin cross-links, and NADH.  When characterizing the skin fluorophores, the penetration depth of light plays an important role.  Usually, skin penetration is about 1 mm at about 700 nm and this depth decreases as the wavelength decreases.  Fluorescence excitation spectra showed three excitation bands, 295, 335, and 370 nm in a mouse skin investigation[64].  Hamzavi *et al* observed that blue vitiligo characterization can be done much better using the autofluorescence spectroscopy compared to the reflectance spectroscopy.[65]  A high correlation between fluorescence and skin pigmentation and redness was also seen.  Thus, once the influence of skin pigmentation and redness was corrected a trend between the 375 nm fluorescence and the age can be observed.[66]

### Research Goals

The goal of our research work is to investigate the applications of (1) Raman spectroscopy to diagnose head and neck squamous cell carcinoma, prostate cancer, and pediatric tumors, and (2) diffuse reflectance spectroscopy to detect acanthosis nigricans skin disorder.  In Chapter 2, the theoretical background is presented and Chapter 3 summarizes the instrumentation of both Raman spectroscopy and diffuse reflectance spectroscopy followed by the tissue processing and data collection.  All the research data

are analyzed using chemometric methods such as principal component analysis and discriminant function analysis which is explained in detail in Chapter 4. Chapters 5, 6, and 7 summarizes the results of Raman spectroscopic investigations of *ex vivo* tissue characterization in head and neck squamous cell carcinoma, prostate cancer, and pediatric tumors, respectively whereas Chapter 8 presents the diffuse reflectance and colorimetric investigation of acanthosis nigricans *in vivo*. Chapter 9 presents the conclusions and future work.

**CHAPTER 2**

**RAMAN SPECTROSCOPY AND DIFFUSE REFLECTANCE SPECTROSCOPY**

The theory of Raman scattering is well worked out and the principles of Raman spectroscopy applied to polyatomic molecules, and solids are well explained in number of books.[3, 67, 68] Similarly, the principles of diffuse reflectance spectroscopy and its biological applications have been described in numerous articles and book.[3, 55-61, 69] Here, a brief description of the principles of these techniques will be given.

Light is an electromagnetic wave. When light is incident on a surface of a bulk material, various processes can happen to a varying degree depending on the nature of the surface and the wavelength of the light used. In this research work, the interest lies on absorption and scattering processes.
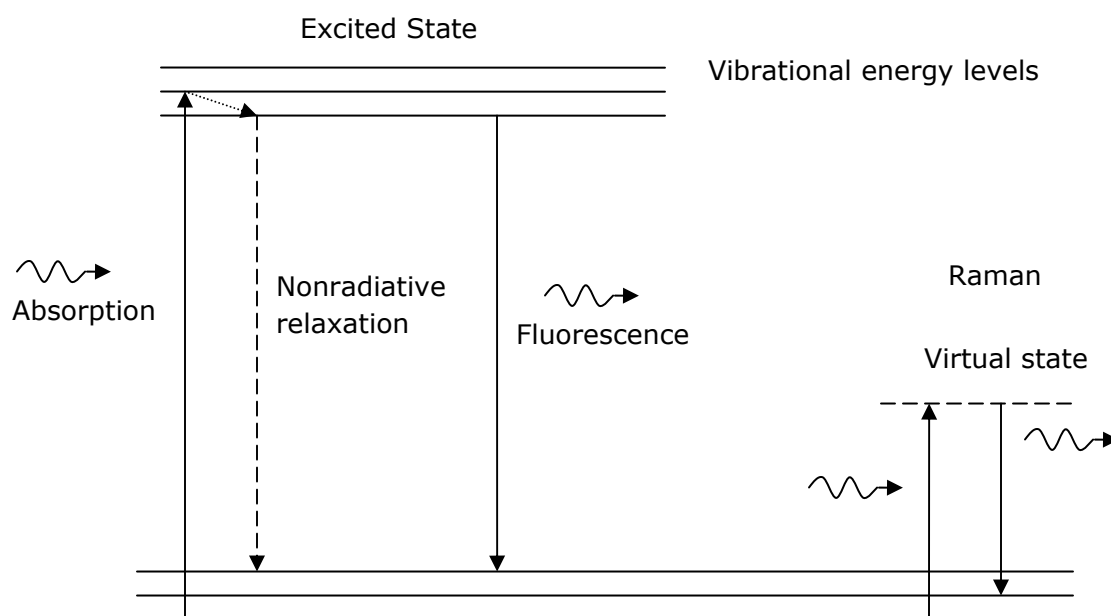
Figure 2.1 Jablonski energy diagram[3]

When a light wave of sufficient energy is incident on a molecule or an atom, it can cause a transition of an electron from the ground state to an excited state. This excitation

process is commonly referred as an electronic transition. In a molecule, the atoms can rotate and vibrate with respect to each other. Thus, for a given electronic state, there are associated discrete vibrational and rotational energy levels, which can be thought of as being packed on top of each electronic level. The excited electrons can relax in radiative or non-radiative processes. If the excited electrons relax without producing photons, then the relaxation process is called non-radiative relaxation.[3]

However, if the excited electron relaxes to the ground state spontaneously by producing another photon, it is called fluorescence. Figure 2.1 schematically describes these relaxation processes. Usually, fluorescence light is shifted to longer wavelengths relative to the excitation light which is commonly referred to as red shift. This is due to the fact that a part of the energy is lost in the excited state due to molecular vibrations. However, if the electrons in the excited state or ground state are in a higher vibrational state due to thermal energy at high temperatures, this may not be the case. Then, the emitted light will have a shorter wavelength than the incident light.

## 2.1 Raman Spectroscopy

The phenomenon of Raman scattering was discovered by Sir Chandrasekhara Venkata Raman in 1928.[40] When a photon is incident on a polarizable molecule it can excite (de-excite) vibrational modes of the molecules, resulting in scattered photons with decreased (increased) energy by the amount of the vibrational transition energies. Raman scattering is an inelastic scattering process as there is an energy exchange between the molecule and the photon. If the scattered photon has the same energy as the incident photon, the scattering is called Rayleigh scattering which is an elastic scattering process.[3]
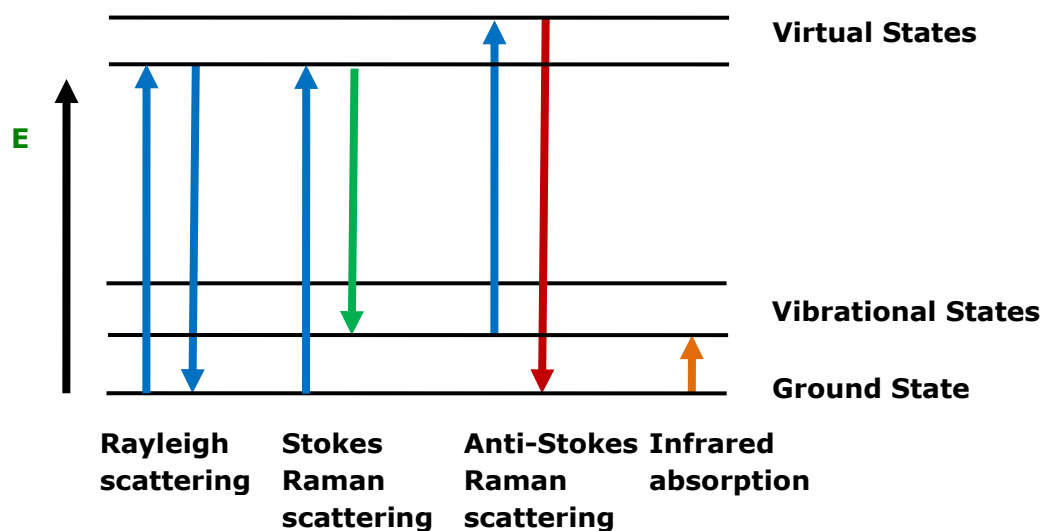
**Figure 2.2** A schematic illustration of scattering and absorption processes

This inelastic scattering is further distinguished as Stokes scattering or anti-Stokes scattering depending on the loss or gain in energy of the scattered photon. In Stokes scattering, the electron is excited to a virtual state. Since, this virtual state is unstable the electron may quickly decay into a higher vibrational instead of the initial state as shown in green arrow in Fig. 2.2. Here, the energy is absorbed by the molecule. Thus, the emitted photon has a longer wavelength or in other words the scattered photon is red shifted. It is common to use the term "red shifted" or "blue shifted" if the energy of the scattered light is lower (longer wavelength) or higher (shorter wavelength). If the scattering results in an excitation from an initial excited vibrational state to a virtual state and subsequent de-excitation to a ground state or a lower excited state than the initial state, then the emitted photon has more energy than the absorbed photon (blue-shifted), and this is called anti-Stokes scattering. However, the number of electrons in the excited vibrational state depends on the Boltzmann distribution.[67]

It is important to distinguish scattering and absorption processes (Fig. 2.2). Infrared absorption involves molecules making direct transitions from one vibrational state to a higher vibrational state by absorbing a photon of exact energy equal to the difference

between the vibrational states ($\sim$ 12–400 meV). In other words, the incident photon energy exactly coincides with the energy difference between vibrational states. This is shown schematically in Fig. 2.2 along with the difference between Stokes and anti-Stokes Raman scattering and Rayleigh scattering.
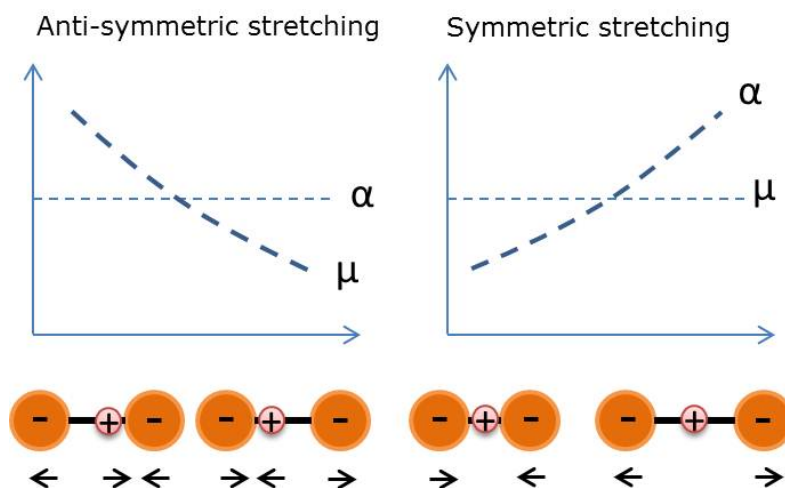


Figure 2.3 Graphical representations of Raman active and inactive modes.[67]

The basic necessary condition for observing the Raman scattering is that the polarizability ($\alpha$) must change during vibrations of the molecules. The bigger the change, the greater the effect is. However, for infrared absorption, a change in dipole moment ($\mu$) causes the absorption[68] and it is a necessary condition. Usually, for Raman scattering, symmetric stretching causes the most intense scattering whereas for infrared absorption asymmetric stretching causes the greatest. In centro-symmetric molecule, both Raman scattering and infrared scattering cannot occur at the same time. Thus, both these techniques can be used to get complimentary information about a particular form of vibration in a molecule.[70, 71]

If a given non-linear molecule contains N atoms in this molecule, then there are 3N-6 possible vibrations (degrees of freedom) except for a linear molecule where there are 3N-

5 possible vibrations.    Molecular vibrations are periodic motions of the atoms in a molecule.[68]

The classical explanation of the Raman scattering is as follows. When an electromagnetic wave interacts with a molecule, the electron cloud around molecule distorts, thus, an induced dipole moment can occur.   This induced dipole moment is a function of the polarizability and the incident electric field, $E$.[71]

$$P = \alpha E \text{-----------------------------------------------------------------------------------------}(2.1.1)$$

The electric field of the incident electromagnetic wave can be written as,

$$E = E_0 \cos(2\pi \upsilon_0 t) \text{---------------------------------------------------------------------------}(2.1.2)$$

$\upsilon_o$ is the frequency of the electromagnetic wave and $E_0$ is its amplitude.   Thus, the induced dipole moment can be expressed as

$$P = \alpha E_0 \cos(2\pi \upsilon_0 t) \text{--------------------------------------------------------------------}(2.1.3)$$

The perturbation of the electron cloud depends on relative location of the atoms of the molecule.[71]   Hence, the polarizability of the molecule depends on the instantaneous position of the atoms in it.   Thus, by using the Taylor expansion, the polarizability can be approximated as

$$\alpha = \alpha_0 + \frac{\partial \alpha}{\partial Q} dQ \text{-----------------------------------------------------------------------}(2.1.4)$$

where, $dQ$ is the displacement of the atoms about their equilibrium position.   For a given vibrational mode, $dQ$ can be written as

$$dQ = Q_0 \cos(2\pi \upsilon_{vib} t) \text{--------------------------------------------------------------------}(2.1.5)$$

where $\upsilon_{vib}$ is the vibrational frequency of the molecule and $Q_o$ is the maximum displacement about the equilibrium.   By substituting, the polarizability is written as,

$$\alpha = \alpha_0 + \frac{\partial \alpha}{\partial Q} Q_0 \cos(2\pi \upsilon_{vib} t) \text{-----------------------------------------------------------}(2.1.6)$$

Hence, induced dipole moment can be expressed as

$$P = \alpha_0 E_0 \cos(2\pi\upsilon_0 t) + \frac{\partial\alpha}{\partial Q} Q_0 E_0 \cos(2\pi\upsilon_0 t)\cos(2\pi\upsilon_{vib} t)$$

--------(2.1.7)

$$= \alpha_0 E_0 \cos(2\pi\upsilon_0 t) + \frac{\partial\alpha}{\partial Q} \frac{Q_0 E_0}{2} \{\cos[2\pi(\upsilon_0 - \upsilon_{vib})t)] + \cos[2\pi(\upsilon_0 + \upsilon_{vib})t]\}$$

This equation illustrates that there are three frequencies which contribute to the induced dipole moment. They are $\upsilon_o$, $(\upsilon_o - \upsilon_{vib})$, and $(\upsilon_o + \upsilon_{vib})$. These three frequencies, respectively, are the scattered frequency, which coincides with the incident light frequency (Rayleigh scattering), the Stokes and the anti-Stokes scattering frequency.[67]

The energy difference between the incident photon and the Raman scattered photon is defined as the Raman shift. Since energy and frequency are proportional, this is also known as the vibrational frequency of the scattered molecule (cm$^{-1}$). Raman shift is given by

$$\overline{\Delta\nu} = \frac{1}{\lambda_{incident}} - \frac{1}{\lambda_{scattered}}$$

---------------------------------------------------------(2.1.8)

where $\overline{\nu}$ is called the wave number in units of cm$^{-1}$. For Raman spectra, vibrational transitions appear in the $10^4$ to $10^2$ cm$^{-1}$ region.

Quantum mechanically, vibrational energy levels of a molecule can be determined by solving the Schrödinger equation for a molecule. In a molecule, the chemical bond between the atoms can be considered as a spring (with spring constant $K$) and the motion of the atoms then can be described by Hooke's law. In treating diatomic molecule as a single particle with reduced mass $\mu$, the Schrödinger equation for such a quantum mechanical system (a simple harmonic oscillator) can be written as,

$$\frac{d^2\psi}{dq^2} + \frac{8\pi^2\mu}{h^2}\left(E - \frac{1}{2}Kq^2\right)\psi = 0$$

---------------------------------------------(2.1.9)

The Eq. 2.1.9 can be solved for the energy eigenvalues or frequencies of the system, which are given by $E_n = h\nu\left(n + \dfrac{1}{2}\right)$ and $\nu = \dfrac{1}{2\pi}\sqrt{K/\mu}$. Here $n$ is called the vibrational quantum number. For a quantum mechanical simple harmonic oscillator, the allowed transitions are $\Delta n = \pm 1$, whereas for anharmonic systems $\Delta n = \pm 2, \pm 3, \ldots$ can be observed which are commonly known as overtones.[67]

### 2.2.1 Fourier Transform Raman Scattering

Fourier transform Raman scattering is regular Raman scattering with a different mechanism for collecting and analyzing the scattered light. Fourier transform Raman methods utilize Michelson interferometers and Fourier transform processors. Once the scattered light from the target is collected and the elastically scattered light is removed, it goes through a Michelson interferometer. In Michelson interferometer, the incident laser beam is split into two beams where one incident on a fixed mirror and the other on a varying mirror. The reflected beam from these two mirrors, combine at the beam-splitter producing a sinusoidal interference pattern. Then, this signal is detected through an infrared detector and Fourier transformed to get a Raman spectrum.[72]

### 2.2.2 Confocal Raman Micro-spectroscopy

Confocal Raman micro-spectroscopy allows one to obtain inference-free Raman spectra in two or three dimensional images of small specimens which are embedded in a strong scattering media. The confocal microscopy configuration consists of two pinhole apertures to remove the residual scattered light generated from out of focus points in the sample. When this is utilized in conjunction with a spectrometer to analyze the Raman scattered signal, the improvement in spatial resolution and depth discrimination can be achieved. These are the major advantages for using this technique.[72]

### 2.2.3 Infrared Absorption Spectroscopy

Infrared (IR) absorption spectroscopy is another valuable vibrational spectroscopic method which is used to identify molecular structures.[68] The main drawback of using IR spectroscopy in biological tissues is the high absorption of water and $CO_2$ in the fundamental frequency range. The water absorption bands lie in the 400-1800 $cm^{-1}$ wavenumber region, and thus interfere with bands due to biomolecules of interest. In fact, IR spectra need to be corrected for absorption due to water molecules and $CO_2$ in air. This makes it difficult to use IR spectroscopy in *ex vivo* or *in vivo* fresh biological tissue characterization for cancer diagnostic purposes. Further, as the Fourier transform IR spectrometers are based on the use of Michelson interferometer, it is difficult miniaturize them to make them hand-held and portable. Nevertheless, both Raman and IR spectroscopy are needed to determine the complete set of vibrational modes of molecules with center of symmetry, as the Raman and IR activity of modes in such molecules is mutually exclusive.[70]

### 2.3 Diffuse Reflectance Spectroscopy

When a laser beam hits a sample, reflection, scattering, absorption, and transmission through the sample can occur. The reflection of light obeys the law of reflection, i.e. the angle of incidence is same as the angle of reflection. The transmission of light obeys Snell's law, $n_1 \sin \theta_1 = n_2 \sin \theta_2$ where $\theta_1$ and $\theta_2$, are angles of incidence and refraction respectively, and $n_1$ and $n_2$ are refractive indices of the two media. When the light travels in multiple media having different indices, its reflection and transmission behavior is described by Fresnel equations.[73] If the light is incident close to the normal incidence, then the reflection ($R$) and transmission ($T$) coefficients can be described by the following equations;

$$R = \left( \frac{n_1 - n_2}{n_1 + n_2} \right)^2 \text{----------------------------------------------------------------(2.3.1)}$$

$$T = 1 - R = \frac{4n_1 n_2}{\left( n_1 + n_2 \right)^2} \text{-----------------------------------------------------------(2.3.2)}$$

Reflection can be either specular reflection or diffuse reflection. Specular reflection is defined as reflection off of smooth surfaces such as a glossy surface in which an incoming ray is reflected into a single outgoing direction obeying the law of reflection as shown in Fig. 2.4.



Specular Reflection
(Smooth surfaces)

Diffuse Reflection
(Rough surfaces)
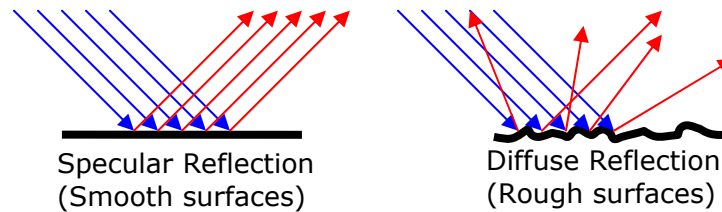
Figure 2.4 Specular reflection and diffuse reflection

When parallel light hits on a rough surface, the direction of the reflected light may be different although each ray follows the law of reflection. This is due to different orientations of the surface, and thus, the normals for the incident ray are different. This reflection is called diffuse reflection. The diffuse reflectance is defined as the probability of photon reemission per unit surface area from a scattering medium.[74] This is also known as the relative diffuse reflectance and we will use these measurements to find the optical properties of biological tissues.

The absorbers or the chromophores in the targeted tissue capture energy of certain wavelengths ($\lambda$). The main chromophores of skin are melanin, DNA, amino acids, hemoglobin, etc.[75] In a tissue, the absorption at certain wavelengths will be higher than

the other wavelengths depending on the chromophores present.  To arrive at the absorption spectrum, we first consider the analytical form of a diffuse reflectance spectrum.

The diffuse reflectance spectrum of skin at a given site, $R(\lambda)$, can be calculated from the equation below,[76]

$$R(\lambda) = \frac{S(\lambda) - S_{dark}(\lambda)}{S_{ref}(\lambda) - S_{dark}(\lambda)} \text{-----------------------------------------------------(2.3.3)}$$

where, $S(\lambda)$ is the measured spectral signal of the reflected light at each wavelength, $S_{dark}(\lambda)$ is the dark correction obtained from a spectrum of a dark (0% reflectance) standard, and $S_{ref}(\lambda)$ is the reference reflectance standard obtained from a white (100% reflectance) standard.[76]

The absorption spectrum $A(\lambda)$ of the tissue can then be calculated as the negative of the logarithm base 10 of $R(\lambda)$.[76]

$$A(\lambda) = -\log_{10}(R(\lambda)) \text{----------------------------------------------------------------(2.3.4)}$$

Melanin concentration is calculated from the slope of this absorption spectrum in the 620-720 nm range.[77]  The absorption spectrum in this spectral range includes both melanin and other scattering components.

Scattering is a change in the direction of incident light caused by variations in the refractive index between layers, cellular components, and the extracellular matrix such as collagen, elastin fibers, erythrocytes, and melanosomes.  Scattering component can be explained using the scattering coefficient $\mu_s$ (cm$^{-1}$), defined as the probability of photon scattering in a medium per unit path length.[74]  The inverse of the scattering coefficient is known as the mean scattering length.

$$\mu_s = a \text{ x } \lambda^{-b} \text{ ------------------------------------------------------------------------(2.3.5)}$$

Here, $a$ and $b$ represent constants which are characteristics of the tissue.  The value of $a$ can be on the order of $10^4$–$10^5$ and whereas $b$ may vary between 1.2 and 1.5.[78]  This

scattering coefficient also depends on the scattering cross section and the scatterer number density.[3]

Hemoglobin can exist as oxyhemoglobin or deoxyhemoglobin depending on whether it has a bound oxygen. The apparent concentrations of oxyhemoglobin and deoxyhemoglobin can be calculated from the melanin-corrected absorption curve using the known extinction coefficients for these molecules and the intensity of the absorption curve at the maxima of absorption for these two molecules: 560 and 580 nm.[77] The hemoglobin concentrations can be estimated using absorption coefficients of biological tissues at different wavelengths as follows.[3]

$$\mu_a(\lambda_1) = \ln(10)\varepsilon_{ox}(\lambda_1)C_{ox} + \ln(10)\varepsilon_{de}(\lambda_1)C_{de} \text{-----------------------------------------}(2.3.6)$$

$$\mu_a(\lambda_2) = \ln(10)\varepsilon_{ox}(\lambda_2)C_{ox} + \ln(10)\varepsilon_{de}(\lambda_2)C_{de} \text{-----------------------------------------}(2.3.7)$$

Here $\varepsilon_{ox}$ (cm$^{-1}$M$^{-1}$) and $\varepsilon_{de}$ are molar extinction coefficients (probability of photon interaction with a medium per unit path length) which can be obtained from experimental curve of molar extinction coefficient vs. wavelength for each type of oxy and deoxyhemoglobin, and $C_{ox}$ and $C_{de}$ are their molar concentrations. Here, $\mu_a$ (cm$^{-1}$), absorption coefficient is defined as the probability of photon absorption per unit path length in a medium. The inverse of the absorption coefficient is commonly known as the mean absorption length. This coefficient depends on the absorption cross section and the number density of the absorbers. Note that extinction coefficients include both absorption and scattering. Once the concentration of oxy and deoxyhemoglobin are determined, the oxygen saturation ($SO_2$) and the total concentration ($C_{Hb}$) can be calculated as follows.[3]

$$SO_2 = \frac{C_{OX}}{C_{OX} + C_{de}} \text{-----------------------------------------------------------------}(2.3.8)$$

$$C_{Hb} = C_{OX} + C_{de} \text{--------------------------------------------------------------------------}(2.3.9)$$

# CHAPTER 3

## INSTRUMENTATION AND DATA COLLECTION

In this chapter, the Raman spectroscopy instrumentation and the diffuse reflectance spectroscopy/colorimetry apparatus used in this work as well as data collection and analyses are described. In addition, the tissue preparation for the Raman spectroscopic measurements and collection of data from acanthosis nigricans patients are also described.

### 3.1 Instrumentation

### 3.1.1 Raman Spectroscopy

The basic elements a of a Raman spectrometer include a source of excitation wavelength (a laser source), a band-pass or a line filter to filter out the unwanted laser lines, a grating to disperse the collected scattered light, a notch filter to eliminate the intense elastically scattered (Raleigh) light, and a detector to record the spectrum. Often a position sensitive charge coupled detector (CCD) is used to record the entire dispersed light (spectrum) simultaneously. There are other detectors in use, such as a photomultiplier tube, but the recording the spectrum with it, is a time consuming process as the only one point of the spectrum is measured at a given time. The CCD increases the speed of data acquisition. The spectrometer control and data acquisition is accomplished by a computer equipped with appropriate software. A schematic of a Raman spectrometer is shown in Fig. 3.1.
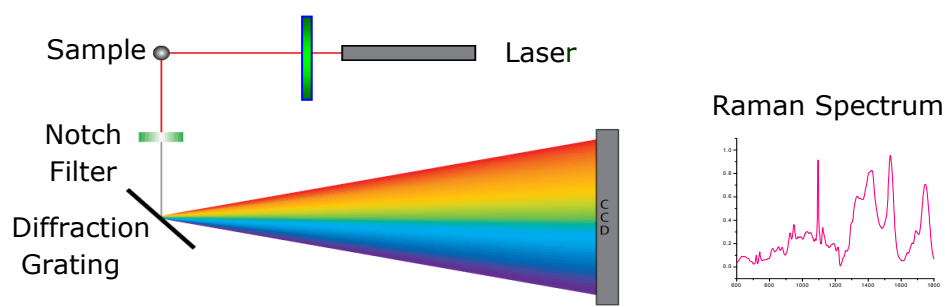


Figure 3.1 Schematic diagram of a Raman instrument[79]

The performance of a CCD detector depends on quantum efficiency (QE) and its operating temperature.  The dark current (noise) of a detector depends on its operating temperature, and thus, liquid nitrogen cooled detectors have a superior performance characteristics compared to their air cooled counterparts.  The dark current originates from thermal agitation.[80]  Hence, in liquid $N_2$ cooled CCD detectors, the dark current is very low, less than 1 electron/pixel/hour.  There is a type of noise associated with dark current, which is called dark noise.  This dark noise occurs due to statistical variations in dark current.[80] Thus, dark noise can be minimized by cooling.  The QE is ratio of number of photoelectrons generated to the number of photons strikes on a photoactive surface of the detector.  As CCD chip is made of silicon, the QE is very high, about 45-50% at 750 nm and depends on several factors such as energy band gap of the semiconductor sensing light.

For the work reported in this thesis, we have used two different Raman spectrometers to record the spectra from frozen and archived human tissues.  One is a Renishaw RM1000 Raman microscope-spectrometer equipped with a 785 nm near-infrared laser excitation source and this instrument is housed in the Children's Hospital of Michigan, Detroit, MI (see Fig. 3.2).  The other is a Jobin-Yvon Horiba Triax550 Raman spectrometer system equipped with a 514.5 nm argon-ion laser excitation source, and it is housed in the Department of Physics and Astronomy, Wayne State University (see Fig. 3.3).  Both these instruments are capable of working with 514.5 nm and 785 nm excitations laser wavelengths.

### *Renishaw RM1000 InVia Raman Microscope*

The Renishaw Raman system is a commercial system and consists of a Leica microscope, air cooled 578x385 pixel Charge-Coupled Device (CCD), ProScan II motorized xy stage and a computer equipped with Wire2 software to control the spectrometer and

acquire the Raman spectra. The microscope is furnished with three objectives 5x, 20x, and 50x.
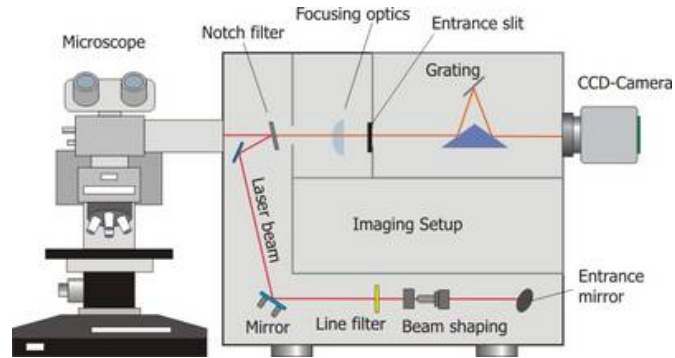


Figure 3.2 A schematic of the Renishaw RM1000 Raman spectrometer[81]

We used the 50x objective with a 0.5 numerical aperture (NA) for all the Raman measurements taken with this instrument. NA is defined as the product of the refractive index, $n$ ($n$ of the medium between the lens and the specimen) and sin ($a$), where $a$ is half of the cone of the collected light.[82] It is an important parameter in deciding the type of objective one wants to use for their Raman measurements. A high NA objective collects much more scattered light compared to a low NA objective,[83] thus, reducing the sample exposure time.

A schematic of the Renishaw Raman spectrometer is shown in Fig. 3.1. We used the 1200 lines/mm grating to disperse the scattered light. The maximum power of the GaAs laser source is about 500 mW with stability less than 1 $cm^{-1}$. The power at the sample is about 30 mW and the resolution of the Raman signal is about 3.5 $cm^{-1}$. With a 50x objective the instrument focuses the laser onto an area of 2 x 20 $\mu m^2$ at the sample. As the focal length of the 50x objective is 1 mm, it was chosen to avoid accidental contacts with the samples.

### *Jobin-Yvon Horiba Triax550 Raman Spectrometer System*

The Jobin-Yvon Horiba Triax550 Raman spectrometer is a triple grating spectrometer equipped with 2400, 1200, and 150 grooves/mm gratings which can be selected by the software. The spectrometer is coupled to an Olympus BX41 optical microscope and the laser source with optical fibers. The CCD detector (2000x800 pixels) is cooled with liquid nitrogen and the detector operating temperature is ~136 K. The spectrometer is controlled with a computer equipped with Labspec software for control and acquisition of Raman spectra. We used a 100x long working distance (3.4 mm) objective with 0.8 numerical aperture for recording the Raman spectra. Again, the choice of long working distance objective was to avoid any contact between the objective lens and the tissue specimen.
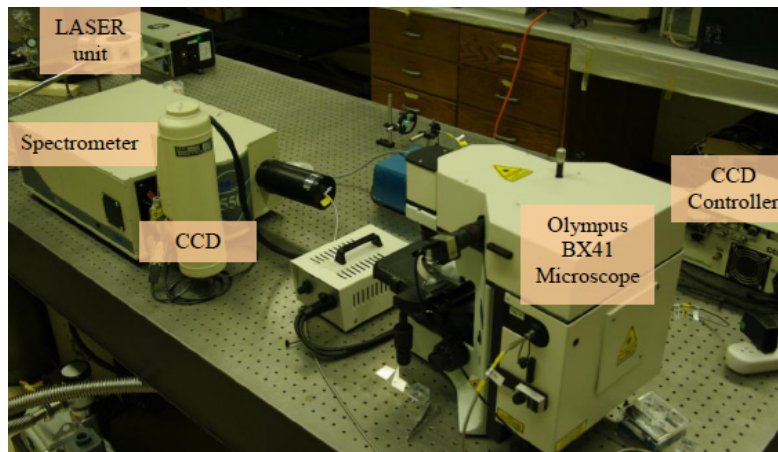


Figure 3.3 A physical set-up of the Triax550 Raman spectrometer

Triax550 can disperse the signal 1.55 nm/mm at 514.5 nm with a 1200 g/mm. The resolution is of 0.03 nm.[84] The focal length of this spectrometer is 550 mm and the input aperture of the monochromator (F number or the focal length divided by the effective aperture diameter) is F/6.4. If the focal length is longer, then the resolution becomes better. It is note that the light collecting power increases with the inverse square of the F number.

The spectrometer can record a spectral range of ~1400 cm$^{-1}$ without moving the grating (static scan). The microscope focuses the laser beam to a circular spot of ~1 μm in diameter on the sample. The Raman intensity is directly proportional to the forth power of the scattered wavelength.[67] Thus, the Raman intensity increases as the wavelength decreases. This is one of the advantages of using a shorter wavelength excitation source. However, the fluorescence background effect is larger for 514.5 nm compared to 785 nm excitation wavelength. This can be minimized by acquiring data from a smaller sample volume or by quenching, i.e. delivering laser light onto the specimen for a certain time period before collecting data[85] (Note that quenching can be done only for the samples which are not prone to damage due to long laser exposure time).

Both these Raman instruments are calibrated using the known silicon peak (520.5 cm$^{-1}$) prior to each measurement. This Si calibration procedure is described in more detail in section 3.3.1.

### 3.1.2 Diffuse Reflectance Spectroscopy and Colorimetry

The diffuse reflectance spectroscopy apparatus consists of a HL-2000 Ocean Optics deuterium tungsten halogen lamp for skin illumination, a broadband spectrophotometer (USB 2000 light detector; BWTEK, Inc., USA) capable of detecting absorbance in the wavelength range 350-850 nm (visible to near infrared region), a bifurcated fiber optic probe for light delivery and scattered light collection, and a computer. The output end of the bifurcated fiber bundle which is placed in contact with the skin was 2.5 mm in diameter. The absorbance spectra are calculated from the measured scattered light spectrum by a custom Labview 8.0 (Labview Inc.) program (Johnson and Johnson, NJ, USA). A schematic of the USB 2000 spectrometer is shown in Fig. 3.4.
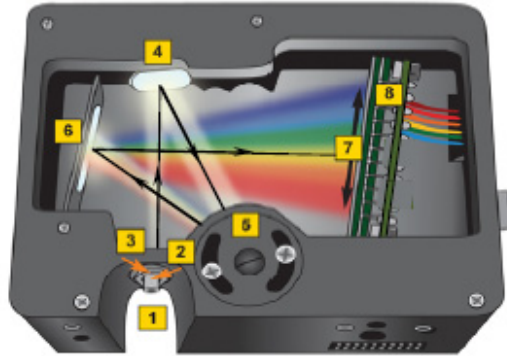
Figure 3.4 A schematic of the interior of the USB 2000 spectrometer. The key optical components are representing through 1-8: the entrance fiber, the entrance slit, an optical filter, the collimating mirror, the grating, the camera mirror, and the detector array[86]

The USB2000 spectrometer consists of a holographic 1200 diffraction grating and a one dimensional CCD detector array with 2048 pixels. The spectral resolution is about 0.5 nm. The F number of the optical fiber is F/2.3 and matches with F number of the collimating mirror.[86] Prior to data acquisition, the spectrophotometer is also calibrated using pure black and pure white tile standards to obtain a dark spectrum and a high-reflectance spectrum, respectively.

### *Colorimetry*

The colorimeter apparatus consists of a CM-2600d spectrophotometer (Konica Minolta CM-2600d, Osaka, Japan) which utilizes a Xenon arc lamp as a source of light for illuminating the skin and a computer. Scattered light is collected by the CM-2600d through an integrating sphere whose internal surface is coated with a barium sulfate coating to make the light diffuse uniformly.[87]
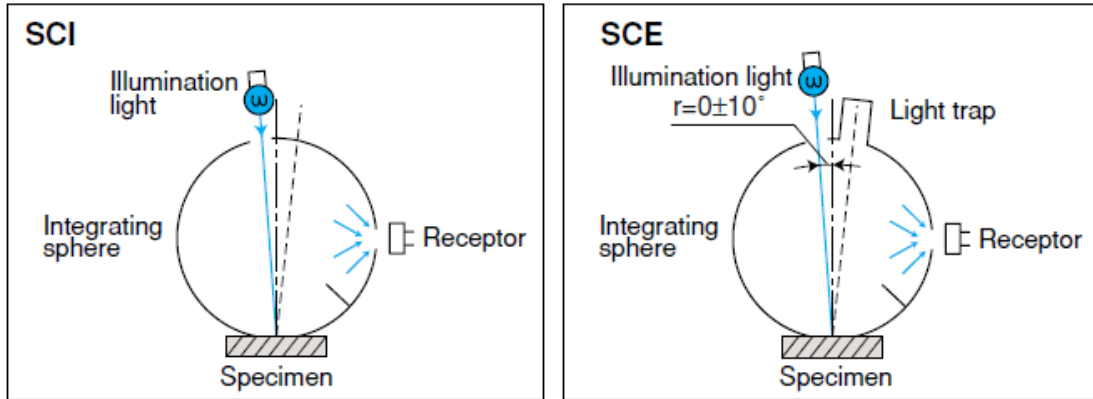
Figure 3.5 A schematic of the integrating sphere system with SCI and SCE options[87]

Figure 3.5 shows the optical geometries, where the reflected light (in all directions) is collected (SCI-specular component included) or excluding the specular reflection (SCE-specular component excluded).  This instrument uses the standard tristimulus color analysis method utilizing the L*a*b* color system.  The primary colors are blue (450 nm), green (525 nm), and red (610 nm).  The numerical parameters L*, a*, and b* represent, respectively, a color's darkness to lightness, its green to red color component, and its blue to yellow color component.[87]  A circular patch of skin 8 mm in diameter is illuminated by the CM-2600d instrument.  This instrument can record both SCI and SCE measurements simultaneously (see Fig. 3.5).  Similar to diffuse reflectance spectroscopy, the colorimeter is also corrected for detector dark current and calibrated with a standard white target prior to measurements on every patient.

The apparatus of both these instruments are shown in Fig. 3.6 together with the size of the each probe.
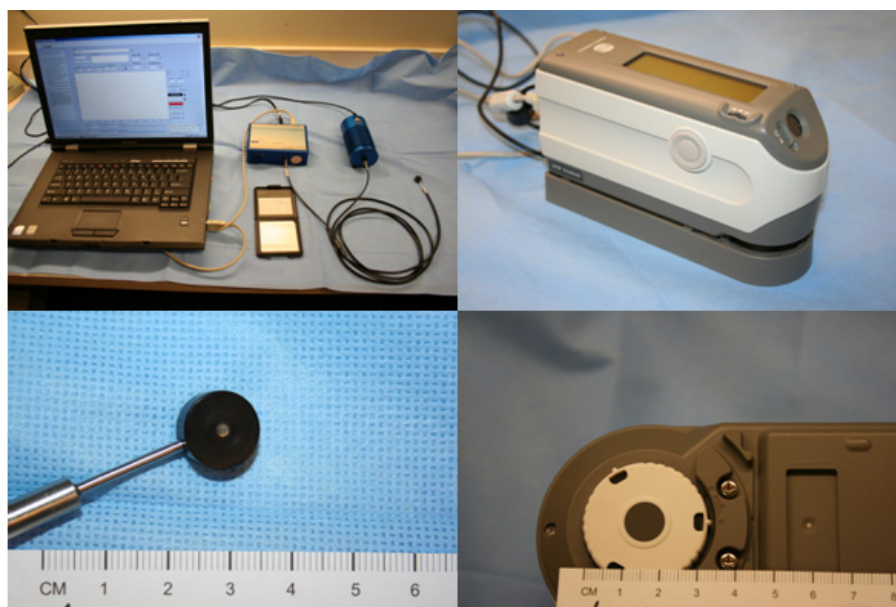
Figure 3.6 Left: DRS setup with spectrometer, light source, calibration plate, bifurcated fiber optic cable, and laptop; a close-up of DRS probe showing 2.5 mm fiber core. Right: Colorimeter and a close-up of colorimeter 8 mm aperture[88]

## 3.2 Tissue Processing

Raman investigations on head and neck squamous cell carcinoma (HNSCC), prostate cancer, and pediatric tumors were carried out using human tissue specimens. All these human tissue investigations are conducted under the established rules of Wayne State University Human Investigation committee (HIC), Institutional Review Board (IRB), and HIPAA (Health Insurance Portability and Accountability Act) regulations (IRB protocol numbers are 0705004871 and 0810006496). In pediatric tumor investigation, we have studied frozen and formalin fixed paraffin processed (FFPP or deparaffinized) tissues from the same patients. For the Raman spectroscopic evaluations of HNSCC and prostate cancer specimens, only the FFPP tissues were examined as there were no frozen tissues available. As acanthosis nigricans research was done *in vivo*, tissue processing was not necessary.

For the investigation of HNSCC and prostate cancer, the samples were obtained from Karmanos Cancer Center and Harper University Hospital in Detroit, MI and for pediatric tumors, the tissue specimens were obtained from the Children's Hospital of Michigan.

### 3.2.1. Pediatric Tumors

The goal of this work was to explore the feasibility of applying the Raman spectroscopy for diagnosis using the deparaffinized tissues. It is known that archiving the tissues using formalin fixing and paraffin processing leads to alteration of the biochemical content of the tissues. Thus, a comparison of the Raman spectra of FFPP or deparaffinized archived tissues with the corresponding fresh tissues would give information about the nature of the alterations. We have studied the frozen tissues of normal adrenal gland, neuroblastoma, and ganglioneuroma from three different patients, and subsequently preserved the same tissues using formalin fixing and paraffin embedding, and again investigated them after deparaffinizing/FFPP. The details of the procedure used are explained below.

*Frozen Tissue Specimens*

The specimens were thawed for about 20 minutes and then the desired frozen sections were cut with a microtome using optimal cutting temperature (OCT) solution. In each tissue specimen from a given patient, a 5 µm thick section was cut and stained with Haematoxylin and Eosin (H&E) for pathological examination and additionally three adjacent 10 µm thick parallel sections were cut for Raman measurements. Then, the remaining tissue specimen, keeping the same orientation, was put in formalin for about 24 hours and subsequently preserved in paraffin wax at Karmanos Cancer Center, Detroit, MI.

*Deparaffinized/FFPP Tissues*

Once the tissues are embedded in paraffin wax these tissue blocks were processed again to remove paraffin wax from the tissue using several ethanol methanol baths. It is important to remove all the wax from the specimens because paraffin gives a strong Raman bands and any residual paraffin left in the specimens interferes with the Raman bands of

the specimens' leading to misinterpretation.  Similar to frozen tissue sectioning, four parallel tissue layers, one for H&E staining and the rest for the Raman spectroscopic investigation were cut from each specimen.  It is important to note that these four tissue sections are adjacent to the four sections that were cut in the frozen state prior to preserving the tissue in paraffin.

### 3.2.2 Head and Neck Squamous Cell Carcinoma and Prostate Cancer

For head and neck squamous cell carcinoma and prostate cancer investigation, only formalin-fixed paraffin-embedded tissue blocks were used. The archived specimens embedded in paraffin were processed before recording their Raman spectra.  Similar to the frozen vs deparaffinized study, for each specimen two adjacent parallel tissue sections were cut and deparaffinized, where one of the sections was stained using H&E for histological examination and the other unstained section was put on a stainless steel substrate for Raman spectroscopic studies.  We used stainless steel substrates because it gives less background signal and no characteristic peaks of its own compared to glass, quartz, or sapphire.



**Unstained section**          **H&E stained section**
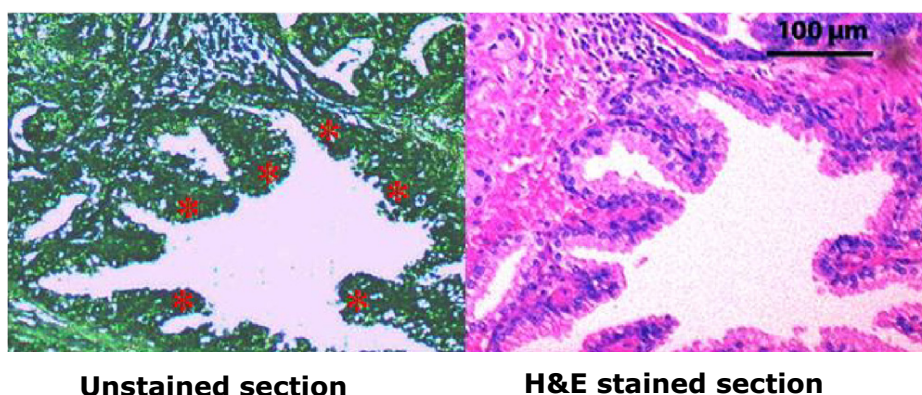
Figure 3.7 Pathological images of H&E stained tissue section and unstained tissue section[89]

When collecting data, pathologically identified regions of the H&E stained tissue were used as a guide to identify the corresponding regions in the unstained tissue section as shown in Fig 3.7.  For the pediatric tumor project Drs. R. Rabah (University of Michigan-Ann Arbor) and J. M. Poulik (Children's Hospital of Michigan) made the ink markings identifying appropriate pathology regions on the H&E slides.  Dr. C-H. Chang (Children's Hospital of Michigan) also served as an external reviewer of the pathology diagnosis.  For the prostate cancer project, Drs. W. Sakr and S. Sethi (Karmanos Cancer Center) marked the desired regions and Dr. D. Shi (Karmanos Cancer Center) also reviewed the H&E slides.

Staining is an important and necessary process in histopathology examination. Hematoxylin and Eosin are commonly used dyes for tissue staining.  Haematoxylin gives a blue color to the nucleus whereas eosin dye makes the rest of the content (cytoplasm, collagen, etc.) pinkish in color, thus, providing contrasts to identify the morphology and diagnose a state of the disease.  Pathologists rely on being able to see the morphology for disease diagnosis.  However, stained tissues produce strong fluorescence from eosin dye and not suitable for Raman spectroscopic studies.[90]  Often, the fluorescence is so strong that the Raman signal cannot be discerned or the detector gets saturated even for brief exposure.  Thus, we only make Raman measurements on unstained frozen or deparaffinized tissues.

### 3.2.3 Acanthosis Nigricans

Spectroscopic investigation of acanthosis nigricans skin disease is done in collaboration with Department of Dermatology, Henry Ford Hospital, Detroit, MI under the leadership of Drs. I. Hamzavi and H. Lim.  In this research work, we test patients' skin sites *in vivo* using a diffuse reflectance spectrometer and a colorimeter.  Consequently, tissue processing is not required as in the Raman investigations.  However, it is a necessary requirement that all the patients be in an insulin resistant condition, meaning elevated

fasting insulin levels with normal fasting glucose levels.  The patients with type 1 diabetes, women who were lactating, pregnant, or planning to become pregnant, and those who were hypersensitive to oral metformin medication were excluded from this study.

In this AN study, eight female patients, ages 12 years and older with AN as diagnosed by a dermatologist were enrolled.  During the course of this research, all patients were consulted by an endocrinologist for treatment of their hyperinsulinemia with diet control, weight reduction and/or oral metformin.  Patients' consents were obtained through an informed consent form and all study-related procedures were explained before enrolling them in the study.

### 3.3 Data Collection and Processing

In this section, the collection, processing and analyses of data using statistical analysis are explained. The methods used in both Raman investigations and the diffuse reflectance studies are described.

### 3.3.1 Raman Data Collection

The Raman spectrometers were calibrated prior to each recording session, using well known silicon peak at 520.5 $cm^{-1}$.  For this purpose a small piece of the silicon wafer was used, and any correction needed was applied using the spectrometer control software.  The Renishaw Raman system has an automated "health check" routine that performs an calibration and alignment check and makes necessary adjustment/alignment so that silicon peak appears at the expected position with optimum intensity.

When targeting a particular region of the tissue, all three objectives of the microscope (5x, 20x, and 50x) are used.  First the desired region of the unstained tissue section is located using the H&E slide with 5x objective.  After final focusing and adjustments, a Raman spectrum is taken using the 50x objective.  In order to achieve a

better signal to noise ratio, three accumulations of 20 s each averaged to obtain a spectrum. The range Raman shift range of 400-1800 $cm^{-1}$ is usually considered as the "biological window" because it contains most of the useful fundamental Raman active modes of biological molecules. However, we collected the data in the range of 500-1900 $cm^{-1}$ for fluorescence background subtraction and for the statistical analysis. While only the 600-1800 $cm^{-1}$ region was used for spectral identification, the extra coverage was to minimize artifacts at the boundaries of the spectral region which can arise from the computerized background subtraction of the spectra.

For the Raman spectral data obtained from the Jobin-Yvon Horiba Triax550 Raman spectrometer system, a 40 second integration time was used and spectra were averaged over five accumulations which gave the optimal signal to noise ratio of a Raman spectrum. Care was taken to ensure that each spectrum was collected from a new location on the tissue. The spectra were recorded in the 428–1812 $cm^{-1}$ region that involved no motion of the grating (static scan), thus, avoiding any spurious Raman shifts because of backlash errors associated with grating motions. The spectrometer does not have an automated "health check" routine and thus the alignment and calibration using silicon were done manually.

### 3.3.2 Raman Data Processing

The raw Raman spectra (shown in Fig. 3.8 (a)) often contain a large fluorescence background and occasionally some unexpected peaks (noise) may appear due to stray light or cosmic. Therefore, before analyzing the data, each spectrum is examined manually and the unusable spectra are removed from the data set. This is followed by the removal of obvious noise peaks and the background fluorescence of the selected Raman spectra.

De-noising and background subtraction is done using a MatLab algorithm[91] which removes fluorescence on a spectrum-by-spectrum basis. This procedure utilizes the

following approach. First, the noise of the raw spectrum is removed and smoothened using a medium filter followed by a wavelet function. Then, the background fluorescence is removed using two steps of the adaptive minmax method.
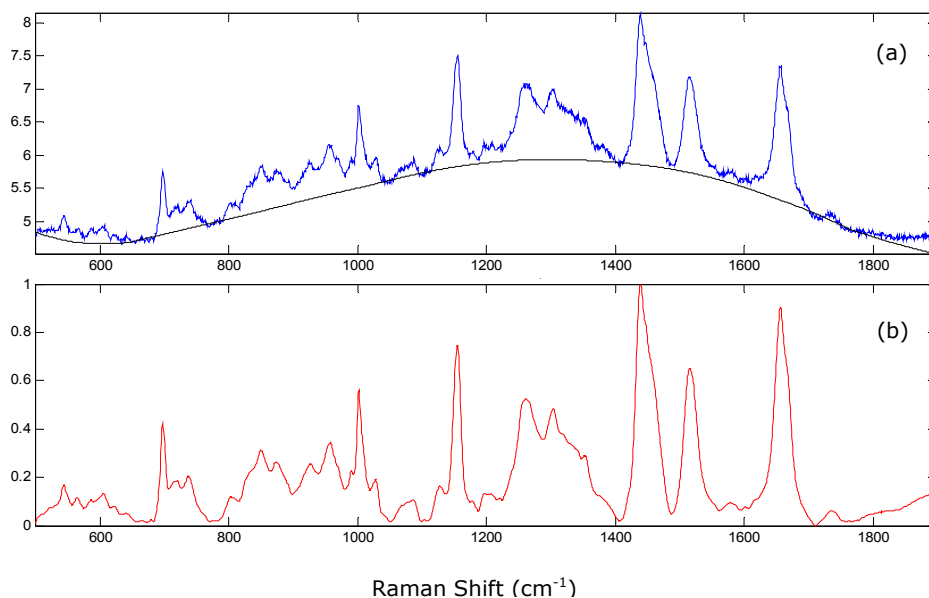


Figure 3.8 (a) Raw spectrum, where black line shows the adaptive minmax fit (b) De-noised, fluorescence subtracted and normalized spectrum

The first step is to perform constrained and unconstrained polynomial fits using two different orders. The different orders are determined using F/S ratio, the ratio of maximum fluorescence to maximum subtracted Raman signal selecting the minimum values of Raman intensities from the original spectrum for the polynomial fit. In the second step, this minmax algorithm will take the maximum values of the fits and remove the background from the desired spectrum.[91]

Once the raw spectrum is de-noised and background subtracted, it is normalized by dividing Raman intensities by the highest Raman peak in the spectrum. Thus, the final Raman intensities vary from 0 to 1 as shown in Fig. 3.8 (b). The artifacts of background subtraction below 600 cm$^{-1}$ and above 1800 cm$^{-1}$ can be clearly seen in Fig. 3.8 (b). These regions are excluded from the analysis as mentioned earlier.

### 3.3.3 Diffuse Reflectance Data Collection

Diffuse reflectance and colorimetry measurements were obtained from four areas on the individual patient; two considered to be healthy tissue and used as a control and two areas of an acanthosis nigricans lesion on the neck. The two control sites were the inner forearm approximately 10 cm from the wrist and the flat part of the upper shoulder several inches from the neck on the clavicle. Spectra from the acanthosis nigricans lesion were collected from the median and lateral areas of neck.[88] DRS and colorimetry data were collected from eight patients with acanthosis nigricans over an eight month period. Three colorimetry and ten DRS absorbance spectra measurements were taken from each site. The higher number of DRS measurements was necessary due to the smaller area of skin sampled by the smaller DRS probe area.



Figure 3.9 Left: DRS probe applied to patient's posterior neck; Right: Colorimeter measuring patient's posterior neck[88]

Since the probing pressure of the human skin can change the diffuse reflectance spectra,[92] care was taken to insure uniform pressure on the skin over the course of the study. Figure 3.9 shows how the posterior neck measurements are obtained using the DRS and the colorimeter. All measurements were taken at the same room temperature and humidity. For documentation and comparison, photographs of the lesion area were also taken during every visit using a digital camera with a cross-polarized filter. A cross-polarizer was used to remove the specular reflection of the skin surface.

While the colorimetry signal is analyzed purely in terms of the intensity of the various color components: L*, a*, and b*, the DRS data analysis software models the physics of the light-skin interaction to calculate biochemical concentrations from the reflected spectrum. Analysis of the DRS data provides a quantification for the contributing skin chromophores, including melanin, oxyhemoglobin, and deoxyhemoglobin. Nevertheless, one can use the intensities of the raw spectral data from DRS ranging from 350-850 mm to characterize the skin condition, and acanthosis nigricans. However, the raw spectra of the colorimeter were not obtainable.

### 3.3.4. Diffuse Reflectance Data Processing

To lessen the scatter of the measurements between-patients due to the inherent differences in skin coloring and composition of the patients, both colorimetry and DRS data were normalized for every patient prior to analysis. This was done by first calculating the average of the patient's forearm control measurements. The scatter of the forearm control measurements about this average allowed us to characterize the anticipated scatter of the lesion measurements which was expected to be even greater due to the heterogeneity of the lesion tissue. All subsequent measurements were then divided by that patient's average forearm control to insure that data obtained from the neck control or neck lesion tissue were really differences from that patient's normal skin coloring or composition.[88]

**CHAPTER 4**

**STATISTICAL ANALYSIS/CHEMOMETRIC METHODS**

The Raman spectra consist of multivariate data. The intensities are function of Raman shifts (variables) for a given molecule or a type of tissue. If a large number of spectra are collected on a few different unknown samples (such as normal, pre-cursor lesions, and cancerous tissues) one would like to know how similar to one another these spectra are, and thus, be able to group them for identifying the samples. Of course, the identification of a sample based on its spectrum is possible provided some "standard" data exists for the samples being analyzed. Thus, when many observed variables (Raman intensities) seem to describe a particular behavior of an entity, interpreting data using mean, standard deviation, standard error, etc. is not sufficient and also cannot determine the variables responsible for that particular behavior of the entity. New method needs to be developed to find the independent variables which account for most of the variance in the data and determine the behavior of an entity. Principal Component Analysis (PCA) and Discriminant Function Analysis (DFA), chemometric methods provide means to achieve this purpose and often used in reduction/grouping of Raman spectroscopic data. In this chapter we briefly describe the principles of these methods which are used in analyzing the data presented in the thesis.

**4.1. Principal Component Analysis**

PCA is one of the multivariate methods of analysis and it allows the number of variables in a multivariate data set to be reduced, while retaining most of the variance present in the observed variables. According to this method, the original data set is transformed into a new data set[93] with a reduced number of orthogonal variables which carry maximum variance of the original data set. These new variables are called "principal components" (PC scores or eigenvectors). The PCs could be then used as predictor or

criterion variables in subsequent analyses such as DFA. PCA is useful when there is large data on a large number of variables and it is anticipated that some of the variables are correlated with one another, possibly because the measurements are on the same entity (such as a given type of tissue). When the data is plotted against the principal components, one observes trends in the plot which are basically the independent groups in the data. In case of biological tissues, the trends in the PCs plots represent pathological conditions of a diseased tissue where each trend is determined from a unique combination of the original predictors of the trend or the pathological state. One can obtain these reduced variables in five steps as follows. For simplicity, let us assume that there are three variables x, y, and z in our data set.

1. **Adjusting the data**: Each variable or dimension is subtracted from its mean, i.e., all the x values are subtracted by the mean of the $x$ ($\bar{x}$). Thus, the mean of the data set is zero. $d_i = x_i - \bar{x}$, where $i$ varies from 1 to $n$. Here $n$ the total number of data cases (assume $n=6$) and similarly we do the same for y and z variables. The mean-subtracted data which we call "AdjustData" matrix look like;

$$AdjustData = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ x_3 - \bar{x} & y_3 - \bar{y} & z_3 - \bar{z} \\ x_4 - \bar{x} & y_4 - \bar{y} & z_4 - \bar{z} \\ x_5 - \bar{x} & y_5 - \bar{y} & z_5 - \bar{z} \\ x_6 - \bar{x} & y_6 - \bar{y} & z_6 - \bar{z} \end{pmatrix}$$

2. **Calculating the covariance matrix**: For three variables or dimensions, the covariance matrix is 3x3.

$$C = \begin{pmatrix} \text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\ \text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z) \end{pmatrix},$$ where covariance elements are defined as

$$\text{cov}(a,b) = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{(n-1)} \text{ for } n\text{=6.}[93]$$

3. **Calculating the eigenvectors and the eigenvalues of the covariance matrix**: By construction, eigenvectors are unit orthogonal vectors which provide information about the trends in the data, and thus characterize the data set. For three variables/dimensions data, there will be three eigenvalues $\lambda_i$ and the corresponding three eigenvectors which are determined from the characteristic equation, $|C - \lambda I| = 0$ where $I$ is the identity matrix. The eigenvectors (*EVs*) can be calculated by solving the matrix, $CEV = \lambda EV$ with the condition $\sum_i EV_i^2 = 1$. These *EVs* are unit vectors.[93]

4. **Choosing eigenvectors and forming a feature vector**: By ordering the eigenvectors by the highest to lowest eigenvalue, one can determine the order of significance of these vectors. The eigenvector with the maximum eigenvalue is the first principle component of the data set. Feature vector is a matrix vector with the eigenvectors one needs to keep without losing much information of the data set.

$$FeatureVector = \begin{bmatrix} EV_{11} & EV_{21} & EV_{31} \\ EV_{12} & EV_{22} & EV_{32} \\ EV_{13} & EV_{23} & EV_{33} \end{bmatrix}$$

Suppose, the first two eigenvectors are enough to represent the variance in the data set of three variables/dimensions, then, the feature vector is a 3x2 matrix.

5. **Obtaining new data set**: First, take the transpose of the two vectors, the feature vector and the mean-adjusted data vector. Then, the new data set is the multiplication of these two transposed vectors.

$$FinalData = FeatureVector^T \, AdjustData^T$$

Now, the new data are in terms of the chosen eigenvectors with the maximum information. Note that one can obtain the original data by adding the original mean to the resulted matrix of the multiplicative of the feature vector and the final data.[93]


## 4.2 Discriminant Function Analysis (DFA)

DFA technique classifies the data into independent groups present in the data by minimizing the variations within the groups and maximizing the variations between the groups.[94] The groups in our case are the pathological states of the tissue. In the data analysis of all my research studies, I have used the output of the PCA as the input to DFA. The number of discriminant functions generated, are one less than the desired number of groups in the data set. The output from the DFA can be used to find the independent groups in the data.

DFA uses the reverse process of multivariate analysis of variance (MANOVA).[95] In MANOVA the question asked is whether the groups are significantly different based on a set of linearly combined discriminant variables. If they are, then the set of linearly combined discriminant variables can be used to predict group membership. On the other hand, the DFA is usually concerned with actually putting the data into groups (classification) and testing how well (or how poorly) data are grouped. In DFA, first a set of discriminant functions are calculated and tested for their significance, and then classification is performed.[94] In order to carry out DFA, we assume the data set satisfy certain conditions:

- Total number of cases ($n$) should be greater than the number of variables ($p$)
- Number of groups ($g$) are greater than 2
- A group consists of at least 2 cases ($n_i \geq 2$ in $i^{th}$ group)
- All the discriminating variables are independent and measured at the same interval level

- Each group is normally distributed with similar covariance matrix[94]

### 4.2.1 Calculating Discriminant Functions

The discriminant function is a linear combination of the discriminating variables where the coefficients ($u_i$) are determined by satisfying certain conditions which are explained in section 4.2.1.1.

$$f_{km} = u_0 + u_1 x_{1km} + u_2 x_{2km} + .............. + u_p x_{pkm}$$ -------------------------------------(4.2.1)

Here, $f_{km}$ is the value of the canonical discriminant function for case $m$ in the group $k$. $x_{pkm}$ represents the value of the discriminating variable $x_i$ for case $m$ in the group $k$. When calculating the $u_i$ for the first discriminant functions, the group mean has to be different as possible. The $u_i$ for the second discriminant function are determined by maximizing the group means plus no correlations with the first discriminant function $u_i$s and so forth.[94]

### 4.2.1.1 Scatter of the Data: Within-Group and Between-Groups

Let the total scatter or covariance matrix be T such that $t_{ij}/(n-1)$ represent each element of the T matrix;

$$t_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (x_{ikm} - x_{i..})(x_{jkm} - x_{j..})$$ ----------------------------------------------------(4.2.2)

Here, $x_{i..}$ or $x_{j..}$ is the total mean of the corresponding variable $i$ or $j$. The scatter of the groups can be measured by the matrix W, where $w_{ij}/(n-g)$ represents each element in the matrix W.

$$w_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (x_{ikm} - x_{ik.})(x_{jkm} - x_{jk.})$$ -------------------------------------------------(4.2.3)

Here, $x_{ik.}$ or $x_{jk.}$ is the mean of the variable $i$ or $j$ of the corresponding group $k$. Similarly, the between group scatter is B = T-W, ie. $b_{ij} = t_{ij} - w_{ij}$. These two matrices

capture the variances in the data, W and B.  Thus, by solving the following equations simultaneously by setting $\sum v_i^2 = 1$ one can find the coefficients of the DFs.[94]

$$\sum b_{1i}v_i = \lambda \sum w_{1i}v_i$$
$$\sum b_{2i}v_i = \lambda \sum w_{2i}v_i$$

------------------------------------------------------------------(4.2.4)

$$\sum b_{pi}v_i = \lambda \sum w_{pi}v_i$$

where $\lambda$ is the eigenvalue and $v_i$ are the set of $p$ coefficients.  Thus, $u_i$'s can be find through,

$$u_i = v_i\sqrt{n-g} \quad \text{and} \quad u_0 = -\sum_{i=1}^{p} u_i x_{i..}$$ ------------------------------------------------(4.2.5)

The largest eigenvalue carries the greatest discrimination.  Thus, the DF1 is the function with the largest eigenvalue and DF2 carries the second most variances in the data set, and so forth.


### 4.2.1.2 Unstandardized Coefficients and Standardized Coefficients

These $u_i$ values are unstandardized coefficients as they are calculated using the original data values.  Unstandardized coefficients make the origin of the discriminant function axes coincides with the grand centroid (In grand centroid, the discriminating variables are averaged over all the cases) by incorporating the absolute contribution of a variable in finding DFs.  One can convert these coefficients into standardize coefficients ($c's$) as follows.[94]

$$c_i = u_i\sqrt{\frac{w_{ii}}{n-g}}$$ ----------------------------------------------------------------------(4.2.6)

The standardized coefficients provide the relative importance of a variable. Consequently, we can use these standardized coefficients to find the variables contribution towards the DFs. i.e. which variable contributes to which discriminant function and how large the contribution is, and larger the magnitude greater the contribution.

### 4.2.1.3 Structure Coefficients

Structure coefficients show the contribution of the each variable in each discriminant function.  Total structure coefficients can be calculated by taking the product-moment between a particular variable and the DF.  In fact, one can find which variable contributes the most in each DF.  The structure coefficients vary from -1 to +1.  If the structure coefficient is close to zero, the DF and that particular variable are hardly correlated.  On the other hand, if the absolute value of the coefficient is close to 1, the variable and the DF are highly correlated and carry the same information.[94]

Within a group, if one needs to find the correlation between a DF and a variable, then within-group structure variables can provide that information.  The within-group structure coefficients ($S_{ij}$) can be determined using the following equation.

$$ s_{ij} = \sum_{k=1}^{p} \frac{w_{ik} c_{kj}}{\sqrt{w_{ii} w_{kk}}} \text{------------------------------------------------------------------(4.2.7)} $$

where $C_{kj}$ is the standardized canonical discriminant function coefficient for variable $k$ on function $j$.

### 4.2.2. Test of Significance

In order to find the significance of the discriminant functions, relative percentage and canonical correlation coefficient are taken into consideration.

### 4.2.2.1 Relative Percentage and Canonical Correlation Coefficient

When there are two or more discriminant functions, by comparing relative percentages one can find the power of the DFs.  Relative percentages can be calculated by adding all the eigenvalues and then dividing the results into individual eigenvalue.

The canonical correlation coefficients, $r*$ can be calculated from the eigenvalues as follows.

$$r_i^* = \sqrt{\frac{\lambda_i}{1 + \lambda_i}} \text{ --------------------------------------------------------------------------(4.2.8)}$$

for $i^{th}$ discriminant function. These canonical correlation coefficients ranging between 0 and 1 provide information about the relationship between groups and discriminant functions.[94] A value closer to 1 indicates an existence of strong correlation between the groups and the discriminant function. If the value is lower the correlation is weaker.

### *4.2.2.2 Wilk's Lambda*

Without using the discriminant functions, one can find the ability of the variables which can discriminate the groups through Wilk's lambda, $\Lambda$, which measures the group differences over the discriminating variables.[94]

$$\Lambda = \prod_{i=k+1}^{g-1} \frac{1}{1 + \lambda_i} \text{ -----------------------------------------------------------------------(4.2.9)}$$

Here, $k$ indicates the number of discriminant functions calculated using the equation 4.2.1. If $\Lambda$ equals to one, there are no group differences whereas $\Lambda$ equals to zero indicates a high discrimination between groups. In order to test the significance of Wilk's lambda, chi-square or F distributions are used. The formula for chi-square is;

$$\chi^2 = -\left[ n - \left( \frac{p+g}{2} \right) - 1 \right] \log_e \Lambda_k \text{ ------------------------------------------(4.2.10)}$$

The degrees of freedom for chi-square are $[(p\text{-}k)(g\text{-}k\text{-}1)]$.[94]

### 4.2.3 Classification of the Data

Classification of the original data into various classes can be performed either using discriminating variables or discriminant function scores. Similar to discriminant functions, the classification functions can be calculated for each group as a linear combination of the

variables. These classification functions ($G_k$) are determined to maximize the variance between the groups and minimize the variance within the groups.

$$G_k = d_{k0} + d_{k1}X_1 + d_{k2}X_2 + .............. + d_{kp}X_p \text{ -----------------------------------------(4.2.11)}$$

Here, $X_i$ can be either the discriminating variables or the DF scores. The coefficients, $d$'s can be calculated using the equations below.

$$d_{ki} = (n-g)\sum_{j=1}^{p} a_{ij}X_{jk} \text{ and } d_{k0} = -0.5\sum_{j=1}^{p} d_{kj}X_{jk} \text{ ----------------------------------(4.2.12)}$$

$a_{ij}$ are the elements of $W^{-1}$ matrix. Once the G's are obtained, then the classification of the data can be done using the Mahalanobis distance functions.[94]

### 4.2.3.1 Mahalanobis Distance Functions

This is a method of calculating the distances of each individual case from its group centroid. Then, the data are categorized into the closest group. The formula is as follows.

$$D^2(X|G_k) = (n-g)\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}(DF_i - DF_{ik.})(DF_j - DF_{jk.}) \text{ -------------------------(4.1.13)}$$

$D^2(X|G_k)$ is the squared distance from a specific case $X$ to the centroid of group $k$. $DF_i$ represents the DF scores, but one can also use the original discriminating variables, $X_i$. A particular case is classified into a group with a smallest $D^2$ value. If the $D^2$ value is high, it may fit to a different group with similar profile as itself.[94]

### 4.3 Other Statistical Methods

The group identification can also be achieved through cluster analysis where a collection of statistical methods are used to identifies groups of samples that have similar characteristics. Like discriminant analysis, cluster analysis is also used for data reduction. In cluster analysis there is no prior knowledge about which elements belong to which clusters and the clusters are determined through an analysis of the data. The clustering

algorithms are broadly classified into two categories: hierarchical and non-hierarchical algorithms.

### 4.3.1 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) is a clustering method which utilizes tree diagram or dendrogram which groups different cases as branches and shows the relationship among each data case. Typically, all the data in a cluster assumes to have equal belonging to its cluster. Thus, if a data case has a unique probability to cluster into a certain group then one has to use alternate methods.[96] There are two approaches for HCA. They are agglomerative (non-hierarchical) and divisive (hierarchical) procedures.[97] The agglomerative procedure uses the closest and the far away cases with bottom-up clumping of clusters, meaning, start with *n* clusters and merge them to form the others, whereas divisive approach is a top-down method which forms one cluster and split it to make the other clusters. It is not necessary to have the same scale for all the independent variables. Hence, hierarchical cluster analysis is much flexible compared to K-mean cluster (see section 4.3.2) and initiating the number of clusters is not necessary for this method.

However, the choice of initial cluster centers depends on the case order. Therefore, one has to perform random orders with different permutations to verify the stability of a given solution. Thus, this may not be suitable for a dataset with large number of cases.

### 4.3.2 K-mean Cluster Analysis

K-Mean Cluster Analysis is an non-hierarchical clustering algorithm which produces disjoint clusters. This technique is suitable for data set which is composed of a number of distinct classes and it classifies the data using the distances of data cases and the means of different clusters or k clusters.[98] The number of groups has to be initiated by the user, which in turn is the number of k clusters. The choice of initial cluster centers depends on

the case order as in HCA. Thus, this may not be a good choice for a dataset with large number of cases. Also, only variables with similar scales (variables such as years and money are not in the same scale) can be used to avoid any misleading results.[99] If the variables are in different scales, either standardizing the data or using an alternate clustering method such as HCA is necessary.

K-mean cluster method does not calculate the distances between all the cases as in hierarchical cluster analysis or DFA, thus, it is a fast and efficient method compared to HCA. When performing a K-mean cluster, each spectrum is assigned to a nearest cluster and then the means of the classification clusters are updated accordingly. Finally, the spectra are grouped into the nearest cluster and the final cluster means represent the average values of the clustering variables for the spectra grouped into each cluster.

### 4.3.3 Support Vector Machines

Support vector machines (SVM) are a group of supervised learning methods that can be applied to classification. These methods are mostly used for pattern recognition.[100] This is a linear machine learning method which later modified to use kernel methods. Kernel method can create possible patterns by using similar features in the data. Thus, a kernel based learning algorithm consists of two parts, which includes a general purpose learning machine and a problem specific kernel function. In the general purpose learning portion, solution is a linear combination of training points with consists of non-negative coefficients. The latter part is the non-linear mapping on data using kernel functions, the inner products in some feature space which returns the value of the dot product between the spectra of the two arguments. SVM uses the hyperplane/decision surface which is one less number of dimensions in the data to distinguish two classes by maximizing the margins (minimal distance to the hyperplane) of the data set.[101]

60
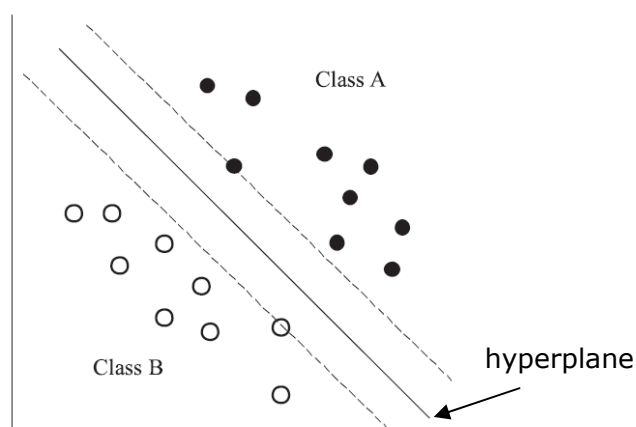


Figure 4.1 Separation of data using SVM[101]

As shown in the Fig. 4.1 the distance between hyperplane and the nearest data point should be large as possible. Here, the decision function is the hyperplane. Data in the decision function and the training algorithm appear only within dot products which is also called duality, a property of SVM. This duality limitation can be resolved by creating a net of simple linear classifiers as in neural network or mapping data into a richer feature space, where, they are linearly separable. Hence, the kernel functions can utilize infinite dimensions. If the dot products are replaced with kernels, one can get a kernel matrix. Then, this kernel matrix can be used to extract the eigen functions which act as features in the data.

Unlike PCA, SVM is not an unsupervised method. Thus, SVM needs prior knowledge of target and many dimensions/variables may complicate the analysis compared to PCA. On the other hand, too many features may lead to over fitting. One study has observed when PCA or DFA is combined with SVM the success rate in pattern recognition is near 100%.[101]

### 4.3.4 Neural Networks

Neural networks (NNs) are other methods which are useful for classification problems. However, for better results of the classification, this method requires a lot of

training data. The algorithms for NNs utilize radial basis function (RBF) or multilayer perceptron (MLP) procedures to create a predictive model based on the variables.[102] There are three major layers; input layer, hidden layer, and the output layer. MLP is the most popular procedure which consists of multiple hidden layers.[103]
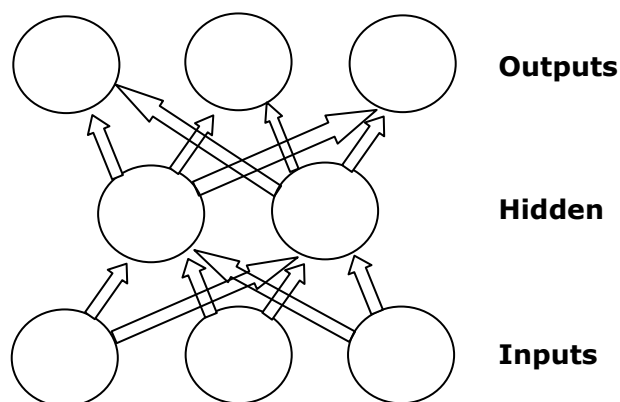


Figure 4.2 Structure of neural network

The structure of NN is known as the feedforward architecture (see Fig. 4.2). Because, the network flows in the forward direction: input layer to the output layer. No feedback loops are used. The weights associated with the nodes are repeatedly adjusted by the training algorithm to minimize the errors and to increase the accuracy of the predictions. The hidden layer in each procedure (RBF or MLP) predicts the output. Usually, the hidden unit function is a Gaussian function which consists of three different standard deviations.[104] In RBF procedure, the network training depends on the number of hidden layer units, centers and sharpness of the Gaussians.

One of the limitations of NN is that the weights associated with the nodes are difficult to interpret and also it is unable to identify the process/processes which lead to the relationship between dependent variables and independent variables. In such cases, it's better to use a different statistical method such as PCA or PLS (see section 4.3.5), etc.

### 4.3.5 Partial Least Squares Regression

Partial least squares (PLS) regression is another regression method which can be utilized when the predictor variables are correlated with each other.[105]   This may be the best alternate to PCA method, when the number of variables is greater than the number of spectra in a dataset.  In fact, the features of PCs and multiple regressions are combined in PLS.  Based on one or more independent variables a model is constructed in PLS, and then one or more dependent variables are tested using this model.  As in PCA, PLS solves an eigenvalue problem to obtain the eigenvectors in a given dataset.

## CHAPTER 5

## RAMAN SPECTROSCOPIC STUDY OF HEAD AND NECK SQUAMOUS CELL CARCINOMA

Head and neck squamous cell carcinoma is the most common malignancy of the head and neck region.  This chapter illustrates the results of Raman spectroscopic investigation of head and neck squamous cell carcinoma of tongue tissues.

### 5.1 Introduction to Head and Neck Squamous Cell Carcinoma

Head and neck squamous cell carcinoma (HNSCC) represent 9% of the cancers.[4] Although there are rapid advances in the fields of therapeutics and diagnostics, the long-term survival of HNSCC patients has not significantly improved in the past 30 years.[106]  The five-year survival in advanced stage cancer is less than 40% and the survival rate for early stage disease is about 80%.[8]  Currently the tumor-node-metastasis (TNM) staging is one of the most important factors clinically used for estimation of survival.  It is important to have reliable estimates for patient survival and it may also impact the treating physician's choice of therapy.  TNM staging is taken into consideration when recommending treatment options. Early stage patients are offered single modality treatment such as radiation or surgery, while multimodality treatment is offered for advanced stage cancer.  As HNSCC is a heterogeneous disease, TNM staging only provides a rough estimate.[8]  Not all patients with the same tumor stage respond in a uniform manner to the available treatment options.

### Motivation

Raman spectroscopy has been extensively applied to detect cancers in a wide range of organs including breast, cervical,[107, 108] skin,[109] and brain[110] tumors.  So far, only a few studies have been performed on HNSCC[49, 50] and most of these studies are on frozen tissue specimens or biopsies.  The studies on nasopharyngeal biopsy tissues showed the spectral features of three Raman bands varied significantly when tissue changed from normal to

malignant state,[49] and another study on frozen tissues from the larynx showed that the characteristic Raman spectral features for normal, carcinoma, and papilloma tissues are different.[50]

In this exploratory study, our primary objective was to investigate the formalin fixed deparaffinized tissues of invasive SCC, normal, and carcinoma *in situ* of tongue. The purpose of this study was to investigate: (*i*) whether Raman spectroscopy when combined with chemometric methods could become a potential diagnostic tool to detect HNSCC in its early stages, and (*ii*) whether there are reproducible and identifiable HNSCC-specific biomarkers in the Raman spectra.

## 5.2 Materials and Methods

### 5.2.1 Sample Preparation

In this exploratory study, we have investigated a total of 17 formalin-fixed deparaffinized tissues from 17 different patients, of which 7 were normal, 4 with carcinoma *in situ*, and 6 with invasive squamous cell carcinoma. The tissue blocks were processed at the University Pathology Research Services, Karmanos Cancer Center. The samples were prepared as described in Chapter 3. The images were taken from the tissue sections to ensure that the Raman spectra were collected from the same regions on the unstained tissue sections as the ones marked by the pathologists on the stained sections. Standard tissue processing methods were used to process the tissues and to remove the wax from them.

### 5.2.2 Pathological Examination

The H&E stained sections were examined separately by three experienced pathologists from Detroit Medical Center, Detroit, MI who marked normal, carcinoma *in situ*,

and invasive SCC regions on every H&E stained slide.  In all the tissue samples used in our study, there was no discrepancy in the evaluations by the three pathologists.

### 5.2.3 Raman Spectroscopic Measurement

For this study, a Jobin-Yvon Horiba Triax 550 Raman spectrometer system with an argon-ion laser (514.5 nm) was used record the spectra (see Chapter 3). The spectral resolution of the Raman spectrometer was about 4 cm$^{-1}$ and the laser power at the spot was ~6 mW.  The spectrometer was calibrated using the known peak from silicon.  The Raman spectra were collected in backscattering configuration with a 100x long working distance (3.4 mm) objective with a numerical aperture of 0.8 (spot size ~ 1 μm diameter).  To improve signal to noise ratio, spectra were recorded using 40 second integration time and averaged over 5 accumulations.  Each spectrum was collected from a new location on the tissue.

### 5.2.5 Raman Data Processing

For statistical analysis, we considered Raman data from 550–1800 cm$^{-1}$ which contains fundamental vibrations of the biomolecules providing information about their molecular structure and conformation.  The Raman spectra often contain fluorescence background emanating from the aromatic chromophores and other noise.  The data were preprocessed for reducing the noise using adaptive minmax/wavelets method[91] and then normalized with respect to the peak with highest Raman intensity in the spectrum.  In order to obtain a reliable statistical analysis, approximately 130 spectra from each category (totaling 401 Raman spectra) from 17 different tissue specimens were recorded.
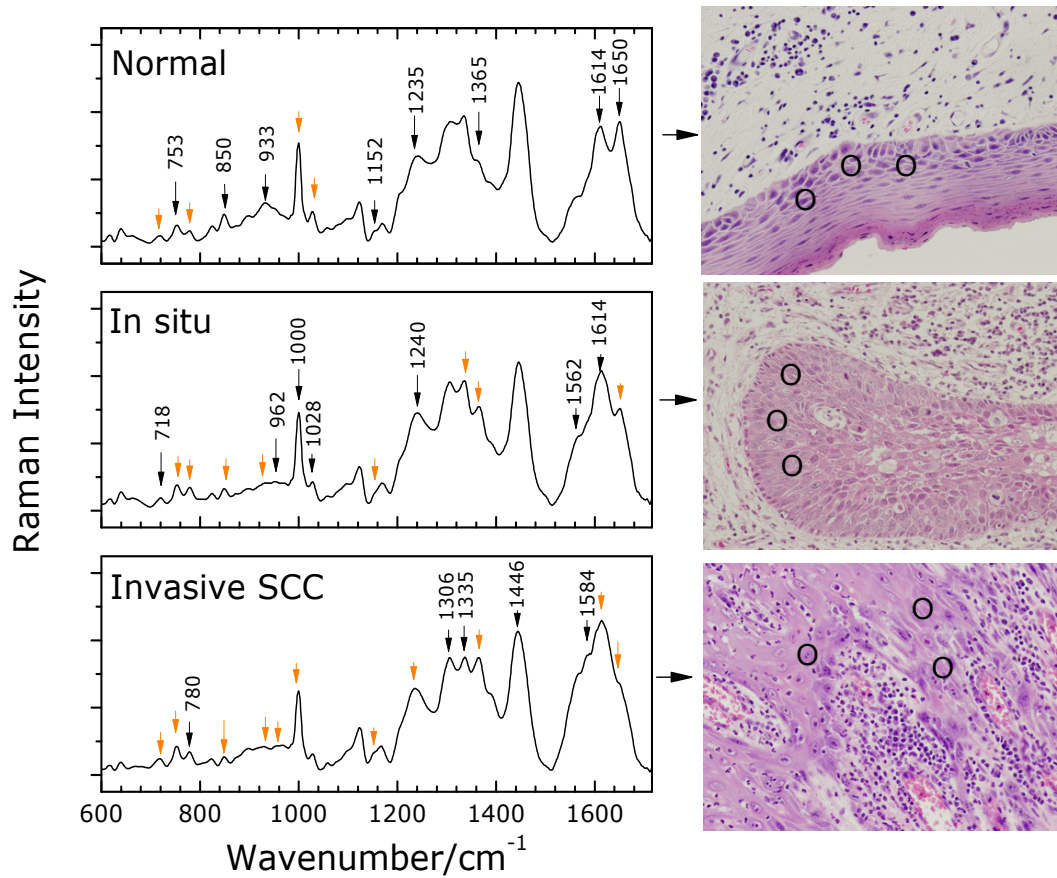
Figure 5.1 Mean Raman spectra of (a) normal, (b) carcinoma *in situ*, and (c) invasive SCC. The three images on the right represent corresponding pathology images. Circle marks represent the approximate locations where the spectra were taken (images are taken from 20)[111]

Figure 5.1 shows the mean Raman spectra of the normal, carcinoma *in situ*, and invasive SCC tissues along with their corresponding pathology images (right panels). The marked circles indicate the representative regions from where the Raman spectra were collected. The Raman shifts of some of the bands whose intensities change when tissue changes from normal to invasive SCC are also marked in the spectra.
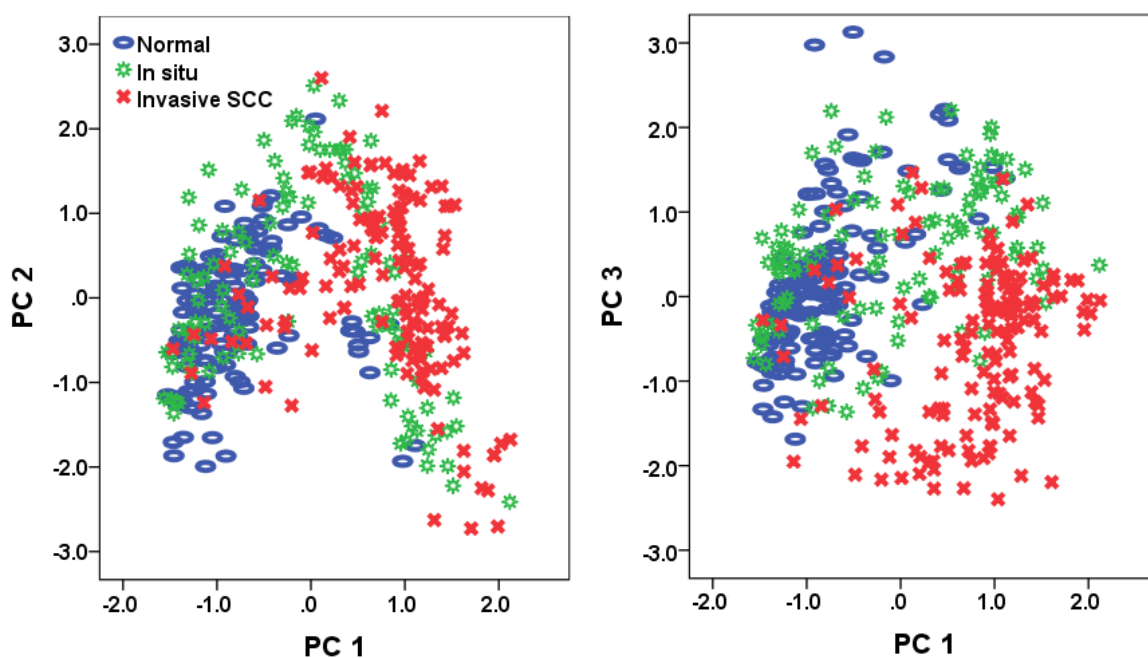
Figure 5.2 PCA plots for full spectrum analysis (a) PC 2 vs. PC 1, (b) PC 3 vs. PC 1[111]

A total of 544 Raman intensities separated by 2 $cm^{-1}$ intervals from the 600-1716 $cm^{-1}$ spectral region (lower and upper edges of the measurement range were discarded to minimize any artifacts due to background subtraction) were used as statistical variables. PCA was performed on the whole data to reduce the number of original variables to fewer independent variables. The initial 544 intensity variables were reduced to 16 variables [principal components (PCs)] which captured 98.2% of the variance in the Raman data.[111]

The first three PCs contained the maximum variance of the Raman data and when plotted against each other, one can see two trends (Fig. 5.2) due to the normal and the invasive SCC data. These PCs were used as input variables for the DFA to classify the data into the three pre-assigned groups. This is achieved by maximizing the variance between the groups while minimizing the variance within the groups. Figure 5.3 shows the groupings in a plot of discriminant functions (DFs). Most of the data are very well separated into distinct groups representing normal, carcinoma *in situ,* and invasive carcinoma.

## 5.3 Results

The Raman spectra of tissue samples are quite complex as the tissues contain a diverse set of small and large biomolecules. The observed spectral intensity variations and frequency shifts shown in Fig. 5.1 are due to molecular compositional and conformational differences in the tissues. The vibrational frequencies associated with different functional groups and backbone chains, for example, in proteins, saccharides and nucleic acids,[42] often overlap, thus making it difficult to assign a specific observed band in the Raman spectrum to a specific functional group of a particular molecule in the tissue. Thus, the frequency assignments have varied significantly among different published studies in the literature. Nevertheless, some functional groups associated with specific molecules often give rise to relatively narrow and well resolved bands in the Raman spectra. These bands can provide direct information about the biochemical composition of the tissue and can be used to study the changes that occur due to the onset of cancer.

### 5.3.1 Chemometric Analysis of the Raman Spectra

The physical significance of the variations in the band intensities can be understood through chemometric analysis methods which correlate the spectral differences with the pathological variations in tissues and thus help in interpreting the spectral differences in terms of changes in biochemical composition. The classification predictions based on Raman spectroscopy measurements performed by a leave-one-out validation test were cross-validated with the "gold standard" results of pathology. The comparison of the two outcomes is given in the cross-validation table (Fig. 5.3). From the data presented in Fig. 5.3, it is encouraging to note that Raman spectroscopy can detect HNSCC about 89% of the time in the tissues we have studied; with 91% accuracy in carcinoma *in situ* and normal tissues.
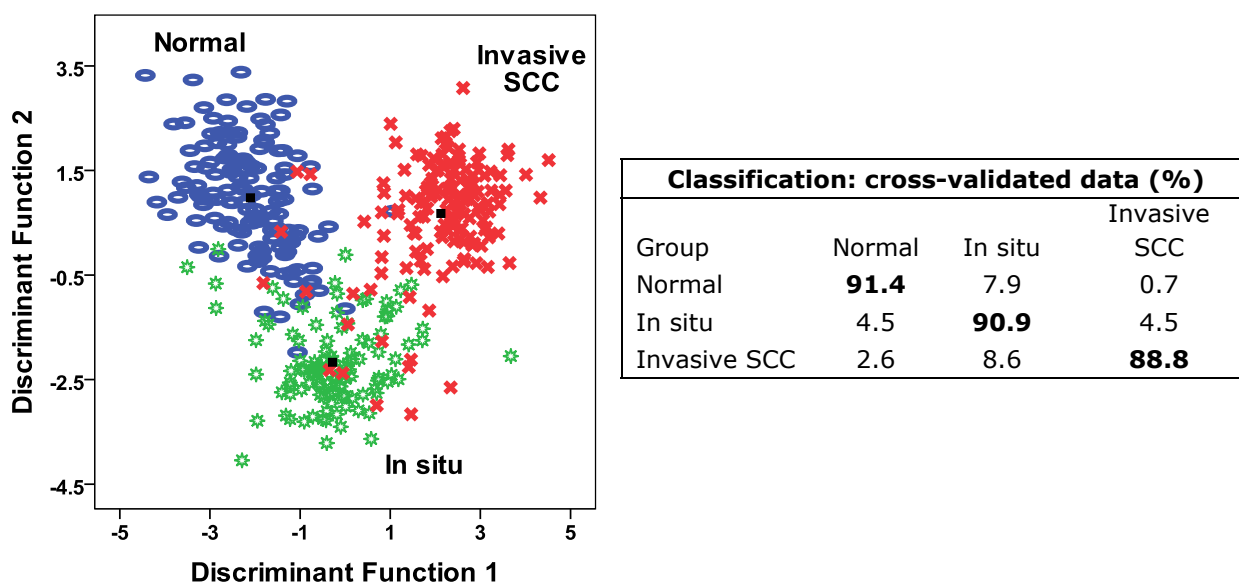
Figure 5.3 DF plot for the analysis using full spectrum, and the corresponding classification results. Ellipse (blue), star (green), and cross (red) symbols represent normal, carcinoma in situ, and invasive SCC, respectively[111]

Biomarkers play very critical role in drug development and treatment as they are the indicators of the normal and cancerous physiologies of an organ. Raman spectroscopy has the potential to determine the biomarkers of a disease. In this exploratory study, we have attempted to determine the Raman bands which play a significant role in differentiating the pathological states of tissue by performing statistical analysis using 12 Raman bands that show largest variation in their intensity (marked on Fig. 5.4). For another DFA, we calculated the sum of the intensities under these bands and used these sums as the input variables for the calculation.
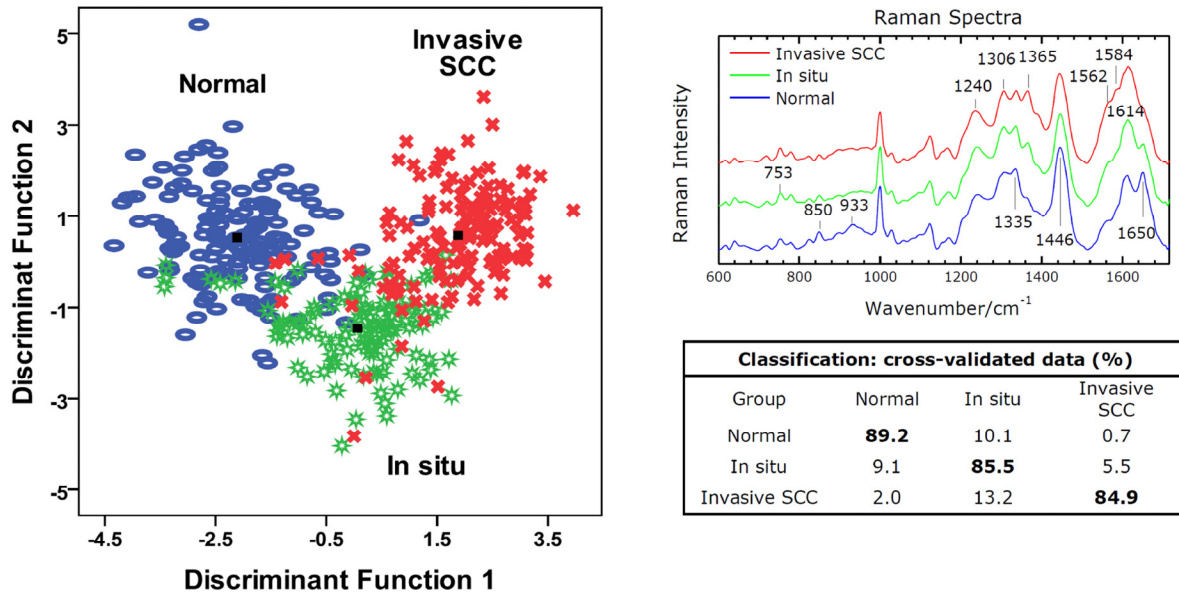
Figure 5.4 DF plot for the analysis using 12 selected Raman bands (shown on the Raman spectra), and the corresponding classification results. Ellipse (blue), star (green), and cross (red) symbols represent normal, carcinoma in situ, and invasive SCC, respectively[111]

Figure 5.4 shows the DFA plot representing normal, carcinoma *in situ*, and invasive SCC states of the tissues as well as the cross-validated classification results using leave-one-out method. It is interesting to compare these results with the full spectrum analysis as done in Fig. 5.3. Comparison of these results with full data (544 intensity variables) shows that these 12 Raman bands are quite successful in classifying the data. The classification table shows that 86% of the original carcinoma *in situ* tissue samples and 85% of the invasive carcinoma tissue samples are correctly classified, whereas the normal tissue sample accuracy was only reduced by 2%. The prediction accuracy only decreased by 4% compared with the whole spectrum analysis (Fig. 5.3), indicating that this set of 12 Raman bands carry sufficient biochemical information of a tissue to accurately predict its pathological state.[111]

## 5.3.2 Signatures of Tryptophan and Keratin in Invasive SCC of Tongue

We have investigated the biochemical origin of these 12 prominent Raman bands whose intensities change noticeably when tissue changes from normal to invasive SCC. Interestingly, we can assign these bands to specific amino acid residues and a particular class of proteins.
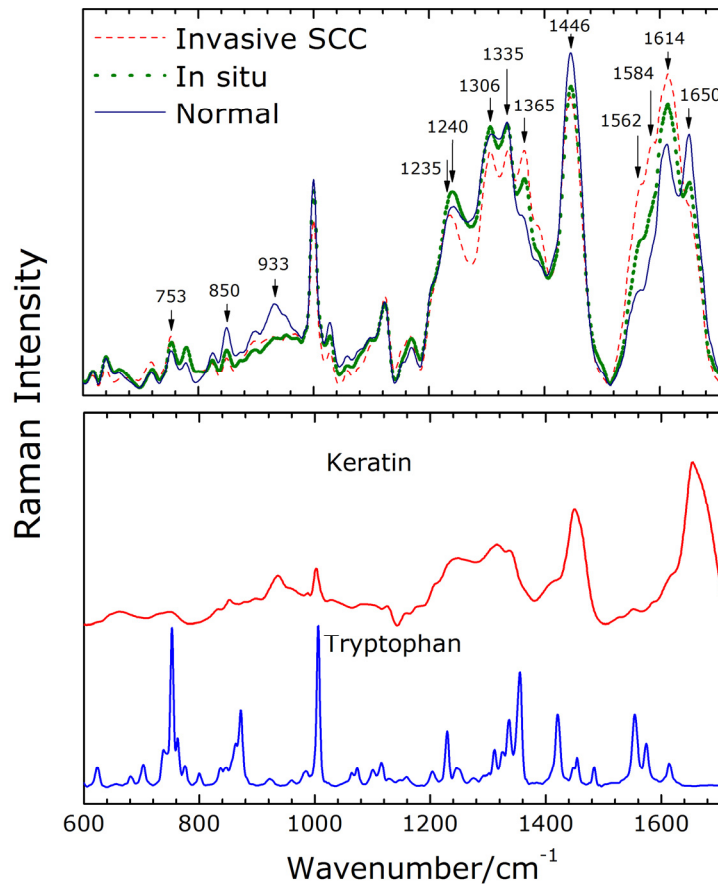


Figure 5.5 Upper panel: The mean Raman spectra for normal (solid blue line), carcinoma in situ (dotted green line), and invasive SCC (dashed red line), lower panel: Raman spectra of keratin and L-tryptophan[111]

In Fig. 5.5, we show the Raman spectra of normal, carcinoma *in situ,* and tumor tissues along with the spectra of keratin (of human white hair) and L-tryptophan (Sigma-Aldrich).

| Raman Shift (cm$^{-1}$) | Peak Assignment |
|---|---|
| 640 | C-C stretching and twisting of proteins-Tyrosine |
| 664 | Ring breathing mode in the DNA bases |
| 718 | C-N nucleotide peak or lipid/DNA |
| **753** | Symmetric breathing of tryptophan |
| 780 | DNA/uracil ring breathing mode |
| 824 | Tyrosine, phosphodiester |
| **850** | Amino acids, keratin |
| 872 | Tryptophan or single bond stretching vibration for amino acids or C-C stretching |
| **933** | C-C stretch, collagen backbone, α-helix, , keratin |
| 1000 | Phenylalanine |
| 1028 | Phenylalanine |
| 1080 | Typical phospholipids, phosphodiester groups in nucleic acids/collagen |
| 1097 | Nucleic acid PO$^-_2$ or C-N stretch |
| 1124 | Lipid or C-N and C-C stretch, proteins, lipids |
| 1152 | C-N stretch, proteins |
| 1167 | C=C stretch lipids |
| 1230 | Tryptophan |
| **1235** | Amide III, keratin |
| **1240** | Amide III, RNA , keratin |
| **1306** | Lipid/protein, keratin |
| **1335** | CH$_3$/CH$_2$ wagging (collagen, nucleic acids), tryptophan, keratin |
| **1366** | Tryptophan or phospholipid (CH$_3$) |
| 1390 | CH-rocking |
| 1420 | COO$^-$ symmetric stretch, CH$_2$ wag, bend, Tryptophan |
| **1446** | CH$_2$ bending mode of proteins & lipids, keratin |
| **1562** | Tryptophan |
| **1584** | Tryptophan, phenylalanine |
| **1614** | Tryptophan |
| **1650** | Amide I, C=C stretch, keratin |

Table 5.1 Raman peaks assignment for the observed bands.[42, 112, 113, 114, 115, 116, 117]  The Raman shifts written in bold face show a significant variation in intensity depending on the pathological state of the tissue

The Raman bands at 753, 1365, 1562, 1584, and 1614 cm$^{-1}$ which show increase in their intensities in invasive SCC tissue can be assigned to L-tryptophan.  While the bands at 850, 933, 1235/1240, 1306, 1335, 1446, and 1650 cm$^{-1}$ which show decrease in their intensities can be attributed to keratin.[118]  A strong correlation between peaks associated with tryptophan and keratin is seen from their Pearson correlation coefficient values (shown in Table A1 in the Appendix A).  Thus, our chemometric analysis strongly suggests that enhancement of tryptophan and down-regulation of keratin may be linked to the progression of SCC of tongue.

It is interesting to note that a recent study by Tankiewicz *et al*[119] has shown an enhancement in the concentration of tryptophan and its derivatives (anthralinic and kynurenic acids) in oral tongue squamous cell carcinoma leading to a proposition that tryptophan may be a factor responsible for the development of cancer.  An earlier study by Crowe *et al* showed a down-regulation of keratin K19 in squamous cell carcinoma cell lines in oral cavity[120] and in another study by Balm *et al* showed the absence of cytokeratin-18 in 66% of the carcinoma of oral cavity.[121]  Our assignment of changes in Raman band intensities to L-tryptophan and keratin are consistent with the above studies.  However, it should be noted that the Raman bands of the pure L-tryptophan are much sharper and the 1420 cm$^{-1}$ peak appears to be absent or shifted in the spectra of the tissues.

The peak broadening in complex systems such as proteins and inhomogeneous system like a tissue compared to pure monomer is expected to occur due to interaction effects.  In amino acids and small oligo-peptides and other molecules, the 1420 cm$^{-1}$ Raman band is assigned to mixed vibrational mode consisting of symmetric COO$^-$ stretching, CH$_2$ wagging and bending vibrations.  Thus, it is possible that the 1420 cm$^{-1}$ is absent in the spectra of tissues because it arises from the symmetric stretching of COO$^-$ group of free amino acid[112-105]. In tryptophan, the Raman band ~1420 cm$^{-1}$ is assigned to one of the well-known vibrational modes of the indole ring.[115-118]  However, as expected, the energy of this

mode changes when present in other medium due to its interaction with the medium's components, e.g., in aqueous solution it shifts[107] to ~1440 cm$^{-1}$. Thus, it is possible that, when L-tryptophan is present in tissue's environment, this mode is either strongly damped or blue-shifted and buried under the intense band structure at 1446 cm$^{-1}$.

### 5.3.3 Monitoring the Progression of SCC with Raman Spectroscopy

It is interesting to note that as normal tissue transforms into carcinoma *in situ*, changes in the intensities of bands at 753, 850, 933, 1235, 1240, 1306, 1335, 1365, 1446, 1562, 1584,1614, and 1650 cm$^{-1}$ are quite significant. These spectral changes are caused by the changes in the biochemical composition of tissues upon normal tissues turning into carcinoma *in situ*. Table 5.1 lists all the observed Raman frequencies along with possible assignments. The frequencies of the peaks that change intensity depending on the state of the tissue are shown in bold in Table 5.1. Further, changes in intensities occur when a carcinoma *in situ* tissue turns into SCC, and thus Raman spectroscopy can distinguish carcinoma *in situ* from normal tissues.

### 5.4 Discussion

Of the available prognostic factors in HNSCC such as TNM stage, surgical margin status, extra-capsular spread, number of positive lymph nodes, perineural and angiolymphatic invasion.[122, 123] TNM staging is currently the most often used clinically. It helps in selecting the aggressiveness of postoperative adjuvant therapy (radiation alone versus chemoradiation). To maximize treatment effectiveness and minimize toxicity, it is important to have reliable survival estimates. Unfortunately, TNM staging provides an estimation of overall survival and is not always correct in predicting treatment response in individual patients.

In this study, we have used Raman spectroscopy which uses characteristic Raman spectral features rather than morphological features to determine the pathological state of tongue squamous cell tissues. It is an objective technique which eliminates inter-observer variability and requires short time (a few minutes) to measure the spectrum of a tissue. In this study, we have identified the spectral features of normal, carcinoma *in situ*, and invasive SCC tissues of tongue that can discriminate between these three tissue types. We have analyzed the Raman spectral data using PCA[93] and DFA[94] chemometric methods. PCA found two distinct trends in the data while DFA was able to group the data into three distinct groups. The classification results were further tested by leave-one-out cross-validation method and compared with the "gold standard" of pathology which clearly predicted the normal, carcinoma *in situ,* and invasive SCC with high accuracy.

We found that the intensities of the Raman bands assignable to tryptophan (753, 1365, 1562, 1584 and, 1614 cm$^{-1}$) increase while those attributable to protein keratin (850, 933, 1235/1240, 1306, 1335, 1446, and 1650 cm$^{-1}$) decrease when tissue changes from normal to carcinoma *in situ* and then to invasive SCC. This suggests that tryptophan and keratin play an important role in SCC of tongue.[111]
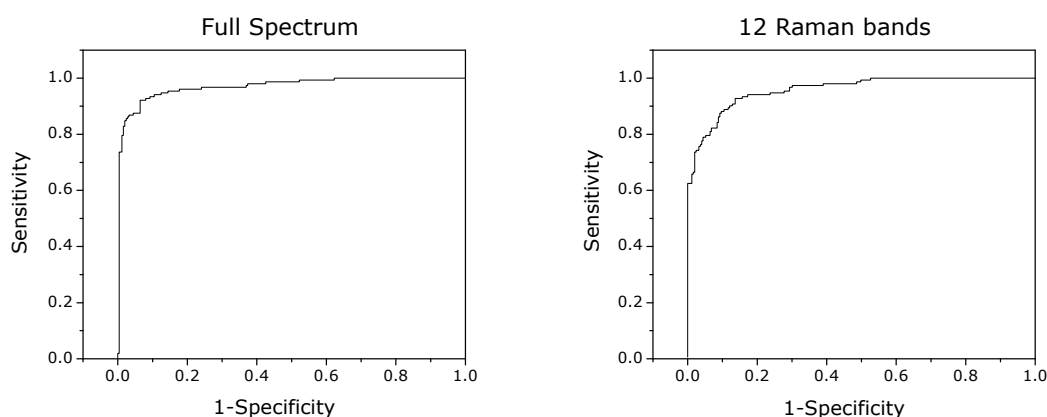


Figure 5.6 ROC curves calculated from the analysis (from left to right) of; full spectra analysis and 12 Raman bands[111]

Based on the discriminant function 1 calculated from the whole data and 12 Raman bands, Receiver Operating Characteristic (ROC) curves (Fig. 5.6) were determined.[124] They are sensitivity vs 1- specificity plots. Sensitivity, calculated from the ratio of carcinoma data to the total data, represents the true positives of a test (for example, in our study it is carcinoma). And 1- specificity describes the false positives of a test (normal but identified as carcinoma) as a ratio of normal to the total data. In the perfect scenario, ROC curve is a step function. If the area is close to 1, then the test result is considered perfect whereas area close to 0.5 is considered bad. In our case, the area under the curve for whole spectrum and analysis using 12 Raman peaks are 0.97 and 0.959. Thus, the capability of differentiation between invasive squamous cell carcinoma and normal using Raman spectroscopy is quite remarkable even with twelve selective Raman peaks that exhibit intensity variations.

### 5.5 Conclusions

Raman spectroscopy is an optical technique capable of providing molecular level information, enabling investigation of functional groups and molecular conformations. Molecule-specific spectral bands can provide direct information about the biochemical composition in a tissue thus allowing one to detect small changes in biochemical compositions which occurs at the inception of cancer. We were able to detect with 91% success rate the normal and carcinoma *in situ* tissues of tongue and with 89% accuracy the invasive squamous cell carcinoma tissues. Thus, this exploratory study shows that Raman spectroscopy can be used in studying the head and neck squamous cell carcinoma. Our study shows increased levels of tryptophan and decreased levels of keratin as the tissue changes from normal to squamous cell carcinoma which is consistent with other biochemical investigations.[119-121]

**CHAPTER 6**

**RAMAN SPECTROSCOPIC STUDY OF PROSTATIC INTRAEPITHELIAL NEOPLASIA**

Prostatic intraepithelial neoplasia is the precursor lesion of prostate cancer.  If this pathological state can be optically identified from different grades of cancer, then, this could lead to the screening of potential prostate cancer patients *in vivo* or *in vitro* using Raman spectroscopy.  In this chapter, we have attempted to explore this possibility using Raman spectroscopy.

### 6.1 Introduction to Prostate Cancer

Prostate cancer is the second most common cancer among men worldwide based on the statistical data collected through 2008.[1]  Ever since screening with the Prostate Specific Antigen (PSA) test began, which measures the level of PSA in patients' blood, its mortality rate has trended downwards due to its early detection and treatment.[125, 126]  Other detection method involves a digital rectal examination to check for growths in or enlargement of the prostate gland in men.  If there is a tumor growth in the prostate, it can often be felt as a hard lump.  To further confirm the tumor, a pathological examination is performed on biopsies of the prostate and the initial grade of the cancer is given by the pathologist.[15]  According to the current policy, the age limit to obtain PSA has been lowered to 40 years.  In addition, a decision to perform biopsies is not followed by PSA and DRE alone, other factors such as patient age, PSA velocity, PSA density, family history, ethnicity, etc. are also considered.[127]  One of the common tissue extraction methods is needle biopsy.

### Motivation

Detection and confirmation of prostate cancer is very crucial for its successful treatment in order to improve survival rate.  Sometimes the standard screening programs can provide misleading results leading to wrong or over-treatments and occasionally to fatal

consequences. The interpretations of histological examination of biopsies, considered as the "gold standard" for diagnosis, are often subjective and can vary significantly from one pathologist to another.[128] Hence, it is imperative to detect the state of the disease with a method which is objective and capable of providing results within a very short period of time, in 1-2 minutes. Raman spectroscopy techniques are very well suited for these types of goals and, in addition, they are also capable of probing disease at cellular level. This optical method can provide information about the changes in the concentrations of the constituent biomolecules of tissues and detect the progression or state of the disease. Therefore, this technique can be used to diagnose pathological condition of organs and progression of disease.

Apart from benign epithelia (BE), PIN, or cancerous micro acinar clusters, stromal cells also surround these biomolecular patterns. The interaction between human prostatic stroma and the epithelial cells in normal prostate is somewhat different from the stromal cells in prostate tumors.[129] The normal human prostatic stroma which consists of fibromuscular matrix enclosing the prostatic ducts limits the proliferation of the epithelia unlike the stromal cells in the prostatic tumors, which contains fibroblasts or myofibroblasts. The stroma bordering prostatic tumors is called "reactive stroma" or "carcinoma associated fibroblasts".[129] It is imperative to explore the spectral features of the stromal cells in benign epithelia (BE), PIN, and tumor stages and understand its linkage with cancer and its progression. It appears that the reactive stroma in prostate initiates the carcinogenesis and helps its progression.[130, 131, 132] Our primary goal is to investigate luminal cells of PIN and compare their spectral features with BE and cancerous cells of the prostate tissues. It is important to compare luminal cells alone of each pathological category since basal cells are absent in the epithelium of microacinar structures of the prostate cancer. Our secondary goal is to examine the Raman spectral features of the stroma surrounding BE, PIN, and

cancer in deparaffinized tissues.  To the best of our knowledge, our exploratory Raman study is the first report of such an investigation.

## 6.2 Materials and Methods

### 6.2.1 Raman Spectroscopy

As in the pediatric tumor study, we used a Renishaw RM1000 Raman microscope-spectrometer with a 785 nm diode laser excitation source.  Each Raman spectrum was averaged over three scans with 20 second integration time to gain a good signal-to-noise ratio.  The scattered light was collected through back scattering geometry and dispersed through a 1200 lines/mm grating structure before being sent to the CCD detector.

### 6.2.2 Preparation of Tissues for Raman Spectroscopy

Tissue specimens which were embedded in paraffin wax were obtained from Karmanos Cancer Center and Harper University Hospital in Detroit, MI, and were processed at the University Pathology Services at Karmanos Cancer Center.  In this study, we particularly selected those tissue specimens that were purely either BE, PIN, or cancerous. For each specimen, two parallel sections were cut.  One 5 μm in thickness (H&E stained) was used for pathological examination and the second 10 μm adjacent thick section was used for Raman spectroscopic measurements.  These H&E slides were reviewed by three experienced pathologists and none of the cases were found to be in dispute.  Standard procedure was used to remove paraffin wax from the 10 μm thick tissue specimens before they were used for Raman study.  The wax in the tissue layers was removed by using xylene and ethanol baths.  However, if this was not done properly, the wax residues would likely interfere with the Raman spectrum in the Raman bands at 1063, 1130, 1296, 1436, and 1465 cm$^{-1}$.  In our samples, we did not observe any of these sharp wax bands.  It is assumed that the morphological features do not change across the few micrometer thick

layers, hence the H&E stained layer can be used as a guide layer to collect Raman spectra from the specific sites of the adjacent unstained deparaffinized tissue section.

### 6.2.3 Raman Spectroscopic Measurements

A total of 34 tissue specimens obtained from 33 patients were used in this study. Out of these, 12 specimens were benign, 11 were PIN, and 11 with cancer.  Images were taken from the marked regions corresponding with each H&E slide to insure that the data were taken from the correct regions of interest.  When collecting Raman spectra, the laser beam was focused only on the luminal cells.  The circular regions in Fig. 6.1 show the regions from where the Raman spectra were collected.  In addition, each data location was imaged to make sure that each Raman spectrum is taken from a new luminal cell of the target tissue.

The Raman spectra were collected in the 500-1900 $cm^{-1}$ range as the region of 600-1800 $cm^{-1}$ is commonly known as the "biological window" in the field of biomedical research. Most of the biomolecules show intense Raman excitations across the biological window.  The extra 100 $cm^{-1}$ were recorded in both ends of the Raman spectra to avoid any artifacts which may occur while removing the fluorescence background from each spectrum.  A total of 1220 Raman spectra were collected from the tissue samples in which 207, 202, and 208 were from BE, PIN, and cancer, respectively.  We have collected roughly the same number of spectra from each category so that the variations in the spectral features of BE, PIN, and cancer can be reliably compared.  In addition, 201 Raman spectra were obtained from the stromal areas surrounding the BE, PIN, and cancer (Gleason grade 6).  Here, the Raman measurements of stromal cells in cancer tissues were obtained from the bordering of the micro acinar clusters as shown in Fig. 6.1.
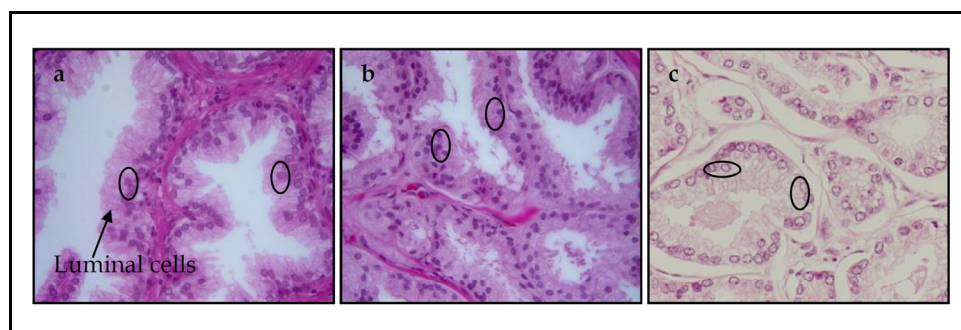
Figure 6.1 Pathology pictures of (a) BE, (b) PIN, and (c) Cancer (images are taken with 40x magnification). Elliptical symbols represent locations of the Raman measurements[133]

### 6.2.4 Raman Data Processing and Chemometric Analysis

While collecting Raman spectra, we made sure that there was no non-standard noise in any spectrum and, if found, we discarded those spectra from our database. The remaining spectra were cleaned from any spurious bands due to cosmic event and noise by using wavelets method.[91] The fluorescence background from each spectrum was removed with a minmax adaptive algorithm that requires no a priori knowledge of the spectra. Finally, each spectrum was normalized with respect to the highest intensity value of the spectrum.

Multivariate/chemometric statistical method, like PCA[93] which determines correlation in the variance, can be used to detect trends in the data set and data can be further analyzed using DFA[94] to find classes of the data. First, we analyzed the data using PCA which reduced the dimensionality of the original data set from 601 variables to 19 new variables, called the eigenvectors. These new variables captured 97% of the variance of the data. Examination of the first two eigenvectors showed distinct trends in the data representing BE, PIN, and cancer. These new fewer variables carrying most of the variance of the data are useful for determination of the groups in the data. To find classes in the data, we have performed DFA[94] using the 19 eigenvectors as the input variables. The classification of each pathological state was done using the leave-one-out method, where each data case is considered a new case and compared with the rest of the data pool.

## 6.3 Results and Discussion

The average Raman spectral features of BE, PIN, and cancer are shown in Fig. 6.2. We found changes in the peak intensities of most of the Raman bands (Raman band assignments are listed in Table 6.1) of the PIN spectrum. These changes are fundamentally related to the changes in the concentrations of the biochemicals of BE and normal luminal cells. Significant changes were found in the region from 600 cm$^{-1}$ to 1145 cm$^{-1}$ which are shown in the lower panel of the Fig. 6.2.
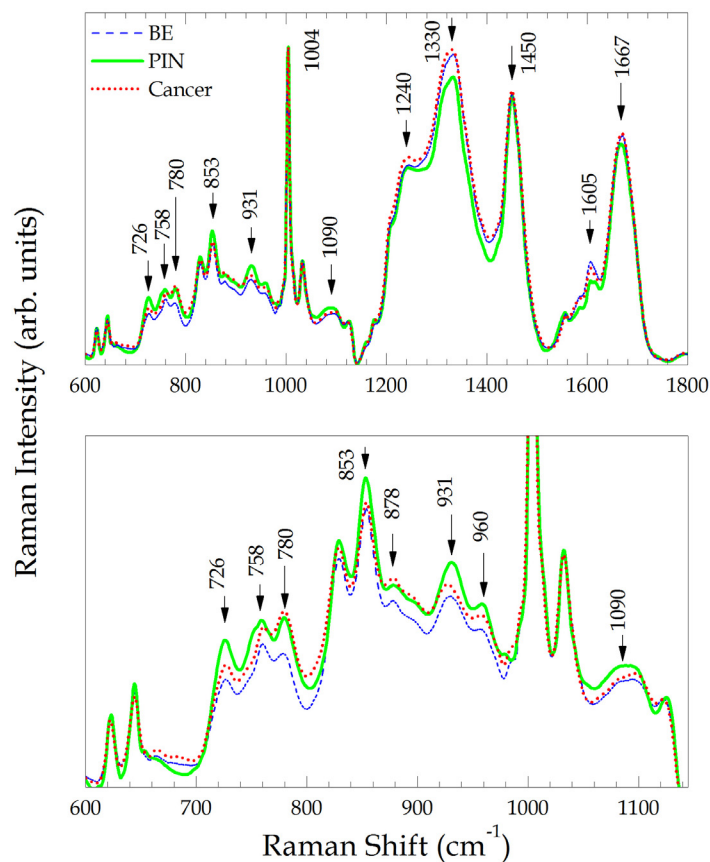


Figure 6.2 Average Raman spectra of BE (Blue dash line), PIN (green), and cancer (red dotted line). The lower panel shows the Raman spectral range of 600-1145 cm$^{-1}$ [133]

Some of the changes are noteworthy: the band at 726 cm$^{-1}$ (assigned to ring breathing mode of DNA/RNA bases) becomes quite intense in PIN, and in addition the

Raman bands at 853, 931 $cm^{-1}$ ($\nu$(C-C) stretching mode of proline, valine, and protein), 960 $cm^{-1}$, and 1090 $cm^{-1}$ (symmetric phosphate stretching vibrations) also showed increases in their intensities when the pathological state of the cell changes from BE to PIN, while the bands at 1605 $cm^{-1}$ and 1667 $cm^{-1}$ (amide I) showed suppression in their intensities. When comparing the average spectral changes of PIN with cancer, we found that the peak intensities at 780 $cm^{-1}$, 1240 $cm^{-1}$ (proline, tyrosine), 1330 $cm^{-1}$, and 1605 $cm^{-1}$ are enhanced when the pathological state of luminal cells changes from PIN to cancer, while the bands at 726, 853, 931, 960, and 1090 $cm^{-1}$ showed a decrease in their intensities which also showed a similar trend like the bands of BE.

| Raman Shift ($cm^{-1}$) | Peak Assignment |
|---|---|
| 726 | A ring breathing mode of DNA/RNA bases |
| 758 | Symmetric breathing of tryptophan |
| 780 | DNA/uracil ring breathing mode |
| 829 | Tyrosine, phosphodiester, O-P-O stretching DNA/RNA |
| 853 | Ring breathing mode of tyrosine, C-C stretch of proline ring, glycogen |
| 878 | Tryptophan, hydroxyproline, C-O-C ring |
| 931 | C-C stretch, α-helix, protein band |
| 960 | Cholesterol, phosphate of HA |
| 1004 | Phenylalanine |
| 1032 | Phenylalanine, proline |
| 1081 | Typical phospholipids, phosphodiester groups in nucleic acids/collagen |
| 1090 | Symmetric phosphate stretching vibrations |
| 1240 | RNA, Amide III, collagen |
| 1313 | Lipid/protein |
| 1330 | Collagen, nucleic acids & phospholipids |
| 1450 | $CH_2$ bending mode of proteins & lipids, methylene deformation |
| 1557 | Tryptophan, tyrosine |
| 1605 | Cytosine, phenylalanine, tyrosine, C=C |
| 1667 | Protein, C=C stretch, amide I |

Table 6.1 Raman peak assignment[41, 42]

The Raman bands at 780 and 878 cm$^{-1}$ showed progressive increase in their intensities when cells change from BE to PIN and then to cancer indicating that the increase in the concentrations of the biochemicals related to these bands is continuous. These Raman bands should be further investigated and if found to show the same trend, then they can be used as diagnostic variables for the detection and progression of prostate cancer.

### 6.3.1 Chemometric Analysis of the PIN, BE and, Cancer

The first three eigenvectors or the principal scores (PCs) are plotted against each other in Fig. 6.3. The left panel is the plot of PC2 vs. PC1 and the right panel represents PC3 vs. PC1. These three principal scores contain 76% of the variance in the data showing different trends for each pathological state. Although some of the data seem to be overlapping with each other, the different trends are still very clear in the data. Largest variance in the data was captured by the PC1 which is shown by the spread of the BE, PIN and cancer data along the PC1 axis. PC3 shows a distinct trend present in the PIN spectral data.
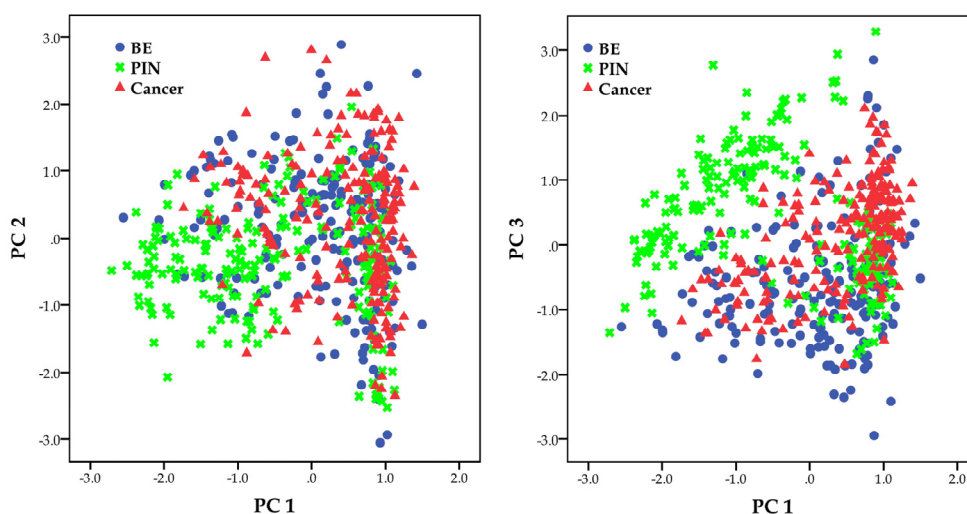


Figure 6.3 PCA results of BE, PIN, and cancer. PC2 vs PC1 is on the left panel and the PC3 vs. PC1 is on to the right[133]
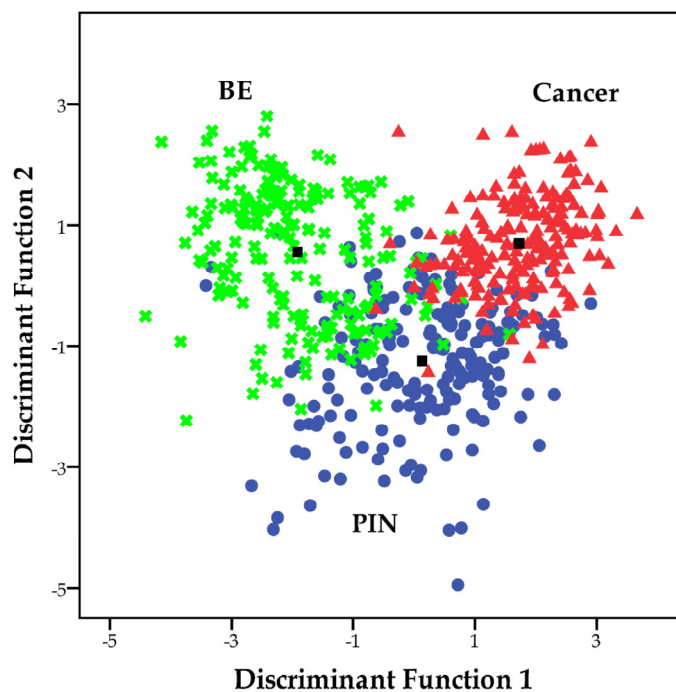
Figure 6.4 DF plot of BE, PIN, and cancer[133]

The average spectra of BE, PIN, and cancer showed distinct variations in the intensities of certain peaks while the spectral features of the individual spectra of each category are expected to be scattered about its average value and must also be different from other categories. As tissue changes from BE to PIN and then to cancer, the spectral variances in the data must exhibit some distinct classes if each of these categories is pathologically different. The classification results are shown in Fig. 6.4 where we clearly observed three distinct classes with their centroids (marked with black squares) far away from each other. It is interesting to note that the PIN class was a very distinct class from other classes and doesn't have much overlap with others. There are only three pathological states, which means, there are only two discriminant functions. The discriminant function 1 shows the maximum variance in the Raman data and the discriminant function 2 contains the rest of the variance.[94] The x-axis shows a continuous change from BE to PIN and then to cancer. The group prediction of the Raman spectroscopy data using DFA was compared

with that of pathological diagnosis, the gold standard for diagnosing cancer. To test the validity of our predicted classifications, we have performed a leave-one-out cross-validation where the group classification for each spectrum with one of the known pathological states is determined while using the remaining data as a training set. The results of cross-validation were compared with the classification results in Table 6.2. We found that the PIN could be predicted with 82% accuracy while the prediction accuracies for BE and cancer were 76% and 91 %, respectively.

|  |  | Group | Predicted Group Membership | | | Total |
|  |  |  | BE | PIN | Cancer |  |
|---|---|---|---|---|---|---|
| Cross-validated | Count | BE | 158 | 20 | 29 | 207 |
|  |  | PIN | 33 | 165 | 4 | 202 |
|  |  | Cancer | 16 | 2 | 190 | 208 |
|  | % | BE | **76.3** | 9.7 | 14.0 | 100.0 |
|  |  | PIN | 16.3 | **81.7** | 2.0 | 100.0 |
|  |  | Cancer | 7.7 | 1.0 | **91.3** | 100.0 |

Table 6.2 Classification results of BE, PIN, and cancer[133]

### 6.3.2 Stroma-Epithelium Raman Spectra Comparison

It is interesting to study the stroma around each pathological state as it could provide useful information about the onset of cancer. The nature of stroma observed was found to depend on its environment.[129] Figure 6.5 shows the average Raman spectra of stroma surrounding BE, PIN, and cancer together with the average Raman spectra of the relevant epithelium in black.
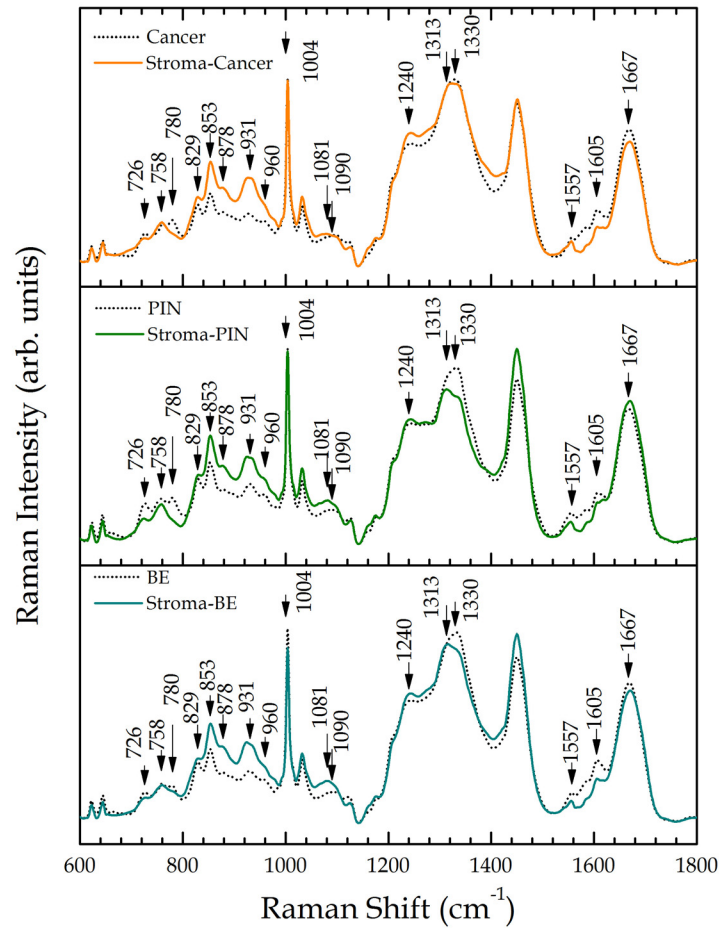
Figure 6.5 Average Raman spectra of stroma in BE, stroma in PIN, and stroma in cancer together with the epithelial cells of BE, PIN, and cancer, respectively which represent in black dotted format[133]

Most of the Raman bands show significant differences between PIN and stroma associated with it, especially the Raman bands at 726, 780, 853, 931, 960, 1081, 1313, 1330, 1605, and 1667 cm$^{-1}$. The Raman bands associated with DNA/RNA (726, 780, 829, and 1330 cm$^{-1}$) are suppressed in all the stromal Raman spectra compared to the luminal cells of BE, PIN, and cancer. In fact, the Raman bands incorporating amino acids/protein/collagen (853, 931, 960, and 1240 cm$^{-1}$) were significantly enhanced. The Raman band at 1081 cm$^{-1}$ which is related to phospholipids was also enhanced in stroma. These observations are in full agreement with the actual biochemical composition of stroma.

The chemometric analysis was performed on the stromal data.  Similar to the epithelia investigation, the PCA created 15 principal scores containing 98% of the variance in the Raman data.  Apparently, the three categories: stroma in BE, stroma in PIN, and stroma in cancer are quite distinct as shown in the DF plot together with the overlapping average Raman spectra of stroma in BE, stroma in PIN, and stroma in cancer (see Fig. 6.6).
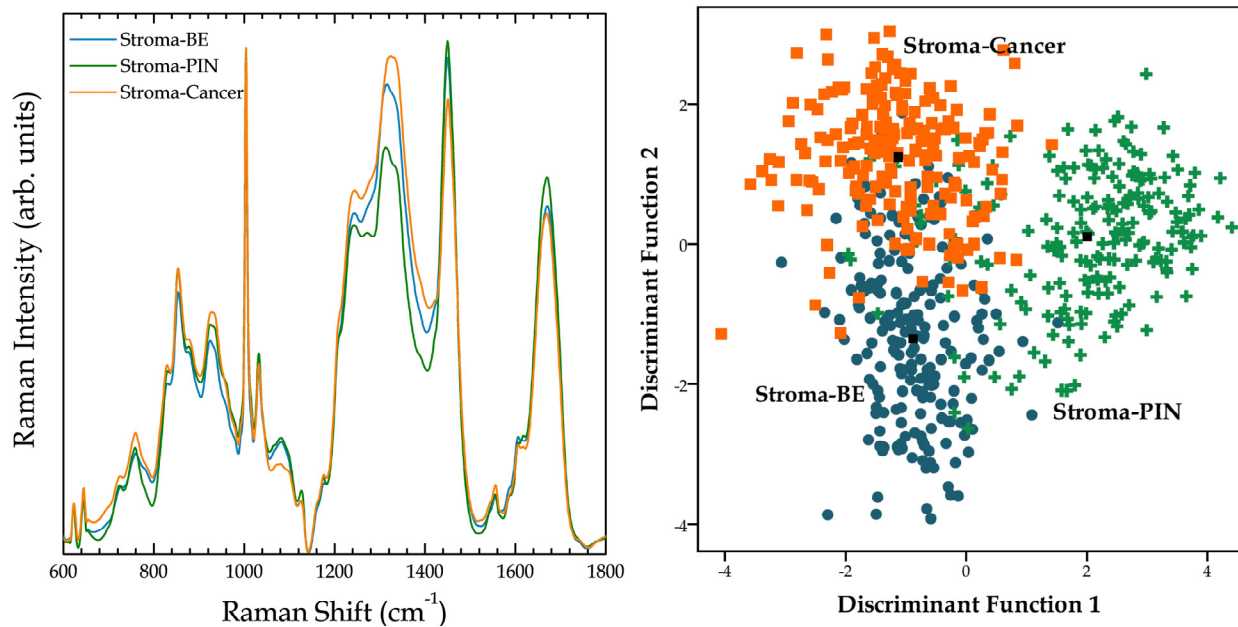


Figure 6.6 DF plot of stroma surrounding BE, PIN, and Cancer.  The average Raman spectra of stromal data for each category are shown to the right of the figure[133]

Table 6.3 shows the predicted group membership using a leave-one out classification method for stromal investigation.  We found that the stroma surrounding PIN could be correctly identified with 83.6% accuracy.  The stroma in BE and cancer were classified with 81.6% and 87.1% accuracy, respectively.

| | | Group | Stroma-BE | Stroma-PIN | Stroma-Cancer | Total |
|---|---|---|---|---|---|---|
| | | | Predicted Group Membership | | | |
| Cross-validated | Count | Stroma-BE | 164 | 3 | 34 | 201 |
| | | Stroma-PIN | 14 | 168 | 19 | 201 |
| | | Stroma-Cancer | 18 | 8 | 175 | 201 |
| | % | Stroma-BE | **81.6** | 1.5 | 16.9 | 100.0 |
| | | Stroma-PIN | 7.0 | **83.6** | 9.5 | 100.0 |
| | | Stroma-Cancer | 9.0 | 4.0 | **87.1** | 100.0 |

Table 6.3 Classification results of stroma in BE, PIN, and cancer[133]

It seems that when comparing stroma enclosing PIN with stroma in cancer, the Raman bands at 853, 931, 1240, and 1330 $cm^{-1}$ were increased while the Raman bands at 1081 $cm^{-1}$ and 1450 $cm^{-1}$ were decreased. In fact, the stroma associated with PIN showed less interference with the changes occurring in the PIN structure whereas the stroma associated with cancer interfered more with the progression of the cancer epithelia as the Raman bands at 1313, 1330, and 1450 $cm^{-1}$ are relatively the same as in cancer and stroma around it. Thus, the stroma associated with BE, PIN, and cancer can be detected with high accuracy using Raman spectroscopy.

When spectral data of all the pathologies are combined and analyzed statistically to find their distinct classes, interestingly we found a clear separation of stroma from BE, PIN, and cancer. In addition, stroma-BE, stroma-PIN and stroma-cancer were also well separated as shown in Fig. 6.7. Note that DFA constructs one less number of discriminant functions than the number of user-defined categories. Thus, this analysis produces 5 discriminant functions due to 6 categories. Figure 6.7 only shows the first two discriminant functions which carry 89% of the information in the Raman spectra. Hence, the overlapping of data in this 2-D plot may not be the same for the other dimensions.
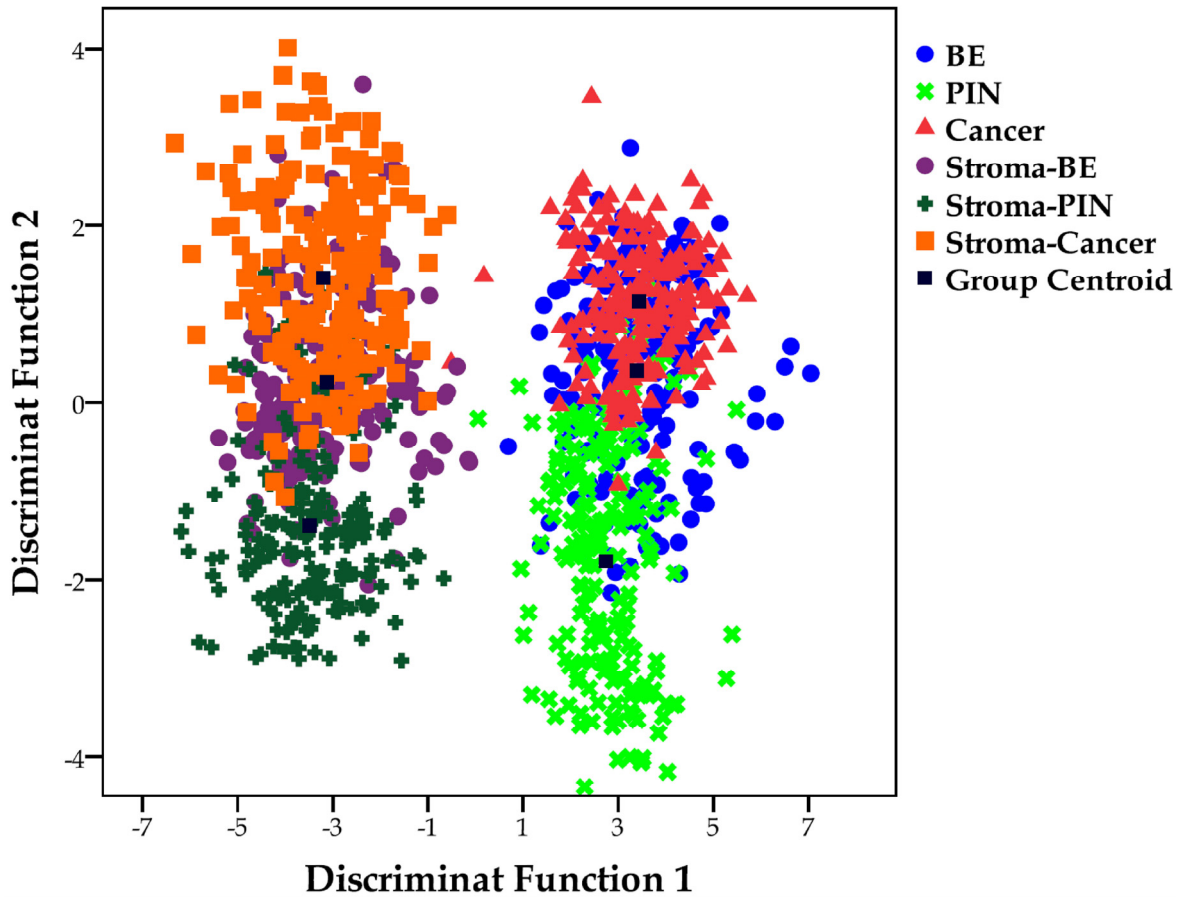
Figure 6.7 DF plot of all the categories: BE, PIN, cancer, and stroma surrounding BE, PIN, and cancer[133]

The results clearly show that as a luminal cell changes from its normal state to PIN before becoming cancerous, the Raman spectral features have already changed and these changes continue as the disease progresses. Moreover, if the laser beam is shined only on the stromal regions far away from its corresponding pathologies (PIN, BE, and cancer), we expect the Raman spectral features to be very different and capable of identifying itself with much higher accuracy.
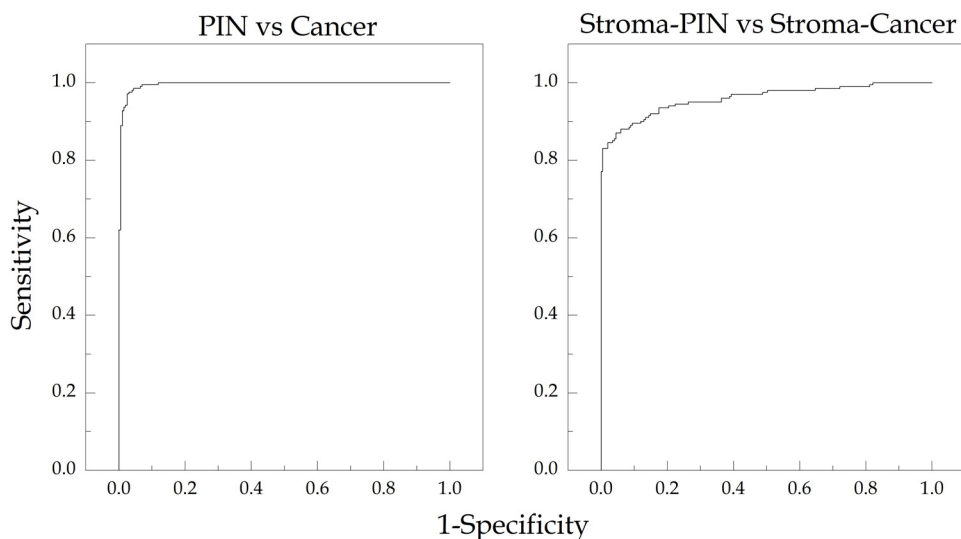
Figure 6.8 ROC curves for PIN vs cancer (left) and stromal PIN vs stromal cancer (right)[133]

We also constructed Receiver Operating Characteristic (ROC) curves[114] for both the analyses, PIN compared to cancer and stromal comparison between stroma around PIN and stroma associated with cancer as shown in Fig. 6.8. This ROC curve is a graph where sensitivity and 1-specificity are plotted against each other. Sensitivity is the ratio of true positives (cancer measurements which are correctly identified as cancer) over the total PIN data and specificity is the ratio of true negatives (PIN measurements which are correctly identified as PIN) over the total PIN data. The sensitivity and specificity for luminal cell investigation of PIN and cancer are 98% and 99%. For the stromal investigation, the sensitivity and specificity are about 90% and 94%, respectively. The area under the curves (AUC) infers the validity of the test. As shown in Fig. 6.8, both the AUC's are more than 0.96 indicating a very good test for PIN and stromal investigation.[133]

## 6.4 Conclusions

In this study, we investigated using Raman spectroscopy on luminal cells from tissues which are purely either BE, PIN, or cancerous. We particularly focused on PIN and

compared its spectral features with BE and cancer. As tissue transforms into PIN, most of the Raman bands of BE showed changes in their intensities and, in addition, an intense Raman band at 780 cm$^{-1}$ (assigned to DNA/RNA) also appeared in PIN's spectrum. As luminal cells become cancerous, some of the Raman bands including 780 cm$^{-1}$ showed progressive increase in their intensities. Additionally we also studied stromal cells around each pathological state; BE, PIN, or cancer. Chemometric analysis of the data shows that the spectral variations in the data are quite pronounced and can be easily classified with very high accuracies into distinct pathological groups. The sensitivity and specificity of luminal cells of PIN and cancer are about 98% for each and for the stroma associated with these pathologies, both the sensitivity and specificity are more than 90%. The current study suggests a further investigation of different grades of PIN pathology: low grade and high grade, including basal cells, and expand Raman spectral database for better prediction accuracy. Since some of the PIN structures may not lead to carcinoma, by spectroscopic investigation one can find the Raman signatures of a PIN structure which can lead to cancer.

# CHAPTER 7

## RAMAN SPECTROSCOPIC STUDY OF PEDIATRIC TUMORS

In this chapter, the investigation of the cellular regions in normal adrenal gland and pediatric tumors, specifically neuroblastoma and ganglioneuroma, using Raman spectroscopy is discussed in detail.  Our investigation is the first report of comparison of frozen and deparaffinized pediatric tissues of the same patients using Raman spectral signatures.

### 7.1 Introduction to Neuroblastoma and Ganglioneuroma

Neuroblastoma and ganglioneuroma are peripheral neuroblastic tumors derived from immature sympathetic nervous system and the adrenal gland.  Neuroblastoma, the third most common type of children's cancer, occurs in embryonic neuroblasts of the sympathetic peripheral nervous system which can spontaneously mature into ganglioneuroma.[17]  It is a solid malignant tumor which can be observed in the abdomen, around the spinal cord near the chest region, or pelvis.  MYCN (v-myc myelocytomatosis viral related oncogene) gene amplification, DNA ploidy, chromosome 1p depletion or loss of heterozygosity, and gain of 17p chromosome are the most vital prognostic markers.[134]  Neuroblastoma can be of favorable or unfavorable histopathology/prognosis based on the Shimada classification.[19, 20] When MKI (mitosis-karyorrhexis index, number of mitoses and karyorrhexis per 5,000 cells) is low or intermediate in neuroblastoma at age less than 1.5 years, these cases belong to the favorable histology group.  In addition, when the differentiating neuroblastoma is diagnosed with low MKI between ages 1.5-5 years, the histology group is also favorable.[135]

If a child is diagnosed with neuroblastoma, it is frequently unamenable to surgical removal at diagnosis and requires chemotherapy followed by surgical debulking.  Prognosis is dependent on the patient age, morphological findings such as degree of ganglionic differentiation, stromal maturation, mitotic/karyorrhectic count and biological molecular

markers.[134] Ganglioneuroma on the other hand is a benign neoplasm that is considered the end result of spontaneous maturation of neuroblastoma in a subset of patients. This is a unique phenomenon that is still incompletely understood. Ganglioneuromas are usually treated by surgical removal, no chemotherapy is needed and the prognosis is excellent. Chemotherapy treatment in neuroblastoma also induces ganglionic maturation and stromal maturation and often neuroblastomas at the time of debulking surgery appear histologically similar to ganglioneuromas.[17]

## Motivation

Although several Raman spectroscopic investigations have been carried out on frozen tissues, the abundant availability of archived formalin fixed paraffin preserved (FFPP) tissues may provide a venue for conducting a thorough investigation for diagnostic purpose. However, the preservation of tissue specimens by embedding them in paraffin wax is known to alter their biochemical composition. In a study using human placenta specimens it has been shown that formalin fixation results in additional peaks as well as shifts in the amide bands resulting from changes in protein conformation and possible cross-links.[136] Thus, for a reliable diagnosis of pathological state in deparaffinized or FFPP tissues using Raman spectroscopy, it is important to identify the biochemical changes which occur in archived FFPP tissues. Once the nature of the loss of biochemical information is known and understood, one can pursue further studies using FFPP tissues using Raman spectroscopy. Further, the pathological evaluations have inherently an element of subjectivity as the outcome of the evaluation is based on the experience and intuition of the pathologist. The information extracted from the pathological evaluations has a high impact in determining the treatment strategy. Hence, it is crucial to develop an objective optical technique such as Raman spectroscopy to diagnose the pathological state of tissue with high reliability. The diagnostic ability of Raman spectroscopy for paraffin embedded archived pediatric

tumors/neoplasm has not yet been fully explored.  In this study, our aim is investigate frozen tissue sections of normal adrenal gland, neuroblastoma, and ganglioneuroma obtained from three different patients, and compared their Raman spectral features with the corresponding deparaffinized sections of FFPP tissues.  The goal of this study is to understand the nature of biochemical changes that occur in FFPP tissues and to assess their feasibility for diagnosis of diseases using Raman spectroscopy.  We have attempted to determine the Raman spectroscopic signatures of each type of tumor, particularly, from the cellular regions of well identified pathology rather than collecting data from random sites in the bulk tissue specimens as was done in the earlier studies.[51-53]

### 7.2 Materials and Methods

For this research work, the Renishaw RM1000 Raman microscope-spectrometer (785 nm excitation laser source) was used to collect the Raman spectra.

### 7.2.1 Normal adrenal gland, Neuroblastoma, and Ganglioneuroma Tissues

For a comparison study of frozen versus deparaffinized specimens using Raman spectroscopy, the frozen specimens were obtained from three patients from the Children's Hospital of Michigan, Detroit, MI.  Patient #1 is a 3 year old infant with adrenocortical tumor present in right adrenal gland.  A normal adrenal gland specimen was obtained from this patient.  The pathology examination verified the absence of any malignant cells in the tissue.  Patient #2 is an 18 month old female infant having neuroblastoma.  The specimen was obtained from the left adrenal gland.  The tumor (12 cm in dimension and weight 0.510 kg) was poor in stroma and was poorly differentiated.  This neuroblastoma specimen has been identified as having a favorable histology, and pathological evaluation shows that this specimen contains nests of tumor cells with centering neuropils.  Patient #3 is a 7 year old child with ganglioneuroma.  The specimen is a spinal mass having differentiating mature

ganglions in a dominant Schwannian stroma environment. The tumor was 3 cm $\times$ 2 cm $\times$ 1.5 cm in size, and clusters of ganglions were observed throughout the specimen.

As explained in Chapter 3, these frozen specimens were thawed for about 20 minutes and then the four frozen sections were cut with a microtome using optimal cutting temperature (OCT) solution. The first cut (5 µm thick section) was stained with Haematoxylin and Eosin (H&E) for pathological examination and the rest (10 µm thick parallel sections) which are adjacent to each other were cut for Raman measurements. Once the tissue sections were obtained, the remaining tissue specimen, keeping the same orientation, was put in formalin for about 24 hours and subsequently preserved in paraffin wax. From this tissue block, four additional parallel sections were cut, one for H&E staining and the rest for Raman spectroscopic investigation. The paraffin was removed from the tissue sections using the standard deparaffinizing procedure. Additionally, we have also studied archived FFPP tissues from nine other patients.

| Cases No. | Patient No. | Pathology | Tissue Source | Raman Spectra |
|---|---|---|---|---|
| 1 | P1 | Normal | Adrenal | 78 |
| 2 | P2 | Normal | Adrenal | 11 |
| 3 | P3 | Normal | Adrenal | 36 |
| 4 | P3 | Ganglioneuroma | Adrenal | 8 |
| 5 | P4 | Ganglioneuroma | Extradural spinal cord mass | 20 |
| 6 | P5 | Ganglioneuroma | Abdominal mass | 40 |
| 7 | P6 | Ganglioneuroma | Retroperitoneal mass | 71 |
| 8 | P7 | Neuroblastoma | Retroperitoneal mass | 37 |
| 9 | P8 | Neuroblastoma | Posterior mediastinal mass | 68 |
| 10 | P9 | Neuroblastoma | Paraspinal mass | 12 |

Table 7.1 The pathologies and the number of Raman spectra obtained from FFPP tissues of nine patients[137]

The details of the pathologies of these patients are given in Table 7.1. Tissue sections were cut from these samples for pathological examination and Raman

spectroscopic investigation as described above.  No frozen specimens were available for these cases.

All the H&E slides were examined by two experienced pathologists who marked the regions of interest on the slides and the images were taken from each pathological condition of the tissues.  We used the pathologically examined/marked H&E slides of the tissue as a guide to find the locations of appropriate pathology in the frozen or deparaffinized tissue sections to collect the Raman spectra.  We assume that the pathological state of the tissue in the 10 μm thick sections is the same as in the adjacent 5 μm thick tissue section used by the pathologists.  We have used three tissues from each state (frozen or deparaffinized) to collect Raman data.

### 7.2.2 Data Collection

For a comparison study of frozen versus deparaffinized specimens, a total of 558 Raman spectra were collected: 205 from normal adrenal gland (103 frozen and 102 deparaffinized), 201 were from neuroblastoma (101 frozen and 100 deparaffinized) and 196 were from ganglioneuroma (102 frozen and 94 deparaffinized).

In the second part of the study, FFPP tissue specimens from nine patients were investigated. The details of the pathologies of these patients and the number of spectra collected are given in Table 7.1.  A total of 381 Raman spectra were collected from the tissue specimens. Each spectrum is an average of three accumulations with 20-second integration time.  The power at the sample was ~20 mW.  The Raman spectra were recorded in the range of 500 - 1900 $cm^{-1}$.  Each spectrum was recorded on a fresh spot of the tissue.

### 7.2.3 Data Processing

Each Raman spectrum was visually examined and any cosmic-ray noise was removed.  The fluorescence background present in the spectrum was removed using a

customized MatLab program which incorporates an adaptive minmax wavelet method.[90] Subsequently, all the Raman spectra were smoothed and their intensities were normalized with respect to the peak of the highest intensity in the spectrum.  To eliminate the influence of the artifacts which can occur during data processing, we excluded a 100 cm$^{-1}$ region of the spectrum at both ends, *i.e.* the region of the spectrum used was from 600 to 1800 cm$^{-1}$, which is commonly known as the "biological window" in Raman spectroscopic studies.

To quantify the spectral differences among the various groups present in the data, we have analyzed the data using principal component analysis (PCA) and discriminant function analysis (DFA), both commonly used chemometric analysis techniques in biomedical research.[93, 94]  PCA is called an "unsupervised" differentiation algorithm and it lowers the dimensionality of the original data and constructs new fewer variables which capture most of the variance in the original data.[93]  It is unsupervised because the identities of the original sets are not provided in the analysis.  These new orthogonal variables basically represent the main independent trends in the original data set.  In our PCA, 591 Raman intensity channels, or variables, separated by $\sim$2 cm$^{-1}$, from 600–1800 cm$^{-1}$ were used as independent variables, and the PCA reduced them to 18 principal components (PCs) capturing approximately 96.7% variance in the data sets.  The PCs were then used as input variables for the DFA, which is supervised, and which expresses the data in terms of a number of discriminant functions (DFs) which is one less the number of groups being classified.[94]

In order to assess the utility of FFPP tissues for diagnostic purpose, the Raman data were collected from the FFPP tissues obtained from nine patients (Table 7.1) and used as training/reference set to test the data obtained from three patients.  For training set, the classification predictions based on Raman spectroscopy measurements with leave-one-out method are cross-validated with the "gold standard" results of pathology.

## 7.3 Results and Discussion

### 7.3.1 Comparison of Frozen and the Corresponding Deparaffinized Tissues

Figure 7.1 shows the mean Raman spectra of frozen and deparaffinized tissue sections, of normal adrenal gland along with the pathology image. Both tissues produce good quality spectra with quite sharp and well defined bands. Clearly, there are noticeable significant differences in the spectra. To highlight the spectral differences, we also show the difference spectrum in Fig. 7.1.
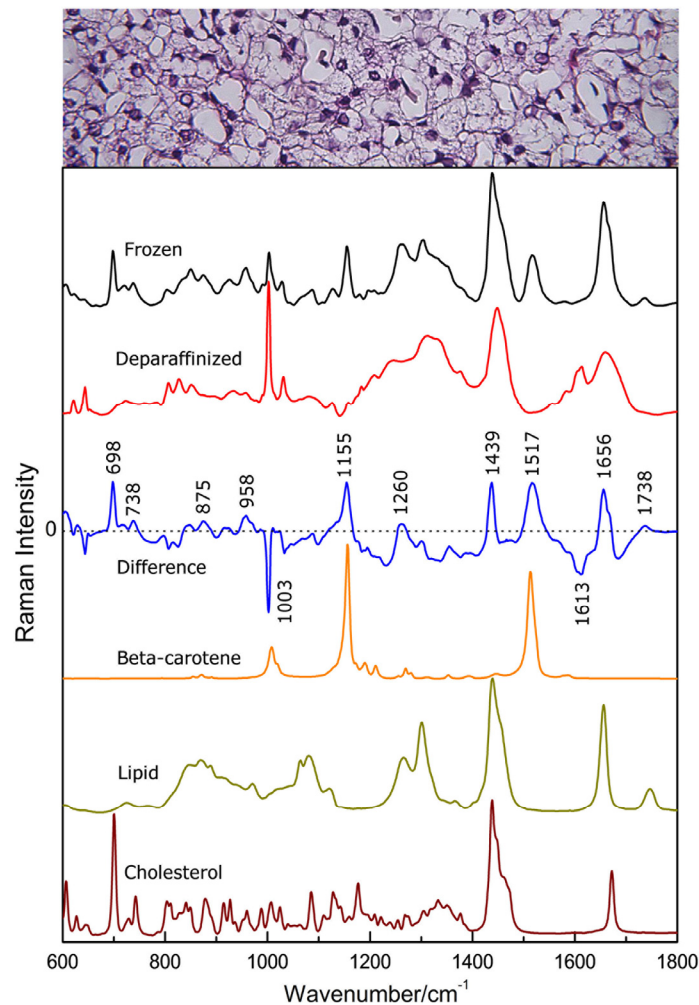


Figure 7.1 The mean Raman spectra of normal adrenal gland: from top to bottom, frozen (black), deparaffinized (red), difference (blue), beta-carotene (orange), lipid (dark yellow), and cholesterol (brown). Pathology image of the normal adrenal gland is taken with 40x magnification[137]

A reduction in the intensity of the bands at 698, 1155, 1439, 1517, 1656, and 1738 $cm^{-1}$ is evident in the deparaffinized tissue compared to the frozen tissue, while the bands around 1003 and 1613 $cm^{-1}$ show enhancement in the intensity.  The Raman bands ~1155 and 1517 $cm^{-1}$ arise from C-C and C=C stretching in carotenoids, which is present in high concentrations in different human organs including the adrenal gland.[138]  The Raman bands at 1656 and 1738 $cm^{-1}$ can be assigned to C=O and C-O stretching in lipids.  In the literature, the band ~1656 $cm^{-1}$ is often assigned to amide I band of proteins.[42]  However, it can also arise from lipids.  Thus, the Raman band at 1650 $cm^{-1}$ is often broad in tissues as it has C=O stretching contributions from proteins, lipids and other molecular groups whose frequencies are slightly different.  The 698 and 1439 $cm^{-1}$ ($CH_2$ bend) bands are two of the strong bands observed in cholesterol and it is generally assigned to "lipids" in the literature, as lipid also has strong band ~1440 $cm^{-1}$.

Figure 7.1 also shows the Raman spectra of β-carotene, fat, and cholesterol.  A comparison of the Raman spectra of frozen and deparaffinized tissue sections of normal adrenal gland clearly indicates that deparaffinizing of the tissue sections with xylene and ethanol removes lipid, cholesterol, and β-carotene.  However, the majority of the spectral features of frozen tissue are preserved in the deparaffinized state.  The other small differences are in the variation of the band intensities and shifts in the frequencies which can be understood in terms of the conformational changes and the intermolecular interactions that occur during the formalin fixing and subsequent deparaffinizing processes. A reaction of proteins with formaldehyde leading to formation of methylene cross-link has been described,[139] and such chemical reactions lead to change in the protein configuration as well as conformation.

Figure 7.2 shows the mean Raman spectra of frozen and deparaffinized tissue sections of neuroblastoma along with the pathology image.  Compared to the normal

adrenal gland (frozen tissue section) the neuroblastoma seems to contain no β-carotene and less lipid/cholesterol.
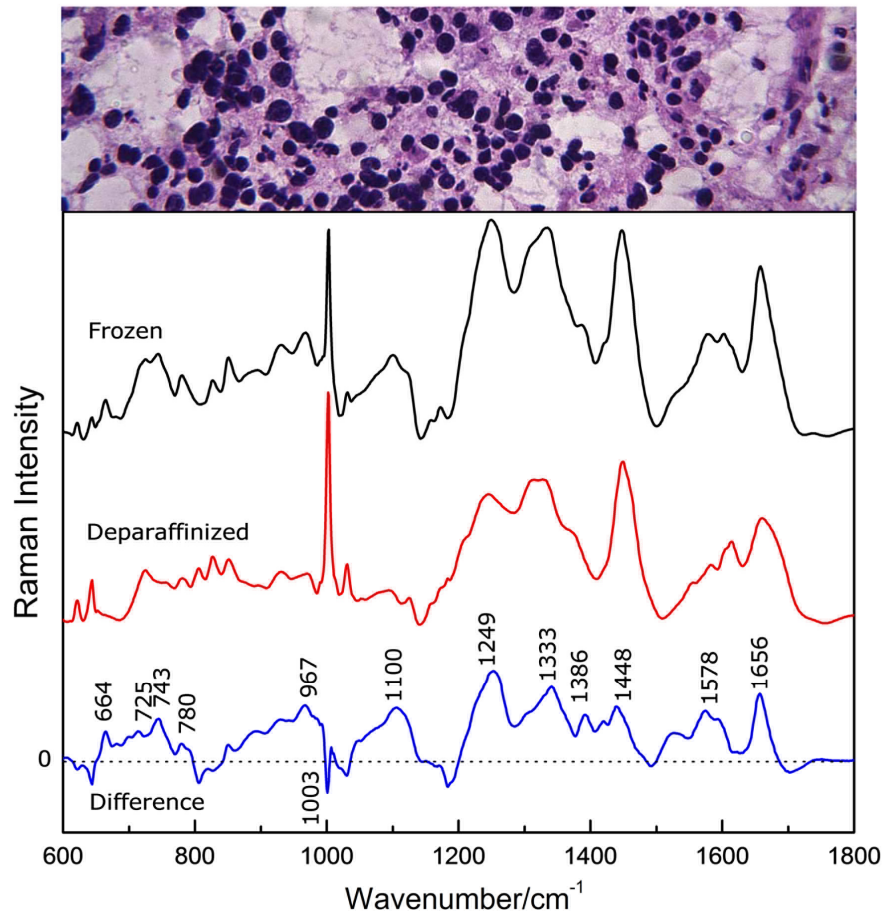


Figure 7.2 The mean Raman spectra of neuroblastoma: frozen (black), deparaffinized (red), and difference (blue)[137]

The intensity of the Raman bands ~1580-1600 cm$^{-1}$ (nucleic acids), 1333 cm$^{-1}$, ~1250 cm$^{-1}$ (amide III/nucleic acids), 1100 cm$^{-1}$ (C-N stretch, nucleic acids), 967, 780, and 743 cm$^{-1}$ (nucleic acids), and ~664 cm$^{-1}$ (C-S vibration of collagen type I or nucleic acids) is higher in neuroblastoma compared to the normal adrenal gland.[140] This indicates that the neuroblastoma cells are rich in nucleic acids which is consistent with the fact that nuclear

and cytoplasmic enlargement can occur during the progression of the tumor[141, 142] in neuroblastic cells.

In this study, the measurements were made from the cellular areas avoiding stromal regions.  The measurements were made specifically from the cellular areas avoiding stromal regions which are rich in collagen. Typical size of neuroblasts being around 10-20 μm (2 to 3 red blood cell diameters),[17] the laser spot is confined to the perimeter of the cell in most of the cases.  A comparison of the Raman spectra of frozen and the corresponding deparaffinized FFPP tissue sections of neuroblastoma indicates that the majority of the spectral features are preserved in the deparaffinized state very similar to that of normal adrenal gland.

The mean Raman spectrum obtained from deparaffinized neuroblastoma (Fig. 7.2) tissue sections is very similar to that of the corresponding frozen tissue but with a somewhat reduced intensity of the overall spectrum.  However, from a visual inspection of the two spectra it is clear that the reduction in the intensity of the bands ~740, ~1250, and ~1580 cm$^{-1}$ is higher than at other frequencies.  These can be attributed to conformational and configurational changes that occur due to the formalin fixation process as described above.
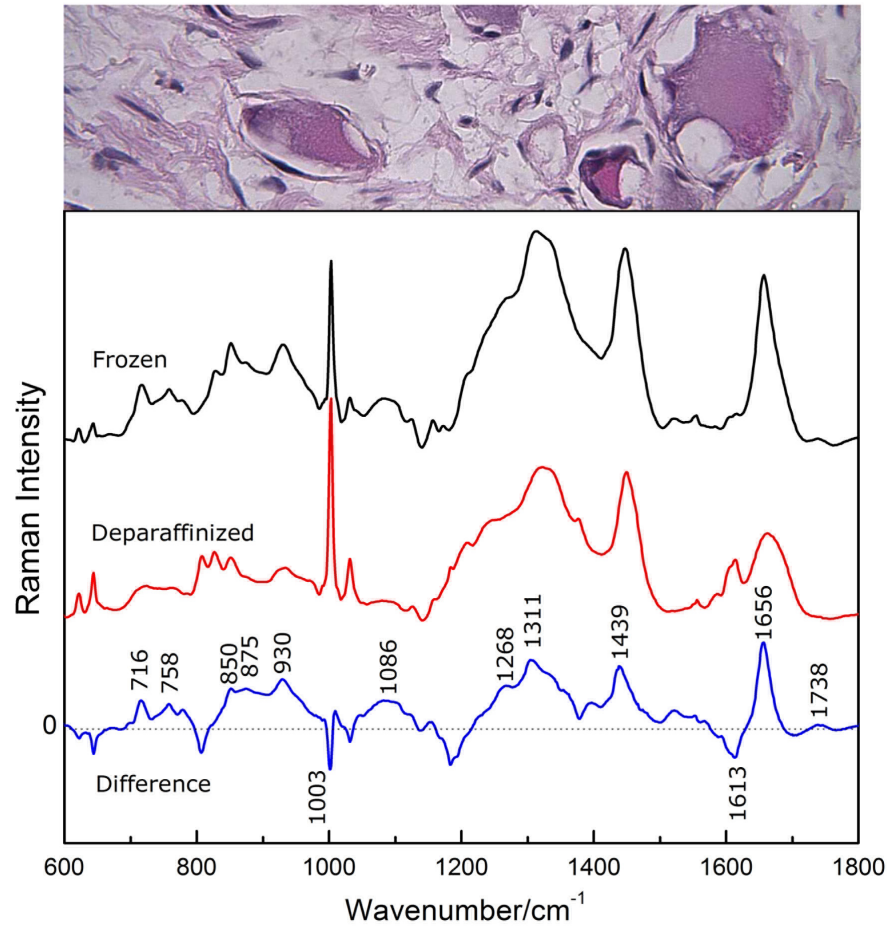
Figure 7.3 The mean Raman spectra of ganglioneuroma: frozen (black), deparaffinized (red), and difference (blue)[137]

Figure 7.3 shows the average Raman spectra of frozen and deparaffinized tissue sections of ganglioneuroma along with the pathology image.  The Raman data were collected from isolated ganglions (much larger than neuroblasts) and thus the laser spot is confined to the perimeter of the cell.  Hence, the Raman bands arise from nucleus and cytoplasm of ganglions. Several Raman spectra from deparaffinized ganglioneuroma tissue sections showed residual paraffin wax peaks although tissues were processed very similar to normal adrenal and neuroblastoma.  This is perhaps due to the intrinsic differences in the molecular interaction with paraffin wax in these tissues.  It should be noted that the Raman spectra with residual wax peaks were not included in the analysis.  A comparison of the

Raman spectra of frozen and the corresponding deparaffinized FFPP tissue sections indicates that the majority of the spectral features are preserved in the deparaffinized state, very similar to neuroblastoma and normal adrenal gland.

We also compared the mean Raman spectra of frozen ganglioneuroma (Fig. 7.3) with that of frozen neuroblastoma (see Fig. 7.2). There are significant differences observed, especially in the spectral region of 1530-1630 $cm^{-1}$ and 1200-1300 $cm^{-1}$. In both of these spectral regions the neuroblastoma has more intensity compared to that of ganglioneuroma. The bands in the 1530-1630 $cm^{-1}$ region can be assigned to amino acid residues such as tryptophan and phenylalanine and the bands in the 1200-1300 $cm^{-1}$ region are assignable to amide III vibrations arising from the protein backbone. Thus, these intensity variations can be interpreted as an indication of the variation in the protein content in these tissues.

### 7.3.2 Chemometric Analysis of Frozen Tissues

To quantify the differences in the Raman spectra of normal adrenal, neuroblastoma, and ganglioneuroma tissues, we have analyzed the data using advanced chemometric methods such as PCA[93] and DFA.[94] Figure 7.4 shows the DF plot of the analysis using the Raman data for the frozen tissue sections. Clearly, the analysis separates the data into three distinct pathological groups representing neuroblastoma, ganglioneuroma, and normal with no overlap. As there are enough significant spectral differences in the spectra of the tissues as discussed above, it is not surprising to see this clear-cut separation. A DF plot of the analysis using the Raman data from the corresponding deparaffinized tissue sections also shows a similar grouping and separation. This is perhaps due to the fact that the each type of tissues in this part of the study came from a single patient and thus lacking any patient to patient variability within a given pathology.
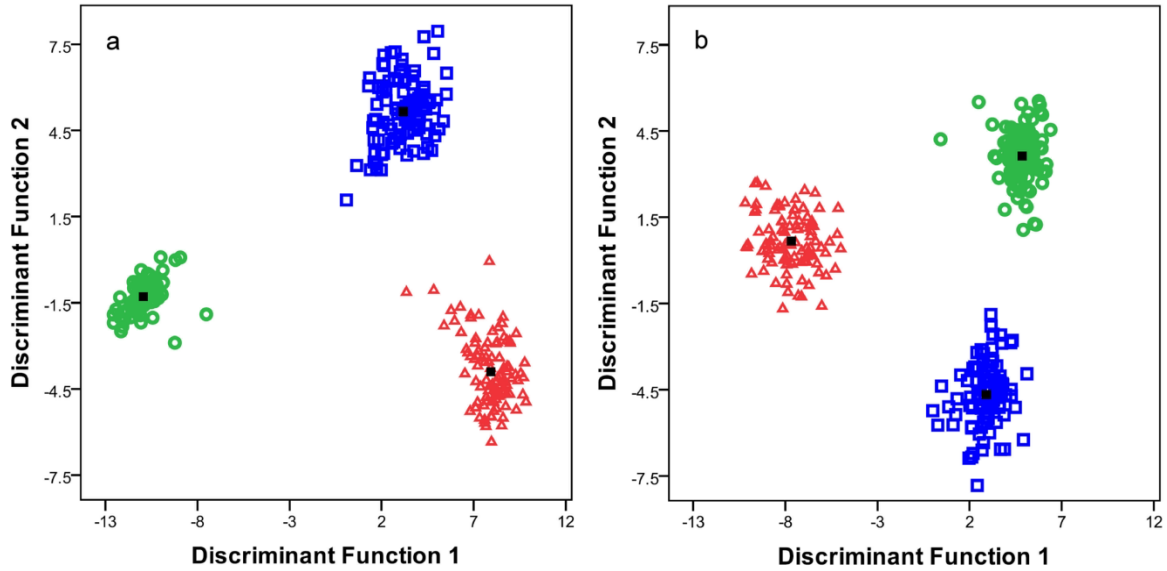
Figure 7.4 DF plots of the frozen tissue data (left) and the deparaffinized tissue data (right): normal adrenal gland (green circles), neuroblastoma (red triangles), and ganglioneuroma (blue squares)[137]

### 7.3.3 Chemometric Analysis of Deparaffinized Tissues

In order to address the variability of Raman spectral differences among patients of a specific pathology we have studied normal adrenal, neuroblastoma, and ganglioneuroma archived tissues from nine different patients (see Table 7.1).  The tissues were available only in the form of FFPP blocks without the corresponding frozen tissues.  Figure 7.5 (a) shows the average Raman spectra of tissues from each distinct pathology as listed in Table 7.1.  It is clear from Fig. 7.5 (a) that the overall shape of the spectra is similar with observable differences in the intensity and broadening of the bands which are quantified by the DF analysis.
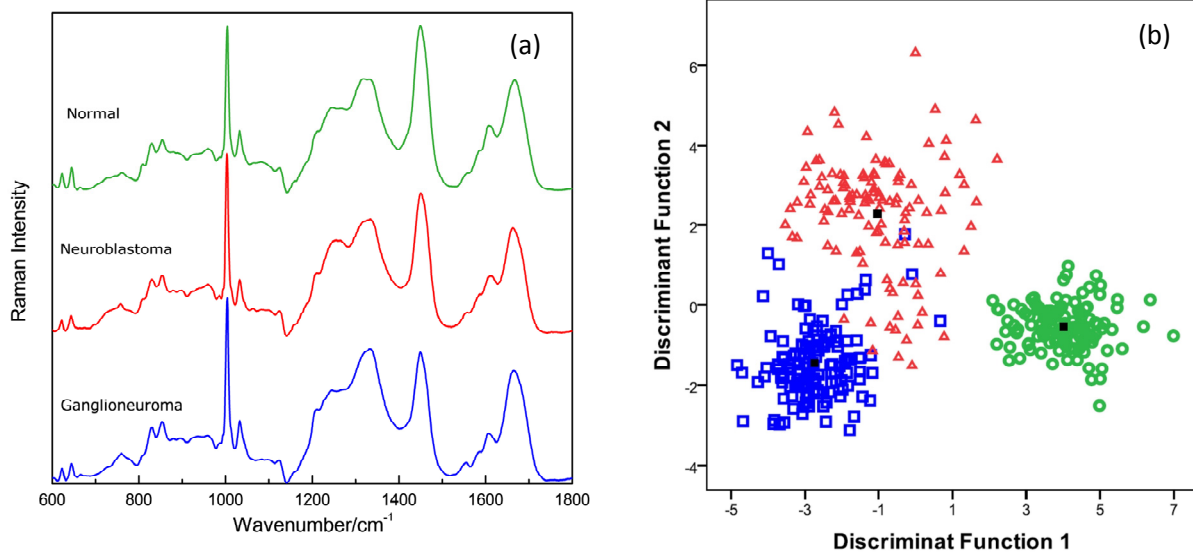
Figure 7.5 (a) The mean Raman spectra of the archived FFPP tissues: normal adrenal gland (green), neuroblastoma (red), and ganglioneuroma (blue) and (b) DF plot of the archived FFPP tissue data[137]

Figure 7.5 (b) shows DF plot of the PCA-DFA analysis using the FFPP tissue data. The chemometric analysis does separate the spectra into three distinct groups belonging to each pathological state of the tissues, although the extent of spread around the group centroids is considerably larger compared to the analysis shown in Fig. 7.4 (b) for 3 patients. This could be attributed to the inter-patient variability in nine patients. The observed overlap between the ganglioneuroma and neuroblastoma can be understood in terms of the similarity of their spectra (Fig. 7.5 (a)) and the presence of some ganglions in neuroblastoma tissues.[17]

According to leave-one out classification method, the accuracies of the training set for normal, neuroblastoma, and ganglioneuroma are 100%, 89.7%, and 95%, respectively. We believe addition of more data to the training set will improve the accuracy of identification. However, the observed overlap between the ganglioneuroma and neuroblastoma may continue to persist because of the presence of some ganglions in neuroblastoma tissues and their histologically similar appearance.

Raman spectral data from the FFPP tissues from the nine patients (Fig. 7.5) were used as the training set in the chemometric analysis to validate the data from the deparaffinized tissues (the test set) obtained from the three patients in the first part of the study. In this external validation, the test data was fed one patient at a time into the statistical analysis of the training set without assigning a particular user defined group. The results are shown in Fig. 7.6.
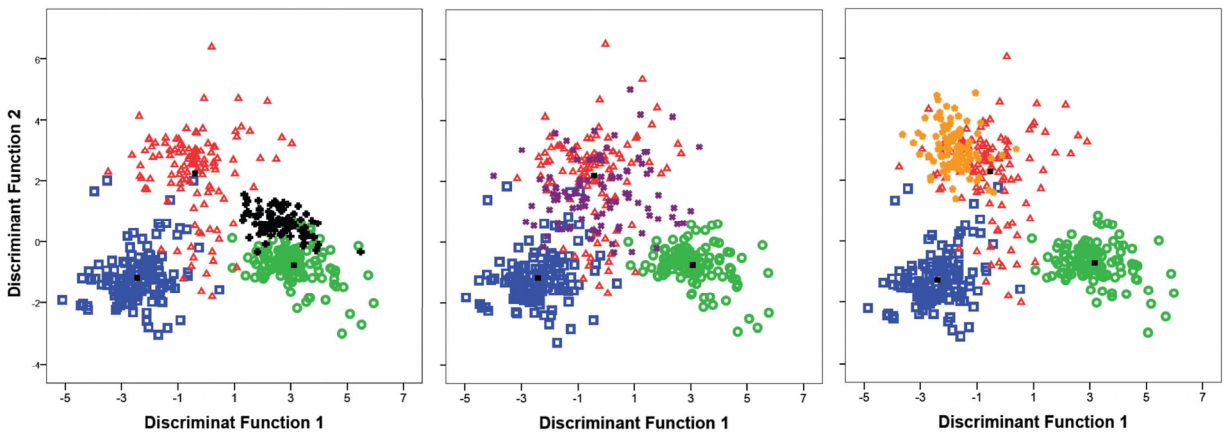


Figure 7.6 DF plots of the archived FFPP training set with three test patients: normal adrenal gland (green circles), neuroblastoma (red triangles), ganglioneuroma (blue squares). From left to right: Patient 1 (black crosses), patient 2 (purple crosses), and patient 3 (orange crosses)

The analysis of normal adrenal test data grouped correctly ~97% of the time with the data from the training set. A similar analysis of neuroblastoma test data correctly identified and grouped with the training set 100% of the time. However, the identification was very poor for the ganglioneuroma with ~10.6%, 8.5%, and 80.9% of the time identifying it as normal, ganglioneuroma, and neuroblastoma, respectively (see Fig. 7.6). However, it is encouraging to note that only a small percentage of the ganglioneuroma data are associated with the normal. The idea behind this external validation is to find the prediction ability of a new patient data. Although, in this study the test data are from only three patients, in an ideal scenario, one can divide the total data set into two and test each

patient of data set 1 against data set 2 and wise versa. This method provides realistic values of sensitivity and the specificity of a particular research data.

As seen in Fig. 7.5, the data from neuroblastoma has much larger spread compared to the normal, and ganglioneuroma. This is, perhaps, the reason for poor identification. Addition of more data to the training set may improve the accuracy of identification. However, the presence of some ganglions in neuroblastoma tissues and their histologically similar appearance may lead to poor categorization in the case of ganglioneuroma. It has been observed that chemotherapy treatment in neuroblastoma also induces ganglionic and stromal maturation and often neuroblastomas appear histologically similar to ganglioneuromas at the time of debulking surgery.[17]

### 7.4 Conclusion

We have investigated neuroblastoma and ganglioneuroma using Raman spectroscopy and compared their spectral characteristics with those of normal adrenal gland. In the first part of the study, the data were collected from frozen and FFPP tissues, obtained from the same tissue block, from three patients. The frozen tissue samples produce a high quality Raman spectra with distinguishable differences compared to the spectra from the corresponding FFPP tissues. The frozen sections of normal adrenal gland tissues show higher levels of carotenoids, lipids, and cholesterol compared to the neuroblastoma and ganglioneuroma tissues. A comparison of Raman spectra of frozen and deparaffinized tissues shows a significant alteration of several biochemicals, including complete removal of carotenoids, lipids, and cholesterol in adrenal tissues. The chemometric analysis of the data from both, the frozen and the deparaffinized, tissues show a clear-cut grouping. Our study shows that the deparaffinized tissues could be utilized for diagnostic research using Raman spectroscopy. However, its prediction accuracy is less than those of the frozen tissues due to removal of some of the diagnostic biomarkers by formalin fixation and subsequent

deparaffinizing processes.  At present, it takes ~4 minutes to record a full Raman spectrum ranging from 500 to 1900 $cm^{-1}$.  However, this time can be reduced by choosing a narrow wavenumber region relevant to the pathology of interest.  As a result, once the database is established, the total detection time required to identify the pathological state of the tissue can be reduced to less than 5 minutes.

In the second part of the study, a Raman spectral data set obtained from deparaffinized tissues from nine different patients was used as a training set against which the data from the deparaffinized tissues from the first part of the study, the test set, was validated using chemometric methods.  The validation process successfully identified and correctly grouped the data with the training set in normal adrenal (>97% of the time) and in neuroblastoma (100% of the time) tissues whereas the validation was not so strong for ganglioneuroma.  This exploratory study indicates that Raman spectroscopy combined with chemometric methods can be successfully used to distinguish neuroblastoma and ganglioneuroma in frozen and deparaffinized tissue sections.  The frozen tissues offer superior diagnostic outcome compared to the deparaffinized tissues because of their high quality spectral features and no alteration and loss of biochemicals which occurs in the process of formalin-fixation, paraffin preservation, and subsequent deparaffinization of tissues for Raman studies.

**CHAPTER 8**

**DIFFUSE REFLECTANCE SPECTROSCOPIC/COLORIMETRIC STUDY OF**

**ACANTHOSIS NIGRICANS**

Diffuse reflectance spectroscopy and colorimetry are non-invasive optical techniques which are beginning to gain acceptance in dermatology. The physics behind both these techniques is the same, although analysis of the data produces different parameters based on different models for each, allowing fundamentally different measurements to be made. One can determine skin color or monitor skin color changes via both diffuse reflectance spectroscopy and colorimetry. In this chapter, the results of our attempts to implement an optical diagnosis of acanthosis nigricans skin condition using both of these optical techniques are discussed.

## 8.1 Why Acanthosis Nigricans?

There are many people who suffer from hyperinsulinemia and hypertension and they are prone to type II diabetes. Acanthosis nigricans (AN), skin thickening and darkening in various parts of the body, is common among these patients. If untreated, 90% of the people who have acanthosis nigricans will eventually get type II diabetes over the next 10 years of their lifetime. There are no optical instruments which can diagnose and quantify this skin disease. Thus, it is imperative to develop new technologies or even to use existing optical instruments to develop the capability of detecting AN optically. These techniques/instruments should be non-invasive as they will be used for the real time screening of patients or children, producing point-of-care diagnoses without any time-consuming laboratory evaluations. This research work is the beginning of such an attempt to investigate whether diffuse reflectance spectroscopy (DRS) or colorimetry or both together can be utilized for optical screening of acanthosis nigricans.

Traditionally, dermatologists have measured skin color through visual observations, but there is an inherent subjectivity to this type of measurement, as the human eye cannot quantify small changes in shades of color. This can be overcome through an optical diagnostic. A diffuse reflectance spectrophotometer or colorimeter provides a modality for characterizing the biophysical basis of color on light absorption and the surface properties of the skin. In DRS, light is delivered onto and into the skin and the reflected light is collected and analyzed with a broadband spectrometer. The reflected light spectrum is used to calculate an apparent absorption spectrum for the skin, which can then be used to calculate the concentration of known skin chromophores, specifically melanin, oxyhemoglobin, and deoxyhemoglobin.[56] These calculations can be very model-dependent and are also highly dependent on the spectral distribution of the excitation source, as in colorimetry. In colorimetry, a white light excitation source illuminates the skin and the reflected light is parameterized by the L*a*b* color system as determined by the Commission International d'Eclairage (CIE) in 1976.[143] Here, L* is the lightness to darkness component and a* and b* are the color components of red (positive) to green (negative) and yellow (positive) to blue (negative), respectively. Thus, colorimetry can be used for complex color analysis with high precision and can accurately determine the spectral reflectance at each wavelength.[87]

Neither colorimetry nor DRS has been fully explored as a way to study the hyperpigmentation of acanthosis nigricans. Few studies have directly compared these two techniques, and none on a single patient pool. Our objective is to evaluate the reliability of colorimetry in quantifying color change and DRS in characterizing the chromophore changes associated with AN. The significance of this lies in the possibility that the degree of a quantifiable color change may be related to the severity of disease states such as hyperinsulinemia when assessed in future studies.

**8.2 Materials and Methods**

The DRS light source (LS-1 Series Tungsten Halogen Light Source HL-2000, Mikropack and Ocean Optics, Inc., USA) emits light in the visible to near-infrared range (360 nm - 2000 nm). This source is connected to one end of a bifurcated fiber optic probe (measurement/illumination diameter of 2.5 mm), which also collects reflected light into a spectrophotometer (BWTEK, Inc., USA). The spectrophotometer measures diffuse reflectance spectra in the range of 350-850 nm. These reflectance/absorbance spectra are used to calculate melanin, deoxyhemoglobin, and oxyhemoglobin chromophore concentrations using the known *in vivo* absorbance characteristics of these complex biomolecules.[78]

The colorimeter instrument (Minolta CM-2600d, Osaka, Japan) consists of an 8 mm diameter measurement/illumination aperture. A pulsed xenon lamp emits an intense white light covering the entire visible spectrum which illuminates the specimen under the aperture uniformly. The colorimeter measures L*, a*, and b* values of the reflected light as defined by the standard CIE system.[87] Figure 8.1 shows the L*a*b* color space.
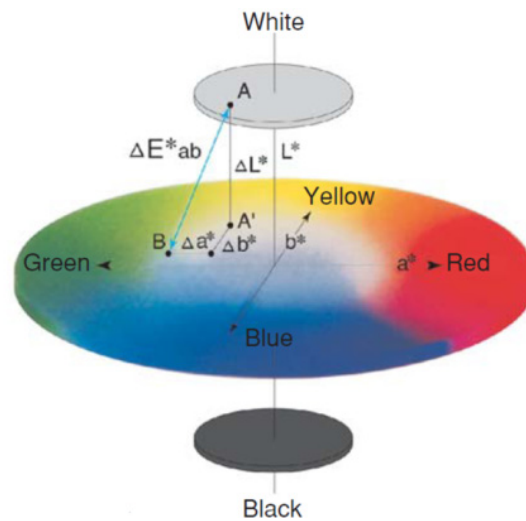


Figure 8.1 L*a*b* color space[87]

In order to calculate L*, a*, and b* values, first, tristimulus values need to be determined. The tristimulus values, XYZ can be determined through the convolution of spectral power distribution of the illuminant (stored in the instrument memory), spectral reflectance of the specimen and the color matching functions ($\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$) as shown in Fig. 8.2.[87]
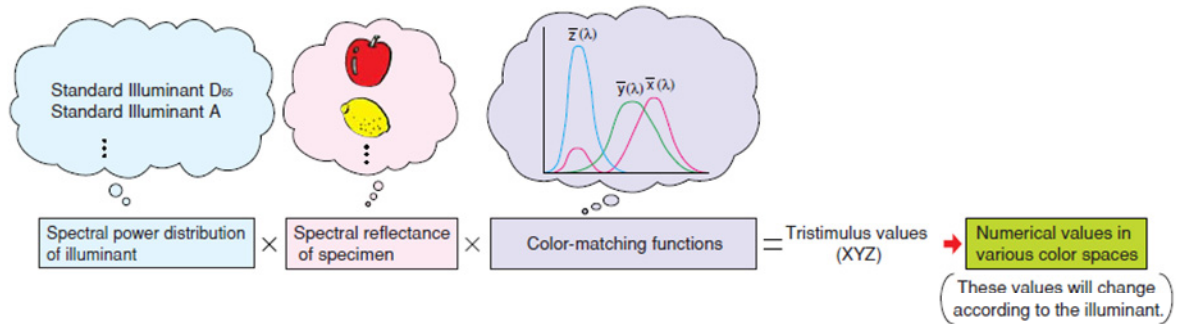


Figure 8.2 Steps to obtain L*, a*, and b* parameters[87]

Once XYX values are obtained the L*, a*, and b* can be calculated using the following equations,

$$L* = 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16$$

$$a* = 500\left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right] \text{-------------------------------------------------(8.1)}$$

$$b* = 200\left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right]$$

Here, $X_n$, $Y_n$, and $Z_n$ are the tristimulus values of a perfect reflecting diffuser.[87]

While the colorimetry signal is analyzed purely in terms of the intensity of the various color components, the DRS data-analysis software modeled the physics of the light-skin interaction to calculate biochemical concentrations from the reflected spectrum.

Analysis of the DRS data provides a quantification for the contributing skin chromophores, including melanin, oxyhemoglobin, and deoxyhemoglobin. The subsection 2.3 in Chapter 2 describes how to extract these chromophore concentrations from the spectral information of the DRS.

Prior to every patient measurement, both instruments were corrected for detector dark current and were calibrated using manufacturer-provided white and black tile standards.

### 8.2.1 Patient Enrollment

Eight female African American patients, ages 12 years and older with AN as diagnosed by a dermatologist were enrolled. It is a necessary requirement that all the patients were in an insulin resistant condition, meaning elevated fasting insulin levels with normal fasting glucose levels. During the course of this research, all patients were consulted by an endocrinologist for treatment of their hyperinsulinemia with diet control, weight reduction and/or oral metformin. Patients' consents were obtained through an informed consent form and all study-related procedures were explained before enrolling. Exclusion criteria for patient enrolment was patients with type 1 diabetes, women who were lactating, pregnant, or planning to become pregnant, and those who were hypersensitive to oral metformin medication.

A skin prosthesis provided by Henry Ford Hospital was used to assess the repeatability of the measurements. In addition, we also obtained measurements from nine healthy volunteers one time to evaluate and compare the DRS and colorimetric measurements of the healthy control group and the AN patients.

### 8.2.2 Data Collection

The pigmented areas were measured from the areas within the AN neck lesion on the posterior neck and the lateral neck. The inner surface of the forearm and the skin over the clavicle (normal neck) were treated as normal controls. The probes of both devices were placed perpendicularly on the skin with light pressure to avoid the skin blanching effect. Due to differences in the diameter of the probe sizes, the colorimeter obtained 3 measurements at 1 spot, and the DRS obtained 10 measurements at 10 different spots. The colorimeter probe is approximately 10 times larger in area than the DRS probe. Photographs of the measurement areas of AN and normal controls were marked so that subsequent measurements were taken at approximately the same location. These photographs were taken every visit to observe and document clinical changes. All measurements were taken in an exam room at a dermatology outpatient clinic at Henry Ford Hospital. Follow-up periods were conducted for 8 consecutive months.

### 8.2.3 Data Analysis Methods

In the first part of the study, we analyzed the eight month data using standard analytical methods such as mean, standard deviation, Pearson correlation, etc. and in the second part of the study we used chemometric methods, principal component analysis (PCA)[93] and discriminant function analysis (DFA)[94] to compare the measurements of both the DRS and colorimetry instruments.

#### 8.2.3.1 Standard Analyses

We calculated the mean and standard deviation values of the 6 model-dependent parameters (L*, a*, b*, melanin, oxyhemoglobin, and deoxyhemoglobin) of each skin site for each patient over the eight month period. Also, we determined the differences of the AN lesion-control sites (normal neck or forearm) for each individual parameter. In determining

a scatter of measurements for all the patients, we wanted to utilize a single parameter to interpret the data. This is because some of the patients exhibited more variance of the monthly measurements from the average than the others. Hence, characterizing the measurement scatter of all the patients around the mean value with a single parameter assists in determining the sensitivity of the measurement. We call this single parameter a "global standard deviation". Global standard deviations for the three skin measurements (AN lesion - normal neck, normal neck - forearm, and AN lesion – forearm) were calculated by measuring the standard deviation of the individual measurements from the average values for all patients, and then averaging those standard deviations to obtain one global standard deviation. In addition, we used the Pearson correlation method to evaluate the correlation between the six parameters (see Table A2). We also calculated the percent difference in these parameters between AN lesion and normal skin as described below:

$$\text{percent difference} = \frac{\text{lesion value - normal value}}{\text{normal value}} \times 100\% \text{ -----------------------------(8.2)}$$

Changes in these calculated percent differences between visits allow an evaluation of absolute changes in skin color.


### 8.2.3.2 Chemometric Approach

It is important to apply a chemometric approach to analyze the spectral data obtained from these AN patients by colorimetry and DRS to evaluate whether a reliable classification of skin as "lesion" skin or "healthy" skin can be made more reliably. A chemometric model which utilizes more of the data at one time compared to a straight-forward calculation of a skin chromophore concentration or the use of individual color parameters (e.g. melanin concentration or L* value related to skin darkness). Another important aspect of the chemometric analysis of the DRS data is its use of only the raw diffuse reflectance (or absorbance) spectra instead of a reliance on calculated model-

dependent melanin, oxyhemoglobin, and deoxyhemoglobin chromophore concentrations to differentiate normal healthy skin from AN lesion. The scattering of light from a medium as heterogeneous as skin, is a complex phenomenon. Thus, any model of this scattering is at best an approximation. Moreover, there were also limitations with using this DRS system for measurement of AN lesions since the original algorithm was designed to evaluate smooth normal skin. The use of the entire raw spectrum eliminates this need for an approximate model and the calculation of any chromophore concentration.

As a result, colorimetry data consisting of the L*, a*, and b* values and the raw absorption spectra of skin were analyzed separately using two chemometric methods known as principal component analysis and discriminant function analysis.[93, 94]

Principal component analysis and discriminant function analysis are statistical methods which have already been described in detail in Chapter 4. In the PCA of the DRS data, the entire broadband DRS spectrum (226 channels spanning 450 nm with a 2 nm channel width) was used as an input data set. The PCA reduced the 226 variables to 4 variables which captured 99.7% of the variance in the data. These new variables were then used as the input to the DFA. The colorimetry data, consisting of only 3 parameters or variables (L*, a*, and b*) were not analyzed with a PCA. The raw spectra were not obtainable from the CM-2600d instrument.

The DFA method classifies the data into the independent groups present in the data by minimizing the variations within the groups and maximizing the variations between the groups.[94] The DFA constructs N-1 discriminant functions (DFs) for discrimination amongst N user-defined groups. Each spectrum can then be classified by N-1 DF scores. The DF scores were then used to allow a classification of an unknown sample as "lesion" or "healthy". Typically each data set possessed only two DF scores, as we were attempting to differentiate between only three user-defined groups: control (forearm), control (neck), and AN lesion. Although median and lateral AN lesion measurements were obtained, a careful

analysis showed no difference in these measurements and no utility in separating these measurements. Therefore, all these lesion measurements were combined in the analysis.

To lessen the between-patient scatter of the measurements due to the inherent differences in skin coloring and composition of the patients, prior to analysis both colorimetry data and DRS data were normalized for every patient. This was done by first calculating the average of the patient's forearm control measurements. The scatter of the forearm control measurements about this average allowed us to characterize the anticipated scatter of the lesion measurements which was expected to be even greater due to the heterogeneity of the lesion tissue. All subsequent measurements were then divided by that patient's average forearm control to insure that data obtained from the neck control or neck lesion tissue were really differences from that patient's normal skin coloring or composition.

## 8.3 Results: Standard Analyses

### 8.3.1 Reproducibility of Measurements

The reproducibility of the colorimeter and the DRS instrument was measured using a skin prosthesis obtained from the Orthopedic Department of Henry Ford Hospital twice daily for 5 days. The percent deviations obtained (pooled standard deviation divided by mean) of the prosthesis and the forearm of patients are shown in Table 8.1.

| Patient# | L* | a* | b* | Melanin | Oxyhemoglobin | Deoxyhemoglobin |
|----------|------|------|------|--------|---------------|-----------------|
| Prosthesis | 0.20% | 0.12% | 0.18% | 10.01% | 19.31% | 20.87% |
| 1 | 0.40% | 1.54% | 0.58% | 2.47% | 23.07% | 11.44% |
| 2 | 0.94% | 1.62% | 0.73% | 2.42% | 20.40% | 8.81% |
| 3 | 0.80% | 1.66% | 1.10% | 2.40% | 32.47% | 8.44% |
| 4 | 0.78% | 1.56% | 0.46% | 2.58% | 26.85% | 5.39% |
| 5 | 0.90% | 2.73% | 1.67% | 2.90% | 33.49% | 10.46% |
| 6 | 0.61% | 1.00% | 0.68% | 1.69% | 37.87% | 9.97% |
| 7 | 0.52% | 1.36% | 0.66% | 2.78% | 36.50% | 11.33% |
| 8 | 1.15% | 0.98% | 2.10% | 2.15% | 36.22% | 9.74% |

Table 8.1 Percent deviations for prosthesis and patients

The colorimeter parameters, L*, a*, and b* values showed a reproducible results, deviating by less than 1% in the prosthesis. Most of the patients also indicated low deviations close to 1%. In contrast, the DRS instrument exhibited significant variance in the values calculated for the chromophores. The melanin percent deviation of the prosthesis is about 10%, whereas oxyhemoglobin and deoxyhemoglobin are approximately 20%. Surprisingly, the DRS measurements on patient skin showed a very low deviation (<3%) for melanin. Deoxyhemoglobin deviation was in the range of approximately 5% to 10%, while oxyhemoglobin deviation ranged from approximately 20% to as high as 38%.

Although it may seem intuitive that an unchanging substrate such as the prosthesis would give reproducible values for melanin concentration, it appears that the absence of the actual chromophores leads to variable results. Also the skin prosthesis is not a homogeneous calibrator. Melanin values on patient skin showed a very good reproducibility, of about 3%. The hemoglobin values showed significant variation in the prosthesis as well as all patients. This may be due to the physiologic effect of blood vessels affecting the hemoglobin concentration. Therefore, the hemoglobin concentration was not a good parameter to monitor the changes in skin color, and as such has little value as diagnostic parameter for use in measuring AN skin.

### 8.3.2 Changes in Pigmentation

In order to determine the changes in the pigmented areas month-to-month basis, the posterior neck lesion measurements were compared to the control data. Figure 8.3 shows both colorimetry and DRS parameter difference of AN lesion and forearm skin (black squares) and also the difference between normal neck and forearm (red stars) for a single patient (Patient 1).
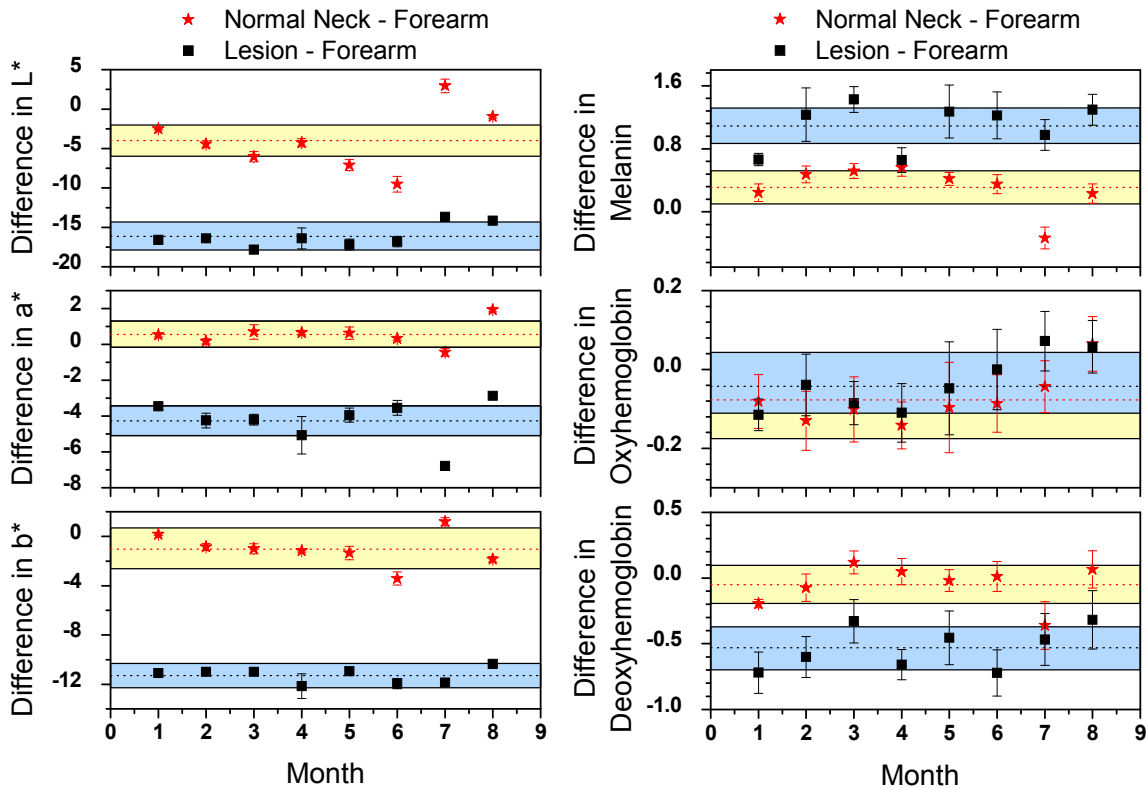
Figure 8.3 Patient 1; comparison between normal neck-normal forearm and AN lesion skin-normal forearm[144]

It is noted that almost all of the measurements fall within one standard deviation of the average and show no discernible trend with time, a result consistent with no change in value over the course of 8 months. The difference of normal neck and forearm values, for all the parameters is relatively close to zero. Thus, normal neck and forearm values are very similar. As a result, the comparison of the difference between normal neck skin and AN lesion skin is plotted in Fig. 8.4. This skin difference is much more realistic for diagnostic purposes as the characteristics of forearm skin may be different from a normal neck area.
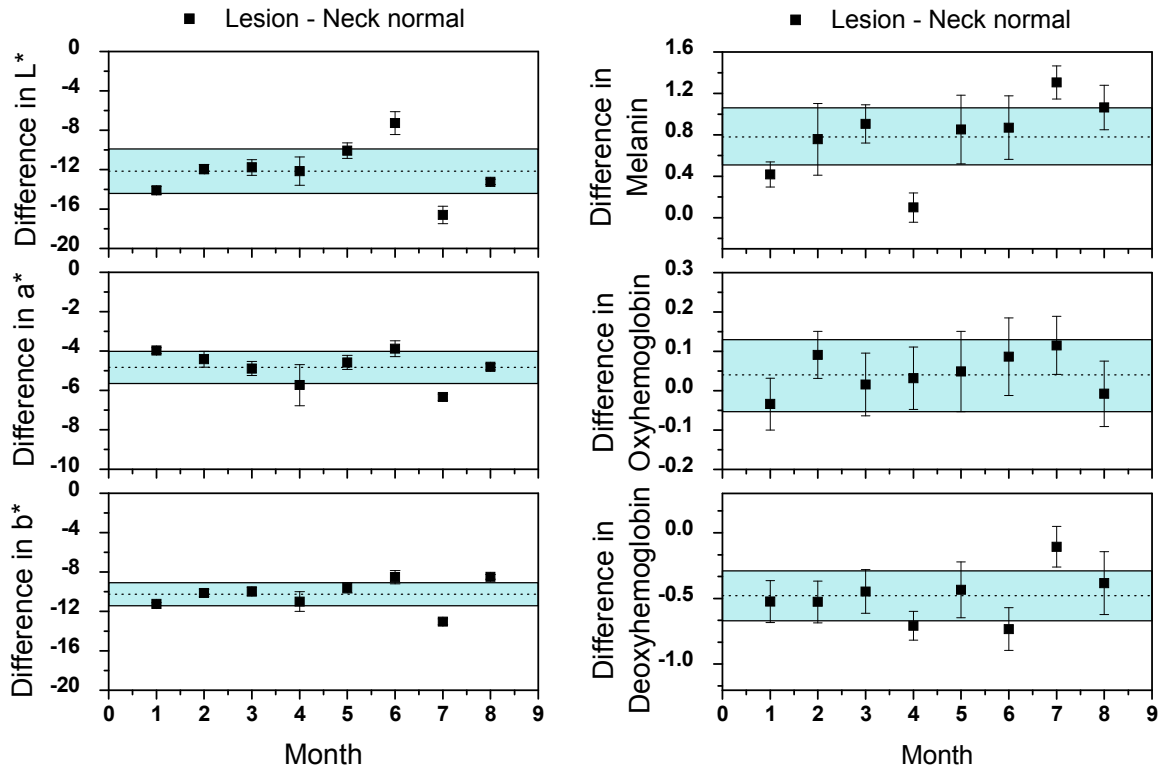
Figure 8.4 Patient 1; comparison between AN lesion skin-normal neck

### 8.3.3 Scatter of the Measurements

In Figs. 8.3 and 8.4, the averaged values show as dashed horizontal lines. This serves as scatter of the measurements in the instruments. Global standard deviations for calculated melanin, oxyhemoglobin, and deoxyhemoglobin of AN lesion-normal neck measurements were 0.28, 0.09, and 0.19, respectively. The global standard deviations for L*, a*, and b* were 2.26, 0.81, and 1.17, respectively. In Fig. 8.4, the shaded regions show the global standard deviations about their mean and represent the spread of this patient's AN lesion–neck normal measurements over the 8-month period. Similarly, shaded regions in Fig. 8.3 show the global standard deviations of normal neck-forearm and AN lesion-forearm for all the six parameters.

As can be seen in Fig. 8.4, the ΔL* values for the lesion tissue relative to normal neck tissue were -12.1. This is approximately 5.4σ from the normal neck average. All the

eight patients showed differences between normal neck and lesion greater than 3σ. In fact, b* values also showed more than a 3σ difference between normal neck and AN lesion for all the 8 patients. However, a* varied from 0.14σ to 8σ and did not show a consistent difference.

Among DRS parameters, only melanin, showed a potential to be an effective discriminator. The average AN lesion-normal neck skin value, 0.78, a difference of approximately 2.8σ, which is above a 95% confidence level. With the exception of one patient (Patient 8), all patients showed more than a 2σ statistical difference for melanin concentrations measured in AN lesion skin compared to normal neck skin. Patient 8 also showed a higher concentration of melanin for lesion skin, but it was not statistically significant enough to reliably differentiate AN. Oxyhemoglobin rarely showed any statistically significant difference between AN lesion tissue and healthy neck tissue (see Figs. 8.3 and 8.4). Deoxyhemoglobin showed a difference in some patients (including patient 1), but not all patients.

### 8.3.4 Evaluation of Lesion Visit-wise

The absolute difference of AN lesion and normal skin were determined through L*, a*, b*, melanin, oxyhemoglobin, and deoxyhemoglobin values obtained from these two instruments (e.g. $\Delta L^* =$ L*(lesion) – L*(normal)). The percent difference were calculated in these parameters between visits, but none of the patients showed any statistically significant change between visits in the percent differences of the six parameters during the eight month testing period. This was also confirmed by periodic insulin level measurements which remained relatively constant for all the patients. In addition, dermatological diagnosis of AN remained unchanged.

Figure 8.5 Dermatological images of the 8 visits of a patient

For a single patient, photographs taken for eight visits were shown in Fig. 8.5. Thus, spectroscopic measurements, insulin measurements, and dermatological observations all confirmed no change in AN condition during this eight month period.

### 8.3.5 Healthy Volunteers' Measurements

Similar to AN patients' measurements, DRS and colorimetry measurements were obtained from the same skin areas including posterior neck from nine healthy volunteers. These measurements were made only once for all the healthy volunteers, but used the same protocols as on the patients, specifically with reference to the number of measurements and sampling locations.
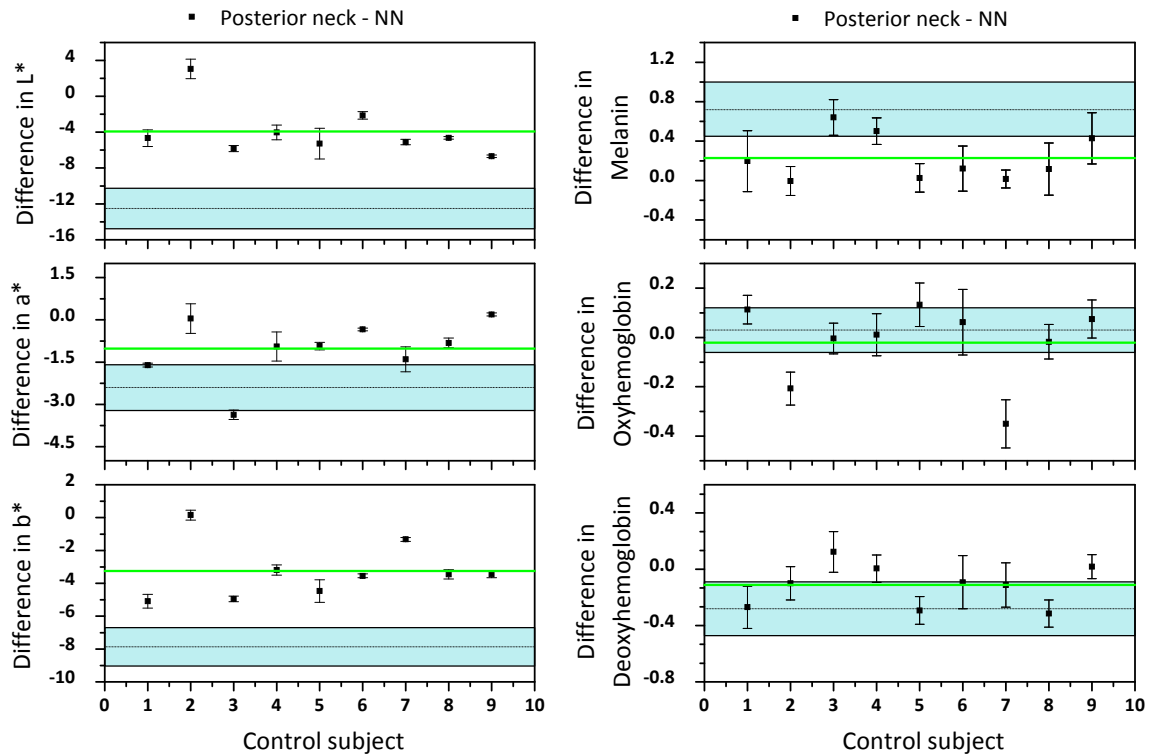
Figure 8.6 Comparison of colorimetry and DRS measurements made on control subjects and diagnosed AN patients. (left) Difference of the colorimetry data (L*, a*, b*) obtained from the posterior neck and normal neck measurements for both control subjects and patients. (right) Difference of the DRS-calculated concentrations (melanin, oxyhemoglobin, and deoxyhemoglobin) obtained from the posterior neck and normal neck measurements for both control subjects and patients. The data points with error bars represent the control subjects and the blue shaded areas represent the patients' standard deviation around the mean

Figure 8.6 compares the difference between the measurements made at the potential lesion site and the neck normal site for patients and controls. Figure 8.6 shows each of the three parameters obtained from DRS and colorimetry showing the difference of posterior neck and normal neck for a total of six graphs. Each graph shows the average value of the patient data (obtained by averaging the value from all 8 patients) with 1σ (blue band around the average dashed line). The uncertainties of individual volunteers for each parameter are also shown in each graph. L*, a*, and b* show a difference of -3.92, -1.02, and -3.25, respectively while DRS parameters, melanin, oxyhemoglobin, and deoxyhemoglobin show 0.22, -0.02, and -0.12 respective differences. The difference

between the control average and patient average for each of the six parameters is also calculated, and expressed in units of standard deviation. These results are shown in Table 8.2, where there is about a 4σ difference in L* and b* values between the average of the healthy volunteers and the patients. Note that a zero difference is interpreted as "healthy". As the control subjects' difference (healthy posterior neck-NN) is rather close to the zero, this gives us confidence that the patients are showing statistically significant differences in AN lesion compared to the control data.

| L* | a* | b* | Melanin | Oxy | Deoxy |
|-----|-----|-----|---------|-----|-------|
| 3.8 | 1.8 | 3.9 | 1.8 | 0.5 | 1.4 |

Table 8.2 Difference in patient average and control subject average for each of the six measured parameters expressed in units of sigma (the standard deviation of the patient measurements)

Note that, if two values are within 1σ of each other there is a 32% possibility that this happened by chance. If two values differ by 2σ there is only about a 5% possibility that this is by chance, and if two values differ by 3σ there is only a 0.3% possibility that it occurs by chance. Therefore, two values that differ by 3σ or more are almost certainly statistically different. L* and b* seem to satisfy this criterion. It can also be seen that none of the control measurements fall within the uncertainly of the patient average, as expected. For the other four parameters, the differences are all less than 2σ, and thus more uncertain. This can be seen in the graphs, where control subject measurements frequently fall into the blue uncertainty band of the patient data.

## 8.4 Discussion: Standard Analyses

In this research work, the reproducibility of the colorimeter and DRS was examined on skin prosthesis and on normal skin of patients with AN. Colorimetry showed a high reproducibility in both prosthesis and *in vivo* skin measurements, with variation below or

near 1% in most cases. The slightly higher variation for *in vivo* measurements can be explained by small month-to-month variations present in normal skin. We can conclude from these measurements that colorimetry values for L* and b* may be reliable modalities to objectively quantify skin color in pigmented areas of skin, while a* is a weak indicator of skin color, and thus not reliable.

When determining the difference between AN lesion skin and normal skin, it was important to use normal neck skin as the control, and not forearm skin. As can be seen in Fig. 8.3, neck skin varied on a month-to-month basis when compared to forearm skin in AN lesion as well as normal neck. This variation may be caused by sun exposure or other factors. This month-to-month change is less apparent in Fig. 8.4, which compares AN lesion skin to normal neck skin.

The scatter of these instruments was determined to distinguish the intensity of the pigmented skin caused by AN from the intensity of the pigment from normal skin. L* and b* color parameters obtained from the AN lesion were shifted by more than a 3σ difference from correspondingly healthy neck tissue for all the patients, which corresponds to less than 0.27% chance that the measurement differed by chance alone. These parameters are therefore sensitive enough to differentiate AN from normal pigmented skin color. However, no parameters obtained from DRS showed sufficient sensitivity to diagnose AN for all the patients. Melanin concentration in the AN lesion showed a significant statistical difference of more than 2σ from the correspondingly healthy neck in all the patients except Patient 8. None of the patients showed a greater than 3σ difference in oxyhemoglobin or deoxyhemoglobin concentration.

By performing a Pearson correlation analysis on the data measured by both instruments, we found a strong negative correlation between L* and melanin values (see Table A2). This means when L* increases, melanin should decrease. We also found that L* and b* decrease while melanin increases in AN compared to normal skin. In addition, the

results of the percent difference from normal for the first four months of the trial are shown in Table A3. We found that L*, b*, and melanin values measured in AN lesions were different from normal skin with statistical significance, as shown by their very low p value (P-value <0.001). This p-value analysis implied that L*, b*, and melanin could be good parameters to objectively quantify skin color in AN lesions, as L* and b* values tended to be lower while melanin values were higher in lesion tissues. Data in subsequent months was similar. Taken together, this suggests that the hyperpigmentation of AN may have a higher melanin content than the normal pigmented skin. This study supports the conclusion that the brown color of AN is caused by hyperkeratosis of the epidermis as well as increased melanin content.[145, 146, 147]

L* and b* seem to be consistent measures of skin color in AN lesions, as the measurements obtained from patients are very different, statistically, than the measurements obtained from the healthy volunteers. Although a* and melanin values are indeed indicative of a change in AN patients (relative to healthy controls), they do not appear to be reliable. Further experiments need to be carried out to draw any conclusions. The oxyhemoglobin and deoxyhemoglobin values do not show a significant difference between patient and healthy volunteers' data, therefore we would conclude, that these values are unreliable for use as an objective color measurement of AN.

A significant limitation of the DRS instrument is that the size of the measurement probe is being small. Therefore, DRS is ideal for skin lesions which are homogeneous or smaller in size where the illuminated spot size needs to be smaller than the integrating sphere aperture of the colorimeter.[76] Another limitation to this study was the fact that AN, as well as other skin conditions, possesses a velvety, rough texture, which causes significant light scattering at the skin surface. A larger DRS probe would be able to capture the light that is reflected from these relatively large surface features.

## 8.5 Results and Discussion: Chemometric Approach

The mean DRS absorbance spectra of the three measurement sites for one patient are shown in Fig. 8.7. Absorbance is a unitless quantity obtained from the diffuse reflectance spectrum. Each of the three spectra shown in Fig. 8.7 is the average of the all the measurements made at that particular site. Regions of large absorbance indicate strong absorption or scattering of the light at that wavelength, sometimes indicative of the presence of a specific chromophore.
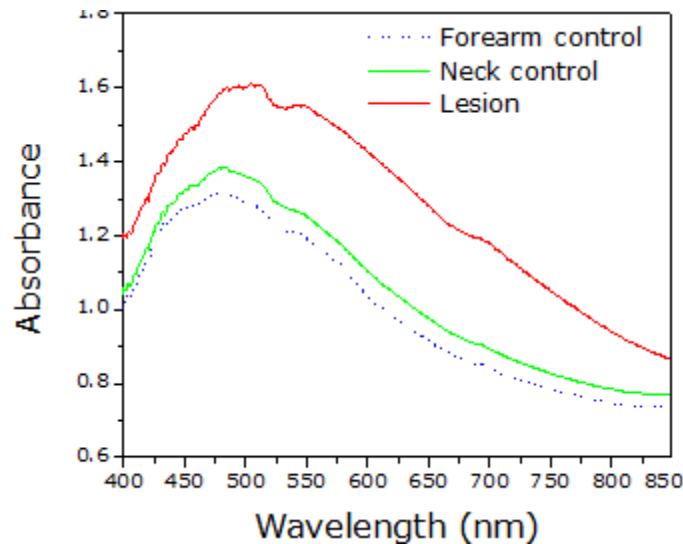


Figure 8.7 Mean absorbance spectra of forearm control, neck control, and lesion. Lesion tissue obviously demonstrates significantly greater absorption/scattering[91]

In Fig. 8.7, a clear increase in the absorbance of the lesion can be seen compared to the normal controls (forearm and neck). This is indicative of an increase in the concentration of the chromophores melanin, oxyhemoglobin, and deoxyhemoglobin, as was shown in our previous analysis of these patients. Aside from this overall increase in absorbance, statistically meaningful and reproducible spectral differences are hard to quantify in the three spectra shown in Fig. 8.7.

Chemometric techniques provided a more reliable way to obtain quantitative classification of the differences in the spectra. Figure 8.8 shows the first two discriminant function scores of the DFA performed on (a) the normalized DRS data (subject to PCA first, as explained above) and (b) the colorimetry data obtained from all eight patients spanning all patient visits.
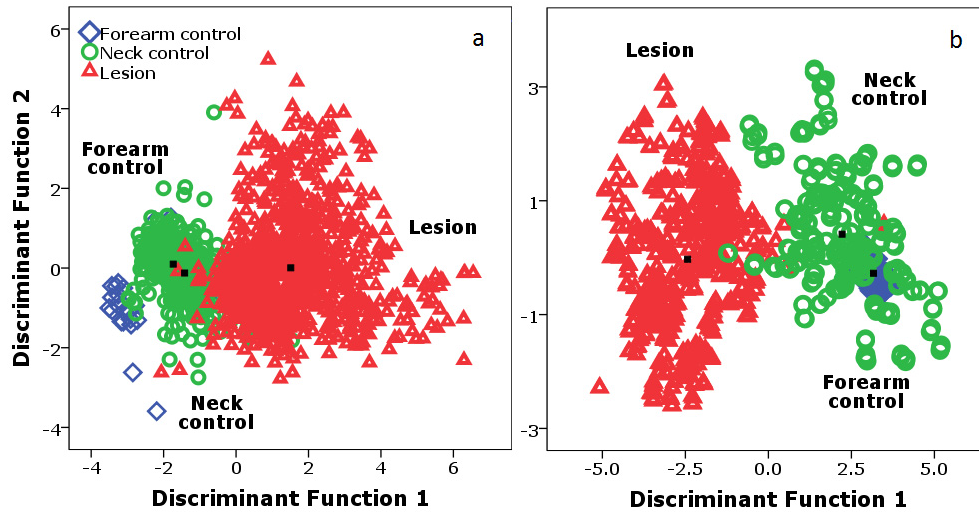


Figure 8.8 Discriminant function plots for (a) DRS and (b) colorimetry data obtained from all patients over the course of this study[91]

In the DFA results, DF1 expresses the maximum variance of the data and DF2 contains the rest of the variance in the data. It is evident that while there is some overlap of individual measurements, the center of mass of the lesion data distribution (shown as the labeled dark square in Fig. 8.8) is well separated from the centers of mass of the distribution for the two controls. The reversal of the DF1 scores for the two modalities (i.e. positive DF1 lesion scores for DRS and negative DF1 scores for colorimetry) has no physical significance and is purely a computational artifact. The fact that all the patients' data can be clustered significantly implies that the variations of the lesion tissue from the control sites are reproducible and similar in all patients.

### 8.5.1 Classification

A leave-one-out (LOO) classification was also performed for both data sets. In a LOO method, each data point is treated as an unknown case and is classified against a data set consisting of all the other data points. The accuracy of that classification is then compared to the known identity of the data point to create a truth table. The truth tables for both spectral methods are shown in Table 8.3. The interpretation of these truth tables is as follows: true positives (TP) indicate a lesion measurement was correctly classified as lesion. True negatives (TN) indicate a normal control measurement was correctly classified as normal. False positives (FP) indicate a normal control measurement was incorrectly classified as lesion. False negatives (FN) indicate a lesion measurement was incorrectly classified as normal.

| DRS (%) | True | False |   | Colorimetry (%) | True | False |
|---|---|---|---|---|---|---|
| Positive | 91.4 | 4.9 |   | Positive | 98.4 | 3.9 |
| Negative | 95.1 | 8.6 |   | Negative | 96.1 | 1.6 |

Table 8.3 Truth tables for leave-one-out classification results of (a) DRS and (b) colorimetry[91]

Colorimetry data showed the highest true positive and true negative results, 98.4% and 96.1% respectively, which confirmed our earlier result that colorimetry seems to be more efficacious for diagnosing AN than DRS as we currently perform it. Impressively, when analyzed with the PCA/DFA the DRS data showed more than 91% sensitivity (TP/TP+FN) and specificity (TN/TN+FP), which was not the case for the standard statistical analyses as reported in subsection 8.3.

**8.5.2 External Validation**

In the previous leave-one-out analysis, an unclassified measurement was classified with a set of discriminant functions constructed using all the other data points, including other data points from that same patient. However, if either of these techniques is to be used for patient screening, it is not realistic to expect the discriminant functions to have been constructed with any data from that patient. Therefore, we performed a DFA excluding one patient at a time from the analysis. Unclassified lesion and neck control data from that patient were then input to the DFA (which contained none of that patient's other data) and were classified according to the library of results from other patients. The truth tables for this analysis are shown in Table 8.4.

| DRS (%) | | | | Colorimetry (%) | | |
|---|---|---|---|---|---|---|
| | True | False | | | True | False |
| Positive | 87.7 | 5.2 | | Positive | 97.2 | 5.4 |
| Negative | 94.8 | 12.3 | | Negative | 94.6 | 2.8 |

Table 8.4 Truth tables for patient exclusion classification results of (a) DRS and (b) colorimetry[91]

As expected the rates of true positives and true negatives declined, although only by a small amount (by 3.7% and 0.3% for DRS and 1.2% and 1.5% for colorimetry) and rates of false positives and false negatives increased (by 0.4% and 3.7% for DRS and 1.5% and 1.2% for colorimetry). These results show a more realistic truth table for the techniques if they were to be used to screen previously unexamined patients for AN.
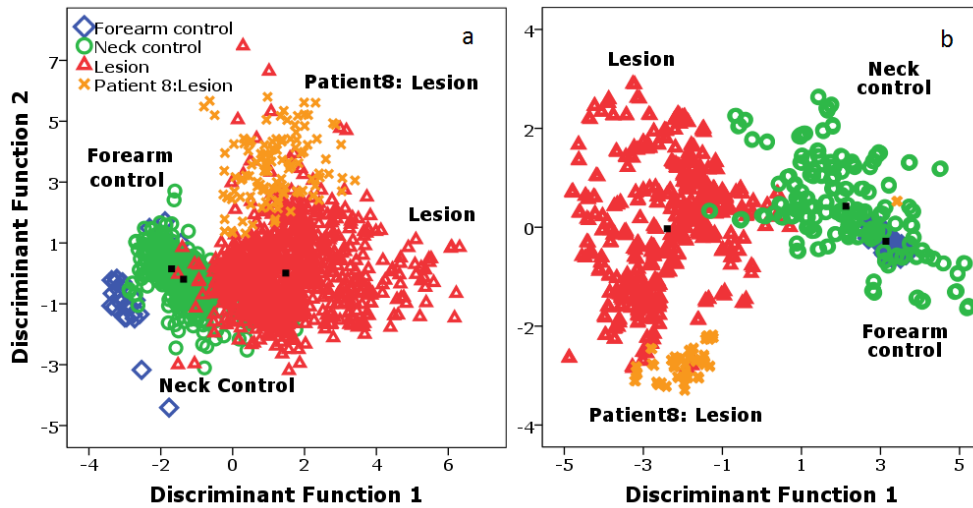
Figure 8.9 DF plots showing (a) DRS and (b) colorimetry data with patient 8's lesion data input as unclassified data into the analysis. This patient's lesion data, which clustered well with itself, was significantly different from the mean of the other patients' lesion data. Nonetheless, it was easily and reliably classified as "lesion" in both analyses[91]

One of the strong benefits of using a chemometric approach is shown in Fig. 8.9 which shows the DF plot of the patient exclusion analysis for patient 8. Patient 8 was the only case in our previous analysis that did not show a 3σ standard deviation between the calculated melanin concentration in the forearm control and the AN lesion indicating that this patient's lesion tissue was very hard to diagnose spectrally. Nonetheless, in Fig. 8.9, when patient 8's lesion data was entered as unclassified data into the DFA (shown as x symbols in Fig. 8.9), almost all of the measurements were easily classified as "lesion" compared to the forearm control or the neck control.

Although forearm and neck controls were both used as normal controls in this AN study, the more realistic site for a normal control is the neck normal (which should have similar properties compared to the lesion measurement site due to tanning, aging, etc.). This can be observed by the neck control data being closer to the lesion data than the forearm control data in Figs. 8.8 and 8.9.

### 8.5.3 ROC Curves

It was vital to know the sensitivity and specificity of the lesion site measurements as compared to the neck control. To investigate this, receiver operating characteristic (ROC) curves were constructed using the DF1 values of the DRS and colorimetry data to act as a "cut point" to discriminate lesion and neck control data (discarding the forearm control data).
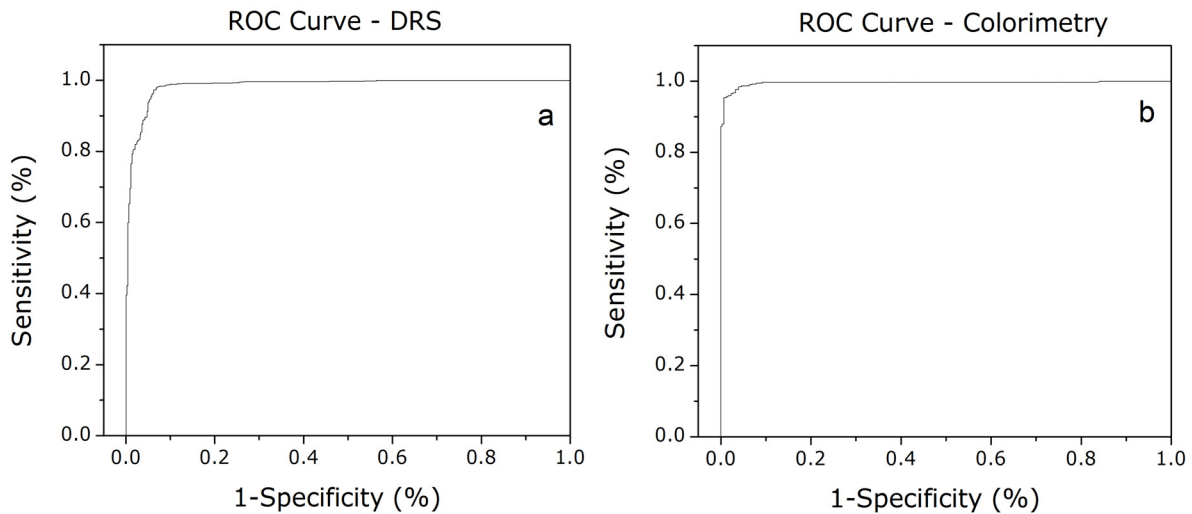


Figure 8.10 ROC curves of a test to differentiate neck control measurements from lesion measurements as shown in Fig. 8.8 on the basis of the DF1 score alone for (a) DRS and (b) colorimetry data. The area under the curves for (a) was 0.985 and for (b) was 0.995[91]

In a ROC curve[114], the sensitivity of the technique (as defined above) is plotted against 1-specificity (as defined above). ROC curves are shown in Fig. 8.10 for DRS (a) and colorimetry (b). The ROC area under the curve (AUC) establishes the usefulness of the test, with an AUC of 1.0 denoting a perfect test, and an AUC of 0.5 denoting a worthless test. The ROC curves of Fig. 8.10 possess an AUC of more than 0.98, indicating a highly reliable test for this AN investigation.

## 8.6 Conclusions

In the standard analytical methods, the average values that differed by more than $3\sigma$ were considered to be valuable predictors of the presence of lesion tissue. Of the three CIE L*, a*, and b* parameters, the luminosity parameter "L*" was found to most reliably distinguish lesion from normally pigmented skin. b* also showed more than $3\sigma$ statistical difference between AN lesion and the skin control sites and a* was considered as a weak indicator. The calculated melanin concentration of the DRS instrument was able to characterize a lesion as compared to normally pigmented skin on the neck, though melanin is a weak indicator of skin change and not a reliable measure to be used independently. However, oxyhemoglobin and deoxyhemoglobin concentrations were not found to be reliable indicators of AN.

Spectroscopic and colorimetric measurements combined with chemometric analysis methods provided sensitive and specific diagnoses of acanthosis nigricans lesions compared to nearby control skin. Analysis of raw DRS absorbance spectra showed clear clustering of normal controls and lesion groups (Fig. 8.8 (a)) for all patients denoting a commonality that may allow diagnosis of previously unmeasured lesions. Colorimetry data also showed the ability to reliably identify AN lesions (Fig. 8.8 (b)). Excluding patients one at a time from a DFA model and then testing that patient's spectral data with the model constructed only from other patient measurements provided a realistic simulation of an acanthosis nigricans screening test. DRS data provided more than 87% sensitivity and 94% specificity and colorimetry data showed more than 95% sensitivity and specificity (Table 8.4) in this type of test. ROC curves also confirmed that the use of a discriminant function analysis on DRS and colorimetry data can provide a sensitive and specific AN test even when only the DF1 score is used to assess skin condition.

Unfortunately, none of the patients showed any improvement in acanthosis nigricans from treatment during the duration of this study which was also confirmed by visual and

photographic observations.  Thus, the changes in skin DRS and colorimetry data that occur during the healing process of AN could not be established.  However, the existing data showed both DRS and colorimetry can be used as a successful diagnostic tool for acanthosis nigricans when combined with chemometric methods such as PCA and DFA.

## CHAPTER 9

## CONCLUSIONS AND FUTURE WORK

This chapter summarizes the conclusions drawn from Raman spectroscopic investigations of head and neck squamous cell carcinoma, prostate cancer, and pediatric tumors, and also summarizes the results of identifying acanthosis nigricans skin disease using diffuse reflectance spectroscopy and colorimetry followed by future work.

### 9.1 Conclusions

### Head and Neck Squamous Cell Carcinoma: Tongue Tissues

We have studied the Raman spectral characteristics of normal, carcinoma *in situ,* and invasive squamous cell carcinoma (SCC) tissues of tongue. There are substantial spectral differences in the observed spectra of these tissues and we were able to identify the spectral features that can discriminate these three tissue types. We were able to assign the observed spectral changes in the intensities of certain Raman bands to specific molecules or category of molecules, and thus, draw inferences on the nature of biochemical compositional changes in these tissues (see Chapter 5). We found tryptophan to increase while keratin to decrease when the tissue changes from normal to invasive SCC. The analysis of Raman spectral data using chemometric methods, principal component analysis (PCA) and discriminant function analysis (DFA), clearly separated the data into three distinct groups consistent with results of pathology. We were able to distinguish with 91% success rate the normal and carcinoma *in situ* tissues, and with 89% accuracy the invasive SCC tissues of tongue. This exploratory study suggests that Raman spectroscopy can be used to distinguish normal, carcinoma *in situ* and invasive SCC.[111]

### Prostatic Intraepithelial Neoplasia

In this study, we investigated the benign, prostatic intraepithelial neoplasia, and difference cancer grades in FFPP tissues using Raman spectroscopy. The tissue specimens used in the study were obtained as radical prostatectomy and each of the diseased tissues contained regions with three pathologies mentioned above.[88] We found that, even with other pathologies present in the tissue, Raman spectroscopy can differentiate each pathological state through careful investigation of epithelia. Thus, the Raman spectra contain signatures from both luminal and basal cells. Further, Raman spectroscopy was able to classify the tissues with Gleason scores of 6, 7, and 8 with more than 85% accuracy. As a part of this study, we also investigated the luminal cells of benign epithelia, prostate intraepithelial neoplasia, and cancer using additional 34 patients' specimens.[133] The reason being, that basal cells are not present in cancerous micro acinar structures. These tissues contain only one pathology, either BE, PIN or cancer (Gleason grade 3). Our aim was to diagnose PIN pathology and compare and contrast its Raman spectral features with that of BE or cancer. The study did show difference in the spectral features, as discussed in Chapter 6. We also examined the Raman fingerprints of stromal cells around each pathological state: BE, PIN, and cancer. The spectral data was analyzed with PCA and DFA, and were able to classify the data into distinct pathological groups with very high accuracy. The sensitivity and specificity of luminal cells of PIN and cancer were about 98% for each and for the stroma associated with these pathologies the sensitivity and specificity were more than 90%.[133]

### Pediatric Tumors: Neuroblastoma and Ganglioneuroma

In this work, we investigated the cellular regions in neuroblastoma and ganglioneuroma using Raman spectroscopy and compared their spectral characteristics with those of the corresponding normal adrenal gland. Thin sections from both frozen and

formalin-fixed paraffin processed (FFPP) tissues, obtained from the same tissue block, were studied in conjunction with the pathological examination of the tissues. Investigation of the spectral data showed that the normal adrenal gland tissues have higher levels of carotenoids, lipids, and cholesterol compared to the neuroblastoma and ganglioneuroma frozen tissues[137]. However, in comparison to the frozen tissues the FFPP tissues show a significant alteration in several biochemicals, including complete removal of carotenoids, lipids, and cholesterol in adrenal gland tissues as shown in Chapter 7.

The quantitative analysis, of the Raman data obtained from the frozen tissues, using PCA and DFA showed a clear-cut classification among pathological groups with high sensitivity and specificity.[137] This study showed that Raman spectroscopy combined with chemometric methods can be successfully used to distinguish neuroblastoma and ganglioneuroma at cellular level in frozen tissue sections. Further, the nature of chemical changes that occur in processing and preservation of these tissues were evident in Raman spectra, and thus suggesting that one can also use the FFPP tissues for optical diagnosis using Raman spectroscopy.

**Acanthosis Nigricans: Diffuse Reflectance Spectroscopy/Colorimetry**

In this study, we investigated acanthosis nigricans (AN) skin diseases, hyperpigmented velvety plaques in skin using diffuse reflectance spectroscopy (DRS) and the colorimetry *in vivo*. The non-invasive optical technique DRS provides information on skin melanin, oxyhemoglobin, and deoxyhemoglobin content through the measured absorption spectrum whereas colorimetry measures the skin color based on the standard Tristimulus values (so called L*, a*, and b* parameters). Measurements were conducted on eight AN patients, spanning eight months of treatment by diet control and/or oral metformin. The study indicated an increase in melanin concentration AN lesion areas compared to the normal control sites. L* and b* show a significant differences between AN

lesion and normal controls, and a* and melanin are weak indicators.[144] In fact, oxyhemoglobin and deoxyhemoglobin didn't show any significant differences what so ever, so they are not reliable for the screening of AN.

PCA and DFA of the spectral data provided more robust results compared to the standard analytical methods in attempting to classify an undiagnosed skin sample as "lesion" or "healthy." PCA-DFA successfully differentiated between AN lesion and normal skin controls with 87.7% sensitivity and 94.8% specificity in DRS measurements, and 97.2% sensitivity and 96.4% specificity in colorimetry measurements.[137] These chemometric methods provided a better diagnostic ability than skin chromophore concentration calculations based on DRS data.

Both spectroscopic measurements, dermatological evaluations, and insulin level measurements showed no change in AN patients between the chronological visits to their doctors. Thus, we could not draw any conclusions regarding suppression of AN as none of the patients healed during the course of this research work.


**9.2 Future Work**

The investigation of the head and neck squamous cell carcinoma, prostate cancer, and pediatric tumors using deparaffinized tissue sections from archived paraffin embedded blocks has lots of advantages. The easy access to these archived tissue blocks reduces the waiting time to obtain a fresh tissue sample from a patient. Since, there are a large number of tissue specimens available in the archives, a large Raman spectroscopic data can be collected for different types of cancers. In order to validate any hypothesis, it is important to have the analysis done with a large data set. However, a prior knowledge of loss of Raman information upon preserving the tissues using formalin fixating and embedding in paraffin is very important. Our study has documented this in the case of pediatric tumors, where, we studied the spectral difference between frozen and their corresponding FFPP

tissues. The nature of changes that occur in preserving the tissues for archival purposes could be different for different types of tissues and the type of cancer. Thus, one needs to investigate the nature of changes that entail tissue archival in different types of cancers using frozen and their corresponding preserved tissues before using the Raman spectroscopy for diagnosis using the archived FFPP tissues. The investigation of frozen tissues seems more realistic and useful as the frozen tissues produce Raman spectra similar to the fresh tissues. By increasing spectral database for the frozen tissues of different cancer types, one can use this technique as an additional tool to obtain complementary information about the state of a disease or cancer to augment the pathological evaluations. Pathological evaluations can be subjective depending on the experience of the pathologists. Raman spectroscopic detection may provide new means for disease/cancer detection, *in vitro* or *in vivo*.

Regarding prostate cancer investigation, it is an established observation that a prostate gland exhibiting PIN cell architecture may or may not progress to carcinoma. So, the fundamental and most important issue here is to determine which cell architecture will lead to cancer. Thus, future studies could focus on finding the unique Raman spectral features of the PIN cell architecture that leads to carcinoma.

Regarding, optical diagnosis of acanthosis nigricans, there are certain draw backs. The main disadvantage is the difficulty in enrolling new patients into this research project. Even if new patients enroll, some of them abandon the research study in the middle of the study period. Another disadvantage is that although the patients are advised not to consume any food during fasting insulin level measurements, this cannot be strictly controlled. If this happens, the measured values have to be discarded from the study. Although, the patients are supposed to control their diet and are advised on exercising to control and get better from this AN skin disorder, as researchers, we do not have the control over the patients' activities but one can minimize these issues with close doctor-

patient relationship and motivating the enrolled patients to remain in the study. One could also conduct this study on other ethnic groups, although, AN is common among African-Americans using other ethnic groups may provide dependence on skin type.

This study also suggests that we can use a different predictor of the patients' insulin level. That would be the hemoglobin A1c test, which does not require fasting. This test provides a reliable measure to know whether the patient's blood glucose level for the previous 3 months is persistently elevated or not. Thus, more meaningful results can be obtained by conducting this research work for about 5 years.

In a near-perfect scenario, one can determine the skin changes during the healing process using diffuse reflectance spectroscopy/colorimetry and correlate the measurements with serum insulin level and insulin sensitivity in AN. Then, this particular skin disorder becomes a diagnostic parameter for insulin-resistant obese AN patients to evaluate the insulin sensitivity. The ultimate objective is to replace the time consuming and resource extensive procedures of blood collection and laboratory evaluations of obese patients having AN by non-invasive DR measurements.

Furthermore, we can strengthen the statistical/chemometric analysis by increasing the size of the data set. For a large enough data set, PCA-DFA analysis can be used to propose a reliable model. Then using this model, one can validate the new data set with the proposed model. The alternate advanced chemometric methods such as partial-least square regression method or support vector machines method could be used to analyze the Raman or diffuse reflectance spectral data to assess the discrimination ability of these techniques compared to the PCA and DFA.

**APPENDIX A**

**SUPPLEMENTARY TABLES**

| Raman Shift (cm$^{-1}$) | | 753 | 850 | 933 | 1235/40 | 1306 | 1335 | 1365 | 1446 | 1562 | 1584 | 1614 | 1650 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 753 | Pearson | 1 | -.064 | -.010 | -.356 | -.484 | -.219 | .712 | -.461 | .809 | .616 | .744 | -.464 |
| | Sig. | | .200 | .836 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 850 | Pearson | -.064 | 1 | .717 | .191 | .357 | .224 | -.517 | .283 | -.501 | -.666 | -.495 | .533 |
| | Sig. | .200 | | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 933 | Pearson | -.010 | .717 | 1 | .062 | .236 | .183 | -.310 | .181 | -.280 | -.444 | -.206 | .293 |
| | Sig. | .836 | .000 | | .217 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 1235/40 | Pearson | -.356 | .191 | .062 | 1 | .915 | .874 | -.086 | .609 | -.475 | -.346 | -.268 | .573 |
| | Sig. | .000 | .000 | .217 | | .000 | .000 | .085 | .000 | .000 | .000 | .000 | .000 |
| 1306 | Pearson | -.484 | .357 | .236 | .915 | 1 | .911 | -.262 | .759 | -.641 | -.537 | -.438 | .649 |
| | Sig. | .000 | .000 | .000 | .000 | | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 1335 | Pearson | -.219 | .224 | .183 | .874 | .911 | 1 | .103 | .681 | -.335 | -.251 | -.122 | .510 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | | .039 | .000 | .000 | .000 | .015 | .000 |
| 1365 | Pearson | .712 | -.517 | -.310 | -.086 | -.262 | .103 | 1 | -.197 | .851 | .843 | .879 | -.474 |
| | Sig. | .000 | .000 | .000 | .085 | .000 | .039 | | .000 | .000 | .000 | .000 | .000 |
| 1446 | Pearson | -.461 | .283 | .181 | .609 | .759 | .681 | -.197 | 1 | -.558 | -.410 | -.501 | .620 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | .000 | .000 | | .000 | .000 | .000 | .000 |
| 1562 | Pearson | .809 | -.501 | -.280 | -.475 | -.641 | -.335 | .851 | -.558 | 1 | .892 | .929 | -.646 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | | .000 | .000 | .000 |
| 1584 | Pearson | .616 | -.666 | -.444 | -.346 | -.537 | -.251 | .843 | -.410 | .892 | 1 | .838 | -.552 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | | .000 | .000 |
| 1614 | Pearson | .744 | -.495 | -.206 | -.268 | -.438 | -.122 | .879 | -.501 | .929 | .838 | 1 | -.670 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | .015 | .000 | .000 | .000 | .000 | | .000 |
| 1650 | Pearson | -.464 | .533 | .293 | .573 | .649 | .510 | -.474 | .620 | -.646 | -.552 | -.670 | 1 |
| | Sig. | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | |

Table A1 Pearson correlation with 2-tailed significance for selected 12 Raman bands in head and neck squamous cell carcinoma (see Chapter 5)

| Parameters | Pearson Correlation Coefficient (*r*) (P-value) | | | | |
|---|---|---|---|---|---|
| | 1st-2nd month | 1st-3rd month | 1st-4th month | 1st–5th month | 1st–6th month |
| ΔL* / Δmelanin | -0.50 (0.31) | -0.48 (0.33) | -0.67 (0.22) | -0.88 (0.12) | -0.97 (0.17) |
| Δa* / Δ HbO$_2$ | -0.32 (0.54) | -0.29 (0.58) | 0.29 (0.63) | 0.76 (0.24) | 1.00 (0.04) |
| Δb* / ΔHb | 0.09 (0.87) | 0.48 (0.34) | -0.97 (0.01) | 0.76 (0.80) | -0.37 (0.76) |

Table A2 Correlations between the L*, a*, and b* color parameters (colorimetry) vs melanin, oxyhemoglobin (HbO$_2$), and deoxyhemoglobin (Hb) (DRS) measured on acanthosis nigricans neck lesions (see Chapter 8)

| Variable | N | Mean | Std Dev | P-value |
|---|---|---|---|---|
| L* 1st month | 8 | -27.56 | 7.86 | <0.001 |
| L* 2nd month | 8 | -26.83 | 7.63 | <0.001 |
| L* 3rd month | 8 | -27.65 | 7.56 | <0.001 |
| L* 4th month | 8 | -29.48 | 7.77 | <0.001 |
| | | | | |
| a* 1st month | 8 | -29.26 | 15.16 | <0.001 |
| a* 2nd month | 8 | -19.49 | 25.88 | 0.071 |
| a* 3rd month | 8 | -17.84 | 27.74 | 0.112 |
| a* 4th month | 8 | -16.31 | 30.60 | 0.176 |
| | | | | |
| b* 1st month | 8 | -47.85 | 12.98 | <0.001 |
| b* 2nd month | 8 | -42.26 | 17.24 | <0.001 |
| b* 3rd month | 8 | -42.17 | 18.46 | <0.001 |
| b* 4th month | 8 | -43.92 | 18.53 | <0.001 |
| | | | | |
| Melanin 1st month | 6 | 33.68 | 10.45 | <0.001 |
| Melanin 2nd month | 8 | 27.50 | 10.37 | <0.001 |
| Melanin 3rd month | 8 | 30.11 | 14.56 | <0.001 |
| Melanin 4th month | 8 | 31.05 | 12.98 | <0.001 |
| | | | | |
| Oxyhemoglobin 1st month | 6 | 64.29 | 114.45 | 0.227 |
| Oxyhemoglobin 2nd month | 8 | 10.92 | 32.92 | 0.380 |
| Oxyhemoglobin 3rd month | 8 | -3.43 | 73.03 | 0.898 |
| Oxyhemoglobin 4th month | 8 | 10.10 | 61.29 | 0.655 |
| | | | | |
| Deoxyhemoglobin 1st month | 6 | -32.08 | 27.05 | 0.034 |
| Deoxyhemoglobin 2nd month | 8 | -44.81 | 26.06 | 0.002 |
| Deoxyhemoglobin 3rd month | 8 | -32.02 | 25.87 | 0.010 |
| Deoxyhemoglobin 4th month | 8 | -44.91 | 35.59 | 0.009 |

Table A3 Percent difference in measured parameters between lesion skin and normal skin[144] in the AN study (see Chapter 8)

**APPPENDIX B**

**STEP BY STEP INSTRUCTIONS FOR DATA PROCESSING**

All the Raman spectral data obtained from Renishaw Raman spectrometer and Jobin-Yvon Horiba Triax550 Raman spectrometer system are in .txt format.

**Denoising and Normalizing**

1. Make a folder, say "Data", copy all the .txt files to the "Data" folder together with the MatLab "adaptminmax" program.

2. Create a text file, "filelist.txt" containing all the names of the Raman spectral data files.

   a. Copy the "Data" folder to the desktop

   b. Open the "command prompt" window

   c. Enter "cd desktop"

   d. Enter "cd Data " (now you are in the "Data" folder)

   e. Enter "dir /b >filelist.txt", make sure you open the "filelist.txt" and delete the name filelist.txt, keep only the names of the relevant files)

3. Run the "adaptminmax" program, enter "y" to see the graphs while processing.

**Creating Mean Raman Spectra**

1. Make a folder on the desktop, say "Mean", and copy all the relevant .txt files to the "Mean" folder together with the matlab "gulaygraph2" program.

2. Create a text file, "N-filelist.txt"

   a. Repeat step 2, from (b) to (d) of "De-noising and Normalizing"

   b. Now enter "dir /b >N-filelist.txt". Similarly, open the "N-filelist.txt" and delete the name "N-filelist.txt"

3. Run the "gulaygraph2" program.

**Creating a one txt file with all the Raman Data**

1. Make a folder on the desktop, say "All-Data", and copy all the .txt files to the "All-Data" folder together with the MatLab "combinespectra81706" program.

2. Create a text file, "filelist.txt"

   a. Repeat step 2 of "De-noising and Normalizing"

   b. Now enter "dir /b >filelist.txt". Similarly, open the "filelist.txt" and delete the name "filelist.txt"

3. Run the "combinespectra81706" program.

4. This will create a .txt file "allspectra.txt" with variables in columns.


**PCA using SPSS**

1. Open "SPSS" software

2. Click "file", enter "read a text"

3. Change the variable names of V1 - V601 to VV600, VV602,…..VV1800

4. Go to "View variable" mode and insert three variables in the first raw.

5. Change the first variable to "string",  change the name to "Spectra"

6. Change name of the $2^{nd}$ variable to "Diagnosis"  and the $3^{rd}$ to "Groups"

7. Click "Value" and create group names and numbers, say 1-Normal, 2-Neuroblastoma, and 3-Ganglioneuroma

8. Go to the Data mode and copy paste the name list to the "Spectra" column

9. Now, enter value 1-3 to the relevant data cases.

10. Click "Analyze/Dimension Reduction/Factor", select all the variables and bring those variables to the window named variables.  Now, select "extraction" as shown by the following window.
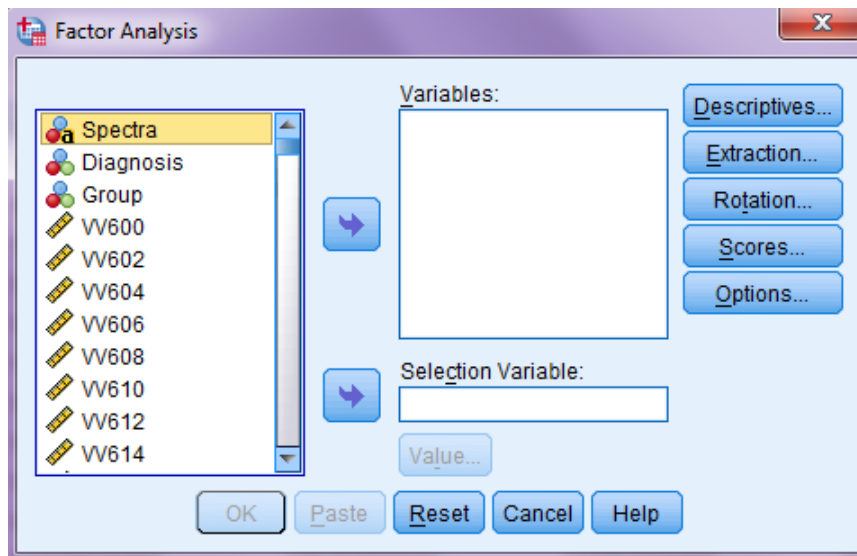
11. Select "Covariance matrix" and press "Continue".  Then click "Scores" and select "Save as variables" and press "Continue".  Now press "OK".  Then, one can see the principal scores at the end of the variables.

**DFA using SPSS**

1. Click "Analyze/Classify/Discriminant", then the following window will appear.
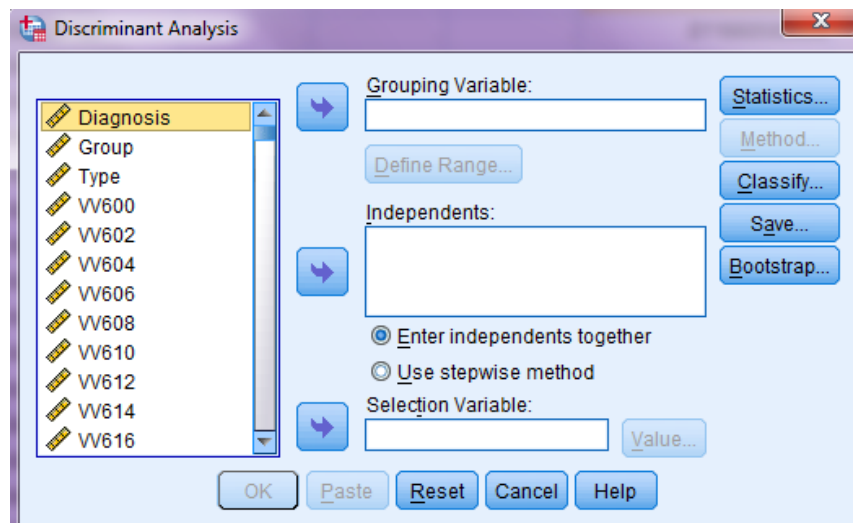


2. Bring the principal components from the left window to the "Independents".  Also bring variable "Group" to the "Grouping Variables", then define the range from 1-3.

Now, select "Statistics", and select "Fisher's" under "Functions Coefficients" and press "Continue".

3. Click "Classify", and select options of "Case wise Results, Summary Table and Leave-one-out Classification" under "Display", and "Combined-groups" under "Plots" and press "Continue".

4. Click "Save" and select the option of "Discriminant Scores" and press "Continue". Finally press "OK".  Then, the software will generate the appropriate tables and plots.

**External Validation**

1. Once you choose which patient data to feed as test data, delete the values of the columns "Diagnosis" and "group" in those data cases before you do the DFA analysis.

2. Perform the DFA as explained above.

3. The results of the selected ungrouped data cases appear in the  classification table as "ungrouped cases"

# REFERENCES

[1] Jemal A., Siegel R., Bray F., Center M.M., Ferlay J., Ward E., and Forman D. (2011). Global Cancer Statistics. *Cancer J Clin.* Vol. 5961, pp. 225-249

[2] Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011

[3] Wang L.V. , Wu H-I. (2007). *Biomedical Optics, Principles and Imaging*, Wiley-interscience

[4] Parkin D.M., Bray F., Ferlay J., and Pisani P. (2005). Global Cancer Statistics, 2002. *CA Cancer J Clin.* Vol. 55(2), pp. 74-108.

[5] Lin W-J., Jiang R.S. , Wu S-H., Chen F-J., and Liu S-A. (2011). Smoking, Alcohol, and Betel Quid and Oral Cancer: A Prospective Cohort Study Journal of Oncology, doi:10.1155/2011/525976

[6] Mirken B. More evidence that marijuana prevents cancer. Jan 2010

[7] Robinson R.A. (2010). *Head and neck pathology*, Wolters Kluwer, Lippincott Williams & Wilkins, USA

[8] Vokes E.E., Weichselbaum R.R., Lippman S.M., Hong W.K. (1993). Head and Neck Cancer *N Engl. J Med.* Vol. 328(3), pp. 184-194

[9] http://www.carsforprostatecancer.org/our-cause/learn-more-about-prostate-cancer/

[10] Henson D.E. and Albores-Saavendra J. (2001). *Pathology of Incipient Neoplasia, Third Edition*, Oxford University Press

[11] Bostwick, D.G. and Qian, J. (2004). High-grade prostatic intraepithelial neoplasia. *Modern Pathology*, Vol. 17, pp. 360–379

[12] Kandaswami C., Lee L.T., Lee P.P, Hwang J.J., Ke F.C., Huang Y.T., and Lee M.T. (2005). The antitumor activities of flavonoids. *In Vivo*, Vol. 19, pp. 895-909

[13] Fleshner N.E., Kapusta L., Donnelly B., Tanguay S., Chin J., Hersey K., Farley A., Jansz K., Siemens R., Trpkov K., Lacombe L., Gleave M., Tu D., and Parulekar W.R. (2011). Progression from High-Grade Prostatic Intraepithelial Neoplasia to Cancer: A Randomized Trial of Combination. *J Clin Oncol*, Vol. 29, pp. 2386-2390

[14] Hricak H., Scardino P.T. and Reznek R.H. (2009) *Prostate Cancer*, Cambridge University Press

[15] Humphrey, P.A. (2004). Gleason grading and prognostic factors in carcinoma of the prostate. *Modern Pathology*, Vol. 17, pp. 292–306

[16] Teba F., Martin R., Gomez V., Herranz L. M., and Santamaria L. (2007). Cell proliferation and volume-weighted mean nuclear volume in high-grade PIN and adenocarcinoma compared with normal prostate. *Image Anal. Stereol.*, Vol. 26, pp. 93-99

[17] Cheng N-K.V. (2005) *Neuroblastoma*, Susan L. Cohn Editors, Springer

[18] Janes-Hodder H. and Keene N. (2002). *Childhood Cancer: A Parents' Guide to Solid Tumor Cancers*, 2th ed. O'Reilly Media Inc

[19] Shimada H., Chatten J., Newton W.A., Sachs N., Hamoudi A.B., Chiba T. (1984). Histopathologic prognostic factors in neuroblastic tumors: definition of subtypes of ganglioneuroblastoma and an age-linked classification of neuroblastomas. *J Natl Cancer Inst* Vol. 73, pp. 405–16

[20] Shimada H., Ambros I. M., Dehner L.P., Hata J-I., Joshi V.V., Roald B., Stram D. O., Gerbing R.B., Lukens J.N., Mathhay K.K., and Castleberry R.P. (1999). The International Neuroblastoma Pathology Classification (the Shimada System). *American Cancer Society*. Vol. 86, pp. 364-372

[21] Cotton C.A., Peterson S., Norkool P.A., Takashima J., Grigoriev Y., Green D.M., and Breslow N.E. (2009). Early and Late Mortality After Diagnosis of Wilms Tumor. *J of Oncology*. Vol. 27(8), pp. 1304-1310

22 Kim S. and Chung D.H. (2006). Pediatric solid malignancies: Neuroblastoma and Wilms' tumor. *Surgical Clinics of North America*. Vol. 86 (2), W. B. Saunders Company

23 Flegal K.M., Carroll M.D., Ogden C.L., and Curtin L.R. (2010). Prevalence and Trends in Obesity Among US Adults, 1999–2008. *Journal of the American Medical Association* Vol. 303(3), pp. 235–241

24 Kahn C.R., Flier J.S., and Bar R.S. (1976). The syndromes of insulin resistance and acanthosis nigricans. Insulin-receptor disorders in man. *N Engl J Med.* Vol. 294(14), pp. 739-745

25 Hermanns-Le T., Hermanns J.F., and Pierard G.E. (2002) Juvenile acanthosis nigricans and insulin resistance. *Pediatr Dermatol.* Vol. 19(1), pp. 12-14

26 Kuroki R., Sadamoto Y., and Imamura M. (1998). Acanthosis nigricans with severe obesity, insulin resistance and hypothyroidism: improvement by diet control. *Dermatology.* Vol. 198(2), pp. 164-166c

27 Guran T., Turan S., Akcay T., and Bereket A. (2008). Significance of acanthosis nigricans in childhood obesity. *J Paediatr Child Health.* Vol. 44(6), pp. 338-341

28 Kobaissi H.A., Weigensberg M.J., Ball G.D., Cruz M.L., Shaibi G. Q., and Goran M.I. (2004). Relation between acanthosis nigricans and insulin sensitivity in overweight Hispanic children at risk for type 2 diabetes. *Diabetes Care* Vol. 27(6), pp. 1412-1416

29 Stoddart M.L., Blevins K. S., Lee E.T., Wang W., and Blackett P.R. (2002). Association of acanthosis nigricans with hyperinsulinemia compared with other selected risk factors for type 2 diabetes in Cherokee Indians: the Cherokee Diabetes Study. *Diabetes Care.* Vol. 25(6), pp. 1009-1014

30 Stuart C.A., Gilkison C.R., Smith M.M., Bosma A.M., Keenan B.S., and Nagamani M. (1998). Acanthosis nigricans as a risk factor for non-insulin dependent diabetes mellitus. *Clin Pediatr (Phila).* Vol. 37(2), pp.73-79

[31] Type 2 diabetes in children and adolescents. American Diabetes Association (2000). *Diabetes Care.* Vol. 23(3), pp. 381-389

[32] Hardin D.S. (2006). Screening for type 2 diabetes in children with acanthosis nigricans. *Diabetes Educ.* Vol. 32(4), pp. 547-552

[33] Cruz P.D. Jr. and Hud J.A. Jr. (1992). Excess insulin binding to insulin-like growth factor receptors: proposed mechanism for acanthosis nigricans. *J Invest Dermatol.* Vol. 98(6 Suppl), pp. 82S-85S

[34] Katz A.S., Goff D.C., and Feldman S.R. (2000). Acanthosis nigricans in obese patients: Presentations and implications for prevention of atherosclerotic vascular disease. *Dermatology Online Journal* Vol. 6(1)

[35] Yosipovitch G., DeVore A., and Dawn A. (2007). Obesity and the skin; Skin physiology and skin manifestations of obesity. *J Am. Acad. Dermatol*. Vol.56, pp. 901-916.

[36] What is Elastin and How is it Involved in Skin Laxity? PRlog (Press release). 2010

[37] http://dynamicnaturesite.blogspot.com/2009/07/skin-care-study-of-skin-structure.html

[38] Panno J. (1999). *Cancer, The Role of Genes, Lifestyles & Environment*, Facts On File, Inc.

[39] Allsbrook W.C., Mangold K.A., Johnson M.H., Lane R.B., Lane C.G., and Epstein J.I. (2001). Interobserver reproducibility of Gleason grading of prostatic carcinoma: General Pathologists. *Human Pathology*, Vol. 32, No. 1, pp. 81-88

[40] Raman C.V. (1928). A new type of secondary radiation. *Nature*, Vol. 121, No. 3048, pp. 501-502

[41] De Gelder J., Gussem K.D., Vandenabeele P., and Moens L. (2007). Reference database of Raman spectra of biological molecules. *J Raman Spec*. Vol. 38, pp. 1133-1147

[42] Movasaghi Z., Rehman S., and Rehman I.U. (2007). Raman spectroscopy of biological tissues. *Applied. Spectroscopy,* Vol. 42, pp. 493-541

[43] Notingher I. (2007). Raman spectroscopy cell-based biosensors. *Sensor,* Vol. 7, pp. 1343-1358

44 de Jong B. W.D. Schut T.C.B.,Maquelin K., Kwast T.V.D., Bangma C.H., Kok D-J., and Puppels G.J. (2006). Discrimination between Nontumor Bladder Tissue and Tumor by Raman Spectroscopy. *Anal. Chem,* Vol. 78, pp. 7761-7769

45 Teh S.K., Zheng W., Ho K.Y., Teh M., Yeoh K.G., and Huanga Z. (2009). Near-infrared Raman spectroscopy for gastric precancer diagnosis. *J. Raman Spectrosc.* Vol. 40, pp. 908–91

46 Haka A., Volynskaya Z., Gardecki J., Nazemi J., Shenk R., Wang N., Dasari R., Fitzmaurice M., and Feld M. (2009). Diagnosing breast cancer using Raman spectroscopy:prospective analysis *Journal of Biomedical Optics* Vol. 14 (5), 054023, pp. 1-8

47 Crow P., Barrass B., Kendall C. A., Hart-Prieto M., Wright M., Persad R., and Stone N. (2005). The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *Br. J of Cancer* Vol. 92(12), pp. 2166-2170

48 Crow P., Stone N., Kendall C. A., Uff J. S., Barr J.A.M., and Wright M. (2003). The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *Br. J of Cancer* Vol. 89(1), pp. 106-108

49 Lau D.P., Huang Z., Lui H., Man C.S., Berean K., Morrison M.D., and Zeng H. (2003). Raman spectroscopy for optical diagnosis in normal and cancerous tissues of the nasopharynx-preliminary findings. *Laser in Surgery and Medicine* Vol. 32, pp. 210-214

50 Lau D.P., Huang Z., Lui H., Anderson D.W., Berean K., Morrison M.D., Shen L., and Zeng H. (2005). Raman spectroscopy for optical diagnosis in Larynx-preliminary findings. *Laser in Surgery and Medicine* Vol.37, pp. 192-200

51 Rabah R., Weber R., Serhatkulu G. K., Cao A., Dai H., Pandya A., Naik R., Auner G., Poulik J., and Klein M. (2008). Diagnosis of neuroblastoma and ganglioneuroma using Raman spectroscopy. *Journal of pediatric surgery* Vol. 43(1), pp. 171-176

[52] Wills H., Kast R., Stewart C., Rabah R., Pandya A., Poulik J., Auner G., and Klein M. (2009). Raman spectroscopy detects and distinguishes neuroblastoma and related tissues in fresh and (banked) frozen specimens. *Journal of pediatric surgery* Vol. 44, pp. 386-391

[53] Wills H., Kast R., Stewart C., Sullivan B., Rabah R., Poulik J., Pandya A., Auner G., and Klein M. (2009). Diagnosis of Wilms' tumor using near-infrared Raman spectroscopy. *Journal of pediatric surgery* Vol. 44(6), pp. 1152-1158

[54] Ghatak (2002). *Optics: 2nd edition*, Tata McGraw-Hill Ltd., New Delhi, India

[55] Farrell T.J. and Patterson M.S. (1992). A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *J Med. Phys*. Vol. 19(4), pp. 879-888

[56] Zonios G., Perelman L.T., Backman V., Manoharan R., Fitzmaurice M., Dam J.V., and Feld M.S. (1999). Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Appl. Optics*. Vol. 38(31), pp. 6628-6637

[57] Zonios G., Bykowski J. and Kollias N. (2001). Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy. *J Invest. Dermatol*. Vol. 117(6), pp. 1452-1457

[58] Schwarz R.A., Arifler D., Chang S.K., Pavlova I., Hussain I.A., Mack V., Knight B., Richards-Kortum R., and Gillenwater A.M. (2005). Ball lens coupled fiber-optic probe for depth-resolved spectroscopy of epithelial tissue. *Optics Letters* Vol. 30(10), pp. 1159-1161

[59] Schwarz R.A., Gao W., Daye D., Williams M.D., Richards-Kortum R., and Gillenwater A.M. (2008). Autofluorescence and diffuse reflectance spectroscopy of oral epithelial tissue using a depth-sensitive fiber-optic probe. *Appl. Optics* Vol. 47(6), pp. 825-834

[60] Katika K.M. and Pilon L. (2006). Steady-state directional diffuse reflectance and fluorescence of human skin. *Appl. Optics* Vol. 45(17)

[61] Perelman L.T., Backman V., Wallace M., Zonios G., Manoharan R., Nusrat A., Shields S., Seiler M., Lima C., Hamano T., Itzkan I., Dam J.V., Crawford J.M., and Feld M.S. (1998). Observation of Periodic Fine Structure in Reflectance from Biological Tissue: A New Technique for Measuring Nuclear Size Distribution. *Physical Review Letters* Vol. 80(3), pp. 627-630

[62] Breslin T.M., Xu F., Palmer G.M., Zhu C., Gilchrist K.W., and Ramanujam N. (2003). Autofluorescence and Diffuse Reflectance Properties of Malignant and Benign Breast Tissues. *Annals of Surgical Oncology* Vol. 11(1), pp. 65–70

[63] Brown J.Q. Wilke L.G.,Geradts J., Kennedy S.A., Palmer G.M., and Ramanujam N. (2009). Quantitative Optical Spectroscopy: A Robust Tool for Direct Measurement of Breast Cancer Vascular Oxygenation and Total Hemoglobin Content In vivo. Cancer Res Vol. 69(7), pp. 2919-2926

[64] Kollias N., Zonios G., and Stamatas G.N. (2002). Fluorescence spectroscopy of skin. *Vibrational Spec* Vol. 28, pp. 17–23

[65] Hamzavi I., Shiff N., Martinka M., Huang Z., McLean D., Zeng H., and Lui H. (2006). Spectroscopic assessment of dermal melanin using blue vitiligo as an in vivo model. *Photodermatol Photoimmunol Photomed* Vol. 22, pp. 46–51

[66] Na R., Stender I-M., Henriksen M., and Wulf H.C. (2001). Autofluorescence of Human Skin is Age-Related After Correction for Skin Pigmentation and Redness. *J Invest Dermatol* Vol. 116(4), pp. 536-540

[67] Ferraro J.R., Nakamoto K., and Brown C.W (2003). *Introductory Raman spectroscopy*. Academic Press

[68] Schorader B. (1995). *Infrared and Raman Spectroscopy*. VSH. Publishers, Inc., New York

[69] Wallace M.B., Wax A., Roberts D. N. and Graf R. N. (2009). Reflectance Spectroscopy. *Gastrointest Endosc Clin N Am*. Vol. 19(2), pp. 233–242

[70] Colthup N.B, Daly L.H., and Wibley S.E. (1996). *Introduction to infrared and Raman spectroscopy*. Academic Press

[71] Long D.A. (2002). The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules. ISBNs: 0-471-49028-8 (Hardback); 0-470-84576-7 (Electronic) John Wiley & Sons Ltd

[72] Laserna J.J. (1996). *Modern techniques in Raman spectroscopy*. John Wiley & Sons

[73] Hapke B. (1993). *Theory of Reflectance and Emittance Spectroscopy*. Cambridge University Press, UK

[74] Frei R.W. and MacNeil J.D. (1996). *Diffuse reflectance spectroscopy in environmental problem-solving*. CRC Press

[75] Young A.R. (1997). Chromophores in human skin. Phys. Med. Biol. Vol. 42, pp. 789–802

[76] Stamatas G.N., Zmudzka B.Z., Kollias N., and Beer J.Z. (2004). Non-invasive measurements of skin pigmentation in situ. *Pigment Cell Res.* Vol. 17(6), pp. 618-626

[77] Kollias N. and Baqer A. (1985). Spectroscopic characteristics of human melanin in vivo. *J Invest Dermatol.* Vol. 85(1), pp. 38-42

[78] Stamatas G.N., Zmudzka B.Z., Kollias N., and Beer J.Z. (2008). In vivo measurement of skin erythema and pigmentation: new means of implementation of diffuse reflectance spectroscopy with a commercial instrument. *Br J Dermatol.* Vol. 159(3), pp. 683-690

[79] http://www.doitpoms.ac.uk/tlplib/raman/method.php

[80] Selecting a CCD camera for Spectroscopic Applications, Jobin Yvon Inc, USA

[81] In Via Raman microscope user guide, M-9836-0797, Issue 1.0, Renishaw plc. UK

[82] Inoue S. and Spring K.R. (1997). VideoMicroscopy: the Fundamen*tals, 2nd Edition.* Plenum, New York

[83] Piston D.W. (1998). Concepts in Imaging and Microscopy: Choosing objective lens: The importance of numerical aperture and magnification in digital optical microscopy. *Biol. Bull*. Vol. 195, pp. 1-4

[84] Specialists in Spectroscopy, Optical Spectroscopy Division: TRIAX Series, Jobin Yvon Inc, USA

[85] Renishaw user guide issue 1.0, Renishaw plc, UK

[86] The USB 2000 spectrometer user guide. Ocean Optics. 2011

[87] Preciae color communication, Color control from perception to instrumentation, Konica Minolta 1998

[88] Devpura S., Pattamadilok B., Syed Z.U., Vemulapalli P., Henderson M., Rehse S.J., Hamzavi I., Lim H.W., and Naik R. (2011). Critical comparison of diffuse reflectance spectroscopy and colorimetry as dermatological diagnostic tools for acanthosis nigricans: A chemometric approach. *Biomed. Opt. Express* Vol. 2, pp. 1664-1673

[89] Devpura S., Thakur J.S., Sarkar F.H., Sakr W.A., Naik V.M., and Naik R. (2010). Detection of benign epithelia, prostatic neoplasia, and cancer regions in radical prostatectomy tissues using Raman spectroscopy. *Vibrational Spec,* Vol. 53, pp. 227-232

[90] Lev R. and Stoward P.J. (1969). On the use of eosin as a fluorescent dye to demonstrate mucous cells and other structures in tissue sections. *Histochemistry and Cell Biology* Vol. 20(4), pp. 363-377

[91] Cao A., Pandya A.K., Serhatkulu G.K., Weber R.E., Dai H., Thakur J.S., Naik V.M., Naik R., Auner G.W., and Rabah R.J. (2007). A robust method for automated background subtraction of tissue fluorescence. *J of Raman Spectrosc*, Vol. 38, pp. 1199–1205

[92] Atencio J.A.S., Gullen E.E.O., Montiel S.V., Rodriguez M.C., Ramos J.C., Gutierrez J.L., and Martinez F. (2009). Influence of probe pressure on human skin diffuse

reflectance spectroscopy measurements. *Optical Memory and Neural Networks (Information optics)* Vol. 18(1), pp. 6-14

[93] Jolliffe I.T. (2002). *Principal Components Analysis* (Second edition), Springer-Verlag, New York, USA

[94] Klecka W.R. (1980). *Discriminant Analysis, Series: Quantitative applications in the social sciences*, Sage Publications Inc., Newbury Park

[95] Cooley W.W. and Lohnes P.R. (1971). *Multivariate Data Analysis*. John Wiley & Sons, Inc

[96] Sambamoorthi N. *Hierarchical Cluster analysis: Some basics and algorithms*. CRMportals Inc, NJ

[97] Anderberg M.R. (1973). *Cluster analysis for applications*. New York: Academic Press

[98] MacQueen J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, Vol 1, pp. 281-297

[99] Hartigan J.A. (1975). *Clustering algorithms*. New York: John Wiley and Sons

[100] Trybula W.J. (1998). *Data Mining and Knowledge Discovery*, Kluwer Academic Publishers, Boston, pp. 121–167

[101] Mazanec J., Melišek M., Oravec M., and Pavlovičová J. (2008). Support vector machines, PCA and LDA in face recognition. *Journal of ELECTRICAL ENGINEERING*, Vol 59(4), pp. 203-209

[102] Bishop C.M. (1995). *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford University Press

[103] Haykin S. (1998). *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: Macmillan College Publishing

[104] Smith L. (2003). An introduction to neural networks. University of Stirling, UK

[105] Rosipal R. and Krämer N. (2006). Overview and Recent Advances in Partial Least Squares. In: *Subspace, Latent Structure and Feature Selection Techniques*, Saunders C., Grobelnik M., Gunn S., and Shawe-Taylor J., eds. Berlin: Springer-Verlag

[106] Clayman G.L.L.S., Laramore G.E and Hong W.K. (2000). in *Head and Neck Cancer in Cancer Medicine* , (Eds: Bast R.C.K.D., Pollock R.E., Weichselbaum R.R., Holland J.F., Frei E.), Hamilton, ON: B.C. Decker Inc., pp. 1175-1220

[107] Lyng F.M., Faoláin E.Ó., Conroy J., Meabe A.D., Knief P., Duffy B., Hunter M.B., Byrne J. M., Kelehan P., and Byrne H.J. (2007). Vibrational spectroscopy for pathology from biochemical analysis to diagnostic tool. *Exp. Mol. Pathology* Vol. 82(2), pp. 121-129

[108] Caspers P.J., Lucassen G.W., Wolthuis R., Bruining H.A., and Puppels G.J. (1998). *In vitro* and *in vivo* Raman spectroscopy of human skin. *Biospectroscopy* Vol. 4(5), pp. S31-S39

[109] Lieber C.A., Majumder S.K., Billheimer D., Ellis D.L., and Mahadevan-Jansen A. (2008). Raman microspectroscopy for skin cancer detection in vitro. *Journal of Biomedical Optics* Vol. 13(2), 024013

[110] Krafft C., Sobottka S.B., Schackert G., and Salzer R. (2005). Near Infrared Raman spectroscopic mapping of native brain tissue and intracranial tumors. *Analyst.* Vol.130(7), pp. 1070-1077

[111] Devpura S., Thakur J.S., Sethi S, Naik V.M., and Naik R. (2012). Diagnosis of head and neck squamous cell carcinoma using Raman spectroscopy: tongue tissues. *J of Raman Spectrosc.* Vol. 43, pp. 490-496

[112] Qian W. and Krimm S. (1994). Vibrational analysis of glutathione. *Biolpolymers* Vol. 34, pp. 1377-1394

[113] Sundius T., Bandekar J., and Krimm S. (1989). Vibrational Analysis of crystalline triglycine. *J. Mol. Str.* Vol. 214, pp. 119-142

[114] Ravikumar B., Rajaram R.K., Ramankrishnan V. (2006). Raman and IR spectral studies of L-phenylalanine L-phenylalaninium dihydrogenphosphate and DL-phenylalaninium dihydrogenphosphate. *J. Raman Spectrosc.* Vol. 37, pp. 597-605

[115] Tsuboi M., Ueda T., Ushizawa K., Ezaki Y., Overman S.A., and Thomas G.J. Jr (1996). Raman tensors for the tryptophan side chain in proteins determined by polarized Raman microspectroscopy of oriented *N*-Acetyl-l-Tryptophan crystals. *J Mol. Struct.* Vol. 379, pp. 43-50

[116] Chuang C-H. and Chen Y-T. (2009). Raman scattering of L-tryptophan enhanced by surface plasmon of silver nanoparticles: vibrational assignment and structural determination. *J Raman Spectrosc*. Vol.40, pp. 150-156

[117] Miura T., Takeuchi H., and Harada I. (1989). Tryptophan Raman bands sensitive to hydrogen bonding and side-chain conformation. *J. Raman Spectrosc*. Vol. 20, pp. 667-671

[118] Hsu S.L., Moore W.H., and Krimm S. (1976). Vibrational spectrum of the unordered polypeptide chain: a Raman study of feather keratin. *Biopolymers* Vol. 15, pp. 1513-1528

[119] Tankiewicz A., Dziemianczyk D., Buczko P., Szarmach I.J., Grabowska S.Z., and Pawlak D. (2006). Tryptophan and its metabolites in patients with oral squamous cell carcinoma: preliminary study. *Advances in Medican Sciences* Vol. 51(suppl1), pp. 221-224

[120] Crowe D.L., Milo G.E., and Shuler C.E. (1999). Keratin 19 downregulation by oral squamous cell carcinoma lines increases invasive potential. *J Dent. Res.* Vol. 78(6), pp. 1256-1263

[121] Balm A.J., Hageman P.C., Doornewaard M.H.V., Groeneveld E.M., and Ivanyi D. (1996). Cytokeratin 18 expression in squamous cell carcinoma of the head and neck. *Eur. Arch. Otorbinolaryngol.* Vol. 253, pp. 227-233

[122] Bernier J., Domenge C., Ozsahin M., Matuszewka K., Lefebvre J., Greiner R.H., Giralt J., Maingan P., Rolland F., Bolla M., Cognetti F., Bourhis J., Kirkpatrick A., and Glabbeke M.V. (2004). Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl. J Med*. Vol. 350(19), pp. 1945-1952

[123] Cooper J.S., Pajak T.F., Forastiere A.A., Jacobs J., Campbell B.H., Saxman S.B., Kish J.A., Kim H.E. , Cmelak A.J., Rotman M., Machtay M., Ensley J.F., Chao K.S.C., Schultz C.J., Lee N., and Fu K.K. (2004). Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* Vol. 350(19), pp. 1937-1944

[124] Mason S.J. and Graham N.E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q. J. R. Meteor. Soc*. Vol. 128, pp. 2145-2166

[125] Oesterling J.E. (1991). Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *J Urol.* Vol. 145, No. 5, pp. 907-923

[126] Shroder F.H., Hugosson J., and Roobol M.J. (2009). Screening and prostate-cancer mortality in a randomized European study. *New Eng J Med.* Vol. 360, pp. 1320-1328

[127] Carroll P., Alnerstein P.C., Greene K., Babalan R.J., Carter H.B., Gann P.H., Han M., Kuban D.A., Sartor A.O., Stanford J.L. and Zletman A. (2009). In: *Prostate-Specific Antigen Best Practice Statement,* http://www.auanet.org/content/guidelines-and-quality-care/clinical-guidelines/main-reports/psa09.pdf

[128] Allsbrook W.C., Mangold K.A., Johnson M.H., Lane R.B., Lane C.G., and Epstein J.I. (2001). Interobserver reproducibility of Gleason grading of prostatic carcinoma: General Pathologists. *Human Pathology*. Vol. 32(1), pp. 81-88

[129] Bowsher W. and Carter A. (2006). *Challenges in prostate cancer* (Second edition), Blackwell Publishing, ISBN 978-4051-0752-5, Massachusetts, USA

130 Olumi A.F., Grossfeld G.D., Hayward S.W., Carroll P.R., Tlsty T.D., and Cunha G.R. (2001). Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res*. Vol. 59, pp. 5002–5011

131 Hayward S.W., Wang Y., Cao M., Hom Y.K., and Zhang B. (2001). Malignant transformation in a nontumorigenic human prostatic epithelial cell line. Cancer Res. Vol. 61, pp. 8135–8142

132 Niu Y.-N. and Zia S.-J. (2009). Stroma–epithelium crosstalk in prostate cancer. *Asian Journal of Andrology* Vol. 11, pp. 28–35

133 Devpura S., Thakur J.S. , Sethi S., Naik V.M., Sarkar F., Sakr W., and Naik R. (2012). Diagnosis of prostatic intraepithelial neoplasia in luminal cells using Raman spectroscopy. In: Intraepithelial Neoplasia (Ed Supriya Srivastava, INTECH)

134 Riley R.D., Heney D., Jones D.R. Sutton A.J., Lambert P.C., Abrams H.R., Young B., Wailoo A.J., and Burchill S.A. (2004). A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin. Cancer Res*. Vol. 10, pp. 4-12

135 Sheithauer B.W., Woodruff J.M., and Erlandson R.A. (1997), *Tumors of the peripheral nervous system: Atlas of tumor pathology*, Armed forces institute of pathology, Washington, D.C.

136 Faolain E.O., Hunter M.B., Byrne J.M., Kelehan P., McNamara M., Byme H.J., and Lyng F.M. (2005). A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vibrational Spectrosc.* Vol. 38, pp. 121-127

137 Devpura S., Thakur J.S., Sethi S, Naik V.M., and Naik R. (2012). Raman spectroscopic investigation of frozen and formalin-fixed paraffin processed tissues of pediatric tumors: neuroblastoma and ganglioneuroma. *J of Raman Spectrosc* (submitted, under review)

138 Kaplan L.A., Lau J.M. and Stein E.A. (1990). Carotenoid composition, concentrations, and relationships in various human organs. *Clin, Physiol. Biochem*. Vol. 8(1), pp. 1-10

139 Kiernan J.A. (2000)*.* Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do. *Microscopy Today* Vol. 00-1, pp. 8-12

140 Deng H., Bloomfeild V. A., Benevides J. M., and Thomas G. J. (1999). Dependence of the Raman signature of genomic B-DNA on nucleotide base sequence. *Biopolymers* Vol. 50, pp. 656-666

141 Kobayashi C., Monforte-Munoz H., Gerbing R.B., Stram D.O., Matthay K.K., Lukens J.N., Seeger R.C., and Shimada H. (2006). Enlarged and prominent nucleoli may be indicative of MYCN amplification. *Cancer.* Vol. 103(1), pp. 174-180

142 Tornoczky T., Semuen D., Shimada H., and Ambros I.M. (2007). Pathology of peripheral neuroblastic tumors: Significance of prominent nucleoli undifferentiated/poorly differentiated neuroblastoma. *Pathology Oncology Res.* Vol. 13(4), pp. 269-275

143 Elsner P. (1995). Chromametry: Hardware, measuring principles and standardization of measurements. In: Berardesca E, Elsner P, Maibach HI, eds. *Bioengineering of the skin : cutaneous blood flow and erythema*. Boca Raton: CRC Press pp. 247-252

144 Pattamadilok B., Devpura S. Syed Z.U., Vemulapalli P., Henderson M., Rehse S.J., Hamzavi I., Lim H.W., and Naik R. (2012). Quantitative skin color measurements: comparison between colorimetry and diffuse reflectance spectroscopy in acanthosis nigricans patients. *Photodermatology, Photoimmunology, Photomedicine.* (accepted and in production)

145 Schwartz R.A. (1994). Acanthosis nigricans. *J Am Acad Dermatol.* Vol. 31(1), pp. 1-19

146 Rendon M.I., Cruz P.D. Jr, Sontheimer R.D., and Bergstresser P.R. (1989). Acanthosis nigricans: a cutaneous marker of tissue resistance to insulin. *J Am Acad Dermatol.* Vol. 21(3 Pt 1), pp. 461-469

147 Darmstadt G.L., Yokel B.K., and Horn T.D. (1991). Treatment of acanthosis nigricans with tretinoin. *Arch Dermatol.* Vol. 127(8), pp. 1139-1140

**ABSTRACT**

**RAMAN SPECTROSCOPY AND DIFFUSE REFLECTANCE SPECTROSCOPY FOR DIAGNOSIS OF HUMAN CANCER AND ACANTHOSIS NIGRICANS**

by

**SUNEETHA DEVPURA**

**August 2012**

**Advisor**     : Dr. Ratna Naik

**Co-Advisor**  : Dr. Steven J. Rehse

**Major**       : Physics (Condensed Matter)

**Degree**      : Doctor of Philosophy

Cancer and diabetes are common chronic diseases in today's world causing numerous deaths in adults as well as children.  Most common types of cancers in adults include prostate, lung, breast, colorectal and head and neck squamous cell carcinoma, while among children; leukemia, and brain and central nervous system cancers are quite common.  In each of these cases, early detection of the cancer or disease dramatically increases the chances of successful treatment.  In recent years, there has been much interest in using Raman spectroscopy and diffuse reflectance spectroscopy as analytical optical spectroscopic methods for early diagnosis of diseases.  Raman spectroscopy can be used to measure changes in the bio-molecular composition of a tissue specimen, and diffuse reflectance spectroscopy can measure chromophores of the skin.  In this research, archived (formalin-fixed paraffin processed) tissues of head and neck squamous cell carcinoma, prostate, and pediatric tumors have been investigated using Raman spectroscopy.  We have utilized statistical methods such as principal component analysis (PCA) and discriminant function analysis (DFA) to analyze the spectral output and distinguish between normal and

cancerous tissues.  The results show cancerous tissues can be successfully distinguished from normal tissues in three cancer types in *ex vivo*.  However, due to loss of biochemical in the tissue processing (paraffinizing and deparaffinizing procedure), the prediction ability of the archived tissues are less compared to frozen tissues as observed in the pediatric tumor investigation.

We also investigated the diagnostic capability of diffuse reflectance spectroscopy and colorimetry on a skin disease, acanthosis nigricans *in vivo*.  The aim is to quantify and characterize the skin color change associated with acanthosis nigricans skin disease in insulin-resistant obese individuals.  We observe both the instruments can be utilized to detect acanthosis nigricans with more than 87% sensitivity and 94% specificity when combined with advanced chemometric methods.

# AUTOBIOGRAPHICAL STATEMENT

## SUNEETHA DEVPURA

| EDUCATION | | | |
|---|---|---|---|
| **INSTITUTION AND LOCATION** | **DEGREE** | **YEAR** | **FIELD OF STUDY** |
| Wayne State University, Detroit, MI | Ph.D. | May 2012 (GPA: 3.97) | Physics (Condensed Matter) |
| Wayne State University, Detroit, MI | M.Sc. | May 2010 (GPA: 3.96) | Physics (Condensed Matter) |
| University of Toledo, Toledo, OH | Ph.D. Student | Fall 2006-Summer 2007 (GPA: 4.0) (transferred to Wayne State University) | Physics (Condensed Matter) |
| University of Colombo, Colombo, Sri Lanka | B.Sc. (First Class Honors) | 2004 | Engineering Physics, minor in Math |

### *PROFESSIONAL EXPERIENCE*

**2010-2012**: Principal Investigator of the Protocol: Raman Spectroscopy for Cancer Detection in Saliva Samples

**2008-2012**: Principal Investigator of the Protocol: Raman Spectroscopy for Cancer Detection in Biological Tissues

**2008-2012**: Principal Investigator of the Protocol: Raman Spectroscopy for Cancer Detection using Pediatric Tumor Tissues

**2008- Aug 2011**: Graduate Research Assistant, Dept. of Physics and Astronomy, Wayne State University, MI, USA

**2007-2008**: Graduate Teaching Assistant, Dept. of Physics and Astronomy, Wayne State University, MI, USA

**2006-2007**: Graduate Teaching Assistant, University of Toledo, OH, USA

**Dec 2004-2006**: Instructor (Teaching Assistant), University of Colombo, Colombo, Sri Lanka

**Aug-Nov 2004:** Undergraduate Research Trainee, Rileys Ltd, Sri Lanka

### *AWARDS*

**Thomas C. Rumble University Graduate Fellowship award:** Fall 2011 to May 2012, Prestigious and highly competitive award for senior outstanding graduate students

**2nd prize for excellent poster presentation award:** 2012, 3rd Annual Graduate Research Exhibition, Wayne State University

**1st prize for excellent poster presentation award:** 2011, 2nd Annual Graduate Research Day, Dept. of Physics and Astronomy, Wayne State University

**2nd prize for excellent poster presentation award:** 2010, 2nd Annual Graduate Research Day, Dept. of Physics and Astronomy, Wayne State University

**P. P. Jayawickrama Memorial Scholarship in Physics:** 2001, University of Colombo**,** Sri Lanka for academic excellence in the first year of undergraduate studies

**Arthur Lambert Rupasinghe Scholarship in Physics:** 2002, University of Colombo**,** Sri Lanka for academic excellence in first and second year of undergraduate studies

### *PEER-REVIEWED PUBLICATIONS*

1. **Devpura S.**, Thakur J.S., *et al.* (**2012**). Raman spectroscopic investigation of frozen and formalin-fixed paraffin processed tissues of pediatric tumors: neuroblastoma and ganglioneuroma. *J of Raman Spectroscopy* (submitted, under review)

2. Pattamadilok B., **Devpura S.**, *et al.* (**2012**). Quantitative skin color measurements: comparison between colorimetry and diffuse reflectance spectroscopy in acanthosis nigricans patients. *Photodermatology, Photoimmunology, Photomedicine* (accepted and in production)

3. **Devpura S.**, Thakur J.S., *et al.* (**2012**) Diagnosis of Head and Neck Squamous Cell Carcinoma using Raman Spectroscopy: Tongue tissues. *J of Raman Spectroscopy* Vol. 43, 490-496, doi: 10.1002/jrs.3070

4. **Devpura S.**, Pattamadilok B., *et al.* (**2011**). Critical comparison of diffuse reflectance spectroscopy and colorimetry as dermatological diagnostic tool for acanthosis nigricans: A chemometric approach. *Biomedical Optics Express* Vol. 2(6), 1664-1673

5. **Devpura S.**, Thakur J.S., *et al* (**2010**) Detection of benign epithelia, prostatic intraepithelial neoplasia, and cancer regions in radical prostatectomy tissues using Raman spectroscopy. *Vibrational Spec.* Vol.53, 227-232

**CHAPTER OF BOOK: Suneetha Devpura,** *et al.* (**2012**) Diagnosis of Prostatic Intraepithelial Neoplasia in Luminal Cells Using Raman Spectroscopy, Intraepithelial Neoplasia, S. Srivastava (Ed.), ISBN: 978-953-307-987-5, InTech.