Abstract

Background: The main goal of this study was to determine the features from the Raman spectral data associated with skin roughness and to detect roughness from normal skin by using a proper classification method. 

Material and Methods: The Raman spectral dataset was constructed from two classes of spectral data, 70 spectra of normal intact skin and 70 spectra of irritated rough skin. Roughness irritation was induced by sodium dodecyl sulfate (SDS) non-ionic surfactant applied daily on rat skin for a week. The spectra were obtained from upper legs and dorsal regions. Some features related to specific bond vibrations of water, lipid and protein structures were extracted from the spectra. T-test statistical analysis was performed to determine whether the specified feature could discriminate two classes of spectral data. The reported efficient features from t-test analysis were applied to well-known classification methods such as Linear Discriminate Analysis (LDA) and K-nearest neighbors algorithm (KNN) for classification. Classification performance was calculated using k-fold cross validation method for selecting the proper classifier and features. The statistical analysis of water content and lipid structures between two classes showed a significant difference by p-value<<0.01, whereas alterations in features related to proteins were not remarkable between two classes of data. Water content and lipid structures were the appropriate features for skin roughness detection.

Results and Conclusion: The results from LDA and KNN for each extracted feature showed a maximum 80% accuracy and 76% specificity in classification of spectral data. In order to improve the efficiency of detection, combination of extracted features were applied to LDA and KNN classifier, which resulted in 85% accuracy and 87% sensitivity in classification.

Keywords: Raman Spectroscopy, T-test statistical analysis, feature extraction, classification, K-fold Cross validation.

Introduction

Skin is a protective layer of the body that accomplishes multiple defensive tasks. It limits penetration rate of external harmful substances such as chemical and microbes. It is also a barrier to excessive loss of water, plasma and proteins of living organs. Skin consists of three sub layers: epidermis, the outer layer of the skin, dermis the middle layer and hypodermis which is the most inner layer. Epidermis is a stratified epithelium, primarily consists of cornified cells called keratinocytes that form the strength of the skin.

Dermis, mainly consists of collagen proteins that cause skin resilience. Hypodermis contains a large amount of lipid cells. Epidermis is the most important layer to fulfill defensive tasks. Skin barrier function resides almost entirely in the epidermis and in particular in its superficial layer, stratum corneum(SC). The structure of stratum corneum consists of keratinocytes that are embedded in a matrix of lamellar lipids. This structure of lipids and keratinocytes are recognized as factors provide dominant role in barrier function of skin [1,2,3]. Impairment of skin barrier function is often demonstrated by an altered integrity of the stratum corneum and usually arises from direct damage to dysfunction of SC lipids.

The most common dermatological disorder is dry and flaky skin. This disorder arises from impairment in SC cell (corneocyte) maturation that causes a
defect in desquamation of corneocytes. This impairment finally cause decrease in water holding capacity [4-6].

Today, different non-invasive approaches are utilized for in-vivo monitoring of the skin barrier integrity. Novel methods such as in-vivo confocal Raman micro spectroscopy offer the possibility for precise and detailed characterization of the skin constituents together with their structures. This method has been well established as a powerful non-invasive method for studying molecular structures. It is also a unique tool in medicine for non-invasive and real-time analysis of biological tissues [7,8].

Raman Spectroscopy is an optical method based on inelastic light scattering that provide information about molecular composition as well as molecular interactions of the investigated sample[1,8]. This information is gained through gathering photons emitted from vibrated molecular bonds of the investigated sample and has been used in dermatology to study skin and its metabolism to diagnoses skin diseases. It has been demonstrated that changes in molecular structure and composition, following a disease, are reflected in Raman spectra.

The initial dermatological studies of Raman Spectroscopy were mainly about skin constituents structures and contents such as lipids, proteins and water [9,10]. Following the initial studies, most prevalent malignant and benign skin lesions such as Melanoma, Basal Cell Carcinoma, Pigmented Nevi, Squamous Cell Carcinoma, Kaposi Sarcoma and Eczema were studied using Raman spectroscopy [11-19]. Among the studies of Raman spectroscopy on skin, different methods were applied for automatic feature extraction and classification of spectral dataset in the process of skin lesion detection. These methods were applied base on well-known machine learning algorithms such as Principle Component Analysis (PCA), Independent Component Analysis (ICA), Support Vector Machine (SVM), Wavelet, Partial Least Square (PLS) and neural network. Only few studies were involved in extracting features regarding relevant Raman peaks and features associated with a specific skin constituent structure and content for detection and study of skin lesions[20]. Caspers [21] and their colleague studied water content and structure in malignant and benign tumors. They suggested a relevant Raman feature to measure water content over the Raman spectrum and concluded that water content might be changed due to a tumor. From this feature, it was concluded that water content will decrease in seborrheic keratosis, while this feature will not change substantially following pigmented nevi, melanoma and basal cell carcinoma. Some other Raman features were also applied to determine molecular concentrations such as urea and water [21,22]. The maintenance and protection role of carotenoids for living cells was elucidated in another study [23]. Carotenoids are pigmented cells can be found in the stratum corneum layer of the skin and play important role in anti-oxidant system.

Relationship between carotenoid concentration and cutaneous malignancy was demonstrated using Raman Spectroscopy[11]. Relevant extracted Raman feature for carotenoids concentration demonstrated substantially lower concentration in basal cell carcinoma, actinic keratosis and their perilesional skin.

In this in-vivo study, we investigated normal intact skin and dry and rough skin. Our main goal was to examine Raman features in two classes of Raman spectra and discern the influential defect in skin constituents that cause dry and roughness of skin.

We inspected the Raman features associated with lipid, water and protein contents and structure together with some other constituent like carotenoids. The most altered features were identified by T-test statistical analysis. This statistical method revealed the efficient features that show most significant differences between two classes of spectral dataset. Finally, we utilized classification methods to measure performance of roughness detection. This measure was the basis for decision-making.

In this method we applied extracted Raman feature to describe skin structures in two different skin conditions.

**Material and Methods**

**Experiments**

Experiments were performed on skin of Wistar male rat; 2 months old, 200+50 gr weights. A total of 140 Raman spectra were gathered from dorsal part of trunk and legs. Raman spectra were consisted of two groups. The first group was normal intact skin dedicated spectra and consisted of a total of 70 spectra. The second was the dry and rough impaired skin induced by applying non-ionic surfactant sodium dodecyle sulfate 20%. This non –ionic detergent is an irritant which has been widely used in dermatological studies leading irritancy for dry and rough skin [24, 25, 26, 27]. This detergent had been applied to the same anatomical regions on a daily basis for a week. After a week skin roughness
and dryness was observed. A total of 70 Raman spectra were also gathered from irritated skin of the same anatomical regions.

**Instrumentation**

Raman spectra were recorded with a ThermoNicolet Raman spectrometer manufactured by Almega corporation. The incidence laser type was a continuous Nd:YLF laser at wavelength 532nm, with 40mW power and focused at 100µm. In total, 32 scans were accumulated with a resolution of 4 cm\(^{-1}\) for each measurement. The total registration time was approximately 3 minutes for each spectrum.

**Preprocessing of spectral data**

Before applying the Raman spectra for this study, some preprocessing steps were performed on the raw spectra. These preprocessing steps included baseline correction, smoothing and normalization with respect to maximum Raman intensity. These tasks were carried out using the codes written in MATLAB software. Baseline correction was implemented by forth order polynomial fit on each spectrum and savitzky golay low pass filter was applied for smoothing.

**Intuitive Data analysis and Discussion**

A typical Raman spectrum of normal skin after preprocessing is shown in Figure 1. This spectrum shows Raman characteristics of epidermis and dermis constituents discussed in the introduction section. Raman characteristics correspond with a specific bond vibration associated with a specific molecular structure. As we discussed in the introduction, epidermis and mainly its superficial layer - stratum corneum, mostly consist of keratinocytes which account for the strength of skin. These protein cells have the corresponding Raman vibrational peak at 2940cm\(^{-1}\) that is obviously seen in this figure. These cells are surrounded by bilayer lipid cells. Lipids in the stratum corneum layer have the dominant stretching vibration of C=O bond that can be seen at 1750cm\(^{-1}\), and C-C vibrations at 1100cm\(^{-1}\) shown in Figure 1. Figure 2.a shows C-C acyl chain configurations which forms the skeleton of the lipid cell. Figure 2.b shows Raman peak of each configuration. Skeletal C-C acyl chain in lipids have Raman peak at 1095 cm\(^{-1}\), in case the structure of C-C chain is gauche as the formation in soft skin and if the acyl chain has all-trans form, as the formation in rough skin, hair and nail, the Raman vibration would shift to the 1030cm\(^{-1}\) and 1130cm\(^{-1}\) vibrational regions. There is another Raman attributed mode to lipid which is due to the lateral intermolecular structure of proximity lipid cells. This Raman characteristic is seen at 2890cm\(^{-1}\) due to asymmetric stretching of C-H bonds of adjacent cells and 2850cm\(^{-1}\), which is due to symmetric stretching of C-H bond.

Collagen attributed Raman characteristic are mainly caused by Amid I amino acid vibration which can be seen at 1654cm\(^{-1}\) and Amid III amino acid with the corresponding Raman mode at 1270cm\(^{-1}\). These protein cells reside mostly in the Dermis layer of the skin. Therefore, we concluded we gathered the related information of both epidermis layers by the thickness of about 100-200µm and a part of dermis. Collagen is also rich in proline and hydroxyl amino acids that can be identified at 921, 859 and 816cm\(^{-1}\) spectral region. Table 1, shows the corresponding Raman modes of skin structures, which is utilized for skin structure studies.

Table 1: Raman vibrational modes of skin constituents

![Figure 1: A typical Raman spectrum of Rat skin](https://example.com/figure1.png)

![Figure 2: C-C acyl chain configurations](https://example.com/figure2.png)

<table>
<thead>
<tr>
<th>Raman Shift(cm(^{-1}))</th>
<th>Protein</th>
<th>Fat</th>
<th>Other Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>829</td>
<td>δ(CH(_3))</td>
<td>v(O-H) Water</td>
<td></td>
</tr>
<tr>
<td>2942</td>
<td>δ(CH(_2))</td>
<td>δ(CH(_3))</td>
<td></td>
</tr>
<tr>
<td>2890</td>
<td>δ(CH(_2))</td>
<td>δ(CH(_3))</td>
<td></td>
</tr>
<tr>
<td>2859</td>
<td>δ(CH(_2))</td>
<td>δ(CH(_3))</td>
<td></td>
</tr>
<tr>
<td>1745</td>
<td>v(C-O)</td>
<td>v(C-O)</td>
<td></td>
</tr>
<tr>
<td>1654</td>
<td>v(C=O) Amide I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1414</td>
<td>δ(CN), δ(CH(_2))</td>
<td>δ(CH(_3)) edocvoring</td>
<td></td>
</tr>
<tr>
<td>1304</td>
<td>v(C(_2))</td>
<td>δ(CH(_2)) Twisting</td>
<td></td>
</tr>
<tr>
<td>1279</td>
<td>v(C(_2)), v(NH(_2)), Amide III</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1089</td>
<td>v(CC), v(CC), amide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1030</td>
<td>v(CC), v(CC), skeletal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>937</td>
<td>v(CC), Proline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>853</td>
<td>δ(C=CH(_3)), Aromatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>820</td>
<td>δ(C=C), Amid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Collagen provides water, v(C=O), v(CH\(_2\)), v(CH\(_3\)), v(C\(_3\)), v(C\(_4\)), v(C\(_5\)), and v(C\(_6\)) for hydration and hydration of the skin, which is associated with various skin properties such as firmness, elasticity, and moisture content. The amino acid structure in collagen provides a rich source of hydrogen bonding, which contributes to the overall hydration of the skin tissue. The highly branched and flexible nature of collagen allows it to maintain its structure and function in a hydrated environment.

Keratin provides v(C=O), v(CH\(_2\)), v(CH\(_3\)), v(C\(_3\)), v(C\(_4\)), v(C\(_5\)), and v(C\(_6\)) for the formation of the skin matrix, which provides strength and elasticity. The amino acid structure in keratin provides a rich source of hydrogen bonding, which contributes to the overall strength and function of the skin tissue. The highly branched and flexible nature of keratin allows it to maintain its structure and function in a hydrated environment.

![Collagen and Keratin](https://example.com/collagen-keratin.png)
Extracted features
In order to find out the altered structure or content between the specified normal and impaired skin, we clarified the relevant variables that revealed skin constituent concentration or molecular structure. The variables have been used to study water, lipids and proteins of skin in following.

Lipids Intra and inter molecular structure
As it is mentioned, there are two main characteristics for lipid structure in skin superficial layer. The first was related to intramolecular lipid cell structure, which is called skeletal lipid structure and its structure is assigned with the Raman vibrational mode in 1000-1200cm$^{-1}$ spectral region.

The parameter, which we defined to assess the intramolecular lipid state, was the wave number where the C-C stretching vibration occurred. This stretching vibration wave number varies between 1130cm$^{-1}$, for all trans configuration C-C structure which is ascribed to skin stiffness, and 1100cm$^{-1}$ that reflect the gauche transformation of C-C chain in well moisturized normal skin. Figure 2 displays the Raman vibrational modes of lipids in skin, hair and nail.

The second characteristic of lipid is related to the interaction of two adjacent lipid cells that determine the intermolecular structure of lipid cells. The ratio of the symmetric and asymmetric methylene C-H stretching modes intensities in lipids is sensitive to the packing order of the acyl chains. This finding was introduced by Gaber and peticolas in their study of biomembrane lipid cells.[28]. They developed a relevant parameter to interpret the intermolecular lipid cells structure shown in equation (1), (2). If the ratio of Raman intensity at 2890 cm$^{-1}$ to 2850cm$^{-1}$ is 2.2, the $S_{lat}$ will become one, revealing that the phospholipid cells are in the crystal state, and the $S_{lat}$ parameter will approach zero if the ratio decline to 0.7. We have utilized this parameter to distinguish the phospholipid layer state. This parameter approaches one in stiff and rough skin as the state of nail and hair phospholipids and approaches zero if the skin is soft and moisturized.

$$S_{lat} = \frac{I_{2890 \text{ Raman Intensity}}}{I_{2850 \text{ Raman Intensity}}}$$

Water content
The ratio of vibrational intensities of C-H bond ascribed to proteins around 2940cm$^{-1}$ and O-H stretching bands ascribed to water around 3250-3400cm$^{-1}$ is related to hydrogen bonding or hydration between protein and water[29]. Therefore, we determined skin hydration or water content by equation (3). The concept of this equation is depicted in Figure 3, which propose the ratio for relative water content of the tissue.

$$\text{Water Content} = \frac{I_{2940 \text{ Raman Intensity}}}{I_{3250 \text{ Raman Intensity}}}$$

Structure and concentration of proteins
We have utilized the Raman intensities at 1650cm$^{-1}$ and 1350cm$^{-1}$ as measures for collagen protein evaluation. These intensities are normalized values between zero and one.
Carotenoids

Carotenoids are thought to play a significant role in skin’s anti-oxidant defense system and may help prevent malignancy. Carotenoids are lipophilic molecules which are abounded the skin lipids to act as chain-breaking anti-oxidants. In previous studies, total carotenoid concentration was represented by Raman intensity at 1524cm\(^{-1}\) and 1159cm\(^{-1}\) [30]. Raman intensity at 1570cm\(^{-1}\) was ascribed to carotenoids in our study. There were not prominent Raman peak at 1159cm\(^{-1}\).

Statistical analysis

We have introduced some parameters related to skin constituent’s content or structure, which will alter in case of impairment. In order to distinguish the disorder and related affected malfunctioning skin molecular structure, a t-Test statistical analysis was performed for each parameter.

Classification methods

Efficient reported parameters from t-test were applied for training and testing of two classifier systems, LDA and KNN. K-fold cross validation by K=5 was chosen to examine the performance of the classifiers. The k factor in the k nearest neighbor was optimal for k>10, so we choose k=10 for this factor. The best and efficient method was decided on the basis of error, sensitivity and specificity parameters.

Results

Statistical analysis results

The results of statistical analysis for the extracted parameters of the two classes are summarized in the Table 2. Significant level (\(\alpha\)) was considered at 0.05.

The water contents were different between normal and irritated rough skin. Box plot of this parameter is shown in Figure 4-a, reveals water content in rough skin was decreased in compare to normal skin.

![Figure 4: parameters values box plot a) Water content, b) \(S_{lat}\)](image)

The same box plot is depicted for \(S_{lat}\) parameter in Figure 4-b. \(S_{lat}\) parameter is approaching one, asserting that interaction of lipid’s C-H bond vibration is mostly symmetric which suggests the stiffness of lipid layers in irritated rough skin spectrum class. The conclusions from the parameters analysis could be seen on Raman spectra. A number of Raman spectra of two classes in the spectral region 2800-3500cm\(^{-1}\) are shown in Figure 5. Figure 5.a , normal intact skin and Figure 5.b is the spectra of irritated rough skin. This region is dedicated to water content and lipids inter-molecular structure. The ratio of Raman intensities in the region 3250-3450cm\(^{-1}\) to 2940cm\(^{-1}\) (water content parameter) in figure 5.b is decreasing compared to figure 5.a. Parameter \(S_{lat}\) is also inferred from this spectral region. It is vividly seen that the asymmetric vibration of the lipid C-H bond is disappeared in the spectra of class 2(rough skin), but the symmetric vibrations still exists.

```
<table>
<thead>
<tr>
<th>parameter</th>
<th>p.value</th>
<th>Result of hypothesis test</th>
<th>conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Content</td>
<td>&lt;&lt;0.01</td>
<td>Null hypothesis rejected</td>
<td>Water content is altered in rough skin in comparison to normal skin</td>
</tr>
<tr>
<td>Intramolecular lipid structure</td>
<td>&lt;&lt;0.01</td>
<td>Null hypothesis rejected</td>
<td>Intramolecular Lipid structure is changed</td>
</tr>
<tr>
<td>Intermolecular lipid</td>
<td>&lt;&lt;0.0</td>
<td>Null hypothesis</td>
<td>Intermolecular lipid is changed</td>
</tr>
</tbody>
</table>
```
Examining the skin spectra of these two classes in spectral region of 1000-1200cm\(^{-1}\) ascribed to C-C vibration of lipid cells reflects the results obtained from t-test analysis. Some spectra of these classes are shown in Figure 6. The stiffness of the skin in class 2 is concluded since the C-C Raman vibational mode shift from 1100 to higher wave numbers. The dotted line in figure 6 shows the concept.

Classifiers results

Four parameters were affected because of skin roughness. These parameters were utilized to classify and measure classification specificity, sensitivity and error. In order to examine impact level of each parameter to distinguish the impairment, each parameter was applied to classifiers separately. Classification result with each parameter is shown in Figure 7. The parameters labeled from 1 to 4 are water content, \(S_{\text{lat}}\), lipid’s C-C structure and carotenoids, respectively.

Combinations of the selected parameters were employed for classification, in order to improve classification performance. LDA had a better performance compare to KNN classifier (Figure 7).

Raman intensity at 1650cm\(^{-1}\) and 1370cm\(^{-1}\) were the other parameters chosen to study skin proteins. The statistical analysis did not show any substantial difference between these two classes. As a result, this structure was not affected due to the impairment.
combination of all four features had the best performance with error rate 15%, specificity of 81% and sensitivity of 87%. It is seen that classification with the parameters that include water content feature has better performance; hence, it is the prominent feature that is affected due to this impairment.

**Conclusion**

In this study, we digitized related Raman vibrational modes with a parameter value that reflects the characteristics of its related molecular structure. The advantages of this method over the previous machine learning tools are more precision in fully probe into molecular structures and interpretation of the assumed alteration. The statistical analysis of water content and lipid structures between two classes showed a significant difference by p-value<<0.1, whereas alterations in features related to proteins were not remarkable between two classes of data. These results reveal that skin roughness leads to changes in water content and lipid structures.

Classifiers performance was best for the series of features included water content, suggested that water content feature was the most altered feature amongst two classes. However, the combination of all four extracted features presented better performance and therefore it was better feature to access skin roughness and dryness.

**References**


