Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 190 (2018) 409–416

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Near-infrared Raman spectroscopy for estimating biochemical changes associated with different pathological conditions of cervix

Amuthachelvi Daniel *, Aruna Prakasarao, Singaravelu Ganesan

Department of Medical Physics, Anna University, Sardar Patel Road, Chennai 600025, India

ARTICLE INFO

Article history:
Received 13 April 2017
Received in revised form 23 July 2017
Accepted 8 September 2017
Available online 18 September 2017

Keywords:
Raman spectroscopy
Artificial Neural Network
Biochemical modeling
PCA-LDA

ABSTRACT

The molecular level changes associated with oncogenesis precede the morphological changes in cells and tissues. Hence molecular level diagnosis would promote early diagnosis of the disease. Raman spectroscopy is capable of providing specific spectral signature of various biomolecules present in the cells and tissues under various pathological conditions. The aim of this work is to develop a non-linear multi-class statistical methodology for discrimination of normal, neoplastic and malignant cells/tissues. The tissues were classified as normal, premalignant and malignant by employing Principal Component Analysis followed by Artificial Neural Network (PC-ANN). The overall accuracy achieved was 99%. Further, to get an insight into the quantitative biochemical composition of the normal, neoplastic and malignant tissues, a linear combination of the major biochemicals by non-negative least squares technique was fit to the measured Raman spectra of the tissues. This technique confirms the changes in the major biomolecules such as lipids, nucleic acids, actin, glycogen and collagen associated with the different pathological conditions. To study the efficacy of this technique in comparison with histopathology, we have utilized Principal Component followed by Linear Discriminant Analysis (PC-LDA) to discriminate the well differentiated, moderately differentiated and poorly differentiated squamous cell carcinoma with an accuracy of 94.0%. And the results demonstrated that Raman spectroscopy has the potential to complement the good old technique of histopathology.

1. Introduction

Cancer is one of the major economic burdens worldwide and is one of the leading causes of mortality. Cancers originating from the epithelial layer account for nearly 85% of all cancers. The most common cancers affecting women worldwide are cancer of breast, colorectal, lung and cervix [1]. The morbidity due to cervical cancer is the second most common and the fourth leading cause of cancer mortality among women globally and generally more common in developing countries [1]. The five-year survival rate of patients after the diagnosis and treatment at stage I is 80–93%, stage II is 58–63%, stage III is 32–35% and 15% for stage IV. [2] Hence early diagnosis and appropriate treatment are the two indispensable measures to effectively combat the mortality due to cervical cancer. Yet the early diagnosis possesses a great challenge since cervical cancer shows no signs or symptoms at early stage and only at advanced stages, symptoms begin to appear.

Generally, Papanicolaou (Pap) smear followed by colposcopy and colposcopy-directed biopsy is the common protocol for clinical diagnosis of cervical neoplasia in practice. Pap smear is the cytology-based screening test for cervical cancer to detect neoplastic/malignant cells, the potential precursors of cervical cancer. The white-light colposcopy is the common gynaecology follow-up procedure for abnormal Pap smears. Histopathological study is still regarded as the golden standard for cervical cancer diagnosis though it is subjective [3]. All these diagnostic modalities are based on either morphological or structural changes at cellular or tissue level. It is of immense importance to develop advanced screening and diagnostic techniques associated with biochemical changes at molecular level for improving existing diagnostic sensitivity and specificity of early cervical cancer detection and understand the reasons for transformation of normal into malignant cells and tissues.

In recent years optical techniques have attracted considerable attention for molecular level diagnosis of cells and tissues. One of the optical techniques is Raman spectroscopy, a vibrational spectroscopic technique that measures inelastic light scattering processes. This technique provides specific spectroscopic fingerprints of biomolecules which could be utilized for cell or tissue characterization. The technique has been used with promising results for both *in vitro* and *in vivo* tissue diagnosis of cancer in a number of organs such as skin [4], breast [5], oral [6,7], larynx [8], esophagus [9], gastric [10], colon [11], lung [12], bladder [13], prostate [14], cervix [15–18], and brain [19].

Many researchers have already reported discrimination of cervical cancer and neoplasia from normal samples [15–18].To name a few:

* Corresponding author.
E-mail address: amuthaleo@yahoo.com (A. Daniel).

https://doi.org/10.1016/j.saa.2017.09.014
1386-1425/

Anita Mahadevan-Jansen et al. has reported a sensitivity of 82% and a specificity of 92% from 36 cervical tissue biopsies from 18 patients and distinguished neoplasia from benign tissue [20]. Lyng and her coworkers analyzed the Raman spectra of formalin-fixed paraffin preserved histological samples of normal, neoplastic and invasive carcinoma tissues [21]. SK. Sharma has reported an *in vitro* study of seven cervical tissues of normal and invasive carcinoma. They had both the fingerprint and high wavenumber region [22]. Raman spectra of cervical cancer and normal cell lines were statistically analyzed by Jess et al. [23]. CM Krishna has also carried out extensive study of cervical tissues by Raman spectroscopy [24,25]. Further, Mahadevan-Jansen et al. and Huang et al. carried out *in vivo* investigations for differentiating neoplastic cervix from normal cervix. Huang and his group have also illustrated the feasibility of HW Raman spectroscopy for *in vivo* identification of cervical neoplasia with a diagnostic sensitivity of 93.5% and a specificity of 97.8% [26]. In addition to the above mentioned studies several other works were also carried out for the tissues under *in vitro* and *in vivo* conditions by NIR Raman spectroscopy and have indicated that this technique has a promising role in clinical medicine as a non-invasive tool for detection of early cancer. Yet more light has to be thrown for deeper understanding of changes at molecular level and for better classification accuracy.

In this context, the aim of the present study was to evaluate the possibility of classifying normal, neoplasia and malignant tissues and for deeper understanding of the biochemical changes accompanying the transformation of normal cells into neoplastic and then malignant cells, biochemical modeling based on Raman spectroscopic signatures was also carried out. Further, statistical techniques including Principal component analysis (PCA) for dimension reduction followed by classification algorithm based on Artificial Neural Network (ANN) was also performed. The predictive performance of the aforementioned statistical algorithm was also tested. In addition, to classify within the malignant samples, principal component analysis followed by linear discriminant analysis (PC-LDA) was performed to validate the histopathological staging.

2. Materials and Methods

2.1. Sample Preparation

Freshly excised cervical tissue samples from normal (n = 64), neoplasia (n = 36) and malignant (n = 145) were obtained from the Aringar Anna Cancer Hospital and Research centre (Kancheepuram, India) with previous consent of the patient and ethical clearance from the hospital. These samples were stored in normal saline and transported in liquid nitrogen container.

2.2. Raman Instrumentation and Spectral Acquisition

Raman spectra were recorded with a LabRam microspectrometer (Horiba Jobin Yvon, Lille, France). The excitation beam from a diode laser 784.12 nm was focused on the sample using a 50× long working distance objective (Numerical Aperture = 0.5). Rayleigh elastic scattering was removed by an edge filter. For these experiments, the confocal hole was set to 400 μm. The Raman Stokes signal was dispersed with a holographic grating (600 grooves/mm) and spectra were recorded using a Peltier cooled charge-coupled device (CCD) as a detector (1024 × 256 pixels). Every tissue sample was mounted on a quartz slide and the spectra were acquired in the wavenumber region of 600–1800 cm^{-1} for a time duration of 60 s per window and the spectra were acquired two times to remove cosmic showers.

2.3. Data Processing

Every Raman instrument has a unique spectral responsiveness; hence, to correct this, an intensity calibration factor was applied to every raw spectrum. Then, the spectra were baseline corrected and smoothened by Savitzky-Golay filter using the Labspec 5 software provided by Horiba Jobin Yvon. The schematic diagram of the process is given in Fig. 1. Following this the Raman spectra were statistically analyzed by PCA and ANN under Matlab platform. The dimension of the dataset was reduced by PCA and the thus obtained 16 Principal Components (PCs) which have accounted for 98.5% of the variance were given as input to ANN.

2.3.1. Artificial Neural Network (ANN)

ANN is a technique that is commonly applied to solve data mining applications. The structure of neural network provides an opportunity to the user to implement parallel concept at each layer level. Another significant characteristic of ANN is fault tolerance. ANNs are well suited in situations where information is noisy and uncertain. ANN are an information processing methodology that differs drastically from conventional methodologies, in that it employs training by examples to solve problem rather than a fixed algorithm [27]. They could be divided into two types based on the training method: Supervised training and unsupervised training [28].

The most popular neural network algorithm is back-propagation algorithm. Although many types of neural networks can be used for classification purposes, the focus is on the feed forward multilayer networks or multilayer perceptrons which are the most widely studied, used and popular neural network classifiers as they are effective, easy-to-learn model for complex and multi-layered networks. Its greatest strength is in providing non-linear solutions to ill-defined problem. A neural network can have a multitude of designs; in this study we have employed the feed forward two layer network that includes a) inputs, b) a hidden layer and c) an output layer. The input units are the first layer of neurons or units which simply represent the obtained PCs.

2.3.2. Biochemical Modeling

In earlier studies, the biochemical model was developed for bladder, heart, breast and cervix [29,30]. Similarly, in this study, biochemical modeling was carried out for the different pathological conditions of cervix, where the spectra of the raw chemicals were recorded, vector normalized and were fitted to a linear combination of major biochemicals by non-negative least squares technique (NNLS).

$$\mathbf{S} = \mathbf{CB} + \mathbf{E}$$

where **S** is the measured spectra of the tissue, **B** is the matrices of spectral components, **C** is the concentration of the various components which has to be computed and **E** gives the residual. Therefore by employing NNLS it is aimed to achieve a linear 'best fit' of the spectral components with minimum residuals. The NNLS was carried out on the assumption that the linear combination of the selected biochemical's spectral components describes the entire measured spectra and the signal intensity scales linearly with the concentration of biochemical components in the tissue. The developed biochemical model determines the relative concentration profiles of major tissue biochemical constituents which accompany the genesis and progression of the disease. This model fits the measured tissue spectra with the basis spectra obtained from the selected biochemicals with least residuals. The major biomolecules selected for fitting were actin, DNA, histone, collagen, glycogen, protein - human serum albumin, Phosphatidylcholine, glycerol triolate, cholesterol, and β-carotene and may be attributed as the major components present in normal, neoplastic and malignant cervix. The matrix **C** gives the relative concentration of every biochemical component in the particular tissue. Thus the molecular changes associated with the different pathological conditions could be gleaned. The fitting coefficients for each pathology rendered using NNLS model were further constrained to a coefficients sum of 100% to estimate the