Optik 157 (2018) 565–570

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Optical screening of nasopharyngeal cancer using Raman spectroscopy and support vector machine

Saranjam Khan [a,*], Rahat Ullah [a,*], Shaheen Shahzad [b], Samina Javaid [b], Asifullah Khan [c]

[a] National Institute for Lasers and Optronics, Nilore, Islamabad 45650, Pakistan
[b] Islamic International University Islamabad, 44000, Pakistan
[c] Pattern Recognition Lab, DCIS, Pakistan Institutes of Engineering and Applied Sciences (PIEAS), Nilore, Islamabad 45650, Pakistan

ARTICLE INFO

Article history:
Received 3 October 2017
Accepted 18 November 2017

Keywords:
Raman spectroscopy
SVM
Nasopharynx cancer
Biomolecules

ABSTRACT

In this study we present the application of Raman spectroscopy combined with support vector machine (SVM) for the characterization of Nasopharyngeal Cancer (NPC) in the human sera. In total, 14 pathological samples from confirmed NPC patients and 15 normal sera samples from healthy individuals have been collected and used in this study. Raman system with 785 nm laser has been used for recoding Raman spectra of both NPC and normal sera samples. Subtle intensity variations have been observed at certain Raman bands between both groups. Principal Component Analysis (PCA) has been used for features transformation and for highlighting spectral differences, while SVM is employed for discriminating cancerous non-cancerous samples. SVM has the ability to transforms linearly inseparable data to high dimensional space, by using mathematical function (kernel) where it may be linearly separable. An SVM model with different kernel functions i.e. polynomial kernel and Gaussian radial basis function (RBF) have been used for classification of diseased samples from the normal one. The performance of the model has been evaluated using k-fold (k = 5) cross validation method. Best performance efficiency has been achieved with RBF.

1. Introduction

Nasopharyngeal carcinoma (NPC) is a broader term used for a group of malignant epithelial tumors with different etiopathogenesis and a diverse array of histopathological manifestations [1]. NPC usually develops around the ostium of the Eustachian tube in the lateral wall of the nasopharynx [2]. It is a multistep carcinogenic process prompted by an interaction between chronic infections with oncogenic herpes virus Epstein Barr virus (EBV) which infects about 95% of the adult population globally [3,4]. Multiple genetic and epigenetic changes accumulate during the development of NPC, which leads to progression of the clone of cells, possessing growth advantages over other cells [5]. The incidence of NPC showed that it is more common among men than in women with the ratio of 3:1 [6]. The carcinoma can occur in all age groups, but has a bimodal age distribution peaks at 50–60 years of age as well as a small peak, observed during late childhood [7]. Globally about 86,500 new cases of nasopharynx cancer reported annually with approximately 50,000 deaths occur due to

* Corresponding authors.
E-mail addresses: k.saranjam@yahoo.com (S. Khan), rahat_khalil@yahoo.com (R. Ullah).

https://doi.org/10.1016/j.ijleo.2017.11.097
0030-4026/

this neoplasm per annum [8]. According to the International Agency for Research on Cancer report in 2008, more than 80% of patients with nasopharynx cancer are in Asia, and only 5% of these cancers are reported in Europe [9].

The symptoms of NPC are manifold, including epistaxis, nasal obstruction or blocked nose, hearing loss, otalgia, headache and involvement of cranial nerve [10]. Initially the symptoms are subtle but approximately 70% of patients have masses of neck and 60–96% has cervical lymph node adenopathy [11]. For controlling the disease and proper management of the treatment plan, an accurate and early diagnosis is of the prime importance. Different clinical tests are performed in routine for the diagnosis of cancer. The most important among these tests are histopathological analysis, biopsy and imaging such as ultra sound, computer tomography (CT) and magnetic resonance imaging (MRI) depending on the condition of the patients [12].

ELISA is commonly used for the screening of NPC based on the detection of anti-bodies produce in response to EBV antigens and serves as markers of remission and relapse. However, due to low sensitivity and specificity this screening test is not satisfactory [13]. CT scan is also used for the early diagnosis of NPC but early metastatic lymph nodes are often too small to be accurately characterized by CT, resulting in frequent false negative results. Conventional screening methods like CT, MRI and positron emission tomography (PET) are not suitable for identification of early neoplasia or trivial lesions because the early cancer symptoms are not so obvious and cannot diagnosed with these techniques. Due to these limitations of the existing diagnostics modalities, there is tremendous need for diagnostic method that accurately diagnoses the disease at its early stage.

Different types of optical spectroscopic techniques are currently being investigated for the early and improved diagnosis of cancer [14–17]. Due its selective nature Raman spectroscopy gained too much interest in the field of biomedical diagnosis. Raman spectroscopy describes an inelastic scattering of light photons after an interaction with matter. A shift in the energy (frequency) termed as Raman shift of scattered photons has been occurred as result of interaction with functional groups of the target molecules [18]. Raman shift provides finger print from which the molecular composition of the sample can be determined.

Each biomolecule present in the sample produces a characteristics Raman peaks, in the Raman spectrum. Shift in the position of the Raman peak has been occurred as a result of any change in the molecular structure. Based on this information the molecular composition of the samples and any change in the composition can be determined. The changes occurred in the molecular composition of the samples can be utilized for diagnostics purpose. Many studies have previously been reported on the application of Raman spectroscopy for the diagnosis of cancer as well as infectious diseases. This technique has been used for the analysis of NPC in the human nasopharynx biopsy samples [19,20]. Likewise it is also used for the analysis of hypopharyngeal [21], head and neck cancer [22] etc. In comparison to routinely used techniques, Raman spectroscopy is less invasive, fast, economical, and also have better chemical selectivity.

As mentioned before, each biomolecule present in sample (sera) produced its own characteristic peak, but some time different molecules contribute to the same peak. Similarly the spectral differences exist between the diseased and normal sera samples are normally very small. Based on these spectral variations one cannot distinguish between normal and the diseased samples by visual observation. This shortcoming of Raman spectroscopy has been overcome by the development of machine learning models, such as SVM [23,24] and Random Forest (RF) [25]. These statistical models investigate the entire Raman spectral data and classify them on the basis of variations in the position of Raman shifts and intensities. This article presents the application of Raman spectroscopy together with SVM for the analysis of nasopharyngeal cancer.

Machine learning techniques are normally used in a situation where the data are not separable. These algorithms prepare the machine to perform the desire task. SVM is also considered as an effective machine learning algorithm, introduced by Burges and Vapnik [26,27]. Due to its discriminative nature, SVM gain much interest in the field of biomedical diagnosis where, most of the time the data are not linearly separable [28]. Data separation is performed by transforming the data to a new higher dimensional space, using transformation functions (kernel functions). In the current study SVM model with two different kernel i.e. Gaussian radial basis function and polynomial have been used.

2. Material and methods

2.1. Sample collection and serum extraction

In total 29 sera samples where, 14 were collected from NPC confirmed patients and 15 (matched age group) from healthy individuals have been used in this study. The ages of these patients were range from 45 to 65 years. All these sera samples were collected directly from Nuclear Medicine, Oncology and Radiotherapy Institute (NORI), Islamabad, Pakistan at different days (depending on patients' availability) with the full consent of patients prior to chemotherapy. In all concerned patients the disease were initially confirmed on the basis of routine screening tests. Blood samples from all subjects were collected by using sterile syringe and were stored in clot activator tubes (HebeiXinle, Sci&Tech Co. Ltd., China). For spectroscopic analysis sera were extracted from all blood samples using Hettich Centrifuge D-7200. The entire experimentation was carried out after obtaining a written permission from the ethical committee of Nuclear Medicine, Oncology & Radiotherapy Institute (NORI), Islamabad Pakistan. During entire experimental procedures ethical standard [29,30], were strictly followed.