

# Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

## ***In vivo* Raman spectroscopy of human uterine cervix: exploring the utility of vagina as an internal control**

Rubina Shaikh  
Tapas Kumar Dora  
Supriya Chopra  
Amita Maheshwari  
Deodhar Kedar K.  
Rekhi Bharat  
C. Murali Krishna

# *In vivo* Raman spectroscopy of human uterine cervix: exploring the utility of vagina as an internal control

Rubina Shaikh,<sup>a</sup> Tapas Kumar Dora,<sup>b</sup> Supriya Chopra,<sup>b</sup> Amita Maheshwari,<sup>c</sup> Deodhar Kedar K.,<sup>d</sup> Rekhi Bharat,<sup>d</sup> and C. Murali Krishna<sup>a,\*</sup>

<sup>a</sup>Chilakapati Lab, ACTREC, Kharghar, Navi-Mumbai 410210, India

<sup>b</sup>Tata Memorial Center, Radiation Oncology, ACTREC, Kharghar, Navi Mumbai 410210, India

<sup>c</sup>Tata Memorial Hospital, Gynecology Oncology, Parel, Mumbai 400012, India

<sup>d</sup>Tata Memorial Hospital, Department of Pathology, Parel, Mumbai 400012, India

**Abstract.** *In vivo* Raman spectroscopy is being projected as a new, noninvasive method for cervical cancer diagnosis. In most of the reported studies, normal areas in the cancerous cervix were used as control. However, in the Indian subcontinent, the majority of cervical cancers are detected at advanced stages, leaving no normal sites for acquiring control spectra. Moreover, vagina and ectocervix are reported to have similar biochemical composition. Thus, in the present study, we have evaluated the feasibility of classifying normal and cancerous conditions in the Indian population and we have also explored the utility of the vagina as an internal control. A total of 228 normal and 181 tumor *in vivo* Raman spectra were acquired from 93 subjects under clinical supervision. The spectral features in normal conditions suggest the presence of collagen, while DNA and non-collagenous proteins were abundant in tumors. Principal-component linear discriminant analysis (PC-LDA) yielded 97% classification efficiency between normal and tumor groups. An analysis of a normal cervix and vaginal controls of cancerous and noncancerous subjects suggests similar spectral features between these groups. PC-LDA of tumor, normal cervix, and vaginal controls further support the utility of the vagina as an internal control. Overall, findings of the study corroborate with earlier studies and facilitate objective, noninvasive, and rapid Raman spectroscopic-based screening/diagnosis of cervical cancers. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.19.8.087001](https://doi.org/10.1117/1.JBO.19.8.087001)]

Keywords: cervical cancer; Raman spectroscopy; *in vivo* spectroscopy; principal component-linear discrimination analysis; leave-one-out cross validation.

Paper 140121PRR received Mar. 12, 2014; revised manuscript received Jun. 29, 2014; accepted for publication Jul. 14, 2014; published online Aug. 7, 2014.

## 1 Introduction

Carcinoma of the uterine cervix is the third most common cancer among females worldwide.<sup>1</sup> Around 80% of the global burden is contributed by developing countries including India.<sup>2</sup> Annually, more than one million new cases are diagnosed in India with >50% mortality, which is primarily attributed to late diagnosis.<sup>2</sup> The papanicolau test (Pap test) has been used as a preliminary screening tool. An abnormal Pap smear is followed by colposcopy-guided biopsy for confirmatory diagnosis. Histopathology is the gold standard for cervical cancer diagnosis. However, conventional screening and/or diagnostic tools have been known to suffer from disadvantages, such as tedious methodology, long output duration, and interobserver variability, besides patient discomfort.<sup>3,4</sup>

Optical spectroscopic techniques, such as fluorescence, infrared (IR), and Raman, have shown potentials in classifying normal and tumor groups in various cancers.<sup>5–8</sup> Several *ex vivo* Raman studies have demonstrated the classification of normal, precancerous, and cancerous cervical tissues.<sup>9–12</sup> The first *in vivo* Raman study of cervical cancer was reported in 1998.<sup>13</sup> A subsequent study has demonstrated similarities between *in vivo* and *in vitro* cervix spectra.<sup>14</sup> Studies were then extended by exploring the classification of high-grade dysplasia and benign conditions.<sup>15</sup> Successful classification among normal

ectocervix, normal endocervix as well as low-grade dysplasia and high-grade dysplasia has been reported.<sup>16</sup> Recent studies suggest that the inclusion of parameters, such as hormonal and menopausal status, would improve the classification efficiency.<sup>17,18</sup> The utility of higher wavenumber Raman spectra in precancer detection has been reported.<sup>19</sup> The efficacy of simultaneous fingerprint and high-wavenumber Raman spectra in enhancing early detection of cervical precancer has also been demonstrated.<sup>20</sup> More recent studies have demonstrated the utility of confocal Raman spectroscopy in improving precancer detection, noninvasive assessment of menopausal-related hormonal changes, and the effect of Vagifem treatment on postmenopausal women.<sup>21–22</sup> Despite an abundance of literature, further careful validations on diverse populations and larger cohorts are required for the translation of this technology into clinics. Cervical cancer subjects in developing countries are very often present at advanced stages and, in such cases when the majority of the cervix is diseased, there is no normal cervix site to acquire control spectra. Therefore, the healthy cervix of subjects having other gynecologic cancers (uterine or ovarian) undergoing hysterectomies have been used as controls,<sup>23</sup> in which subject accrual is often a major constraint. Vagina and cervix (ectocervix) are known to have similar histology;<sup>24</sup> malignancy-associated changes/cancer field effects have not been reported in cervical cancers. Therefore, as an

\*Address all correspondence to: Chilakapat Murali Krishna, E-mail: [mchilakapati@actrec.gov.in](mailto:mchilakapati@actrec.gov.in) or [pittu1043@gmail.com](mailto:pittu1043@gmail.com)

alternative, vagina can be utilized as an internal control, especially in screening camps where colposcopy may not be available. In the present study, we have investigated the feasibility of the classification between normal and cervical cancers in the Indian population and also the utility of the vagina as an internal control. Spectral data were analyzed by multivariate statistical tools, principal component-linear discriminant analysis (PC-LDA) followed by a leave-one-out cross validation and an independent test prediction. The feasibility of classification between all control groups, i.e., the normal cervix sites and vaginal sites of cancerous and noncancerous subjects, was also investigated. It was followed by the analysis of normal cervix, vagina, and tumor spectra. The findings of the study are discussed in the paper.

## 2 Methods and Materials

### 2.1 Sample Details

Ninety-three (93) subjects diagnosed with gynecological cancers and referred for treatment at Advanced Center for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Center, participated in the study. Informed and written consents were taken from each subject prior to the spectral recording. Nonpregnant subjects in the age group of 30 to 70 years, with no history of a hysterectomy, were recruited. Details of menstruation status were obtained through a questionnaire. Among the 93 subjects, 87 (93%) were postmenopausal and only 6 (7%) were premenopausal subjects.

#### 2.1.1 Normal and cancer subjects

To explore the classification between normal and cervical cancers, 61 subjects (154 spectra) were enrolled in the study. Out of 154, 80 spectra were acquired from a cervical tumor (T) of 31 cervical cancer subjects and 74 spectra from an uninvolved normal cervix (N) of 30 subjects. Sample details are given in Table 1(a).

#### 2.1.2 Exploring the utility of the vagina as an internal control

To explore the utility of the vagina as an internal control, 230 spectra from 66 subjects were utilized. Among the 230 spectra, 74 were acquired from normal cervix sites (N) of 30 normal subjects, 64 spectra were acquired from normal vaginal sites (VN) of 20 normal subjects and 92 spectra from normal vaginal sites (VT) of 36 cervical cancer subjects [Table 1(b)].

#### 2.1.3 Evaluating internal controls

To evaluate the utility of a vagina as an internal control, a total of 409 spectra were utilized. Out of 409, 181 spectra were acquired from 63 cervical cancer subjects (T), 74 spectra from normal cervix sites (N) of 30 normal cervix subjects, 154 spectra were acquired from uninvolved vaginal sites of 56 subjects with or without cervical cancers (V) [Table 1(c)].

Cuscus speculum was inserted in vagina to observe cervix that was cleaned with a saline solution. Multiple *in vivo* Raman spectra (3 to 6) were recorded from gross tumors normal cervix or vagina of normal and tumor subjects under clinical supervision. Raman spectra were acquired by placing the probe perpendicular to the surface of the site. To avoid contamination among subjects, prior to spectral recording, the probe was

**Table 1** Details of samples: (a) classification of normal (N) and cervical tumor (T); (b) exploring internal controls; and (c) evaluating internal controls.

Categories	Number of subjects	Number of spectra
(a) Tumor Vs normal: sample details		
Normal cervix (N)	30	74
Tumor (T)	31	80
Total number of cases	61	154
(b) Exploring internal controls: sample details		
Normal cervix (N) <sup>a</sup>	30	74
Normal vagina of normal cervix cases (VN) <sup>a</sup>	20	64
Normal vagina of normal cervix cases (VT)	36	92
Total number of cases	66	230
(c) Evaluating internal controls: sample details		
Tumor (T) <sup>b</sup>	63	181
Normal cervix (N)	30	74
Vagina (V) <sup>b</sup>	56	154
Total number of cases	93	409

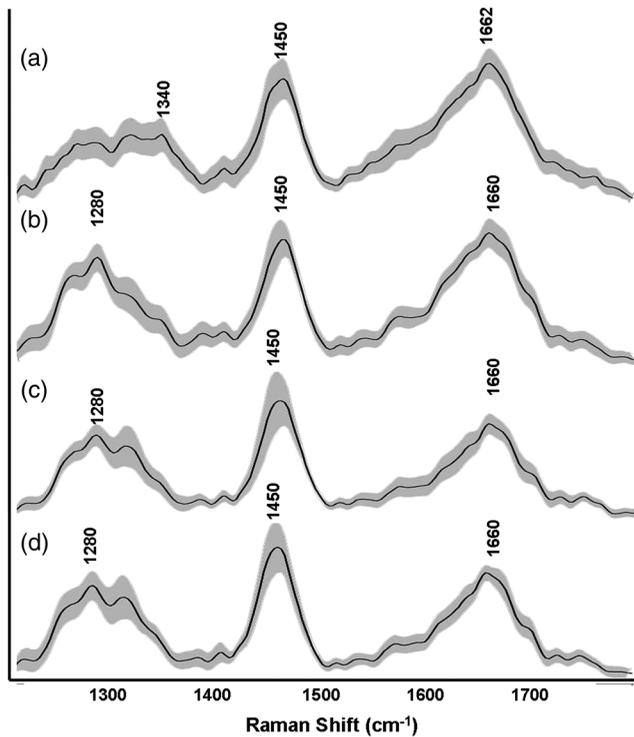
<sup>a</sup>20 cases were common.

<sup>b</sup>56 cases were common.

disinfected with CIDEX (Johnson and Johnson, Mumbai, India) solution and wrapped in parafilm.

### 2.2 Raman Spectroscopy

Spectra were recorded with HE-785 commercial Raman spectrometer (Jobin-Vyon-Horiba, France). Details of the instrument were described elsewhere.<sup>25</sup> Briefly, this system consists of a diode laser (PI-ECL-785-300-FC, Process Instruments) of 785-nm wavelength as an excitation source, a high-efficiency spectrograph (HE-785, HORIBA Jobin Yvon, France) with fixed 950 gr/mm grating coupled with charge coupled device (CCD) (CCD-1024X256-BIDD-SYN, Synapse, Longjumeau, France). The commercial RamanProbe (RPS 785/12-5, InPhotonics Inc. (Downey St., Norwood, Massachusetts), consisting of an excitation and collection fibers (NA-0.40) with diameters of 105 and 200  $\mu\text{m}$ , respectively, was used to couple the excitation source and the detection system. The estimated spot size and depth of penetration as per the manufacturer's specifications were 105  $\mu\text{m}$  and 1 mm, respectively.<sup>26</sup> This instrument has no movable parts and the spectral resolution as per manufacturer's specification was  $\sim 4 \text{ cm}^{-1}$ . As per the protocol followed in previous studies, to maintain the focus during all measurements, a detachable spacer with a length of 5 mm was attached at the tip of the probe.<sup>25,27</sup> Spectra were acquired at 80-mW laser power, integrated for 5 s, and averaged over three accumulations.



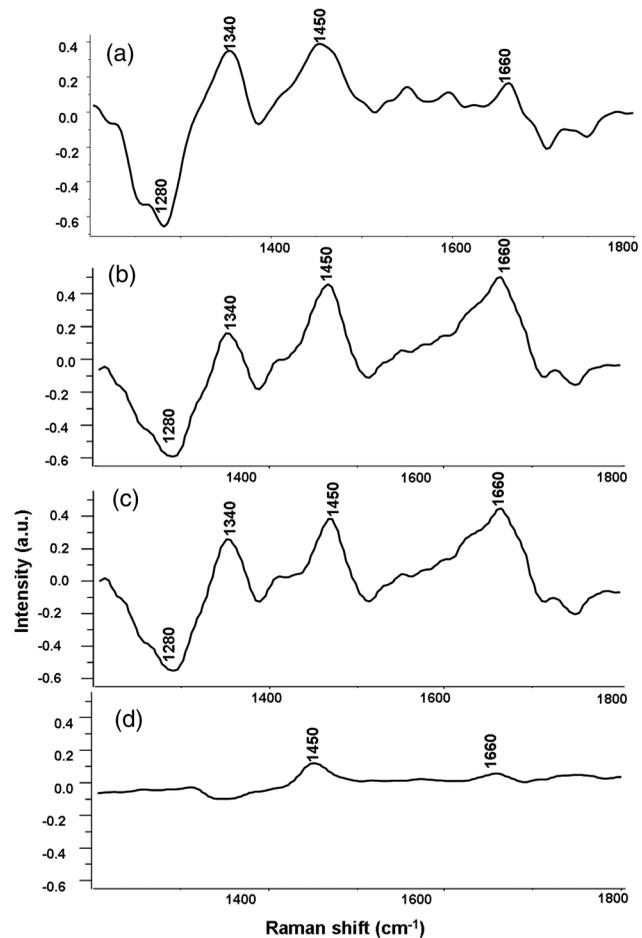
**Fig. 1** *In vivo* mean Raman spectra of (a) cervical tumor, (b) normal cervix, (c) vagina of normal cervix subjects, and (d) vagina of cervical tumor subjects.

### 2.3 Data Analysis

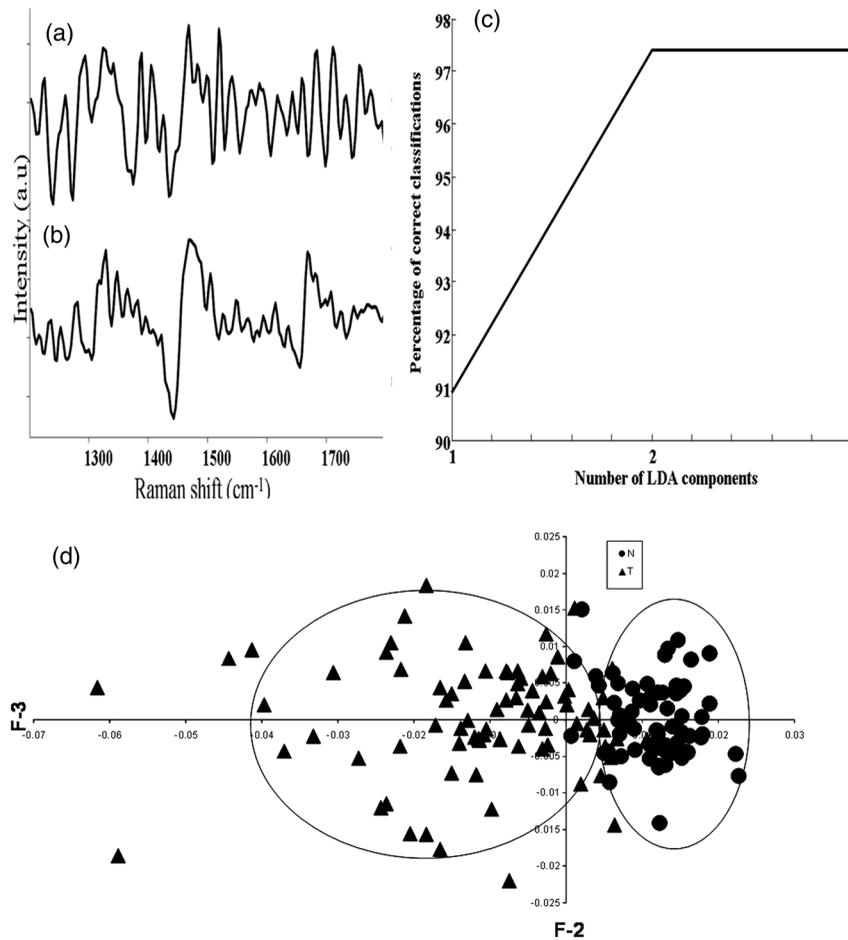
*In vivo* Raman spectra were preprocessed by correcting for a CCD response with a National Institute of Standards and Technology certified SRM 2241 material followed by background subtraction to remove the spectral contaminations from fibers and optic elements.<sup>28</sup> To remove the interference of the slow-moving background, the first derivative of the pre-processed Raman spectra was computed (Savitzky–Golay, window size 3).<sup>29,30</sup> The primary objective of the first derivative correction was to construct a spectral profile depicting point-by-point deviation of the spectral intensity over a moving window of three points. This transformation provides a peak profile in the spectra irrespective of optical response-related intensity. Since our previous studies have demonstrated the efficacy of the 1200 – 1800  $\text{cm}^{-1}$  region in classifying normal and malignant oral cancers<sup>27,28</sup> and as it is less influenced by fiber signals, we have employed same region for further analysis. Background-corrected spectra were interpolated and first derivatized, which was followed by a vector normalization. Vector-normalized spectra were subjected to PC-LDA, the multivariate statistical tool.<sup>27,28</sup>

In PC-LDA, principal component analysis (PCA) is first carried out on the entire data set to reduce the dimensions of the data, while preserving the diagnostically significant information for classification. PCA describes data variance by identifying a new set of orthogonal features, so-called principal components (PCs) or eigenvectors. Due to their orthogonal characteristics, the first few PCs are sufficient to represent the maximum data variance. Every eigenvector is associated with the original spectrum by a variable named as a PC score, which characterizes the rank of that particular component against the source

spectrum. Differences between different classes are reflected by PC scores. An unpaired student's *t*-test was used to diagnostically identify significant PCs ( $p < 0.05$ ).<sup>31,32</sup> These PC scores were used as input data for linear discriminant analysis (LDA) algorithms-based classifications. Although PCA aims to identify the features that represent variance in the data, LDA provides data classification based on an optimized criterion, which was objective for more class separability. An LDA transformation matrix was generated, and it further identifies eigenvector or LDA components of this classification criterion.<sup>32</sup> Scree plots depict the variance (% correct classifications) accounted for by the total number of LDA components selected for analysis. The outcomes of PC-LDA were represented in the form of a confusion matrix, where diagonal elements were true positive and nondiagonal elements were false-positive predictions. The confusion matrix aids in understanding the separation within the groups, which were acquired by accounting for the contribution of all selected factors used for analysis. PC-LDA results were also represented in the form of scatter plots, generated by plotting various combinations of scores of the factors. The performance of the PC-LDA diagnostic models was further validated in an unbiased method by leave-one-out cross validation (LOOCV). In the LOOCV methodology, one spectrum were held out from the data set and the remaining data were used to redevelop the PC-LDA model. In our experiments, test prediction was also used to validate the model.



**Fig. 2** Difference spectra: (a) cervical tumor–normal cervix, (b) cervical tumor–vagina of cervical tumor, (c) cervical tumor–vagina of normal cervix, (d) vagina of cervical tumor–vagina of normal cervix.



**Fig. 3** Classification of tumor and normal: (a) loadings of factor 3, (b) loadings of factor 2, (c) scree plot, (d) scatter plot [cervical tumor (filled triangles) and normal cervix (filled circles)].

This process was repeated until all withheld spectra were classified. Algorithms for these analyses were implemented in MATLAB® (Mathworks Inc., Norwood, Massachusetts) based in-house software.<sup>33</sup>

Average spectra were calculated from the background subtracted spectra prior to derivatization for each group and were baseline corrected by fitting a fifth-order polynomial

function. These baseline-corrected, smoothed (Savitzky-Golay, 3), and vector-normalized spectra were used for spectral comparisons and for computing the difference spectra. The standard deviation and mean were also computed to illustrate the intraclass variability.

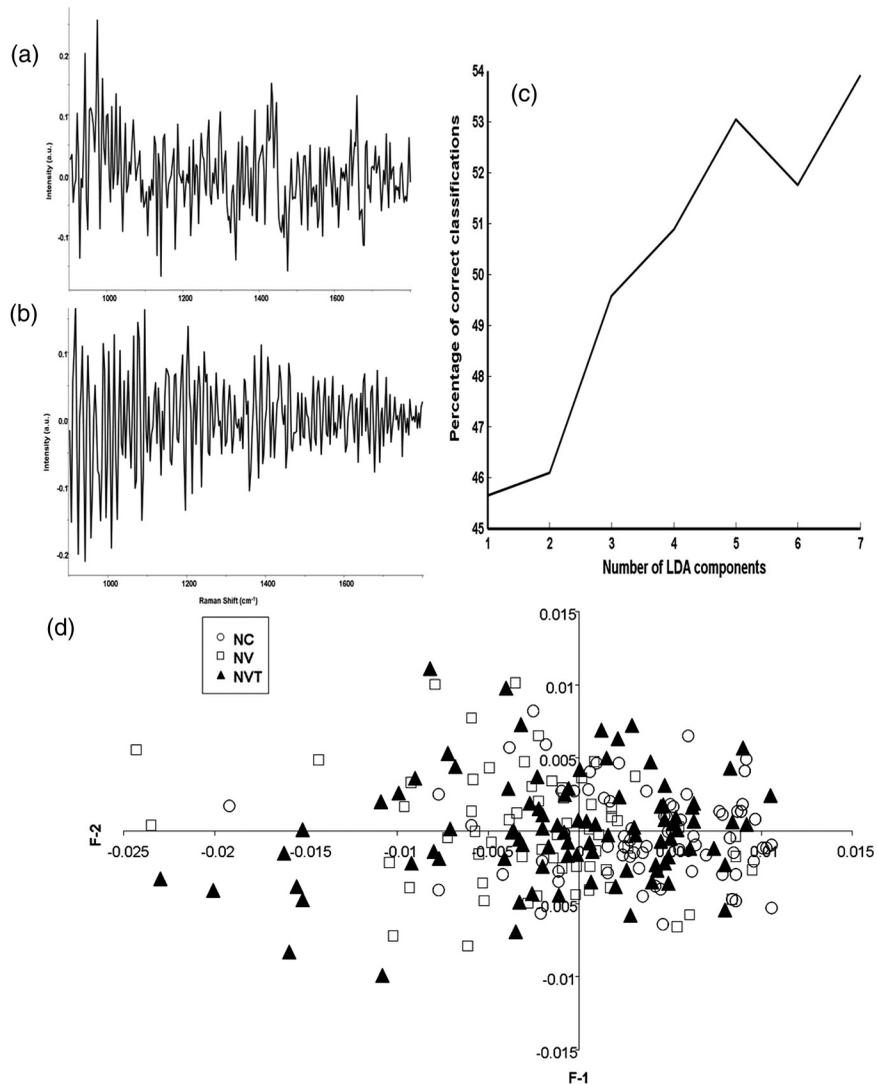
### 3 Results and Discussions

#### 3.1 Raman Spectral Features

Vector-normalized average *in vivo* spectra of cervical tumor (T), normal cervix (N), vagina of normal (VN), and cervical cancer (VT) cases along with their standard deviations are illustrated in Fig. 1. Standard deviation spectra showed minor intensity-related changes within the groups. The mean spectra of a normal cervix exhibit characteristic features of amide III and strong and broad amide I, which can be attributed to the presence of collagenous proteins. Prominent spectral features in a tumor, with respect to a normal spectrum, are strong and sharper amide I, a minor shift in  $\delta\text{CH}_2$  and a distinct band at 1340 cm<sup>-1</sup>, which are indicative of DNA and noncollagenous proteins. These findings corroborate earlier *ex vivo* and *in vivo* Raman spectroscopic studies on cervical cancers,<sup>9,10,24</sup> although *in vivo* vagina spectral features were very similar to normal cervix and exhibited features of collagenous proteins, indicating resemblances in their biochemical compositions.

**Table 2** Principal component-linear discriminant analysis: (a) standard models, and (b) leave-one-out cross-validation of normal cervix (N) and tumor cervix (T) group. (Diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions.)

	Normal (N)	Tumor (T)
(a) Standard model		
Normal (N)	70	4
Tumor (T)	0	80
(b) Leave-one-out cross-validation		
Normal (N)	69	5
Tumor (T)	0	80



**Fig. 4** Exploring internal control-PC-LDA of normal cervix, vagina of cervical tumor, and vagina of normal cervix: (a) loadings of factor 1, (b) loadings of factor 2, (c) scree plot, (d) scatter plot [normal cervix (circles), vagina of normal cervix subjects (squares), and vagina of normal cervix subjects (filled triangles)].

To reveal the spectral differences between cervical tumor and normal cervix, the difference spectra, one of the conventional ways to understand the spectral variations, were employed. Difference spectrum was computed by subtracting the mean normal cervix spectrum from the mean cervical tumor spectrum [Fig. 2(a)]. The positive peaks in the difference spectrum are from a tumor, while negative bands are from normal. Strong positive peaks of proteins, such as amide I ( $1662\text{ cm}^{-1}$ ),  $\delta\text{CH}_2$  deformation ( $1450\text{ cm}^{-1}$ ), and DNA ( $1340\text{ cm}^{-1}$ ) suggest their predominance in tumor spectrum, whereas negative peaks ( $1280\text{ cm}^{-1}$ ) indicate higher collagenous protein in a normal cervix. The difference spectra of T-VT, T-VN, and VT-VN are illustrated in Figs. 2(b)–2(d), respectively. It was observed that the difference spectra of T-VT and T-VN showed a similar profile to that of T-N (Fig. 2). Observed positive peaks at  $1660$ ,  $1450$ , and  $1340\text{ cm}^{-1}$  of the difference spectra were characteristic of a cervical tumor, which indicates increased DNA and protein while negative peaks at  $1280$  and  $1240\text{ cm}^{-1}$  indicate collagenous proteins. The difference spectra of VT-VN showed

minor variations in amide I and  $\delta\text{CH}_2$  peaks, further exhibiting similarities among VT and VN.

### 3.2 Multivariate Analysis

#### 3.2.1 Exploring the classification between normal and cancer groups

To determine the feasibility of classification between normal and cervical tumors, the vector-normalized preprocessed spectra were subjected to supervise the method of PC-LDA followed by LOOCV. Loadings of factors can provide crucial information of biochemical variations among different groups. Loadings of factors 3 and 2 are presented in Figs. 3(a) and 3(b), respectively. The findings corroborate spectral variability suggesting differences in the types of proteins and the content of nucleic acid among normal and tumor groups. The third PC (PC3) has two major bands that correspond to  $\delta\text{CH}_2$  stretch and a vibrational mode at  $1340\text{ cm}^{-1}$  from nucleic acid. In the second component (PC2), the main peak features correspond to amide I,

$\delta\text{CH}_2$  stretch, and nucleic acid. A scree plot depicts the variance or percent correct classifications accounting for the total number of factors selected for analysis. Three factors contributing to  $\sim 97\%$  of the classification were used for analysis, which is shown in Fig. 3(c). The scatter plot has been shown in Fig. 3(d); it depicts exclusive clusters corresponding to normal and cervical tumors. These results are also summarized in Table 2. As seen, 70/74 normal and 80/80 tumor spectra were correctly classified. None of the tumor spectra were misclassified, whereas 4 normal spectra were misclassified as tumors. LOOCV was also executed to evaluate the classification efficiency of the model and has been shown in Table 2(b). Only 5 spectra out of 74 normal cervix sites were misclassified and all tumor spectra were correctly classified. An average classification efficiency of 97% was observed.

### 3.2.2 Exploring internal controls

Several *in vivo* Raman studies on cervical cancers have investigated the role of parameters, such as menopausal status, hormonal status, age, and parity on the classification efficiency of the data.<sup>14-19</sup> In such studies, colposcopy-guided spectra of normal sites were used as controls. As mentioned earlier, in developing countries such as India, cervical cancer subjects present at advanced stages (Stage IIA and above).<sup>4</sup> As the majority of the cervix is involved, no normal cervix sites are left to acquire control spectra. In such cases, spectra acquired from normal cervix of subjects undergoing hysterectomy for other gynecological cancers were used as controls.<sup>24</sup> This could be a major constraint for subject accruals, especially in cancer hospitals where normal hysterectomy cases rarely visit. Moreover, the composition of a vagina and an ectocervix is similar as they contain the inner lining of squamous epithelial cells;<sup>25</sup> no CFE/MAC has been reported in cervical cancers, thus vagina can serve as a good internal control. Hence we have explored the utility of a vagina as an internal control. This approach could be helpful to circumvent interpatient variability as well as it could be useful, especially in screening camps where colposcopy may not be available.

To explore the variations between the control groups, 74 spectra from normal cervix (N), 64 from the vagina of normal subjects (VN), and 92 from the vagina of tumor subjects (VT) were analyzed by PC-LDA using seven factors. Loadings of factors 1 and 2 are shown in Figs. 4(a) and 4(b), respectively. As seen at the loading, no major differences were observed within the control groups. Scree and scatter plots are shown in Figs. 4(c) and 4(d), respectively. A large overlap and/or misclassifications were observed among normal cervix (N), vagina of normal cervix subjects (VN) and vagina of cervical tumor subjects (VT).

The findings of PC-LDA are also shown in a confusion matrix for a standard model and LOOCV in Table 3(a) and 3(b), respectively. In the case of the standard model, 49/74 normal cervix spectra (N), 41/64 vagina spectral sites of normal subjects (VN), and 34/92 vagina spectral sites of tumor subjects (VT) were correctly classified. In the case of LOOCV, 44/74 of normal cervix spectra (N), 25/64 vaginal sites of normal subjects (VN), and 21/92 vaginal sites of tumor subjects (VT) were correctly classified. Whereas 13 and 17 spectra of the normal cervix (N) were misclassified with the vaginal sites of normal subjects (VN) and the vagina of tumor subjects (VT), respectively, and in the case of the vagina of normal subjects (VN), 9 and 30 spectra were misclassified with a normal cervix (N) and the vagina of tumor subjects (VT), respectively. For vaginal sites of tumor subjects (VT), 36 spectra were misclassified with a normal

cervix (N) and 35 with spectra of vagina of normal subjects (VN). The higher misclassifications were observed among the spectra of normal cervix (N), vagina of normal subjects (VN), and vaginal sites of tumor subjects (VT), which were indicative of the biochemical similarities among these groups. The findings suggest that vaginal sites can be used as an internal control on similar lines to the oral and breast cancers wherein, contralateral and uninvolved areas are employed as controls, respectively.<sup>26,34</sup> This approach may also help to circumvent the possible influence of hormonal status, menopausal status, age, and parity. Also, spectral acquisition does not require colposcope at the site. As the spectra of a vagina of normal subjects (VN) and a vagina of tumor subjects (VT) show similarities, we have grouped them together and referred to as vagina spectra (V) in our subsequent evaluations of a vagina as a control. It is also important to note that among 93 subjects, 87 (93%) cases were postmenopausal and only 6 (7%) were premenopausal. Hence, menopausal status may not have an influence on current results.

### 3.2.3 Evaluating internal controls

In order to evaluate the efficacy of Raman spectroscopic methods in discriminating tumor conditions against control groups, spectra from the site of the tumor (T), the normal cervix (N), and the vagina (V) were subjected to PC-LDA. In the first step, 28 spectra from tumor (T), 34 from normal cervix (N), and 24 spectra of the vaginal (V) were employed to build the standard model by utilizing seven factors. Loadings of factors, scree, and scatter plots are shown in Fig. 5. The loadings of factors 1 and 2 exhibited similar patterns as it was observed in the classification of normal and cervical tumors, also corroborating the observations of mean spectral findings. The first PC (PC1) has three major bands that correspond to amide I,  $\delta\text{CH}_2$  stretch, and a band at  $1340\text{ cm}^{-1}$  from nucleic acid. In the second component, PC2 exhibited minor peak features corresponding to amide I and

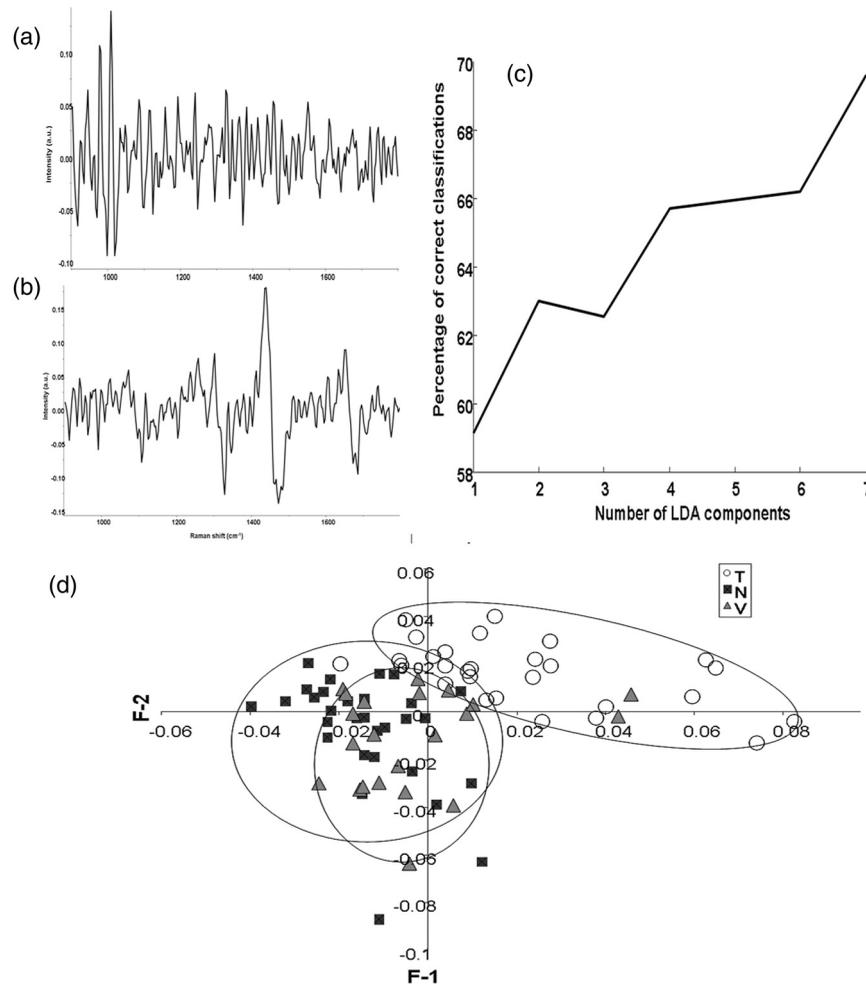
**Table 3** Exploring internal control principal component-linear discriminant analysis: (a) standard model, and (b) leave-one-out cross validation of normal cervix (N), vagina of normal cervix subjects (VN), and vagina of cervical tumor subjects (VT). (Diagonal elements are true positive predictions and ex-diagonal elements are false positive predictions).

	Normal cervix (N)	Vagina of normal cervix subjects (VN)	Vagina of tumor cervix subjects (VT)
(a) Standard model			
Normal cervix (N)	49	13	12
Vagina of normal cervix subjects (VN)	9	41	14
Vagina of tumor cervix subjects (VT)	32	26	34
(b) Leave-one-out cross-validation			
Normal cervix (N)	44	13	17
Vagina of normal cervix subjects (VN)	9	25	30
Vagina of tumor cervix subjects (VT)	36	35	21

$\delta\text{CH}_2$  stretch. The scatter plot exhibited clear classification among the clusters belonging to tumors (T) and controls, i.e., the normal cervix sites (N) and the vagina (V). Spectra from normal cervix sites (N) and from vaginas (V) exhibited very high overlap. Findings of PC-LDA are also shown in the confusion matrix of Table 4(a) and 4(b). As seen, 25/28 cervical tumor, 27/34 normal cervix, and 19/24 vagina spectra were correctly classified. While, in the case of LOOCV, 24/28 cervical tumor (T), 22/34 normal cervix (N), and 8/24 vaginal spectra (V) were correctly classified. 12/34 spectra of normal cervix were misclassified with vagina (V) and 15/24 vaginal (V) exhibiting the misclassifications with normal cervix (N) spectra. Predictive efficiencies of the standard model were evaluated by using 153 tumors (T), 40 normal cervix (N), and 132 vagina (V) as an independent test data set. In this case, 130/153 tumor (T), 28/40 normal cervix (N), and 64/132 vagina spectra (V) were correctly predicted. However, 11/40 normal cervix (N) was misclassified as vagina (V) and 65/132 vagina (V) was misclassified as normal cervix (NC) [Table 4(c)]. Higher misclassifications between the normal cervix (N) and vagina (V) once again suggest similarities between the biochemical compositions. This further supports the applicability of a vagina as internal control. Out of 153 tumors, 16 and 7 spectra were misclassified as the normal cervix (N) and vaginal sites (V), respectively. The observed minor misclassifications of tumor

(T) as normal cervix (N) may be attributed to heterogeneity of tumors. In this study, spectra were recorded at several points and few of the sites could be from islands of the normal in tumors.

In the present study, we have evaluated the feasibility of Raman spectroscopic classification of normal and cervical cancers in an Indian population and also explored the utility of vaginal sites as an internal control. We have observed that PC-LDA of the normal (N) and tumor (T) spectra gave classification efficiencies of 98.5%. On the other hand, PC-LDA of the normal cervix (N), and vagina of tumor subjects (VT), and vagina of normal subjects (VN) showed higher misclassifications, suggesting similarities in biochemical composition among the controls. PC-LDA of tumor (T), normal cervix (N), and vagina (V) showed the classification between tumor and all controls, i.e., normal cervix (N) and vagina (V). While large misclassifications between the controls spectra were observed, this further supports the utility of a vagina as an internal control. Findings of the study corroborate with earlier studies and suggest applicability of Raman spectroscopic methods for objective, noninvasive, and rapid diagnosis of cervical cancers. The study also demonstrates that Raman spectroscopy may be used for improving cervical cancer diagnosis by incorporating internal control, such as the vagina, to circumvent the influence of parameters such as hormonal status, menopausal status, and age, besides the



**Fig. 5** Evaluation of internal control-PC-LDA of tumor, normal cervix, vagina of cervical tumor, and vagina of normal cervix: (a) loadings of factor 2, (b) loadings of factor 1, (c) scree plot, (d) scatter plot [cervical tumor (circles), normal cervix (filled squares), and vagina (filled triangles)].

**Table 4** Evaluation of internal control principal component-linear discriminant analysis of cervical tumor (T), normal cervix (N), and vagina (V), (a) standard model (b) leave-one-out cross-validation and (c) test prediction. (Diagonal elements are true positive predictions and off-diagonal elements are false positive predictions).

	Tumor (T)	Normal (N)	Vagina (V)
(a) Standard Model			
Tumor (T)	25	2	1
Normal (N)	0	27	7
Vagina (V)	Q	5	19
(b) Leave-one-out cross-validation			
Tumor (T)	24	4	0
Normal (N)	0	22	12
Vagina (V)	1	15	8
(c) Test prediction			
Tumor (T)	130	16	7
Normal (N)	1	28	11
Vagina (V)	3	65	64

requirement of colposcope especially for mass screening in make-shift camps.

### Acknowledgments

The authors would like to acknowledge all of the participants of the study. The Raman spectrometer employed in the study was procured from Project No. BT/ PRI11282/MED/32/83/2008, entitled "Development of *in vivo* laser Raman spectroscopy methods for diagnosis of oral precancerous and cancerous conditions," Department of Biotechnology, Government of India. The authors would also like to acknowledge Dr. Surya Pratap Singh, ACTREC for editorial makeover of the manuscript.

### References

- J. Ferlay et al., "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *Int. J. Cancer* **127**(12), 2893–2917 (2010).
- R. Sankaranarayanan et al., "HPV screening for cervical cancer in rural India," *N. Engl. J. Med.* **360**(14), 1385–1394 (2009).
- R. A. Kerkar and Y. V. Kulkarni, "Screening for cervical cancer: an overview," *J. Obstet. Gynecol. India* **56**(2), 115–122 (2006).
- M. López-Gómez et al., "Cancer in developing countries: the next most preventable pandemic. The global problem of cancer," *Crit. Rev. Oncol. Hematol.* **88**(1), 117–121 (2013).
- A. L. N. Francisco et al., "PP021: fluorescence spectroscopy for evaluation of safety margins in individuals with squamous cell carcinoma of the oral cavity surgically treated," *Oral Oncol.* **49**, S100, (2013).
- J. W. Spliethoff et al., "Improved identification of peripheral lung tumors by using diffuse reflectance and fluorescence spectroscopy," *Lung Cancer* **80**(2), 165–171 (2013).
- S. Lee et al., "Improving the classification accuracy for IR spectroscopic diagnosis of stomach and colon malignancy using non-linear spectral feature extraction," *Analyst* **138**(14), 4076–4082 (2013).
- Q. Tu and C. Chang, "Diagnostic applications of Raman spectroscopy," *Nanomed. Nanotechnol. Biol. Med.* **8**(5), 545–558 (2012).
- A. Mahadevan-Jansen et al., "Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers," *Photochem. Photobiol.* **68**(1), 123–132 (1998).
- C. M. Krishna et al., "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix," *Vib. Spectrosc.* **41**(1), 136–141 (2006).
- E. O. Faolain et al., "The potential of vibrational spectroscopy in the early detection of cervical cancer: an exciting emerging field," *Proc. SPIE* **5826**, 25–36 (2005).
- F. M. Lyng et al., "Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool," *Exp. Mol. Pathol.* **82**, 121–129 (2007).
- A. Mahadevan-Jansen et al., "Development of a fiber optic probe to measure NIR Raman spectra of cervical tissue *in vivo*," *Photochem. Photobiol.* **68**(3), 427–431 (1998).
- U. Utzinger et al., "Near-infrared Raman spectroscopy for *in vivo* detection of cervical precancers," *Appl. Spectrosc.* **55**(8), 955–959 (2001).
- A. Robichaux-Viehoever et al., "Characterization of Raman spectra measured *in vivo* for the detection of cervical dysplasia," *Appl. Spectrosc.* **61**(9), 986–993 (2007).
- E. M. Kanter et al., "Multiclass discrimination of cervical precancers using Raman spectroscopy," *J. Raman Spectrosc.* **40**(2), 205–211 (2009).
- E. M. Kanter et al., "Effect of hormonal variation on Raman spectra for cervical disease detection," *Am. J. Obstet. Gynecol.* **200**(5), 512.e1–512.e5 (2009).
- E. M. Kanter et al., "Application of Raman spectroscopy for cervical dysplasia diagnosis," *J. Biophoton.* **2**(1–2), 81–90 (2009).
- J. Mo et al., "High wavenumber Raman spectroscopy for *in vivo* detection of cervical dysplasia," *Analyt. Chem.* **81**(21), 8908–8915 (2009).
- S. Duraipandian et al., "Simultaneous fingerprint and high-wavenumber confocal Raman spectroscopy enhances early detection of cervical precancer *in vivo*," *Anal. Chem.* **84**(14), 5913–5919 (2012).
- S. Duraipandian et al., "Non-invasive analysis of hormonal variations and effect of postmenopausal Vagifem treatment on women using *in vivo* high wavenumber confocal Raman spectroscopy," *Analyst* **138**(14), 4120–4128 (2013).
- S. Duraipandian et al., "Near-infrared-excited confocal Raman spectroscopy advances *in vivo* diagnosis of cervical precancer," *J. Biomed. Opt.* **18**(6), 067007 (2013).
- S. Rubina et al., "*In vivo* Raman spectroscopy on cervix cancers," *Proc. SPIE* **8940**, 89400E (2013).
- B. Young et al., "Wheater's functional histology: a text and colour atlas," Chapter 19 in *Female Reproductive System*, pp. 351–383, Elsevier, USA (2006).
- S. P. Singh et al., "*In vivo* Raman spectroscopic identification of pre-malignant lesions in oral buccal mucosa," *J. Biomed. Opt.* **17**(10), 105002 (2012).
- InPhotonics, Users' Manual, version 2.21, RamanProbe™, Fiber optic sampling probe, p. 6, Norwood, MA (2010).
- S. P. Singh et al., "*In vivo* Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/cancer field effects (CFE)," *Analyst* **138**(14), 4175–82 (2013).
- C. Gobinet et al., "Preprocessing methods of Raman spectra for source extraction on biomedical samples: application on paraffin-embedded skin biopsies," *IEEE Trans. Biomed. Eng.* **56**(5), 1342–1371 (2009).
- A. Nijssen et al., "Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy," *J. Biomed. Opt.* **12**(3), 034004 (2007).
- S. Koljenovic et al., "Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy," *Lab. Invest.* **82**(10), 1265–1277 (2002).
- J. L. Devore, *Probability and statistics for engineering and the science*, Richard Stratton, Boston, MA (1992).
- S. K. Teh et al., "Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue," *Br. J. Cancer* **98**(2), 457–465 (2008).
- A. D. Ghanate et al., "Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario," *J. Biomed. Opt.* **16**(2), 025003 (2011).
- C. Kendall et al., "Evaluation of Raman probe for oesophageal cancer diagnostics," *Analyst* **135**(12), 3038–3041 (2010).

Biographies of the authors are not available.