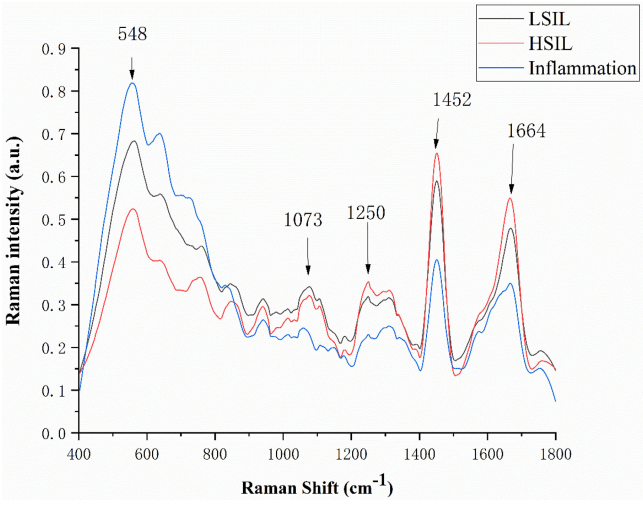
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Feature Fusion Combined With Raman Spectroscopy for Early Diagnosis of Cervical Cancer

Volume 13, Number 3, June 2021

Huiting Zhang
Cheng Chen
Cailing Ma
Chen Chen
Zhimin Zhu
Bo Yang
Fangfang Chen
Dongfang Jia
Yizhe Li
Xiaoyi Lv



DOI: 10.1109/JPHOT.2021.3075958

Feature Fusion Combined With Raman Spectroscopy for Early Diagnosis of Cervical Cancer

Huiting Zhang,[1] Cheng Chen,[1] Cailing Ma,[2] Chen Chen,[1] Zhimin Zhu,[1] Bo Yang,[1] Fangfang Chen,[1] Dongfang Jia,[1] Yizhe Li,[1] and Xiaoyi Lv[3,4]

[1]College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
[2]State Key Laboratory of PPTHIDCA/Department of Gynecology, The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China
[3]College of Software, Xinjiang University, Xinjiang, Urumqi 830046, China
[4]Key Laboratory of Signal Detection and Processing, Xinjiang University, Xinjiang, Urumqi 830046, China

DOI:10.1109/JPHOT.2021.3075958

Manuscript received April 12, 2021; accepted April 22, 2021. Date of publication April 27, 2021; date of current version May 18, 2021. This work was supported in part by the Special Scientific Research Project for Young Medical Science under Grant 2019Q003, in part by the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia under Grant SKL-HIDCA -2019-5, and in part by the Xinjiang Uygur Autonomous Region Science and Technology Branch Project of China under Grant 2019E0282. Corresponding authors: Cheng Chen; Xiaoyi Lv (e-mail: chenchengoptics@gmail.com; xjuwawj01@163.com).

Abstract: Cervical cancer is a serious threat to women's health due to malignant tumours, and early detection can greatly reduce mortality. In this paper, cervical tissue was used as the research object, and Raman spectroscopy analysis of cervical inflammation and precancerous tissues was used to detect cervical cancer. This provides a clinical basis for the use of Raman spectroscopy in analysis of cervical precancerous lesions. In this study, the actual Raman spectrum signal of precancerous cervical tissue was collected, and the PLS and Relief methods were used to extract the signal characteristics of the spectrum. Then, we established and compared KNN and ELM classification models and finally achieved the early diagnosis of cervical cancer. This experiment designed a novel feature fusion method in feature extraction, and we used the first and second derivative features that reflect more peak details of the original spectrum for fusion. The accuracy rate of KNN without feature fusion is 88.17%, and the accuracy rate after fusion is 93.55%. The accuracy rate of ELM without feature fusion is 90.81%, and the accuracy rate after fusion is 93.51%. The results show that the accuracy of feature fusion has been improved to a certain extent, and this method is expected to be used as a new method of spectral data fusion.

Index Terms: Raman spectroscopy, feature fusion, cervical cancer, derivative feature.

1. Introduction

Every year, more than 500000 women worldwide are diagnosed with cervical cancer, and the number of female cervical cancer deaths worldwide is now as high as 300000 [1], [2]. In the early stages of cervical cancer, if effective treatment can be obtained, the patient's 5-year survival rate is as high as 90%. However, the average survival rate of late-stage patients is only 15%–35%

TABLE 1
Experimental Process

Data processing	Improved data processing
a. Background subtraction from original spectral data	a. Background subtraction from original spectral data
b. Data feature extraction (PLS)	b. Data derivation
c. Data classification model (KNN and ELM)	c. Data feature extraction (PLS)
	d. Relief feature selection
	e. Data classification model (KNN and ELM)

[3]. Therefore, the Director-General of the World Health Organization (WHO) issued a global call for action in 2018. This is for the early screening of cervical cancer to eliminate cervical cancer [4], [5]. Cervical inflammation is common among women. Although chronic inflammation can be cured, long-term cervicitis infection can induce cervical cancer [6]. The occurrence of cervical cancer is a relatively slow process. Low-grade squamous intraepithelial lesions (LSIL) (CIN1) are mainly caused by low-risk Human Papilloma Virus (HPV) infection, and high-grade squamous intraepithelial lesions (HSIL) (CIN2 + CIN3) remainly caused by persistent high-risk HPV infection. It takes approximately 5–8 years after HPV infection to cause cervical cancer from CIN. Regardless of the level of CIN, there will be 3 conditions: regression, persistence, and further development of the disease. The lower the grade is, the greater the chance of disease reversal and the lower the risk of disease progression. The higher the grade is, the higher the risk of disease progression. Currently, commonly used diagnostic tools are cervical cytology, HPV testing and colposcopy. Cervical cytology requires a microscopic examination by an experienced cytologist, which cannot be realized in most resource-scarce environments. HPV screening of false negatives is very high. The colposcopy procedure is simple, but its accuracy is affected by human factors [7], [8]. Therefore, there is an urgent need to develop a stable, efficient and fast diagnostic method to provide help for the early diagnosis and pathological research of cervical cancer.

Raman spectroscopy is a nondestructive and highly sensitive optical analysis technique based on inelastic scattering spectroscopy. It has been widely used in the diagnosis of thyroid disease [9], [10], breast cancer [11], prostate cancer [12], hepatitis C [13] and other diseases. Almost all types of diseases, such as cancer and infectious diseases, initially occur at the molecular level [14]. Raman spectra of these changes show that the characteristic peak location, intensity and spectral width may change, which provides the basis for the diagnosis of the disease. This article uses spectral analysis to compare the differences between cervical inflammation, LSIL and HSIL. Raman spectra have a strong background. In this experiment, adaptive iterative weighted least squares (airPLS) were used for background subtraction [15]–[17]. Because spectral data have the characteristics of high dimension, they contain a large amount of redundant information, so this paper adopts partial least squares (PLS) to extract features from spectral data [18], [19] The experiment process is shown in Table 1.

The experimental results show that the accuracy was improved after feature fusion, which shows that the derivative fusion method can be used as a new spectral data fusion method. Wang *et al.* used a support vector machine (SVM) to achieve cervical cancer screening. SVM needs to find the best C and g parameters, and the model runs for a long time. Their research did not provide enough modelling details and contributed little to developing a Raman spectroscopy method to diagnose diseases. The experiment was carried out only once, so the robustness of the model cannot be guaranteed. In addition, there is no innovation in the methodology [20]. This research designed a novel feature fusion method for use in spectroscopy. The spectral analysis of cervical tissue using Raman spectroscopy is expected to be used in clinical screening, and it is expected to become a new method of screening for cervical precancerous lesions.

2. Materials and Methods

2.1 Experimental Materials

In this study, a total of 123 cervical tissue sections were collected from the First Affiliated Hospital of the First Affiliated Hospital of Xinjiang Medical University. According to the clinical diagnosis, 49 samples exhibited inflammation, 29 samples exhibited LSIL, and 45 samples exhibited HSIL. Each formalin-fixed sample was embedded in paraffin and cut into 10-μm-thick cervical tissue sections. According to the standard clinical procedure, we performed the following wax removal method on the sections. The tissue sections were placed in an oven and heated at 60 °C for 40 minutes and then removed immediately and soaked in dewaxing liquid for 15 minutes. Then, the tissue biopsy samples were removed from the wax-removing liquid and cleaned with distilled water to remove residual tissue dewaxing fluid. After wax was removed from the cervical cancer tissue samples, the biopsy samples were placed in a cool, dry, ventilated place. We completed Raman spectroscopy collection in two days.

2.2 Raman Spectral Data Acquisition

Raman spectral data were obtained at ambient temperature (25 $\pm$ 1 °C) by a laser Raman spectrometer (LabRAM HR Evolution Raman spectrometer, HORIBA Scientific Ltd.) with a spectral resolution of approximately 0.35 cm^{-1}. The spectra were recorded at 400 to 1800 cm^{-1} using an excitation wavelength of 532 nm. Using a 50$\times$ objective lens to focus the laser spot, the laser power on the sample surface was approximately 100 mW. The spectral data were acquired in 8 s, and a numerical aperture (NA $=$ 0.5) was used. The laser spot size was 6 μm. To provide the overall sample information, the entire slice was crossed with a span of 100 μm. For each sample, five spectra were recorded at different locations. A total of 615 spectral data points were obtained: 245 for inflammation, 145 for LSIL and 225 for HSIL. During the experiment, all experimental conditions remained unchanged.

2.3 Spectral Data Preprocessing

This experiment obtained Raman spectra containing obvious background because after paraffin embedding and dewaxing, the tissue section usually appears white and diffuse, without fluorescence but with strong scattering [15], [21]. The airPLS background subtraction algorithm is a method to correct spectral baseline drift proposed in recent years. It is mainly to introduce adaptive iteration and sparse matrix technology to quickly fit the accurate background. In addition, as the number of variables increases, its calculation speed is changing in linear relationship, which can quickly obtain the result of the background deduction without having a large variable [22], [23]. This experiment first performs background subtraction on the obtained spectrum data, which effectively removes the fluorescent background. We used airPLS for background subtraction, normalized the derived data, and then normalized the spectrum to eliminate noise interference and improve the convergence speed. This experiment used the mapminmax normalization function. Data analysis was performed in the MATLAB 2018a (Origin, 2018) environment.

2.4 Spectral Analysis

The average Raman spectra processed by background subtraction and normalization are shown in Fig. 1 for the range of 400 to 1800 cm^{-1}. At 548, 640, 721, 830, 938, 1073, 1250, 1309, 1452, and 1664 cm^{-1}, obvious Raman peaks were observed, and all the characteristic peaks of the specific material are shown in Table 2.

By comparing the three spectra, there was a large difference between inflammation and HSIL that was identified. Although the three spectral lines change by roughly the same amount, the intensities of the peaks at 548, 640, 1452, and 1664 cm^{-1} are quite different for the three spectra. Comparing the Raman spectra of early cervical lesions shows that their Raman spectra have many

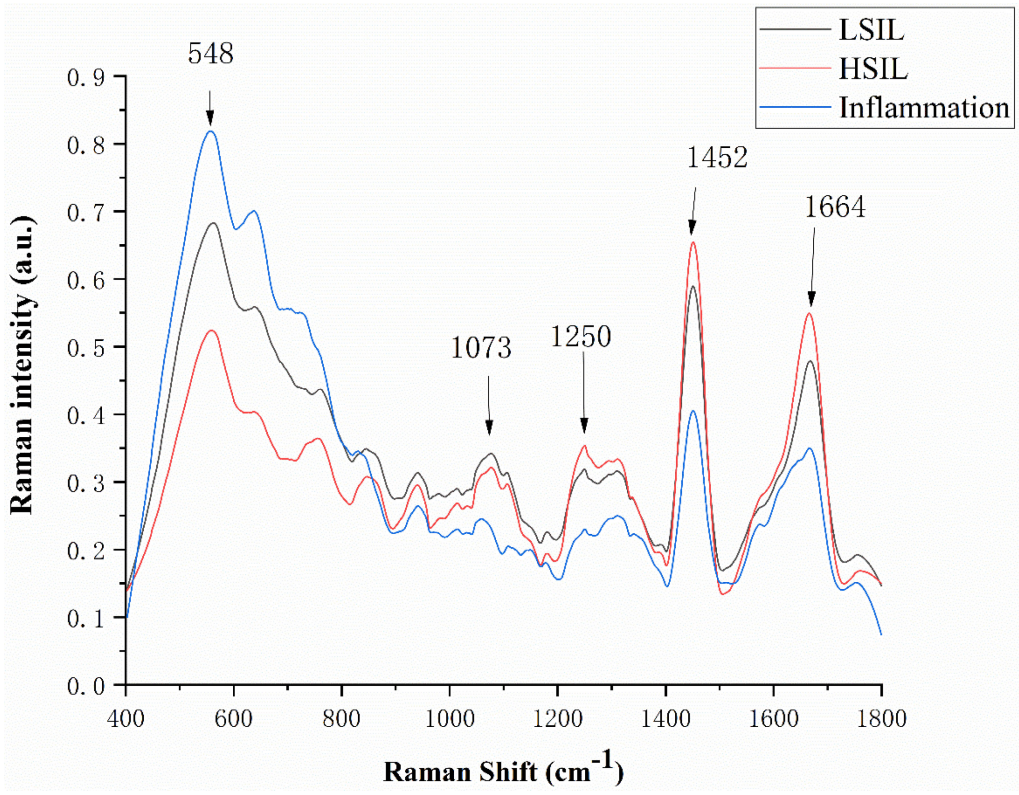


Fig. 1. Normalized average Raman spectrum.

differences, and the intensities of multiple peaks are different. Therefore, based on the differences in the spectral intensities of the three spectra, we can classify them according to the spectral data. There are some differences in the changing trends of the three spectrograms, which lays the foundation for the selection of derivative feature fusion for the difference in changing trends.

2.5 Derivative of the Raman spectrum

The derivative transformation method (differential method) is based on differential analysis or uses mathematical functions to estimate the spectral curve slope. Raman spectral data processing is mainly used to extract different spectral parameters. The derivative Raman spectrum consists of two zero points, positive and negative peaks, which correspond to the peaks, troughs, and Raman spectra on both sides of the inflection point [24]. The curve obtained after the first derivative transformation mainly reflects the transformation rate of the spectral curve, that is, the slope of each point corresponding to the wavelength of the curve. The second derivative spectrum is the result of the second derivative transformation of the original spectrum, so the second derivative is equivalent to increasing each change point on the original spectrum curve based on the first derivative. Studies have shown that the data processing method of derivative spectroscopy (first-order derivative or high-order derivative) has been used to analyse near-infrared spectroscopy and ultraviolet spectroscopy data because of its ability to extract subtle spectral features [25], [26]. The derivation process is as follows. A discrete spectrum x is at wavelength x_k, and the width of the differential window is g:

$$\text{First Derivative}: \quad x_{k,1} = \frac{x_{k+g} - x_{k-g}}{2g}$$

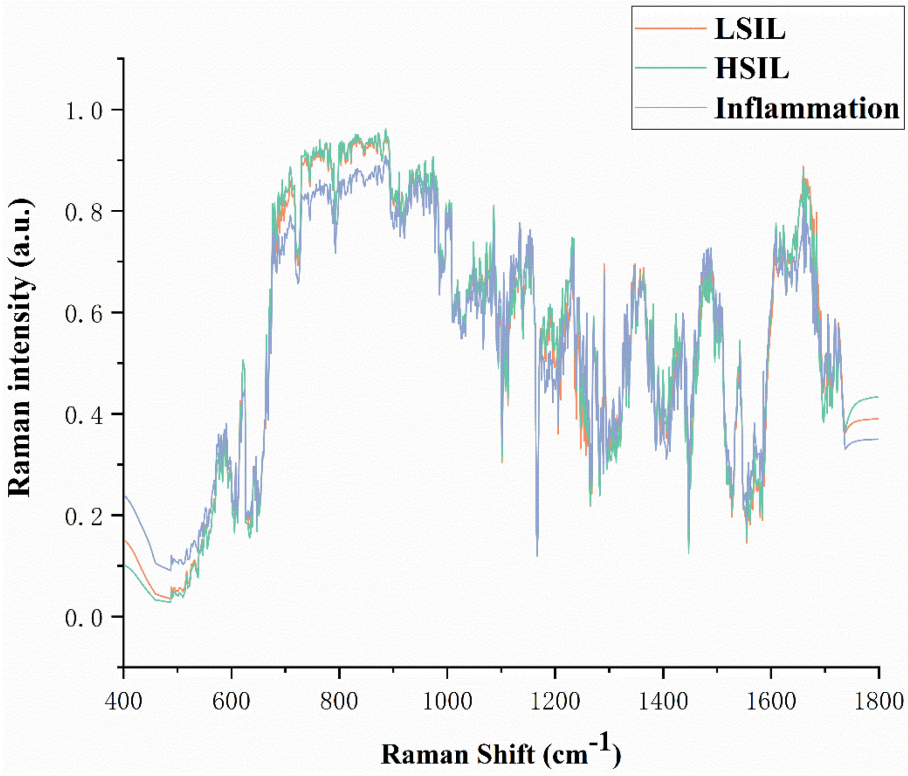


Fig. 2. Normalized first derivative Raman spectrum of the sample.

$$\text{Second Derivative}: \; x_{k,2} = \frac{x_{k+g} - 2x_k + x_{k-g}}{2g}$$

Usually, derivative processing obtains a spectrum with a higher resolution and clearer spectrum variation than the original spectrum. The purpose of second derivative processing is to expand the spectral difference and improve the signal-to-noise ratio so that more effective content can be obtained from the spectrum. As shown in Fig. 3, in the second derivative graph, the original spectrum with insignificant changes has obvious changes in some characteristic bands. These results show that derivative spectroscopy technology has many advantages in reflecting subtle spectral characteristics. The first derivative and second derivative spectra of this experiment are shown in Fig. 2 and Fig. 3, respectively:

2.6 Feature Extraction and Fusion

Because it is difficult to directly distinguish between high-dimensional spectral data, to improve the efficiency of diagnosis, it is very important to reduce the dimensionality and select features of the spectral data before building the model. Multifeatured combination and feature optimization are hot topics in the study of spectral classification. At present, the classification method based on feature selection after fusion has not been used for the early screening of cervical cancer. KIRA *et al.* proposed the Relief feature weight algorithm to calculate feature weights in 1992 [27], [28]. The basic idea is to randomly select a sample through nearest neighbour learning. The distance between samples of the same type with better features is small, and the distance between samples of different types is large. PLS is a supervised linear dimensionality reduction method. It has a good ability to process small samples at high dimension and has been widely used in various research fields. This research used PLS to reduce the dimensionality of the spectral data [18],

TABLE 2
Peak Position and Distribution of the Raman Spectra

Raman bands (cm^{-1})	Assignment
548	Cholesterol
640	C-S stretching & C-C twisting of proteins-tyrosine
721	C-N (membrane phospholipid head)/adenine
830	Tyrosine
938	C-C stretching of the backbone
1073	Carbonate symmetric stretching vibration of calcium carbonate apatite
	Triglycerides (fatty acids)
1250	Amide III
1309	CH_3/CH_2 twisting or bending mode of lipid/collagen
1452	CH_2 deformation in lipids
1664	Amide I

[29]and adopted the Relief method to perform feature selection on the feature fusion data, aiming to establish a more efficient classification model.

3. Results

In this experiment, all samples were randomly divided into a training set and test set at a ratio of 7:3: the inflammation training set contained 34 samples, and the test set contained 15 samples; the LSIL training set contained 20 samples, and the test set contained 9 samples; and the HSIL training set contained 31 samples, and the test set contained 14 samples. Random division was achieved by the MATLAB randperm function. Due to the chance of dividing the dataset randomly, to make the experimental results more convincing, the experimental results are based on three random divisions and averaged. To determine the best classification effect, the 50 principal components extracted from the PLS feature are sequentially sent to the classifier model, and the model results are the average of three runs. For the data obtained after the fusion of the various derivative features, Relief is used for feature selection, and 50 features are still selected for the construction of the classification model.

3.1 KNN Classification Model

KNN is one of the more famous classification recognition algorithms based on statistical methods. It is not only one of the simplest algorithms in machine learning but also one of the most widely used algorithms in machine learning [30], [31]. The idea is that for a given sample data set, a certain distance measurement rule is first established to calculate the distance between the samples to calculate the k nearest sample points to a sample point and then judge which category of the k sample points has more samples to classify the sample points [32]–[36]. In this experiment, the k value is 4, and the highest accuracy rate is 93.55%. Fig. 4 shows the accuracy of the model.

Fig. 4 shows that the construction of the KNN model changes the accuracy of different features. Its accuracy rate is 88.17% in the first 14 features. With the increase in the number of features, the accuracy rate fluctuates relatively greatly, and it is basically stable at approximately 78% in the end. The classification model constructed after 0 + 1-order derivative feature fusion has the highest accuracy rate of 91.94% for the first 23 features, and there is little fluctuation in the follow-up. The classification model constructed after the 0 + 2-order derivative feature fusion has the

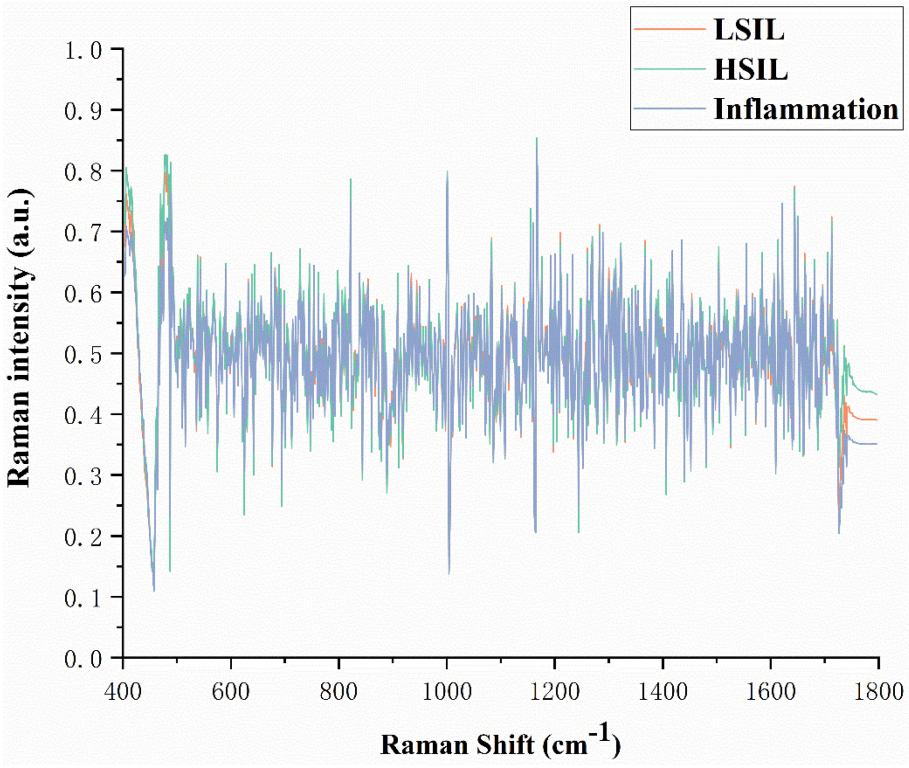


Fig. 3. The normalized and averaged second derivative Raman spectrum of the sample.

TABLE 3
Highest Diagnosis Rate for Each Model of KNN

	0 derivative	0+1 derivative	0+2 derivative	0+1+2 derivative
Cervical inflammation	93.33%	97.78%	93.33%	88.89%
LSIL	86.36%	89.39%	84.85%	92.42%
HSIL	86.67%	90.67%	94.67%	97.33%
Accuracy	88.17%	91.94%	90.86%	93.55%

highest accuracy of 90.86% for the first 33 features. The model has relatively small fluctuations, and the classification effect is relatively stable. The classification model constructed after the fusion of $0 + 1 + 2$-order derivative features fluctuates with the increase in the number of features in the early stage. The accuracy rate is the highest at 93.55% for the first 43 features, and there is a slight decrease in the subsequent results. The specific diagnosis rate of each model is shown in Table 3.

In this paper, the 0-order derivative represents the data after the original spectrum is extracted. The 1-order derivative represents the data after the first derivative of the original data. The $0 + 1$-order derivative fusion refers to the data fusion of the data obtained by the 0-order derivative and the 1-order derivative. The 2-order derivative refers to the data after the second derivative feature extraction of the original data, and the $0 + 2$-order derivative fusion refers to the data obtained after the 0-order derivative and the 2-order derivative are fused.

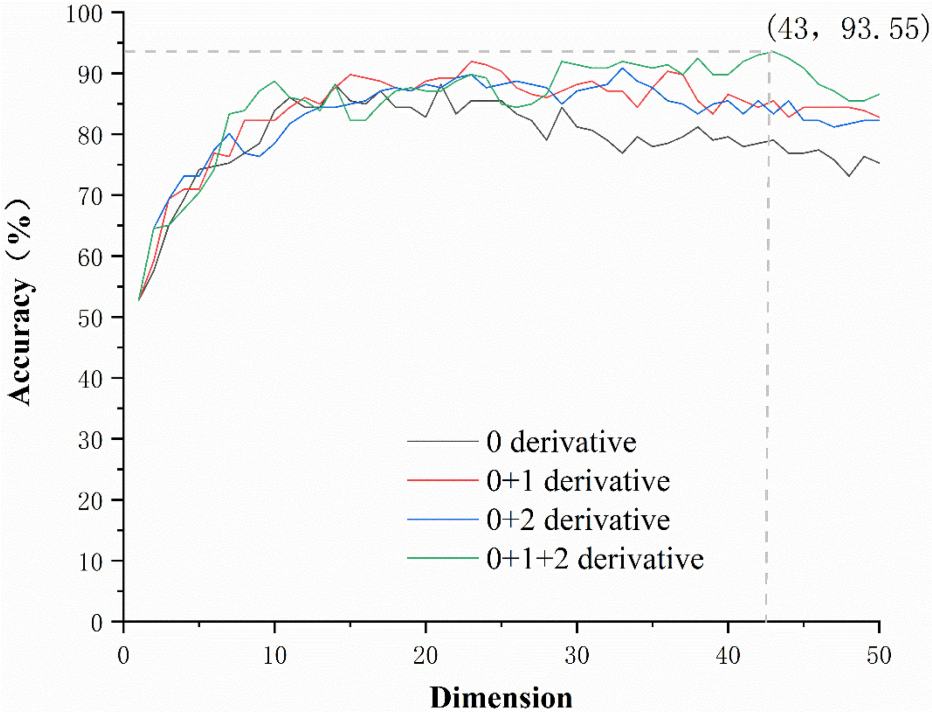


Fig. 4. Accuracy of the first 50 dimensions of each model of KNN (the average of the model results is taken three times).

3.2 ELM Classification Model

The extreme learning machine (ELM) is a new extended network used to train a single hidden layer feedforward neural network. Compared with other existing neural network learning methods, such as backpropagation networks (BP) [37], and typical machine learning algorithms, such as support vector machines (SVM) [10], [38], the ELM performs the random initialization of input weights, and only by solving the equation can the advantage of the primary weight be determined. Therefore, the ELM algorithm has fast and powerful learning capabilities [39]–[41]. The internal state of the ELM has the function of mapping dynamic characteristics so that the system has the function of time-varying characteristics, and the generalization performance of the model is very good. The parameters of the ELM are as follows: the number of hidden neurons is 60, and the transfer function is a sigmoid function. Fig. 5 shows that the accuracy of the ELM model constructed by each feature has changed. Among the fusion models, the 0-order derivative has the highest accuracy of 90.81% in the first 27 features. After feature fusion, 0 + 1-order derivative fusion has the highest accuracy rate of 92.79% in the first 28 features, 0 + 2-order derivative fusion has the highest accuracy of 91.53% when the first 43 features are fused, and the 0 + 1 + 2-order derivative features are fused with the highest accuracy rate of 93.51% in the first 41 features. The accuracy of each model fluctuates with the change in features. The overall accuracy shows a certain increase with the increase in the number of features. After reaching the highest accuracy rate, the accuracy rate fluctuates as the number of features increases and finally stabilizes within a certain range. The specific diagnosis rate of each model is shown in Table 4.

3.3 Result Analysis

This research combines feature selection and feature fusion to construct a classification model. The best model is the KNN model after 0 + 1 + 2 derivative feature fusion. The accuracy rate of

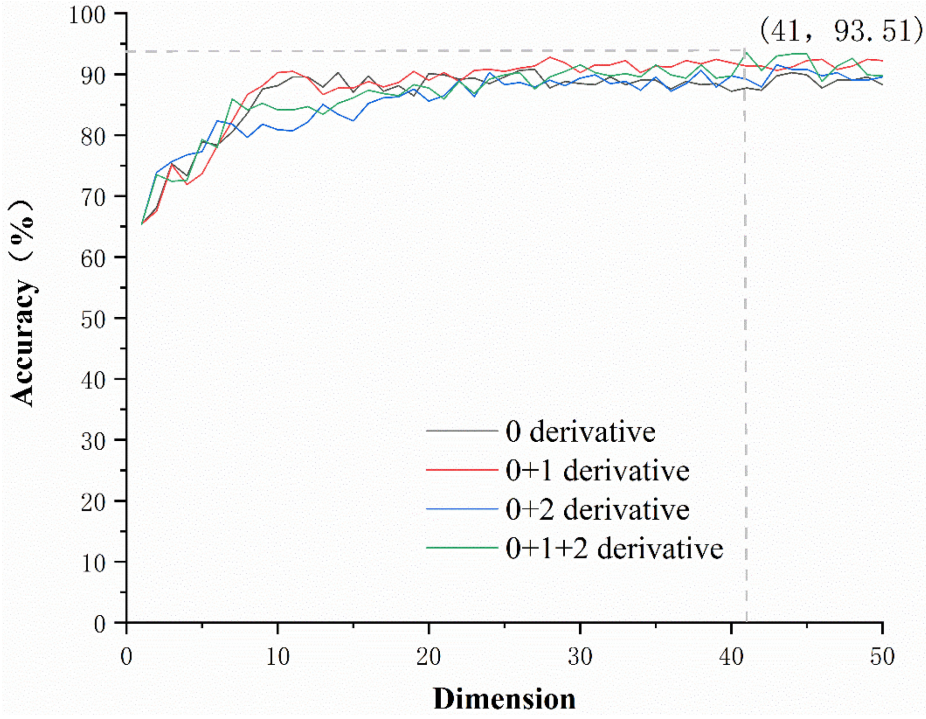


Fig. 5. Accuracy of the first 50 dimensions of each ELM model (the model results are the average of three runs).

TABLE 4
Highest Diagnosis Rate for Each Model of ELM

	0 derivative	0+1 derivative	0+2 derivative	0+1+2 derivative
Cervical inflammation	91.85%	87.41%	93.33%	88.89%
LSIL	83.08%	90.77%	81.54%	91.28%
HSIL	96.89%	97.78%	99.11%	98.22%
Accuracy	90.81%	92.79%	91.53%	93.51%

the first 43 features is 93.55%. The experimental results show that the accuracy of the data has been improved after feature fusion, and the accuracy of the model after $0 + 1 + 2$ derivative fusion is the largest. This shows that the data derivation can retain more effective spectral information to a certain extent, which has great reference value for data modelling.

4. Conclusion

This article attempts to realize the early diagnosis of cervical cancer through Raman spectroscopy combined with machine learning algorithms, which shows the feasibility of the early-stage diagnosis of cervical cancer. In this study, an analysis system was established from the perspective of spectroscopy, and the airPLS-PLS-KNN and airPLS-PLS-ELM models were established, with the highest accuracy rates of 88.17% and 90.81%, respectively. On this basis, to improve the accuracy of model discrimination, a feature fusion method of various orders is proposed. After fusion, the accuracy of the model increased by 5.38% and 2.7%. These improvements may be due to the ability of derivative spectroscopy to extract subtle spectral features, which provides a new idea for

future Raman spectral feature extraction. The sample size used in the model of this research is too small. In subsequent research, we will collect more samples to verify the stability and effectiveness of the model. Because the technology has the advantages of being fast and low cost, it provides a new method for the clinical prevention and treatment of cervical disease.

Conflict of Interest: The authors declare that there are no conflicts of interest related to this article.

References

[1] F. Bray *et al.*, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] P. A. Cohen *et al.*, "Cervical cancer," *Lancet*, vol. 393, no. 10167, pp. 169–182, Jan. 2019.

[3] Y. Jin *et al.*, "Use of autoantibodies against tumor-associated antigens as serum biomarkers for primary screening of cervical cancer," *Oncotarget*, vol. 8, no. 62, pp. 105425–105439, Nov. 2017.

[4] A. Cayon, "WHO Director-General calls for all countries to take action to help end the suffering caused by cervical cancer," PAHO WHO, 2018.

[5] W.-J. Koh *et al.*, "Uterine neoplasms, version 1.2018, NCCN clinical practice guidelines in oncology," *J. Natl. Compr. Canc. Netw.*, vol. 16, no. 2, pp. 170–199, Feb. 2018.

[6] N. Teng *et al.*, "Cervical cancer guidelines. Clinical practice guidelines in oncology," *J. Natl. Compr. Canc. Netw.*, vol. 2, no. 6, pp. 612–630, Nov. 2004.

[7] L. Denny *et al.*, "Screen-and-treat approaches for cervical cancer prevention in low-resource settings: A randomized controlled trial," *JAMA*, vol. 294, no. 17, pp. 2173–2181, Nov. 2005.

[8] C. Muralikrishna *et al.*, "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix," *Vib. Spectrosc.*, vol. 41, pp. 136–141, Jul. 2006.

[9] C. Chen *et al.*, "Exploration research on the fusion of multimodal spectrum technology to improve performance of rapid diagnosis scheme for thyroid dysfunction," *J. Biophotonics*, vol. 13, no. 2, Feb. 2020, Art. no. e201900099.

[10] X. Zheng *et al.*, "Rapid and low-cost detection of thyroid dysfunction using Raman spectroscopy and an improved support vector machine," *IEEE Photon. J.*, vol. 10, no. 6, Dec. 2018, Art. no. 3901412.

[11] H. Nargis *et al.*, "Comparison of surface enhanced Raman spectroscopy and Raman spectroscopy for the detection of breast cancer based on serum samples," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 246, Feb. 2021, Art. no. 119034.

[12] S. Li *et al.*, "Noninvasive prostate cancer screening based on serum surface-enhanced Raman spectroscopy and support vector machine," *Appl. Phys. Lett.*, vol. 105, no. 9, Sep. 2014, Art. no. 091104.

[13] A. Sohail *et al.*, "Analysis of hepatitis c infection using Raman spectroscopy and proximity based classification in the transformed domain," *Biomed. Opt. Exp.*, vol. 9, no. 5, pp. 2041–2055, Apr. 2018.

[14] S. Khan *et al.*, "Analysis of hepatitis b virus infection in blood sera using raman spectroscopy and machine learning," *Photodiagnosis, Photodyn. Ther.*, vol. 23, pp. 89–93, Sep. 2018.

[15] F. Bonnier *et al.*, "In vitro analysis of immersed human tissues by raman microspectroscopy," *J. Raman Spectrosc.*, vol. 42, no. 5, pp. 888–896, May 2011.

[16] G. Yang *et al.*, "Rapid and visual detection of the main chemical compositions in maize seeds based on Raman hyperspectral imaging," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 200, pp. 186–194, Jul. 2018.

[17] F. Chen *et al.*, "Rapid detection of seven indexes in sheep serum based on Raman spectroscopy combined with DOSC-SPA-PLSR-DS model," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 248, Mar. 2021, Art. no. 119260.

[18] T. Mehmood *et al.*, "A review of variable selection methods in partial least squares regression," *Chemom. Intell. Lab. Syst.*, vol. 118, pp. 62–69, Aug. 2012.

[19] N. Zhao, Q. Xu, and H. J. I. A. Wang, "Marginal screening for partial least squares regression," *IEEE Access*, vol. 5, pp. 14047–14055, Jul. 2017.

[20] J. Wang *et al.*, "Raman spectroscopic study of cervical precancerous lesions and cervical cancer," *Lasers Med. Sci.*, vol. 5, Jan. 2021.

[21] S. Li *et al.*, "Identification and characterization of colorectal cancer using raman spectroscopy and feature selection techniques," *Opt. Express*, vol. 22, no. 21, pp. 25895–25908, Oct. 2014.

[22] Z. M. Zhang *et al.*, "An intelligent background-correction algorithm for highly fluorescent samples in raman spectroscopy," *J. Raman Spectrosc.*, vol. 41, no. 6, pp. 659–669, Jun. 2010.

[23] S. Chen and Y. Z. Ling, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst*, vol. 135, no. 5, May 2010, Art. no. 1138.

[24] Y. Lu *et al.*, "Label free hepatitis b detection based on serum derivative surface enhanced Raman spectroscopy combined with multivariate analysis," *Biomed. Opt. Exp.*, vol. 9, no. 10, pp. 4755–4766, Sep. 2018.

[25] L. MICLO *et al.*, "Determination of the ratios of the aromatic amino acid residues by first-or second-derivative UV spectrometry for a simple characterization of peptides," *Int. J. Pept. Protein Res.*, vol. 46, no. 2, pp. 186–192, Aug. 1995.

[26] M. Gallignani, S. Garrigues, and M. J. A. de La Guardia, "Direct determination of ethanol in all types of alcoholic beverages by near-infrared derivative spectrometry," *Analyst*, vol. 118, no. 9, pp. 1167–1173, 1993.

[27] K. Kira, and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. 10th Nat. Conf. Artif. Intell.*, Jul. 1992, pp. 129–134.

[28] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.*, Apr. 1994, pp. 171–182.

[29] R. Rosipal and N. Krämer, *Overview and Recent Advances in Partial Least Squares*, Berlin, Heidelberg: Springer, Nov. 2005. pp. 34–51.
[30] M.-L. Zhang, and Z.-H. J. P. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.
[31] N. Othman *et al.*, "Classification of salivary adulterated NS1 SERS spectra using PCA-Cosine-KNN," in *Proc. Int. Conf. Intell. Inform. Biomed. Sci.*, Shanghai, China, Nov. 2019.
[32] S. Dhanabal, and S. J. I. J. C. A. Chandramathi, "A review of various k-nearest neighbor query processing techniques," *Int. J. Comput. Appl.*, vol. 31, no. 7, pp. 14–22, Jan. 2011.
[33] N. C. Dingari *et al.*, "Development and comparative assessment of Raman spectroscopic classification algorithms for lesion discrimination in stereotactic breast biopsies with microcalcifications," *J. Biophotonics*, vol. 6, no. 4, pp. 371–381, Apr. 2013.
[34] F. Tian, F. Tan, and H. J. V. S. Li, "An rapid nondestructive testing method for distinguishing rice producing areas based on raman spectroscopy and support vector machine," *Vib. Spectrosc.*, vol. 107, Mar. 2020, Art. no. 103017.
[35] H. Zhang *et al.*, "Rapid identification of cervical adenocarcinoma and cervical squamous cell carcinoma tissue based on raman spectroscopy combined with multiple machine learning algorithms," *Photodiagnosis Photodyn. Ther.*, vol. 33, no. 21, Mar. 2021, Art. no. 102104.
[36] R. Stables *et al.*, "Feature driven classification of Raman spectra for real-time spectral brain tumour diagnosis using sound," *Analyst*, vol. 142, no. 1, pp. 98–109, Jan. 2017.
[37] Y. Zhou, Y. Wang, and Q. Yao. "Segmentation of rice disease spots based on improved BPNN," in *Proc. IEEE Int. Conf. Image Anal. Signal Process.*, Jan. 2010.
[38] C. Chen *et al.*, "Raman spectroscopy combined with multiple algorithms for analysis and rapid screening of chronic renal failure," *Photodiagnosis Photodyn. Ther.*, vol. 30, Jun. 2020, Art. no. 101792.
[39] J. Ma, J. Wu, and X. J. A. M. E. Wang, "Fault diagnosis method based on wavelet packet-energy entropy and fuzzy kernel extreme learning machine," *Adv. Mech. Eng.*, vol. 10, no. 1, Jan. 2018, doi: 10.1177/1687814017751446.
[40] J. Ri, and H. J. D. S. P. Kim, G-mean based extreme learning machine for imbalance learning. *Digit. Signal Process.*, vol. 98, Mar. 2020, Art. no. 102637.
[41] Y. Sönmez *et al.*, "Phishing web sites features classification based on extreme learning machine," in *Proc. 6th Int. Symp. Digit. Forensic Secur.*, Antalya, Turkey, Mar. 2018.