

Article

Diagnosis of Breast Cancer Tissues Using 785 nm Miniature Raman Spectrometer and Pattern Regression

Qingbo Li ^{1,*}, Can Hao ^{1,*} and Zhi Xu ²

¹ School of Instrumentation Science and Opto-Electronics Engineering, Precision Opto-Mechatronics Technology Key Laboratory of Education Ministry, Beihang University, Xueyuan Road No. 37, Haidian District, Beijing 100191, China

² Department of General Surgery, Third Hospital, Peking University, Beijing 100083, China; zy1517312@buaa.edu.cn

* Correspondence: qbleebuaa@buaa.edu.cn (Q.L.); HC_LIHUASHI@163.com (C.H.); Tel.: +86-10-8233-8768 (Q.L.)

Academic Editors: Sebastian Wachsmann-Hogiu and Zachary J. Smith

Received: 25 January 2017; Accepted: 16 March 2017; Published: 19 March 2017

Abstract: For achieving the development of a portable, low-cost and in vivo cancer diagnosis instrument, a laser 785 nm miniature Raman spectrometer was used to acquire the Raman spectra for breast cancer detection in this paper. However, because of the low spectral signal-to-noise ratio, it is difficult to achieve high discrimination accuracy by using the miniature Raman spectrometer. Therefore, a pattern recognition method of the adaptive net analyte signal (NAS) weight k-local hyperplane (ANWKH) is proposed to increase the classification accuracy. ANWKH is an extension and improvement of K-local hyperplane distance nearest-neighbor (HKNN), and combines the advantages of the adaptive weight k-local hyperplane (AWKH) and the net analyte signal (NAS). In this algorithm, NAS was first used to eliminate the influence caused by other non-target factors. Then, the distance between the test set samples and hyperplane was calculated with consideration of the feature weights. The HKNN only works well for small values of the nearest-neighbor. However, the accuracy decreases with increasing values of the nearest-neighbor. The method presented in this paper can resolve the basic shortcoming by using the feature weights. The original spectra are projected into the vertical subspace without the objective factors. NAS was employed to obtain the spectra without irrelevant information. NAS can improve the classification accuracy, sensitivity, and specificity of breast cancer early diagnosis. Experimental results of Raman spectra detection in vitro of breast tissues showed that the proposed algorithm can obtain high classification accuracy, sensitivity, and specificity. This paper demonstrates that the ANWKH algorithm is feasible for early clinical diagnosis of breast cancer in the future.

Keywords: Raman spectrometer; breast cancer; pattern recognition; adaptive net analyte signal weight K-local hyperplane (ANWKH)

1. Introduction

Approximately 231,840 new cases of invasive breast cancer and 40,290 breast cancer deaths were estimated to occur among US women in 2015, according to the data provided by the American Cancer Society. Aside from skin cancers, breast cancer is the most common cancer diagnosed among US women, accounting for nearly one in three cancer cases. Breast cancer is also the second leading cause of cancer death among women, after lung cancer [1]. In 2015, approximately 260,000 new cases of breast cancer were diagnosed and 70,000 breast cancer deaths occurred in China. This incidence ranked

first in the female malignant tumors [2]. The data above show that early breast cancer diagnosis is important in implementing timely, effective, and early medical treatments.

In the operation, the traditional pathological diagnostic method is the frozen section pathological diagnosis. However, pathological diagnosis requires waiting a long time and cannot meet the requirements of rapid diagnosis. In addition, the tumor margin cannot be accurately and quickly ascertained, which makes the recurrence rate high and patients need a second operation in a short time. As a form of molecular spectroscopy, Raman spectroscopy can detect cancer-induced changes in the molecular structure and composition of breast tissue. Raman spectroscopy also exhibits characteristics such as sharp peaks, a lack of interference of water, a low amount of samples required, and no chemical treatment of samples required [3–7]. Raman spectroscopy has potential for the early and rapid clinical diagnosis of breast cancers at the molecular level because some chemical changes (e.g., in carbohydrates, lipids, proteins, and nucleic acids) occur during tumor formation.

As a diagnostic technology for breast cancers, Raman spectroscopy has been developed recently. Many investigations on Fourier transform Raman spectroscopy (FTRS), confocal Raman microspectroscopy (CRS), resonance Raman spectroscopy (RRS), surface-enhanced Raman spectroscopy (SERS), and conventional Raman spectroscopy for breast cancer diagnosis have been performed during the development of diagnostic technology [8–15]. SERS, FTRS, CRS, and RRS have been extensively investigated as breast cancer diagnosis tools for allowing reduced fluorescence and high resolution of the Raman spectra [8–11,16–18]. However, due to requiring a large Raman spectrometer or a large desktop microscope, these technologies increase the costs and lessen the portability of clinical diagnosis. Combined with an optical fiber probe, conventional Raman spectrometers are small in size, portable, and low in cost, and hold much promise for in vivo and in situ cancer detection. However, the Raman spectra obtained by the miniature Raman spectrometers exhibit a strong fluorescence background interference and a low spectral signal-to-noise ratio. Both increase the difficulty of data processing and decrease the discrimination accuracy of common data analysis methods. Therefore, considering these key problems for detecting breast cancer, investigating the discrimination analysis method for high classification accuracy is significant.

In this paper, a novel algorithm of adaptive net analyte signal weight K-local hyperplane (ANWKH) is proposed to discriminate cancerous and normal human breast tissue Raman spectra. ANWKH is an extension and improvement of K-local hyperplane distance nearest-neighbor (HKNN) [19]. Although HKNN performs well only for small values of K, it suffers from bias for data with high dimensions [20]. Although AWKH is the improvement of the HKNN method, there is noise, fluorescence, and other mixed component interference. These factors will not only increase the calculation, but also lower the calculation accuracy. It is necessary to improve the AWKH method. ANWKH combines the advantages of the adaptive weight k-local hyperplane (AWKH) and the net analyte signal (NAS) [21]. In this algorithm, NAS was first used to eliminate the interference caused by other non-target factors. The key factors are the biomolecules such as lipids, proteins, and nucleic acids. Other non-target factors are noise interference, fluorescence background, and other components of tissue. Then, the distance between the test set samples and hyperplane was calculated with consideration of the feature weights. The feature weight is estimated by using the ratio of the between-group to within-group sums of squares for the data assigned to the given classes. The HKNN works well only for small values of the nearest-neighbor. However, the accuracy decreases with increasing values of the nearest-neighbor. The method presented in this paper can resolve the basic shortcoming through using the feature weights. The original spectra are projected into the vertical subspace without the target factors. NAS was employed to obtain the spectra without irrelevant information. NAS can improve the classification accuracy, sensitivity, and specificity of the early diagnosis of breast cancer. Consequently, ANWKH is effective for discriminating cancerous and normal human breast tissue Raman spectra.

2. Materials and Methods

2.1. Tissue Specimens

Normal and cancerous samples of human breast tissue were obtained from female patients in Peking University Third Hospital, including four normal tissues and twelve cancerous tissues. The tissues were sourced from female patients aged from 33 years old to 88 years old (mean age is 56 years old). The samples were preserved in liquid nitrogen after the spectra were acquired and sent for the pathological diagnosis as the reference in the spectral analysis. These spectra were supported by pathological analysis (hematoxylin-eosin stain) results.

2.2. Raman Spectral Measurements

The conventional Raman spectra were acquired by an Ocean Optics QE65Pro miniature fiber optic Raman spectrometer at a 785 nm excitation wavelength. Without any chemical treatment, specimens were frozen using liquid nitrogen. They were placed in the glass slide for Raman spectral measurement after thawing at room temperature. The thickness of the sample is about 2 cm. The penetration depth of the probe is about the micron level. The laser power used for acquisition was 30 mW. The integration time was 30 s. All spectra were acquired at a wavelength range from 700 cm^{-1} to 2000 cm^{-1} . 700–2000 cm^{-1} is known as the fingerprint region, which contains complete information about the biomolecules such as lipid, protein, nucleic acids, etc. For objective data, every sample was measured at different locations. Three spectra were also measured at every same location and then averaged to reduce the noise level. In total, 131 sample spectra were obtained from four normal tissues and twelve cancerous tissues on the same environmental conditions on two days (a total of 139 spectra, except for eight outlier spectra samples acquired improperly with other laser powers, which were made by human behaviors' error.). Up to 73 Raman spectra (16 normal and 57 cancerous) were obtained in the first day, and 58 Raman spectra (18 normal and 40 cancerous) were obtained in the second day.

2.3. Spectra Preprocessing Method

Noisy and fluorescence background occurs in the spectra collected by the Ocean Optics QE65Pro Raman spectrometer. Thus, the spectra were preprocessed before use. First, the noise was removed by wavelet transform; then, the fluorescence background was removed by fitting the smoothed spectra to a third-order polynomial function; third, the data sets were normalized to zero mean and unit variance. Data preprocessing can enable clear spectral peaks and optimize the spectral quality.

2.4. Discrimination Analysis Method

In the paper, ANWKH algorithm is proposed. ANWKH algorithm is an improvement and extension of HKNN algorithm. As HKNN performs badly for data with high dimensions, ANWKH resolves the problem and obtains high accuracy by combining the feature of AWKH and NAS. Before the use of the AWKH algorithm, NAS was first employed to eliminate constituents caused by other non-target parameters, such as the noise, background interference, and other components interference. The original spectra are projected into the subspace with various interference factors but without the objective factors to obtain the spectra without irrelevant information. Then, the Euclidean distances between the test set samples and hyperplane were calculated after considering the feature weights estimated by using the ratio of the between-group to within-group sums of squares. The nearest neighbors are selected by the weighted Euclidean distance between the test sample and training set. Finally, the class labels are distinguished according to the Euclidean distances between the test set samples and hyperplane.

The ANWKH algorithm specific procedures are as follows:

Suppose that the training set X consists of L samples with J classes. Each training sample consists of d input features $X_i = (X_{i1}, \dots, X_{id})^T$ with known class label $y_i = c(i = 1, \dots, L; c = 1, \dots, J)$. The goal is to predict the class label of a query with input vector $q = (q_1, \dots, q_d)^T$.

Step 1 The original spectra are reconstructed based on the first f principal components. Then, they are projected into the subspace X_{-k} with various interference factors but without the objective factors. The net spectra x according to the following formulas were then calculated [22]. Each training sample consists of d input features $x_i = (x_{i1}, \dots, x_{id})^T$ with known class label $y_i = c (i = 1, \dots, L; c = 1, \dots, J)$.

$$\begin{aligned} x &= X[I - (X_{-k}^+)^T \times X_{-k}] \\ X_{-k} &= X(I - \beta \times X^+ \times y \times \bar{X}) \\ \beta &= \frac{1}{\bar{X} \times X^+ \times y} \end{aligned} \quad (1)$$

where I is an identity matrix and X_{-k}^+ is the generalized inverse matrix of X_{-k} , which is the subspace with various interference factors but without the objective factors. Besides, y denotes the training sample class label, \bar{X} denotes the mean spectrum of the training sample set, and β is a scalar.

Step 2 Calculating the feature weight w of the training sample, the formulas are as follows [23,24]:

$$\begin{aligned} r_j &= \frac{\sum_i \sum_c I(y_i = c) (\bar{x}_{cj} - \bar{x}_j)^2}{\sum_i \sum_c I(y_i = c) (x_{ij} - \bar{x}_{cj})^2} \\ w_j &= \frac{\exp(r_j)}{\sum_{j=1}^d \exp(r_j)} \quad \forall j = 1, \dots, d \end{aligned} \quad (2)$$

where \bar{x}_j denotes the j th component of the grand class centroid and \bar{x}_{cj} denotes the j th component of class centroid of class c . $I(\cdot)$ denotes the indicator function equals to 1 when $y_i = c$. Otherwise, it is equal to 0. x_{ij} denotes the j th component of the i th training sample.

Step 3 Calculating the weighted Euclidean distance metric D between training samples and the test samples according to the following formula:

$$D(x_i, q) = \sqrt{\sum_{j=1}^d w_j (x_{ij} - q_j)^2} \quad (3)$$

Step 4 In accordance with the Euclidean distance D , we select K nearest neighbors of class c $p_c = (p_{c1}, \dots, p_{cn_c})$ for the given query q . Then, we construct the local hyperplane of class c with p_c as follows:

$$\begin{aligned} LH_c(q) &= \{s | s = \sum_{i=1}^{n_c} \alpha_i V_{i,j} + m_c\} \\ m_c &= \frac{1}{n_c} \sum_{i=1}^{n_c} p_{ci} \\ V_{i,j} &= p_{ci} - m_c \\ \alpha &= (\alpha_1, \dots, \alpha_{n_c})^T \end{aligned} \quad (4)$$

Step 5 Calculating the minimum distance between q and $LH_c(q)$ according to the following formulas:

$$\begin{aligned} J_c(q) &= \min_{\alpha} \sum_{j=1}^d w_j (V_{j,j} \alpha + m_{cj} - q_j)^2 + \lambda \alpha^T \alpha \\ &= \min_{\alpha} (s - q)^T W (s - q) + \lambda \alpha^T \alpha \\ W &= \text{diag}(w_1, \dots, w_d) \end{aligned} \quad (5)$$

where the regularization parameter λ is selected to avoid α being too large. We solve the equation $\frac{\partial J_c(q)}{\partial \alpha} = 0$, then achieve the value α under the minimum distance by $\alpha = (U^T V + \lambda I_{n_c}) / (U^T (q - m_c))$, where $U^T = V^T W$.

Step 6 Evaluating a class label to q by the formula as follows:

$$\text{label}(q) = \text{argmin}_c J_c(q) \quad (6)$$

3. Results and Discussion

3.1. Spectral Preprocessing

From the original spectra without preprocessing (Figure 1), the noises and fluorescence backgrounds exhibited a serious influence on the spectra and decreased the discrimination accuracy. Thus, the noise was removed by wavelet transform. The Symmlet-5 wavelet filter and four-decomposition scales were adopted to smooth the Raman spectra (Figure 2). The fluorescence background was removed by a third-order polynomial function. The Raman spectra of normal and cancerous tissues after preprocessing are shown in Figure 3. The Standard Deviation values are also shown in Section 3.2. The normalization was performed across the samples. In order to eliminate the effect of different excitation light energy and spectral collection efficiency on the whole spectrum, it is convenient to compare the intensity of the spectra and different Raman signals.

The quality of the optimized Raman spectra was improved greatly after data preprocessing (Figures 1 and 3). The Raman peaks of normal and cancerous tissues are pronounced after preprocessing. Thus, preprocessing can improve the discrimination accuracy.

The Raman peaks of normal tissues and cancerous tissues are shown in Figure 3. The Raman peaks of normal tissue occur at 872, 1078, 1305, 1447, 1659 and 1747 cm^{-1} . Meanwhile, the Raman peaks of cancerous tissue occur at 870, 1093, 1278, 1305, 1447, 1466, 1659 and 1747 cm^{-1} .

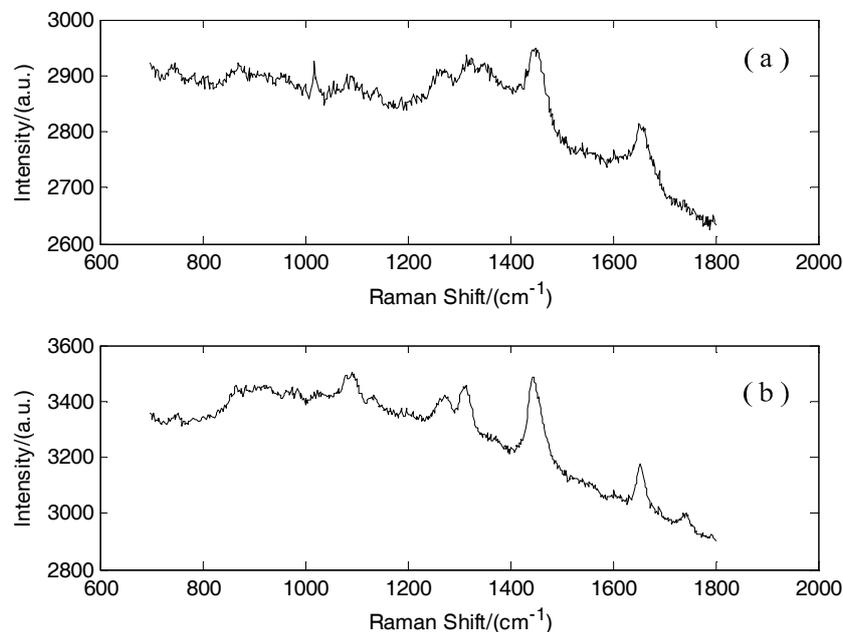


Figure 1. (a) Original Raman spectra of breast cancerous tissue without preprocessing; (b) Original Raman spectra of breast normal tissue without preprocessing.

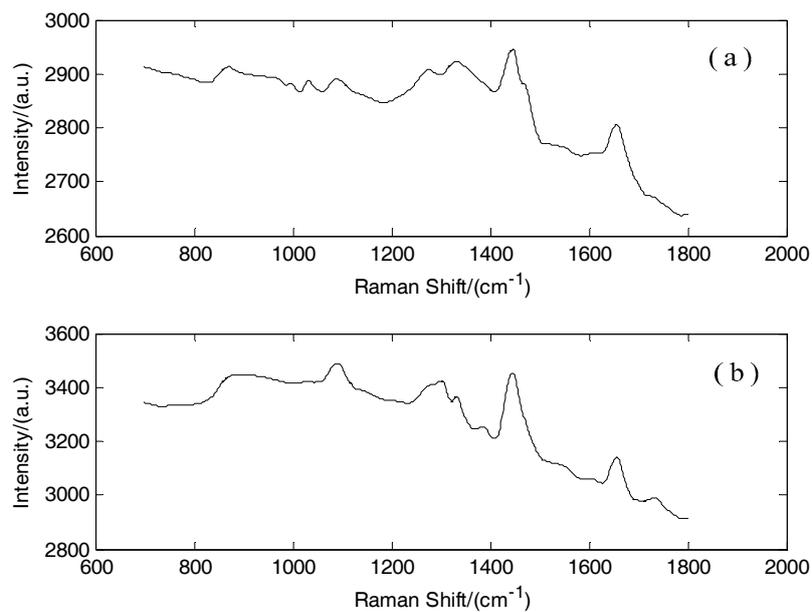


Figure 2. (a) Raman spectrum of cancerous tissue after wavelet transforms; (b) Raman spectrum of normal tissue after wavelet transforms.

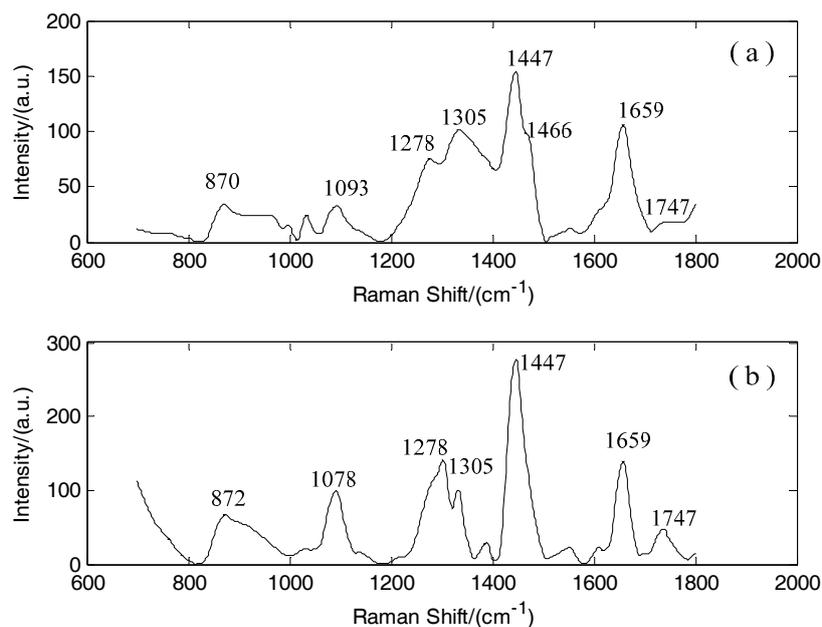


Figure 3. (a) Raman spectrum of breast cancerous tissues after preprocessing; (b) Raman spectrum of breast normal tissues after preprocessing.

According to the research made by using the Raman spectroscopy to measure normal breast tissues and cancerous breast tissues, normal tissue spectra are attributable to lipid molecules, whereas cancerous tissue spectra are attributable to protein molecules. For example, the peak at 1663 cm^{-1} represents amide I, one of the protein molecules. The peak at 1278 cm^{-1} indicates the presence of amide III(C-N stretch), and the peak at 1453 cm^{-1} indicates the presence of CH_2 deformation. The main differences between normal breast tissue spectra and cancerous breast tissue spectra are shown. The peak position representing protein molecules appears at 1278 cm^{-1} in cancerous tissues and nearly disappears in normal tissues. Besides, normal tissue shows a sole prominent lipid peak at 1447 cm^{-1} and 1659 cm^{-1} , where the peak intensities in cancerous tissues decrease obviously compared with

those in normal tissues. Changes have occurred in the configurations, components, and quantities of proteins, lipids, and nucleic acids during tumor formation [25,26]. Normal tissues contain more lipids, whereas cancerous tissues contain more relative proteins. This condition is the basis of breast cancer diagnosis.

Specific assignments of individual peaks can be found in Table 1.

Table 1. Peak positions and assignments of breast tissue.

Peak Position (cm ⁻¹)	Major Assignment
1078	C–C or C–O stretch (lipid)
1278	Amide III(C–N stretch) (protein)
1305	Amide III, α -helix, C–C str and C–H (protein)
1447	Scissoring mode of methylene (CH ₂) (lipid)
1453	CH ₂ deformation (protein)
1653	lipid
1663	Amide I(C=O stretch) (protein)
1747/1750	C=O stretch (lipid)

3.2. Statistical Analysis

The data after preprocessing were separated into two parts. One was the training set and the other was the test set. The STD Dev value of all breast cancerous spectra was 0.0232, and the STD Dev value of all breast normal spectra was 0.1797. Each classifier was learned on the training set and applied on the test set.

The entire spectra were preprocessed at first. Then, the 73 Raman spectra (16 normal and 57 cancerous) obtained on the first day were selected as the training set, and the 58 Raman spectra (18 normal and 40 cancerous) obtained on the second day were selected as the test set. Third, the training and test sets were normalized to zero mean and unit variance. Finally, the test set was classified by ANWKH, AWKH, HKNN, and support vector machine (SVM) classifiers, respectively.

The daily classification results are shown in Table 2.

Table 2. The daily classification results of the test set for ANWKH, AWKH, HKNN, and SVM.

Method	Sensitivity (%)	Specificity (%)	The Positive Predictive Value (%)	The Negative Predictive Value (%)	Accuracy (%)
SVM	92.5	61.1	84.1	78.5	82.75
HKNN	90.0	66.7	85.7	75.0	82.76
AWKH	95.0	72.2	88.4	86.7	87.93
ANWKH	99.2	79.7	91.6	97.9	93.10

The entire data were preprocessed initially and normalized to zero mean and unit variance. Data were then classified by ANWKH, AWKH, HKNN, and SVM classifiers, respectively. Finally, the cross-verification method was used to verify the discrimination accuracy. The results are shown in Table 3.

Table 3. Cross-verification classification results of the test set for ANWKH, AWKH, HKNN, and SVM.

Method	Sensitivity (%)	Specificity (%)	The Positive Predictive Value (%)	The Negative Predictive Value (%)	Accuracy (%)
SVM	96.9	85.3	94.95	90.6	93.89
KNN	97.9	97.1	99.0	94.3	97.71
AWKH	99.0	97.1	99.0	97.1	98.47
ANWKH	99.0	100	100	97.1	99.24

Data processing was conducted two more times. The total 131 Raman spectra were split into two data sets randomly 10 times. Every time, 87 Raman spectra after preprocessing were selected as the training set. The other 44 Raman spectra after preprocessing were selected as the test set. Subsequently, the algorithms were examined. Table 4 shows the average accuracy of the 10 experiments using four different methods with optimal parameters.

Table 4. The average results of random classification for ANWKH, AWKH, HKNN, and SVM.

Method	Sensitivity (%)	Specificity (%)	The Positive Predictive Value (%)	The Negative Predictive Value (%)	Accuracy (%)
SVM	95.1	71.9	90.4	82.6	92.53
KNN	95.9	68.0	89.6	85.0	88.63
AWKH	95.0	74.0	92.2	82.0	93.18
ANWKH	97.1	82.4	94.1	89.8	94.83

As shown in Tables 2 and 3, ANWKH achieved the highest accuracy among the four different classifiers, and AWKH came second. In the experiment, the discrimination accuracy by ANWKH was 93.1%, the sensitivity was 99.2%, the specificity was 79.7%, the positive predictive value was 91.6%, and the negative predictive value was 97.9%. The average results of the random classification for ANWKH, AWKH, HKNN, and SVM are shown in Table 4. It shows that ANWKH values were the highest between the four classifiers on accuracy, sensitivity, specificity, the positive predictive value, and the negative predictive value. The results by ANWKH were much more accurate than the ones by AWKH.

From the experimental results above, ANWKH shows a great advantage for classifying the Raman spectra of breast tissues. The spectra with irrelevant or redundant features can be classified accurately with ANWKH because it eliminates the influence caused by other non-target factors and considers the feature weights. SVM can perform well with large-scale data. However, the choices of the parameters for the kernel are complex and unstable. The HKNN works well only for small values of the nearest-neighbor, but the accuracy decreases with the values of the increasing nearest-neighbor. AWKH can improve the classification accuracy by using the feature weights, which is the improvement of the HKNN method. However, there are noises, fluorescence, and other mixed component interference. These factors will not only increase the calculation, but also lower the calculation accuracy. Consequently, ANWKH is effective and worth studying.

4. Conclusions

Raman spectroscopy, as a sensitive probe on the molecular level, can achieve the early diagnosis of breast cancer and ascertain the tumor margin accurately and quickly at the early stage of tumor formation. The miniature laser Raman spectrometer with a 785 nm excitation is easily used in the clinical diagnosis of breast cancer because of its advantages, such as small size, portability, and low cost. Some disadvantages are strong fluorescence background interference and a low spectral signal-to-noise ratio, which increase the difficulty of data processing and decrease the discrimination accuracy of the data analysis methods. Thus, it is important to investigate the discrimination analysis method for high classification accuracy.

The fundamental experiment results show that the proposed classification algorithm is effective. First, the conventional Raman spectra of breast tissues were acquired by the miniature laser Raman spectrometer at a 785 nm excitation. Then, the preprocessing procedures were investigated. Finally, a novel classification algorithm, ANWKH, was proposed. In this algorithm, NAS was employed to obtain the spectra without irrelevant information. The original spectra are projected into the vertical subspace without the target factors, which can eliminate the irrelevant information caused by other non-target parameters. Then, the distance between the test set samples and hyperplane was calculated after considering the feature weights. The HKNN works well only for small values of the nearest-neighbor.

However, the accuracy decreases with the increasing values of the nearest-neighbor. The method presented in this paper can resolve the basic shortcoming through using the feature weights. The experimental results in vitro indicate that ANWKH achieved high classification accuracy, although the Raman spectra obtained by the miniature laser Raman spectrometer exhibited strong fluorescence background interference and a low spectral signal-to-noise ratio. It is proved that it is viable, rapid and accurate for breast cancer diagnosis in vivo and in situ with a miniature laser Raman spectrometer. In the future, the miniature spectrometer can be used for breast cancer diagnosis in vivo and in situ and can ascertain the tumor margin to relieve female breast cancer patients' pain.

Acknowledgments: This work is supported by the National Natural Science Foundation of P.R. China under Grant No. 61575015.

Author Contributions: Qingbo Li and Can Hao conceived and designed the experiments; Zhi Xu performed the experiments; Qingbo Li and Can Hao analyzed the data; Zhi Xu contributed reagents/materials/analysis tools; Qingbo Li and Can Hao wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carol, E.D.; Stacey, A.F.; Ann, G.S.; Joan, L.K.; Robert, A.S.; Ahmedin, J. Breast Cancer Statistics, 2015: Convergence of Incidence Rates Between Black and White Women. *Cancer J. Clin.* **2016**, *66*, 31–42.
2. Wang, Y. *High Morbidity of Breast Cancer and Gap of Ultrasound Examination Physician*; China Business News: Shanghai, China, 18 October 2016. Available online: <http://newspaper.hwebook.cn/detail.html?id=shdycjrb/dycjrb/20161018/7717a97b-418c-42d7--8eb2-b64b527fef16> (accessed on 28 October 2016).
3. Li, Q.B.; Sun, X.J.; Xu, Y.Z.; Yang, L.M.; Zhang, Y.F.; Weng, S.F.; Shi, J.S.; Wu, J.G. Diagnosis of gastric inflammation and malignancy in endoscopic biopsies based on Fourier transform infrared spectroscopy. *Clin. Chem.* **2005**, *51*, 346–350. [[CrossRef](#)] [[PubMed](#)]
4. Alfano, R.R.; Tang, G.C.; Pradhan, A.; Lam, W.; Choy, D.; Opher, E. Fluorescence spectra from cancerous and normal human breast and lung tissues. *IEEE J. Quantum Electron.* **1987**, *23*, 1806–1811. [[CrossRef](#)]
5. Alfano, R.R.; Liu, C.H.; Sha, W.L.; Zhu, H.R.; Akins, D.L.; Cleary, J.; Prudente, R.; Clemer, E. Human breast tissues studied by IR Fourier transform Raman spectroscopy. *Lasers Life Sci.* **1991**, *4*, 23–28.
6. Pu, Y.; Wang, W.B.; Yang, Y.L.; Alfano, R.R. Native fluorescence spectra of human cancerous and normal breast tissues analyzed with non-negative constraint methods. *Appl. Opt.* **2013**, *52*, 1293–1301. [[CrossRef](#)] [[PubMed](#)]
7. Teh, S.K.; Zheng, W.; Ho, K.Y.; Teh, M.; Yeoh, K.G.; Huang, Z. Near-infrared Raman spectroscopy for early diagnosis and typing of adenocarcinoma in the stomach. *Br. J. Surg.* **2010**, *97*, 550–557. [[CrossRef](#)] [[PubMed](#)]
8. Liu, C.H.; Zhou, Y.; Sun, Y.; Li, J.Y.; Zhou, L.X.; Boydston-White, S.; Masilamani, V.; Zhu, K.; Pu, Y.; Alfano, R.R. Resonance Raman and Raman Spectroscopy for Breast Cancer Detection. *Technol. Cancer Res. Treat.* **2013**, *12*, 371–382. [[CrossRef](#)] [[PubMed](#)]
9. García-Flores, A.F.; Raniero, L.; Canevari, R.A.; Jalkanen, K.J.; Bitar, R.A.; Martinho, H.S.; Martin, A.A. High-wavenumber FT-Raman spectroscopy for in vivo and ex vivo measurements of breast cancer. *Theor. Chem. Acc.* **2011**, *130*, 1231–1238.
10. Yu, C.; Gestl, E.; Eckert, K.; Allara, D.; Irudayaraj, J. Characterization of human breast epithelial cells by confocal Raman microspectroscopy. *Cancer Detect. Prev.* **2006**, *30*, 515–522. [[CrossRef](#)] [[PubMed](#)]
11. Zoladek, A.; Pascut, F.C.; Patel, P.; Notingher, I. Non-invasive time-course imaging of apoptotic cells by confocal Raman micro-spectroscopy. *J. Ram. Spectrosc.* **2011**, *42*, 251–258. [[CrossRef](#)]
12. Haka, A.S.; Volynskaya, Z.; Gardecki, J.A.; Nazemi, J.; Lyons, J.; Hicks, D.; Fitzmaurice, M.; Dasari, R.R.; Crowe, J.P.; Feld, M.S. In Vivo Margin Assessment during Partial Mastectomy Breast Surgery Using Raman Spectroscopy. *Cancer Res.* **2006**, *66*, 3317–3322. [[CrossRef](#)] [[PubMed](#)]
13. Haka, A.S.; Volynskaya, Z.; Gardecki, J.A.; Nazemi, J.; Shenk, R.; Wang, N.; Dasari, R.R.; Fitzmaurice, M.; Feld, M.S. Diagnosing breast cancer using Raman spectroscopy: Prospective analysis. *J. Biomed. Opt.* **2009**, *14*, 054023. [[CrossRef](#)] [[PubMed](#)]

14. Chowdary, M.V.P.; Kalyan Kumar, K.; Mathew, S.; Rao, L.; Krishna, C.M.; Kurien, J. Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: A spectral deconvolution study. *Biopolymers* **2009**, *91*, 539–546. [[CrossRef](#)] [[PubMed](#)]
15. Liu, C.H.; Wang, W.B.; Alimova, A.; Sriramoju, V.; Kartazayev, V.; Alfano, R.R. Monitoring changes of proteins and lipids in laser welded aorta tissue using Raman spectroscopy and basis biochemical component analyses. *Proc. SPIE Int. Soc. Opt. Eng.* **2009**, *7175*, 345–348.
16. Vargas-Obieta, E.; Martínez-Espinosa, J.C.; Martínez-Zerega, B.E.; Jave-Suárez, L.F.; Aguilar-Lemarroy, A.; González-Solís, J.L. Breast cancer detection based on serum sample surface enhanced Raman spectroscopy. *Lasers Med. Sci.* **2016**, *31*, 1317–1324. [[CrossRef](#)] [[PubMed](#)]
17. Yang, J.; Wang, Z.Y.; Zong, S.F.; Song, C.Y.; Zhang, R.H.; Cui, Y.P. Distinguishing breast cancer cells using surface-enhanced Raman scatterin. *Anal. Bioanal. Chem.* **2012**, *402*, 1093–1100. [[CrossRef](#)] [[PubMed](#)]
18. Bitar, R.A.; Martinho, H.S.; Tierra-Criollo, C.J.; Zambelli, R.L.N.; Netto, M.; Martin, M.A.A. Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. *J. Biomed. Opt.* **2006**, *11*, 054001. [[CrossRef](#)] [[PubMed](#)]
19. Okun, O.G. K-local hyperplane distance nearest-neighbor algorithm and protein fold recognition. *Pattern Recognit. Image Anal.* **2006**, *16*, 19–22. [[CrossRef](#)]
20. Vincent, P.; Bengio, Y. K-local hyperplane and convex distance nearest neighbor algorithms. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic NIPS 2001, Vancouver, BC, Canada, 3–8 December 2001; pp. 985–992.
21. Hajian, R.; Mousavi, E.; Shams, N. Net analyte signal standard addition method for simultaneous determination of sulphadiazine and trimethoprim in bovine milk and veterinary medicines. *Food Chem.* **2013**, *138*, 745–749. [[CrossRef](#)] [[PubMed](#)]
22. Li, Q.B.; Yu, C.; Zhang, Q.X. Research on Raman Spectra Quantitative Analysis Model of Ethanol Aqueous Solution Based on Net Analyte Signal. *Spectrosc. Spectr. Anal.* **2013**, *33*, 390–394.
23. Li, Q.B.; Gao, Q.S.; Zhang, G.G. Classification for breast cancer diagnosis with Raman spectroscopy. *Biomed. Opt. Express* **2014**, *5*, 435–2445.
24. Yang, T.; Vojislav, K.; Cao, L.B. Classification by ALH-Fast Algorithm. *Tsinghua Sci. Technol.* **2010**, *15*, 275–280. [[CrossRef](#)]
25. Shafer-Peltier, K.E.; Haka, A.S.; Fitzmaurice, M.; Crowe, J.; Dasari, R.R.; Feld, M.S. Raman microspectroscopic model of human breast tissue: Implications for breast cancer diagnosis in vivo. *J. Ram. Spectrosc.* **2002**, *33*, 552–563. [[CrossRef](#)]
26. Haka, A.S.; Shafer-Peltier, K.E.; Fitzmaurice, M.; Crowe, J.; Dasari, R.R.; Feld, M.S. Diagnosing breast cancer by using Raman spectroscopy. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12371–12376. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).