

# Fast Characteristic of Skin Lesions by Machine-Learning of Raman Spectrum

**Hua Zhang**

Jilin University First Hospital

**Danhua Wang**

Jilin University First Hospital

**Limei Qu**

Jilin University First Hospital

**Ying Xue**

HOOKE Instruments Ltd.

**Xinli Li**

HOOKE Instruments Ltd.

**Bei Li**

Changchun Institute of Optics Fine Mechanics and Physics Chinese Academy of Sciences

**Bin Liu** (✉ [l\\_bin@jlu.edu.cn](mailto:l_bin@jlu.edu.cn))

Jilin University First Hospital <https://orcid.org/0000-0003-3179-6235>

---

## Research

**Keywords:** skin tumor, Raman spectroscopy, machine learning, molecular diagnosis

**Posted Date:** January 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-154353/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

---

1 Fast Characteristic of Skin Lesions by Machine-Learning of  
2 Raman Spectrum

3 Hua Zhang<sup>1</sup>, Danhua Wang<sup>1</sup>, Limei Qu<sup>1</sup>, Ying Xue<sup>3</sup>, Xinli Li<sup>3</sup>, Bei Li<sup>2,3,\*</sup>, Bin Liu<sup>1,\*</sup>

4 <sup>1</sup>The First Hospital of Jilin University, Changchun 130061, China

5 <sup>2</sup>The State Key Lab of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics,  
6 CAS, 130033 Changchun, China

7 <sup>3</sup>HOOKE Instruments Ltd., 130033 Changchun, China

8

9 **Correspondence**

10 Prof. Bin Liu, The First Hospital of Jilin University, 71 Xinmin St, Changchun, Jilin, China. Tel:

11 +8613756661290; E-mail: l\_bin@jlu.edu.cn;

12 Prof. Bei Li, The State Key Lab of Applied Optics, Changchun Institute of Optics, Fine Mechanics

13 and Physics, CAS, 130033 Changchun, Jilin, China. Tel: +8618343116622; E-mail:

14 beili@ciomp.ac.cn.

15 **Abstract**

16 **Background:** The traditional diagnosis of skin lesions mainly relies on dermoscope and  
17 pathological biopsy, of which the former is non-objective and the latter is invasive and time-  
18 consuming. It is necessary to find an objective and non-invasive inspection method for the diagnosis  
19 of skin cancer which is the most common malignant tumor. Herein, we aimed to fast identify the  
20 skin cancers on ultrathin frozen fresh tissue sections by combining Raman spectroscopy detection  
21 and machine learning technology.

22 **Methods and material:** 22 fresh frozen tissue sections including 3 squamous cell carcinomas, 11

---

23 basal cell carcinomas, 2 malignant melanomas, 3 seborrheic keratosis, and 3 melanocytic nevi, were  
24 included and performed Raman detection. To prevent the discrete Raman data distribution affecting  
25 the generalization ability of the learning model, a series of adaptive preprocessing algorithms were  
26 first applied to standardize the raw Raman data of five skin lesions. The processed Raman data were  
27 performed visualized cluster analysis by principal components analysis (PCA) and t-distributed  
28 stochastic neighbor embedding (t-SNE). And, using K-nearest Neighbor (KNN) and support vector  
29 machine (SVM) classifiers, two predictive models for diagnose were established and evaluated in  
30 the training set and test set by the confusion matrixes and receiver operating characteristic (ROC)  
31 curves.

32 **Results:** The mean variance Raman spectrum graph of 5 skin lesion types were acquired after  
33 standardization procession and 4 peak positions with large differences were found. Through  
34 dimensionality reduction by PCA and t-SNE, the visual clustering results of Raman data showed  
35 heterogeneous intra-cluster homogeneity and inter-cluster dispersion. The test accuracies reached  
36 94.56% and 98.94% in KNN and SVM classifiers respectively. The areas under the ROCs of the  
37 two classifiers, in the category dimension and the sample dimension, were all more than 0.99 which  
38 is close to the perfect classification effect.

39 **Conclusions:** Raman spectroscopy is a competitive candidate for the fast and accurate diagnosis of  
40 skin lesions and the molecular information provided may be used in the pathological classification,  
41 predicting immunotherapy responsiveness and stratifying prognostic risk. Furthermore, the  
42 combination of Raman spectroscopy and machine learning methods showed great diagnostic  
43 capabilities with high accuracy is a promising tool for the diagnosis of skin lesions.

44 **Keywords:** skin tumor, Raman spectroscopy, machine learning, molecular diagnosis

45

---

46 **Introduction**

47 Skin cancers, including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and  
48 melanoma (MM), are the most common malignancies worldwide.[1] Among them, BCC and SCC  
49 collectively named nonmelanoma skin cancer (NMSC), account for about 40% of all malignant  
50 tumors.[1, 2] According to the World Health Organization (WHO) in 2014, 2 ~ 3 million and 13  
51 kilo people suffered from NMSC and MM worldwide each year respectively.[3] In the United States,  
52 approximately 3 million patients of NMSC are expected to be treated each year while ~10 thousands  
53 new cases and almost 7 thousands death arise.[4, 5] Moreover, ~ 40% of patients will relapse within  
54 2 years.[6]

55 The diagnosis of skin lesions mainly relies on dermoscope and pathological biopsy. Dermoscope  
56 is a non-invasive in situ diagnostic tool based on visual and morphological recognition. However,  
57 it is non-objective and highly depends on the experience of doctors. Many studies showed the  
58 accuracy of melanoma diagnosis by dermatologists varies between 56% and 82.8%, while up to one  
59 third of melanomas were misdiagnosed as benign lesions.[7, 8] Although pathological biopsy is the  
60 gold standard for diagnosis, its invasion and time-consuming burden on patients and doctors, and  
61 increases a number of unnecessary biopsies. The number needed to treatment for the resection of  
62 one malignant skin lesion was reported 20 ~ 59.[9, 10] Therefore, it is necessary to find a non-  
63 invasive, objective and high-efficient screening and diagnosis method.

64 Raman spectroscopy (RS) has been widely used for molecular detection due to its ultra-fine  
65 capability of acquiring chemical composition, structure and spatial information. In 1997, Gniadecka  
66 et al. observed the changes in lipids (1301 - 1321  $\text{cm}^{-1}$ ) and proteins (2939 - 2948  $\text{cm}^{-1}$ ) of skin  
67 lesions through RS.[11] As more attention were attracted by Raman in medicine[12, 13], RS has

---

68 been widely used in tumor identification and efficiently improved the sensitivity and specificity of  
69 tumor diagnosis.[14-16] However, there are still some challenges for RS for skin lesions. Firstly,  
70 paraffin embedding, formalin fixation as well as dewaxing were widely used in tissue  
71 preparation.[17-19] They were usually accompanied with changes of tissue composition and loss of  
72 tissue fluid, resulting in amounts of untrue and incomplete information acquired.[20, 21] Therefore,  
73 "zero processing" tissue specimens may be the best choice for Raman. Secondly, the Raman spectra  
74 on tissue contains complex and huge information reflecting the intricate lesion progression and  
75 microenvironment. [22, 23] The result is, too many peaks arise and become stupendous to  
76 distinguish them and identify the underlying skin cancers.

77 Artificial intelligence (AI) based on machine learning (ML) has attracted huge attention in recent  
78 years in medical research. A method of non-negative matrix factorization was used and revealed the  
79 diversity of tumor mutation process.[24] A distance-based classifier with a 90%-above recognition  
80 accuracy was established and used to develop more personalized cancer immunotherapy.[25]  
81 Moreover, the accuracies of predicting cancer susceptibility, recurrence and mortality were  
82 increased by 15% - 25% by ML algorithm.[26] Powerful data processing capabilities of AI may be  
83 able to provide good assistance for RS diagnosis.

84 Here in this paper, we aimed to fast identify skin lesions using 22 ultrathin frozen fresh tissue  
85 sections by combining RS detection and ML technology (Fig. 1). To prevent the affections of  
86 discrete Raman data on the generalization of the learning models, a series of adaptive preprocessing  
87 algorithms were first applied to standardize the Raman data before performing both supervised and  
88 unsupervised identification. Linear principal components analysis (PCA) and non-linear t-  
89 distributed stochastic neighbor embedding (t-SNE) dimensionality reduction were used to visualize

---

90 clustering and both showed heterogeneous intra-cluster homogeneity and inter-cluster dispersion.  
91 K-nearest Neighbor (KNN) and support vector machine (SVM) classifiers were used in category  
92 classification. The mean areas under the receiver operating characteristic (ROC) curves of KNN and  
93 SVM in the category and sample dimensions were all greater than 0.99, which is close to the perfect  
94 effect. Besides, the significant differences of tryptophan, phenylalanine and proline in 5 skin lesion  
95 types indicated by RS may be used in the pathological classification, predicting immunotherapy  
96 responsiveness and stratifying prognostic risk.

## 97 **Methods**

### 98 **Sample preparation**

99 During the period from December 2018 to September 2019, patients with skin lesions who passed  
100 the routine examinations and were recommended for surgical resections by more than two  
101 specialists, entered our screening process. According to our inclusion criteria (16 - 65 years old; no  
102 underlying diseases; no history of tumors; no history of drug use, smoking and drinking in the last  
103 month; non-pregnant status), a total of 40 patients were included.

104 After surgical removal of the lesion, we dripped the embedding agent (OCT embedding agent,  
105 Servicebio, Cat: G6059) on the lesion tissue ( $25 \times 25 \times 2$ mm) which was placed on the supporter.  
106 Then the tissue together with the supporter were put on the freezing table (Leica CM1950 incubator  
107 cryomicrotome) and clamped by the microtome holder. Two adjacent sections with thickness of 5  
108  $\mu$ m were obtained with one for Hematoxylin-eosin (HE) staining. While the other one was kept at  
109  $4^\circ\text{C}$  for Raman inspection on the quartz glass slide (Delmon Technology; model  $72\text{ mm} \times 25\text{ mm}$   
110  $\times 1\text{ mm}$ , manufacturer Wuxi Zejia Analytical Instrument Co., Ltd). All the section processes were

---

111 performed by professional technicians from Pathology Department of the First Hospital of Jilin  
112 University.

### 113 **HE staining and pathological diagnosis**

114 The frozen sections attached to the glass slides were fixed with a mixture of 95% ethanol and 5 ml  
115 glacial acetic acid for 1 min and washed with tap water. Then we stained sections with  
116 hematoxylin for 1 - 2 min, and washed with tap water and dilute ammonia for 20 s and 30 s  
117 respectively. After washing for another 20 s with tap water, sections were eosin-stained for 20 s - 1  
118 min, dehydrated and sealed with neutral gum. All the works were completed by professional  
119 technicians from Pathology Department of the First Hospital of Jilin University. The prepared HE  
120 sections were handed over to two or more pathologists for diagnosis.

### 121 **Raman spectroscopy detection**

122 To ensure homogenization, we selected 22 sections with a single pathological result, including 3  
123 SCC cases, 11 BCC cases, 2 MM cases, 3 SK cases and 3 MN cases. After anchoring lesion areas  
124 of the HE sections, the corresponding quartz slides were placed on the stage of the Raman  
125 spectrometer (HOOKE D100, HOOKE Instruments Ltd., China) for detection and spectrum  
126 collection. The laser wavelength for Raman detection is 785 nm. The single spectrum acquisition  
127 condition is 25 mw for 10 s. 50 spectra for each sample and 1100 spectra in total were collected  
128 respectively.

### 129 **Raman spectrum processing**

130 To decrease the noise and interference, it is necessary to train the data before identifying skin lesions  
131 by Raman spectra. In order to reduce impact of the data scale and distribution on the learning model,  
132 we standardized the spectral data and map it to the same dimension. Then, preprocessing operations  
133 such as smoothing and baseline correction were performed. Additionally, to smooth the curve, the  
134 Savitzky-Golay (SG) filtering algorithm was applied to fit the low-frequency component (the signal

---

135 part) and remove the high-frequency component (the noise part) from the Raman spectrum. Finally,  
136 we used adaptive iteratively reweighted penalized least squares algorithm (airPLS) baseline  
137 correction algorithm to remove the background introduced by fluorescence, chip and the tissue slide  
138 itself. As an example, the Raman spectrum standardization results of BCC were shown in Figure 2,  
139 as well as the preprocessing schemes and corresponding algorithms in Table 1.

## 140 **Algorithms analysis**

### 141 **Unsupervised learning**

142 The Raman spectra of skin lesions has 701 dimensions, of which many are noise and redundant  
143 information that have no contribution to classifications. To compress spectral dimensions and reduce  
144 overfittings, a feature dimension reduction method is required. Without knowing data feature  
145 contribution, unsupervised learning used in data compression can improve the usability of the  
146 algorithms and their performance in high dimensions, and is helpful for the visualization of the data.

147 (1) PCA (principal component analysis) is a conversion technique used in unsupervised linear  
148 data, which is the most widely used data compression algorithm[27, 28]. It carries out orthogonal  
149 transformation (a kind of linear transformation, in which the inner product of two vector spaces  
150 remains unchanged during transformation) according to the data characteristics to eliminate the  
151 correlation between each component of the original vectors. The corresponding eigenvectors with  
152 decreasing eigenvalues are obtained by transformation. After orthogonal transformation, the high-  
153 dimensional space Raman spectrum can be expressed as a low-dimensional space.

154 (2) t-SNE (t-distributed stochastic neighbor embedding) algorithm is a non-linear dimension  
155 reduction method. First, it converts the Euclidean distance between two high-dimensional space  
156 data points into similarity probability. Then, the joint probability of the high-dimensional space data  
157 point and the corresponding low-dimensional space analog data point is used to replace the  
158 conditional probability in the random neighborhood embedding algorithm. t-SNE makes the shorter

---

159 distance data points in high dimensional space have larger distances after mapping, so that the points  
160 in the same cluster were gathered more closely, and the points in different clusters farther apart,  
161 which effectively solves the data crowding problem in the low-dimension space.

## 162 **Supervised learning**

163 Compared with unsupervised learning, the training data had eigenvalues and label values. Through  
164 the study of training data, the learning model independently established the connection between  
165 eigenvalues and label values, and predicted label values based on data features. Following are the  
166 methods we used, KNN (K-nearest Neighbor) and SVM (support vector machine).

167 (1) KNN, also known as the nearest neighbor algorithm, is based on an analogous learning  
168 method by comparing a given test tuple with its similar training tuple. For each new data, the  
169 closest K data will be found in the given data tuple, then the K data and the new data will be  
170 initially set to the same category. In this paper, Euclidean distance was adopted. Suppose the

171 Euclidean distance of two points or tuple sum ( $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ ,  
172  $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ ) is:

$$173 \quad \text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (1)$$

174 (2) SVM is a method of classifying linear and nonlinear data and has been widely used in many  
175 clinical predictions[29, 30]. It maps the training data to higher new dimensions and searches for the  
176 best classification plane (decision boundary) which can separate the data into different classes. Two  
177 data tuples can always be separated by the decision boundary as higher as the non-linear mapping  
178 dimension is. For RS data with a lot of eigenvectors, the calculation of the inner product in the high-  
179 dimensional space is too large to solve and remains the core of SVM function. In this article, SVM

180 kernel function selects polynomial kernel: the kernel function of samples  $x_i$  and  $x_j$  is  
181  $\kappa(x_i, x_j)$ ,  $d$  is the degree of the polynomial, when  $d = 1$ , polynomial kernel degenerates to  
182 linear kernel:

$$183 \quad \kappa(x_i, x_j) = (x_i^T x_j)^d \quad (2)$$

---

184 **Liquid chromatography-mass spectrometry (LC-MS)**

185 According to our inclusion criteria and single pathology principle, another three lesion tissues (Table  
186 5, 6) were obtained for LC-MS. After getting the same amounts of tissues from the central of lesions,  
187 chromatographic grade methanol (mass volume ratio 1 g : 2.5 mL) were immediately added in and  
188 vortexed for 1 min. Then tissues were homogenized for 3 min with 2 ~ 3 zirconium dioxide grinding  
189 beads. After grinding 3 min, the homogenates were centrifuged at 14,000 rpm for 10 min at 4 °C,  
190 and the upper aqueous layers were used in LC-MS analysis. The standards of phenylalanine and  
191 tryptophan were dissolved in pure methanol to obtain 2.00 mg/mL stock solutions. Before sample  
192 detection and analysis, the stock solutions were diluted with pure methanol and made into mixed  
193 standards (2000 ng/mL).

194 The high-resolution mass spectrometry (MS) (Q Exactive™, Thermo Fisher Scientific (China)  
195 Co., Ltd.) coupled with electrospray ionization (ESI) was performed in the positive and negative  
196 ion switching scan mode. Parallel reaction monitoring (PRM) was selected in the detection. The  
197 resolution of the equipment is 17500 and scan range 50.0~500.0 m/z. During detection, the spray  
198 voltage was set at 3.2 kV in positive ionization mode, capillary temperature at 300 °C and nitrogen  
199 at 40 Arb. Data collection time was 8.00 min. Analyte information were shown in Table 2. The  
200 Liquid chromatography (LC) (UltiMate 3000 RS, Thermo Fisher Scientific (China) Co., Ltd.) used  
201 T3 column (2.1 × 150 mm 3 μm, waters) with a flow rate of 0.30 ml/min which was maintained at  
202 35 °C. The aqueous phase was 10 mM ammonium formate solution at PH 3.0 adjusted by formic  
203 acid and the organic phase was acetonitrile. The elution gradients were showed in Table 3. The  
204 injection volume was 5 μL for each sample.

205 The chromatogram acquisition and integration were processed by the software Xcalibur 3.0  
206 (Thermo Fisher) and linear regression with  $1/X^2$  as weighting coefficient was performed to get the  
207 standard curves of phenylalanine and tryptophan (Table 5, 6).

208 **Results**

209 **Raman characteristics and molecular information of five skin lesion types**

---

210 In order to avoid the uneven distribution of categories, 50 Raman spectra were collected for each  
211 lesion sample and a total of 1100 Raman spectra were acquired. The 1100 spectra were sequentially  
212 standardized and batched. After processing, the data was trained and tested for the identification  
213 model. The following Figure 3 showed the mean variance map of the preprocessed Raman spectra  
214 of 5 skin lesions, where the solid lines were the mean spectra, and the shaded bars represented the  
215 standard deviations within groups.

216 From the mean variance graph of Raman spectra of 5 skin lesions, 4 peak positions ( $720\text{ cm}^{-1}$ ,  
217  $752\text{ cm}^{-1}$ ,  $853\text{ cm}^{-1}$ ,  $1002\text{ cm}^{-1}$ ) with significant differences were noticed. Their physical origins and  
218 peak intensity disparities were summarized in Table 4 and Figure 4. The peak intensities and spans  
219 of SCC, MM and BCC at  $720\text{ cm}^{-1}$  (nucleic acid band)[31] (Table 4) were higher than those of SK  
220 and MN (Fig. 4A). As nucleic acid is positively correlated with tumor malignancy[32], this result  
221 preliminarily verified the ability of RS in tumor detection. The peak intensities of  $752\text{ cm}^{-1}$  (the  
222 symmetrical respiration of tryptophan)[33] and  $1002\text{ cm}^{-1}$  (the vibration mode of the ring breathing  
223 caused by phenylalanine)[33, 34] (Table 4) in SCC and MM were significantly higher than other  
224 lesions (Fig. 4B, D). In order to verify the reliability of our test results, we used LC-MS to detect  
225 the contents of phenylalanine and tryptophan in SCC, SK and MN. The contents of phenylalanine  
226 (Table 5) and tryptophan (Table 6) in SCC were indeed higher than SK and MN. In addition, the  
227 peak intensity of  $853\text{ cm}^{-1}$  (the stretching of collagen proline ring (C-C))[35] (Table 4) in MN was  
228 much higher than other lesions, followed by SK (Fig. 4C), indicating that proline decreases as the  
229 malignancy of skin lesions increases.

### 230 **Visualized clustering results in t-SNE and PCA**

231 Using the standardized Raman data of 5 skin lesions, two ways were carried out for cluster analysis.  
232 Figure 5 showed that nonlinear t-SNE dimensionality reduction and linear PCA were used to  
233 visualize the clustering results. Figure 5A showed t-SNE dimensionality reduction results which  
234 using two largest contribution dimensions of t-SNE 1 and t-SNE 2 achieved highly nonlinear  
235 distinguishable for these 5 types of skin lesions. Figure 5B showed the result of PCA three-  
236 dimensional visualization. The three largest principal components of PC1, PC2 and PC3 spectra

---

237 were used to achieve linear separable of 5 types of skin lesions. The above two unsupervised  
238 learning methods both showed heterogeneous intra-cluster homogeneity and inter-cluster dispersion.

### 239 **Confusion matrixes and validations of SVM and KNN models**

240 Next, RS data of five skin lesions were learned and analyzed by two common splitters of SVM and  
241 KNN in supervised learning. 20% RS data were tested and confusion matrixes of the recognition  
242 results were showed in Figure 6 (A, B). In the confusion matrixes, the horizontal direction  
243 represented the true category label, the vertical direction labeled the represented category label (the  
244 category label corresponding to the highest predicted probability), and the diagonal value indicated  
245 the recognition accuracy of the corresponding category test data. Calculating the mean value of the  
246 diagonal lines of the confusion matrixes, KNN and SVM test accuracies were 94.56% and 98.94%  
247 respectively. In KNN, 11.1% of SCCs were misjudged as BCCs, 5.6% of SKs were confused with  
248 MMs and 10.5% of MNs were misdiagnosed as SCCs (Fig. 6A). In SVM, 5.3% of MNs were  
249 misjudged as SKs (Fig. 6B).

250 With false positive rate (FPR) as the horizontal axis and true positive rate (TPR) as the vertical  
251 axis, ROC curves of five skin lesion categories were drawn in KNN and SVM, and area under  
252 curve (AUC) was used to measure the excellence of the prediction models (Fig. 6C, D). Macro-  
253 average ROC curves were drawn using the mean value of ROC curves of 5 categories indicating  
254 category dimension prediction. Micro-average ROC curves were drawn using the mean value of  
255 ROC curves of all test samples indicating sample dimension prediction.

256 After calculating, the AUCs of the five categories in KNN classifier were all greater than 0.97,  
257 and the mean AUCs in the category dimension and the sample dimension were both 0.99 (Fig.

---

258 6C). The AUCs of all test samples and categories both were 1 in the SVM classifier (Fig. 6D).

259 These data indicated that KNN and SVM were all close to perfect classifiers.

## 260 **Discussions**

261 Since the specimens in pathology department were all preserved with formalin fixation and paraffin  
262 embedding, many human tissue specimens were tested directly within the wax blocks[17, 18] and  
263 some underwent gentle dewaxing treatment[19] or digital dewaxing of RS signal[46]. Although  
264 some studies demonstrated that the detection of wax or formalin-fix blocks have no effects on the  
265 Raman spectra of specimens and paraffin tissues can be almost completely dewaxed[21, 47, 48],  
266 some studies revealed that the amount of paraffin remained varied in different tissues[21], and  
267 formalin fixation affected Raman information of tissues[20]. Huang Z et al compared the Raman  
268 spectra of fresh human bronchial tissues with formalin-fixed tissues and found that formalin fixation  
269 has a significant effect on the near-infrared Raman spectra of tissues and the diagnostic markers  
270 from the 980-1100 and 1500-1650  $\text{cm}^{-1}$  regions of fixed tissues do not seem to be suitable for in  
271 vivo lung cancer detection[20]. To provide accurate Raman information for in vivo applications, it  
272 may be better to use fresh tissue specimens. Some experiments use frozen or fresh tissue blocks.[49,  
273 50] However, although they were free from the interference of various treatments in vitro, the  
274 influences of residual substances on the surface of the skin cannot be avoided. Meanwhile, the  
275 detection site cannot be accurately guaranteed to be the same pathological type as the HE staining  
276 indicates because of the block thickness. The thickness of the fresh frozen tissue section we used is  
277 5  $\mu\text{m}$ , about 1/2 of cell diameter, and it is adjacent to the HE section, which maybe avoid the above  
278 problems. Moreover, all of our tests were completed within 1 hour with a layer of water film dripped  
279 on the surface of the section during detection to minimize the qualitative change of the tissue after

---

280 resection. Although our RS detections were performed in vitro, the 785 nm laser we use can detect  
281 deep tissues at 200  $\mu\text{m}$ , laying the foundation for the non-invasive detection of RS on human body.  
282 Concerning the vivo testing, different pathological types of the same lesion should be further  
283 detected in next study.

284 So far, most of the Raman identifications for skin lesions are mainly concentrated between the  
285 lesion and the normal tissue,[51-53] the tumor and the adjacent tissue,[54] or mainly applied in  
286 the rapid determination of the resection margin during the operation.[55-57] In our study, five types  
287 of lesions were analyzed simultaneously, including benign, malignancy, MM and NMSC. In order  
288 to get better classification effects, unsupervised learning (PCA, t-SNE) and supervised learning  
289 (KNN, SVM) were used in cluster and classification analysis of 5 types of skin lesions. PCA and t-  
290 SNE both show better clustering based on categories (Fig. 5). With its lower coupling feature, t-  
291 SNE showed its superiority in high-dimensional Raman spectral data dimensionality reduction and  
292 visualization (Fig. 5A). Two classifiers of KNN and SVM showed high test accuracies in the RS  
293 identification of 5 skin lesion types. In KNN, 11.1% of SCC was misjudged as BCC and 5.6% of  
294 SK was confused with MM (Fig. 6A). For SCC, BCC, SK and MM, further surgical resections and  
295 pathological biopsies are necessary, so the above misjudgments are acceptable in the clinic. 10.5%  
296 of MNs were misjudged as SCC (Fig. 6A), which may increase unnecessary biopsies. Compared to  
297 the lower-skilled physicians, KNN has more experience in diagnosis deserving the title of "senior  
298 physician". SVM, with only 5.3% of MNs misjudged as SK (Fig. 6B), did show an almost ideal  
299 classification effect. Moreover, in KNN AUCs of ROC curves in the category dimension and the  
300 sample dimension were both 0.99, and in SVM were both 1 (Fig. 6C, D), indicating that both  
301 classifiers are perfect classifiers of RS in the identification of 5 skin lesion types. All in all, the

---

302 application of machine learning in RS can better identify skin lesion types and provides a better  
303 bridging method for the application of RS in AI diagnostics.

304 How to achieve individualized treatment is still one of the ten challenges facing tumor  
305 immunotherapy.[58] The essential amino acid tryptophan catabolism is recognized as an important  
306 microenvironmental factor that suppresses antitumor immune responses in cancer and regulates T  
307 cell proliferation, activation and anti-tumor effects[36-38]. Phenylalanine is involved in regulating  
308 cell cycle progression,[42] modulating invasion-related signaling/function proteins,[43] and  
309 promoting tumor cell adhesion and spread.[44] The lack of phenylalanine can induce focal adhesion  
310 kinase-dependent apoptosis and mitochondria-initiated apoptosis.[43, 45]. Interestingly, the  
311 different contents of tryptophan and phenylalanine were detected by RS in the 5 skin lesion types  
312 we studied and were highly consistent with LC-MS results (Fig. 4B, D, Table 5, 6). In addition, as  
313 the main component of collagen, the decrease of proline indicates tumor metastasis and poor  
314 prognosis[39-41]. Similarly, lower strengthen of proline signal was found in malignant lesions as  
315 compared to benign lesions in our study (Fig. 4C). These results indicate that RS provides reliable  
316 molecular information related with tumor therapy and progression. Furthermore, AI-aided RS may  
317 be a reliable screening method for immunotherapy responsiveness and individualized therapy. These  
318 conjectures will be examined in our following research.

### 319 **Conclusion**

320 In summary, RS is a competitive candidate for the fast and accurate diagnosis of skin lesions with  
321 ultrathin frozen fresh sections providing high-quality Raman spectra. And, the application of  
322 machine learning methods in Raman spectrum classification showed excellent diagnostic  
323 capabilities for 5 skin lesion types. KNN and SVM predictive models diagnosed 5 skin lesion types

---

324 with almost perfect accuracy. The significant differences of tryptophan, phenylalanine and proline  
325 indicated by RS may imply different progression and treatment responsiveness of 5 skin lesion types.  
326 These results identify that ML-aided RS is a potential tool in clinic diagnosis and screening of tumor  
327 immunotherapy, progression and prognosis.

### 328 **Abbreviations**

329 BCC: Basal cell carcinoma; SCC: Squamous cell carcinoma; MM: Melanoma; NMSC:  
330 Nonmelanoma skin cancer; WHO: World Health Organization; RS: Raman spectroscopy; AI:  
331 Artificial intelligence; ML: Machine learning; PCA: Principal components analysis; t-SNE: t-  
332 distributed stochastic neighbor embedding; KNN: K-nearest Neighbor; SVM: Support vector  
333 machine; ROC: Receiver operating characteristic; HE: Hematoxylin-eosin; LC-MS: Liquid  
334 chromatography-mass spectrometry; MS: Mass spectrometry; ESI: Electrospray ionization; PRM:  
335 Parallel reaction monitoring; FPR: False positive rate; TPR: True positive rate; AUC: Area under  
336 curve.

### 337 **Acknowledgements**

338 Not applicable.

### 339 **Authors' contributions**

340 H. Zhang, Y. Xue, B. Li and B. Liu conceived the study. H. Zhang and D. Wang collected samples.  
341 D. Wang and L. Qu performed tissue section preparation and HE staining. H. Zhang and Y. Xue  
342 performed Raman detections. H. Zhang, Y. Xue and X. Li analyzed the data and established the  
343 predictive models. H. Zhang, Y. Xue and X. Li wrote the paper, and B. Li and B. Liu revised the  
344 manuscript. All authors read and approved the final manuscript.

### 345 **Funding**

---

346 Bethune Medical Engineering and Instrument Center; Grant number: 04025290001

347 **Availability of data and materials**

348 The data that support the findings of this study are available from the corresponding author upon  
349 reasonable request.

350 **Ethics approval and consent to participate**

351 The human samples were approved by the institutional review board at First Hospital of Jilin  
352 University. All the experiments were performed in accordance with relevant guidelines and  
353 regulations. Skin lesion tissues were obtained from patients recommended for surgical resections  
354 at First Hospital of Jilin University after informed consents were obtained from the patients. All  
355 subjects were provided written informed consents in accordance with the Declaration of Helsinki.

356 **Consent for publication**

357 Not applicable.

358 **Competing interests**

359 The authors declare no potential conflict of interest.

360

361 **References**

- 362 1. Cakir BO, Adamson P, Cingi C. Epidemiology and economic burden of nonmelanoma skin  
363 cancer. *Facial Plast Surg Clin North Am.* 2012; 20: 419-422.
- 364 2. Brandt MG, Moore CC. Nonmelanoma Skin Cancer. *Facial Plast Surg Clin North Am.* 2019;  
365 27: 1-13.
- 366 3. Radiation: Ultraviolet (UV) radiation and skin cancer  
367 [<http://www.who.int/uv/faq/skincancer/en/index1.html>]
- 368 4. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence Estimate of Nonmelanoma  
369 Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012. *JAMA Dermatol.* 2015;  
370 151: 1081-1086.
- 371 5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70: 7-30.
- 372 6. Raasch BA, Buettner PG. Multiple nonmelanoma skin cancer in an exposed Australian  
373 population. *International journal of Dermatology.* 2002; 41: 652-658.

- 
- 374 7. Morton C, Mackie R. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. The  
375 British journal of dermatology. 1998; 138: 283-287.
- 376 8. Elmore J, Barnhill R, Elder D, Longton G, Pepe M, Reisch L, Carney P, Titus L, Nelson H,  
377 Onega T *et al.* Pathologists' diagnosis of invasive melanoma and melanocytic proliferations:  
378 observer accuracy and reproducibility study. *BMJ (Clinical research ed)*. 2017; 357: j2813.
- 379 9. Redondo P, Gímenez de Azcarate A, Núñez-Córdoba JM, Andreu EJ, García-Guzman M,  
380 Aguado L, Prosper F. Efficacy of Autologous Melanocyte Transplantation on Amniotic  
381 Membrane in Patients With Stable Leukoderma. *JAMA Dermatology*. 2015; 151.
- 382 10. English DR, Del Mar C, Burton RC. Factors influencing the number needed to excise: excision  
383 rates of pigmented lesions by general practitioners. *Medical journal of Australia*. 2004; 180: 16-  
384 19.
- 385 11. Gniadecka M, Wulf HC, Nielsen OF, Christensen DH, Hercogova J. Distinctive molecular  
386 abnormalities in benign and malignant skin lesions: studies by Raman spectroscopy.  
387 *Photochemistry and photobiology*. 1997; 66: 418-423.
- 388 12. Santos IP, Caspers PJ, Bakker Schut TC, van Doorn R, Noordhoek Hegt V, Koljenovic S,  
389 Puppels GJ. Raman Spectroscopic Characterization of Melanoma and Benign Melanocytic  
390 Lesions Suspected of Melanoma Using High-Wavenumber Raman Spectroscopy. *Anal Chem*.  
391 2016; 88: 7683-7688.
- 392 13. Huser T, Chan J. Raman spectroscopy for physiological investigations of tissues and cells.  
393 *Advanced drug delivery reviews*. 2015; 89: 57-70.
- 394 14. Zhao J, Zeng H, Kalia S, Lui H. Wavenumber selection based analysis in Raman spectroscopy  
395 improves skin cancer diagnostic specificity. *Analyst*. 2016; 141: 1034-1043.
- 396 15. Zhang Y, Moy AJ, Feng X, Nguyen HTM, Sebastian KR, Reichenberg JS, Wilke CO, Markey  
397 MK, Tunnell JW. Assessment of Raman Spectroscopy for Reducing Unnecessary Biopsies for  
398 Melanoma Screening. *Molecules*. 2020; 25.
- 399 16. P. Santos I, van Doorn R, Caspers PJ, Bakker Schut TC, Barroso EM, Nijsten TEC, Noordhoek  
400 Hegt V, Koljenović S, Puppels GJ. Improving clinical diagnosis of early-stage cutaneous  
401 melanoma based on Raman spectroscopy. *British Journal of Cancer*. 2018; 119: 1339-1346.
- 402 17. Grosset A, Dallaire F, Nguyen T, Birlea M, Wong J, Daoust F, Roy N, Kougioumoutzakis A,  
403 Azzi F, Aubertin K *et al.* Identification of intraductal carcinoma of the prostate on tissue  
404 specimens using Raman micro-spectroscopy: A diagnostic accuracy case-control study with  
405 multicohort validation. *PLoS medicine*. 2020; 17: e1003281.
- 406 18. Chen Y, Zheng X, Chen G, He C, Zhu W, Feng S, Xi G, Chen R, Lan F, Zeng H. Immunoassay  
407 for LMP1 in nasopharyngeal tissue based on surface-enhanced Raman scattering. *International*  
408 *journal of nanomedicine*. 2012; 7: 73-82.
- 409 19. Gajjar K, Heppenstall LD, Pang W, Ashton KM, Trevisan J, Patel, II, Llabjani V, Stringfellow  
410 HF, Martin-Hirsch PL, Dawson T *et al.* Diagnostic segregation of human brain tumours using  
411 Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis. *Anal*  
412 *Methods*. 2012; 5: 89-102.
- 413 20. Huang Z, McWilliams A, Lam S, English J, McLean D, Lui H, Zeng H. Effect of formalin  
414 fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial  
415 tissues. *International Journal of Oncology*. 2003.
- 416 21. Fullwood LM, Griffiths D, Ashton K, Dawson T, Lea RW, Davis C, Bonnier F, Byrne HJ, Baker  
417 MJ. Effect of substrate choice and tissue type on tissue preparation for spectral histopathology

- 
- 418 by Raman microspectroscopy. *Analyst*. 2014; 139: 446-454.
- 419 22. Auner GW, Koya SK, Huang C, Broadbent B, Trexler M, Auner Z, Elias A, Mehne KC,  
420 Brusatori MA. Applications of Raman spectroscopy in cancer diagnosis. *Cancer Metastasis Rev*.  
421 2018; 37: 691-717.
- 422 23. Rau JV, Marini F, Fosca M, Cippitelli C, Rocchia M, Di Napoli A. Raman spectroscopy  
423 discriminates malignant follicular lymphoma from benign follicular hyperplasia and from  
424 tumour metastasis. *Talanta*. 2019; 194: 763-770.
- 425 24. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR,  
426 Bolli N, Borg A, Borresen-Dale AL *et al*. Signatures of mutational processes in human cancer.  
427 *Nature*. 2013; 500: 415-421.
- 428 25. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, Crawford JC, Clemens  
429 EB, Nguyen THO, Kedzierska K *et al*. Quantifiable predictive features define epitope-specific  
430 T cell receptor repertoires. *Nature*. 2017; 547: 89-93.
- 431 26. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis.  
432 *Cancer Inform*. 2007; 2: 59-77.
- 433 27. Nerviani A, Boutet M, Tan W, Goldmann K, Purkayastha N, Lajtos T, Hands R, Lewis M, Kelly  
434 S, Pitzalis C. IL-23 skin and joint profiling in psoriatic arthritis: novel perspectives in  
435 understanding clinical responses to IL-23 inhibitors. *Annals of the rheumatic diseases*. 2020.
- 436 28. Hahn V, Knutsdottir H, Luo X, Bedi K, Margulies K, Haldar S, Stolina M, Yin J, Khakoo A,  
437 Vaishnav J *et al*. Myocardial Gene Expression Signatures in Human Heart Failure With  
438 Preserved Ejection Fraction. *Circulation*. 2021; 143: 120-134.
- 439 29. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, Zhu S, Ye Z. Comparison and development of  
440 machine learning tools in the prediction of chronic kidney disease progression. *Journal of*  
441 *translational medicine*. 2019; 17: 119.
- 442 30. Ma X, Wu Y, Zhang L, Yuan W, Yan L, Fan S, Lian Y, Zhu X, Gao J, Zhao J *et al*. Comparison  
443 and development of machine learning tools for the prediction of chronic obstructive pulmonary  
444 disease in the Chinese population. *Journal of translational medicine*. 2020; 18: 146.
- 445 31. Puppels G, de Mul F, Otto C, Greve J, Robert-Nicoud M, Arndt-Jovin D, Jovin T. Studying  
446 single living cells and chromosomes by confocal Raman microspectroscopy. *Nature*. 1990; 347:  
447 301-303.
- 448 32. Naber SP. Molecular pathology--detection of neoplasia. *New England Journal of Medicine*.  
449 1994; 331: 1508-1510.
- 450 33. Jess PR, Smith DD, Mazilu M, Dholakia K, Riches AC, Herrington CS. Early detection of  
451 cervical neoplasia by Raman spectroscopy. *Int J Cancer*. 2007; 121: 2723-2728.
- 452 34. De Gelder J, Scheldeman P, Leus K, Heyndrickx M, Vandenabeele P, Moens L, De Vos P. Raman  
453 spectroscopic study of bacterial endospores. *Analytical and bioanalytical chemistry*. 2007; 389:  
454 2143-2151.
- 455 35. Chen M, Lord R. Laser-excited Raman spectroscopy of biomolecules. VIII. Conformational  
456 study of bovine serum albumin. *Journal of the American Chemical Society*. 1976; 98: 990-992.
- 457 36. Munn DH, Shafizadeh E, Attwood JT, Bondarev I, Pashine A, Mellor AL. Inhibition of T cell  
458 proliferation by macrophage tryptophan catabolism. *Journal of Experimental Medicine*. 1999;  
459 189: 1363-1372.
- 460 37. Curti A, Pandolfi S, Valzasina B, Aluigi M, Isidori A, Ferri E, Salvestrini V, Bonanno G, Rutella  
461 S, Durelli I *et al*. Modulation of tryptophan catabolism by human leukemic cells results in the

- 
- 462 conversion of CD25- into CD25+ T regulatory cells. *Blood*. 2007; 109: 2871-2877.
- 463 38. Botticelli A, Mezi S, Pomati G, Cerbelli B, Cerbelli E, Roberto M, Giusti R, Cortellini A,  
464 Lionetto L, Scagnoli S *et al.* Tryptophan Catabolism as Immune Mechanism of Primary  
465 Resistance to Anti-PD-1. *Front Immunol*. 2020; 11: 1243.
- 466 39. Loayza-Puch F, Rooijers K, Buil LC, Zijlstra J, Oude Vrielink JF, Lopes R, Ugalde AP, van  
467 Breugel P, Hofland I, Wesseling J *et al.* Tumour-specific proline vulnerability uncovered by  
468 differential ribosome codon reading. *Nature*. 2016; 530: 490-494.
- 469 40. Elia I, Broekaert D, Christen S, Boon R, Radaelli E, Orth MF, Verfaillie C, Grunewald TGP,  
470 Fendt SM. Proline metabolism supports metastasis formation and could be inhibited to  
471 selectively target metastasizing cancer cells. *Nat Commun*. 2017; 8: 15267.
- 472 41. Tang L, Zeng J, Geng P, Fang C, Wang Y, Sun M, Wang C, Wang J, Yin P, Hu C *et al.* Global  
473 Metabolic Profiling Identifies a Pivotal Role of Proline and Hydroxyproline Metabolism in  
474 Supporting Hypoxic Response in Hepatocellular Carcinoma. *Clin Cancer Res*. 2018; 24: 474-  
475 485.
- 476 42. Fu YM, Yu ZX, Ferrans VJ, Meadows GG. Tyrosine and phenylalanine restriction induces  
477 G0/G1 cell cycle arrest in murine melanoma in vitro and in vivo. *Nutr Cancer*. 1997; 29: 104-  
478 113.
- 479 43. Fu Y-M, Meadows GG. Specific amino acid dependency regulates the cellular behavior of  
480 melanoma. *The Journal of nutrition*. 2007; 137: 1591S-1596S.
- 481 44. Fu Y-M, Zhang H, Ding M, Li Y-Q, Fu X, Yu Z-X, Meadows GG. Specific amino acid restriction  
482 inhibits attachment and spreading of human melanoma via modulation of the integrin/focal  
483 adhesion kinase pathway and actin cytoskeleton remodeling. *Clinical & experimental metastasis*.  
484 2005; 21: 587-598.
- 485 45. Fu Y-M, Yu Z-X, Pelayo BA, Ferrans VJ, Meadows GG. Focal adhesion kinase-dependent  
486 apoptosis of melanoma induced by tyrosine and phenylalanine deficiency. *Cancer research*.  
487 1999; 59: 758-765.
- 488 46. Tfayli A, Gobinet C, Vrabie V, Huez R, Manfait M, Piot O. Digital dewaxing of Raman signals:  
489 discrimination between nevi and melanoma spectra obtained from paraffin-embedded skin  
490 biopsies. *Applied spectroscopy*. 2009; 63: 564-570.
- 491 47. O Faolain E. Raman Spectroscopic Evaluation of Efficacy of Current Paraffin Wax Section  
492 Dewaxing Agents. *Journal of Histochemistry and Cytochemistry*. 2005; 53: 121-129.
- 493 48. Short MA, Lui H, McLean D, Zeng H, Alajlan A, Chen XK. Changes in nuclei and peritumoral  
494 collagen within nodular basal cell carcinomas via confocal micro-Raman spectroscopy. *J*  
495 *Biomed Opt*. 2006; 11: 34004.
- 496 49. Cicchi R, Cosci A, Rossari S, Kapsokalyvas D, Baria E, Maio V, Massi D, De Giorgi V,  
497 Pimpinelli N, Saverio Pavone F. Combined fluorescence-Raman spectroscopic setup for the  
498 diagnosis of melanocytic lesions. *J Biophotonics*. 2014; 7: 86-95.
- 499 50. Gniadecka M, Philipsen PA, Sigurdsson S, Wessel S, Nielsen OF, Christensen DH, Hercogova  
500 J, Rossen K, Thomsen HK, Gniadecki R *et al.* Melanoma diagnosis by Raman spectroscopy and  
501 neural networks: structure alterations in proteins and lipids in intact cancer tissue. *J Invest*  
502 *Dermatol*. 2004; 122: 443-449.
- 503 51. Lui H, Zhao J, McLean D, Zeng H. Real-time Raman spectroscopy for in vivo skin cancer  
504 diagnosis. *Cancer Res*. 2012; 72: 2491-2500.
- 505 52. Sigurdsson S, Philipsen PA, Hansen LK, Larsen J, Gniadecka M, Wulf HC. Detection of Skin

- 
- 506 Cancer by Classification of Raman Spectra. IEEE Transactions on Biomedical Engineering.  
507 2004; 51: 1784-1793.
- 508 53. Schleusener J, Gluszczyńska P, Reble C, Gersonde I, Helfmann J, Fluhr JW, Lademann J,  
509 Rowert-Huber J, Patzelt A, Meinke MC. In vivo study for the discrimination of cancerous and  
510 normal skin using fibre probe-based Raman spectroscopy. *Exp Dermatol.* 2015; 24: 767-772.
- 511 54. Schut TCB, Caspers PJ, Puppels GJ, Nijssen A, Heule F, Neumann MH, Hayes DP.  
512 Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy. *Journal*  
513 *of investigative dermatology.* 2002; 119: 64-69.
- 514 55. Barroso E, Ten Hove I, Bakker Schut T, Mast H, van Lanschot C, Smits R, Caspers P, Verdijk  
515 R, Noordhoek Hegt V, Baatenburg de Jong R *et al.* Raman spectroscopy for assessment of bone  
516 resection margins in mandibulectomy for oral cavity squamous cell carcinoma. *European*  
517 *journal of cancer (Oxford, England : 1990).* 2018; 92: 77-87.
- 518 56. Matousek P, Stone N. Development of deep subsurface Raman spectroscopy for medical  
519 diagnosis and disease monitoring. *Chemical Society reviews.* 2016; 45: 1794-1802.
- 520 57. Haka A, Volynskaya Z, Gardecki J, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari R,  
521 Crowe J, Feld M. In vivo margin assessment during partial mastectomy breast surgery using  
522 raman spectroscopy. *Cancer research.* 2006; 66: 3317-3322.
- 523 58. Hegde PS, Chen DS. Top 10 Challenges in Cancer Immunotherapy. *Immunity.* 2020; 52: 17-35.

524

525 **Figure 1.** Schematic of machine-learning of skin lesion Raman spectra for fast diagnosis. Ultrathin  
526 frozen fresh sections of skin lesions were acquired and prepared for Raman detection with a 785nm  
527 laser. Raman spectra were processed gradually and analyzed by machine learning methods.  
528 Diagnose information were put out at last.

529 **Figure 2.** BCC Raman single spectrum standardization results. RI, Raman Intensity.

530 **Figure 3.** The mean variance graph of Raman spectra of 5 skin lesions after preprocessing.

531 **Figure 4.** Normalized Raman intensity for bands at 720 cm<sup>-1</sup>, 752 cm<sup>-1</sup>, 853 cm<sup>-1</sup>, and 1002 cm<sup>-1</sup> of  
532 5 skin lesions. RI, Raman Intensity. Data were presented as mean ± SD. \* P < 0.05; \*\* P < 0.01;  
533 \*\*\* P < 0.001. (Student's *t*-test)

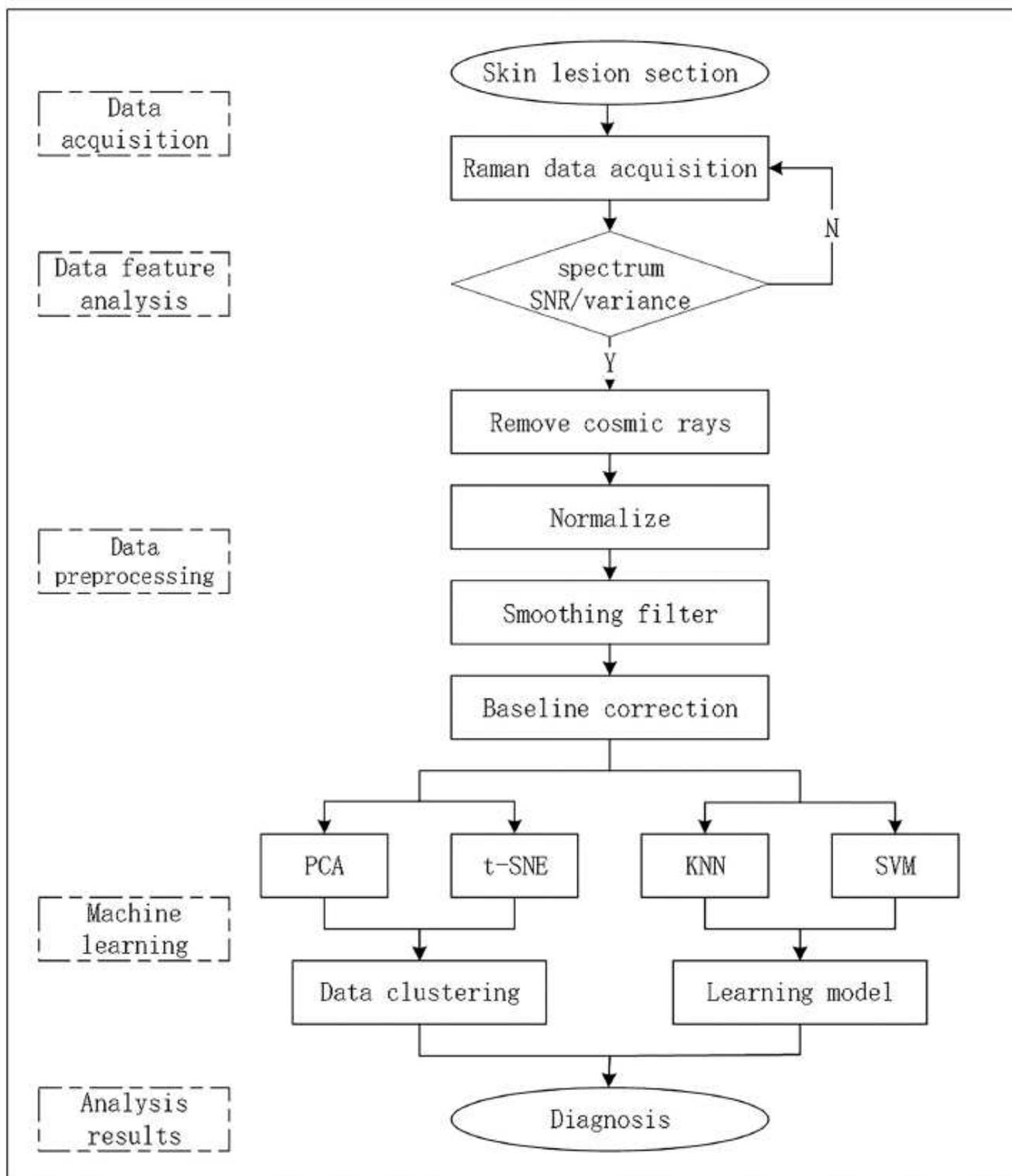
534 **Figure 5.** Visualized clustering results of 5 skin lesion types after dimensionality reduction in t-SNE  
535 (A) and PCA (B).

536 **Figure 6.** Confusion matrix and ROC curves of test results. (A, B) Confusion matrixes of 20% RS

---

537 data test results in KNN (A) and SVM (B). (C, D) The ROC curves for the recognition probabilities  
538 of 5 skin lesion categories in KNN (C) and SVM (D).

# Figures



**Figure 1**

Schematic of machine-learning of skin lesion Raman spectra for fast diagnosis. Ultrathin frozen fresh sections of skin lesions were acquired and prepared for Raman detection with a 785nm laser. Raman

spectra were processed gradually and analyzed by machine learning methods. Diagnose information were put out at last.

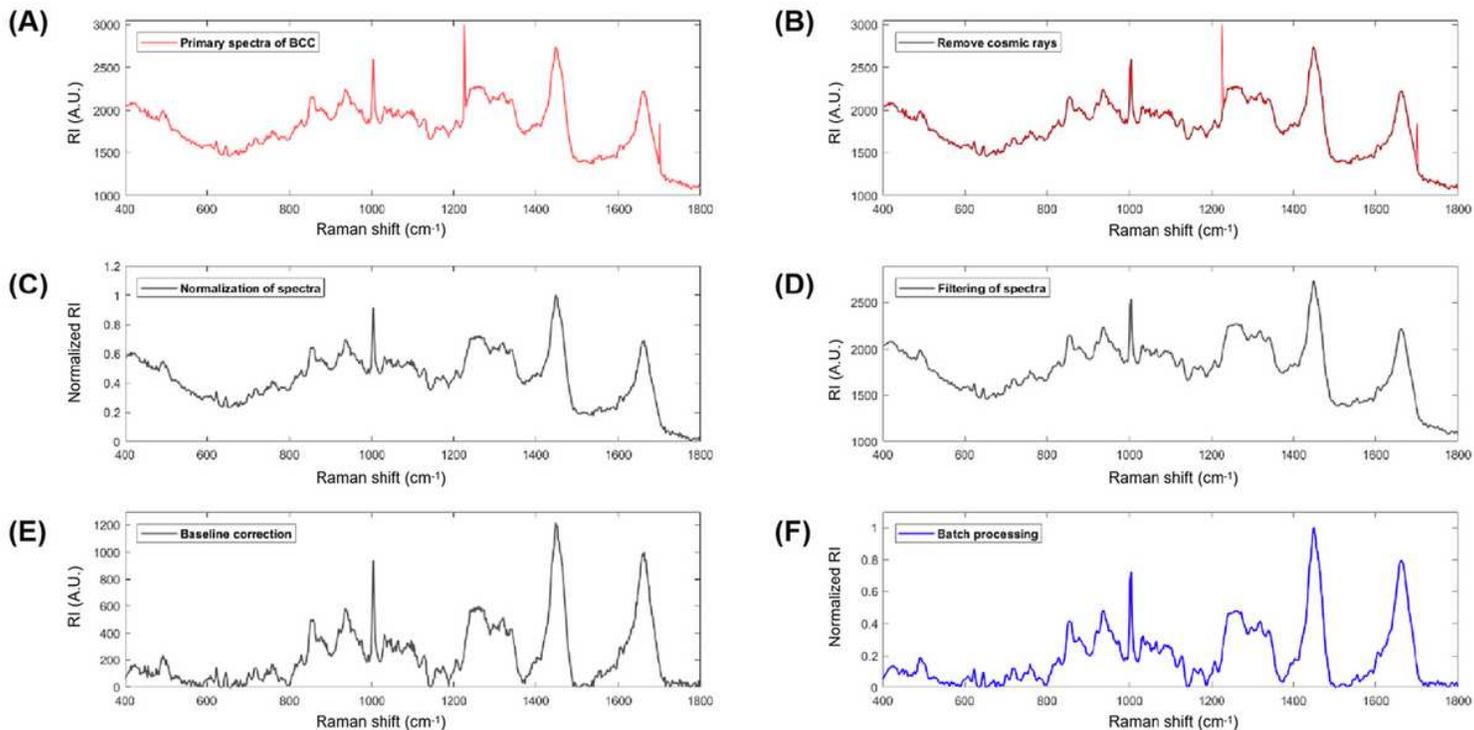
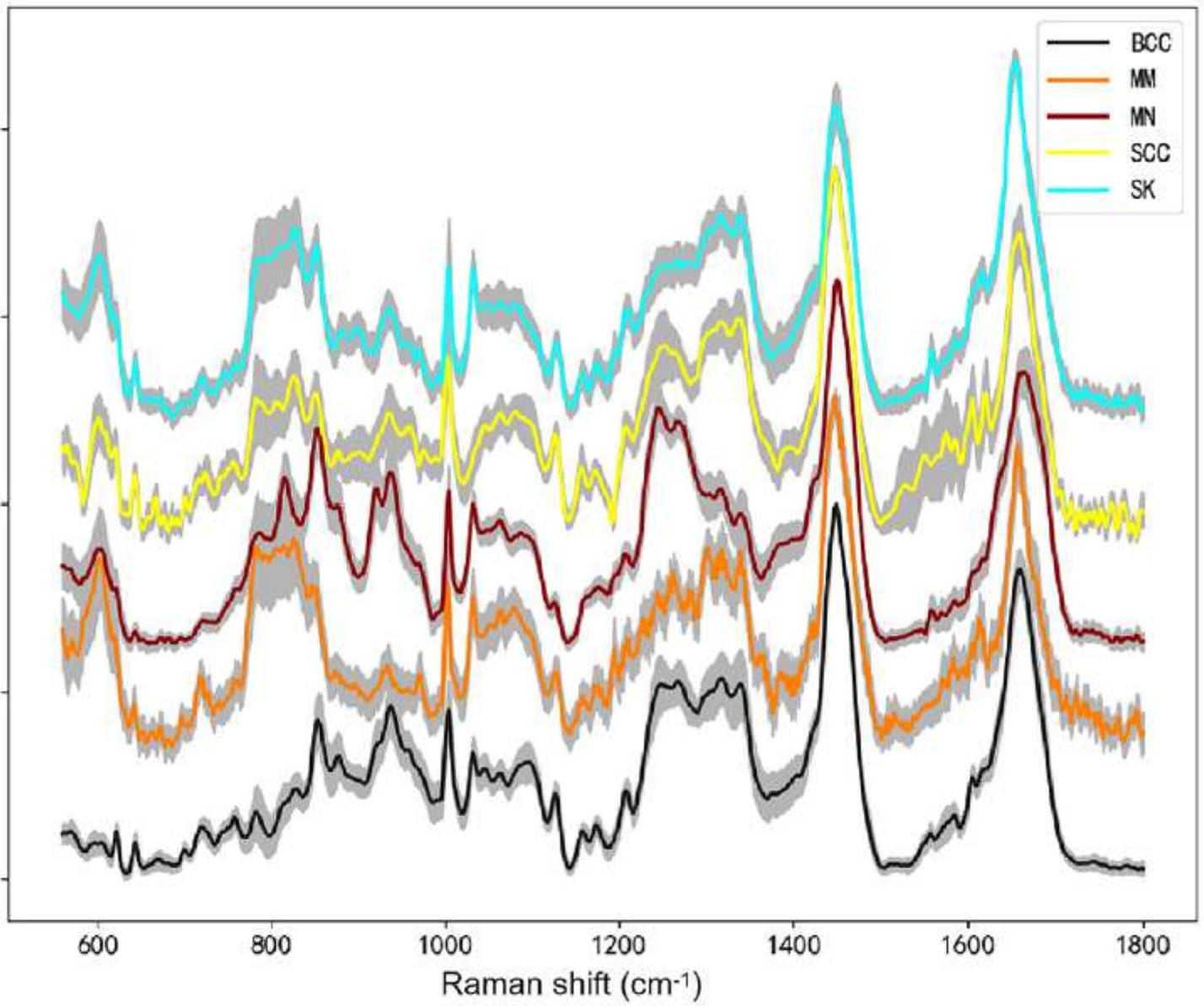


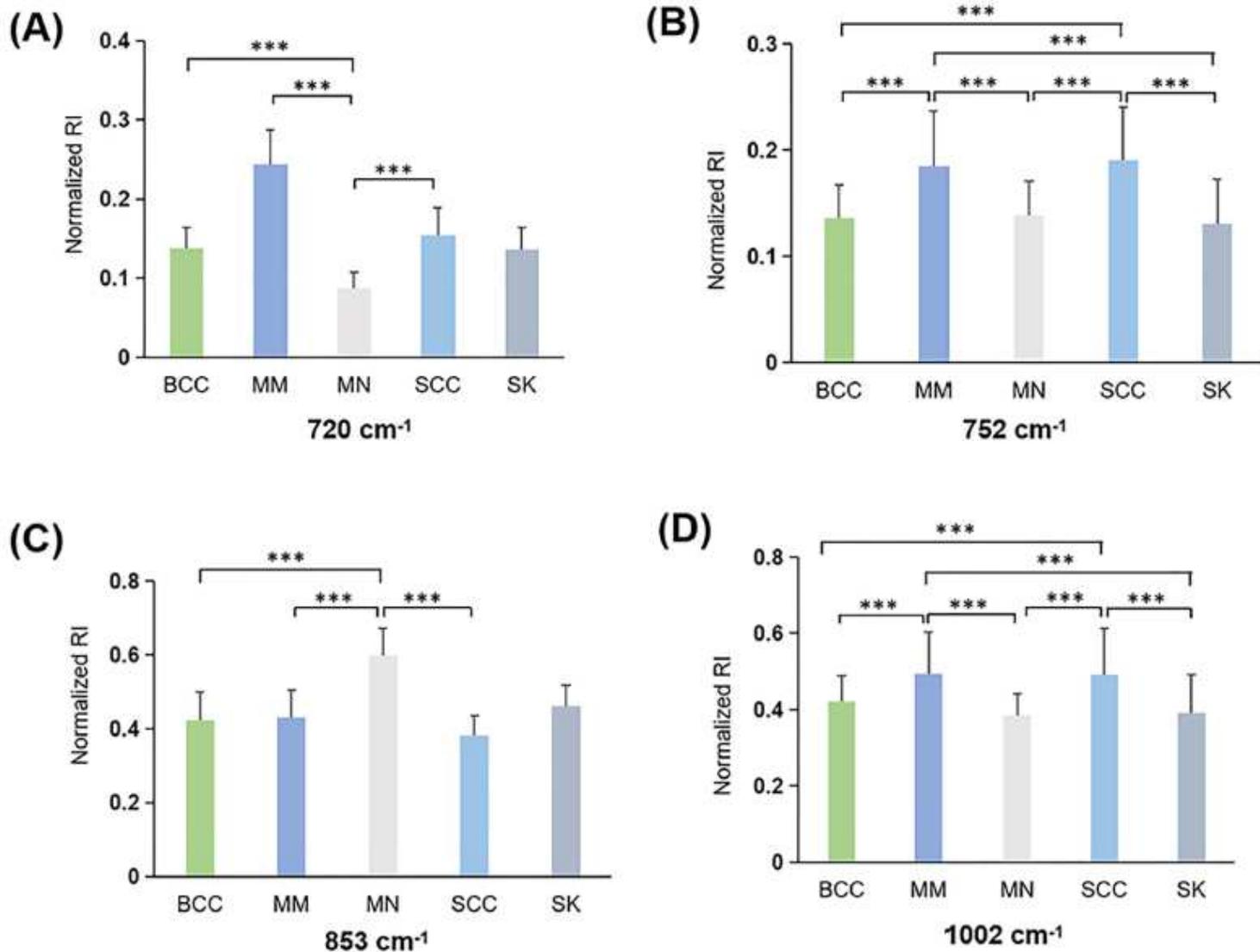
Figure 2

BCC Raman single spectrum standardization results. RI, Raman Intensity.



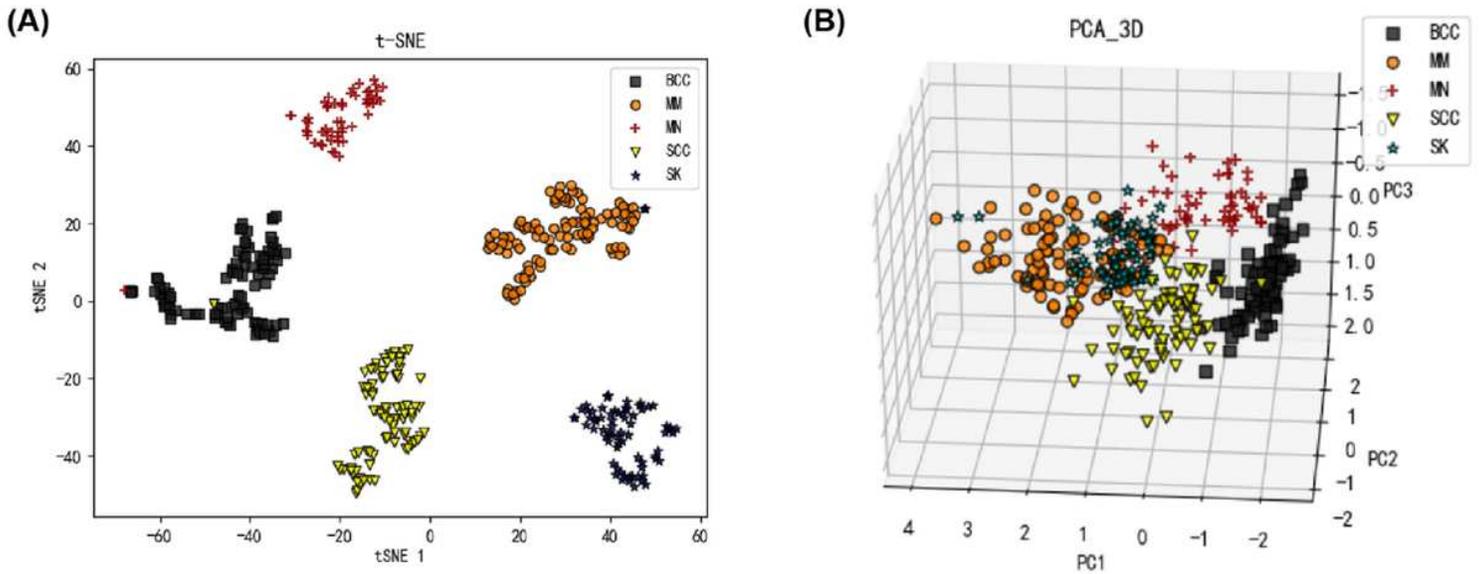
**Figure 3**

The mean variance graph of Raman spectra of 5 skin lesions after preprocessing.



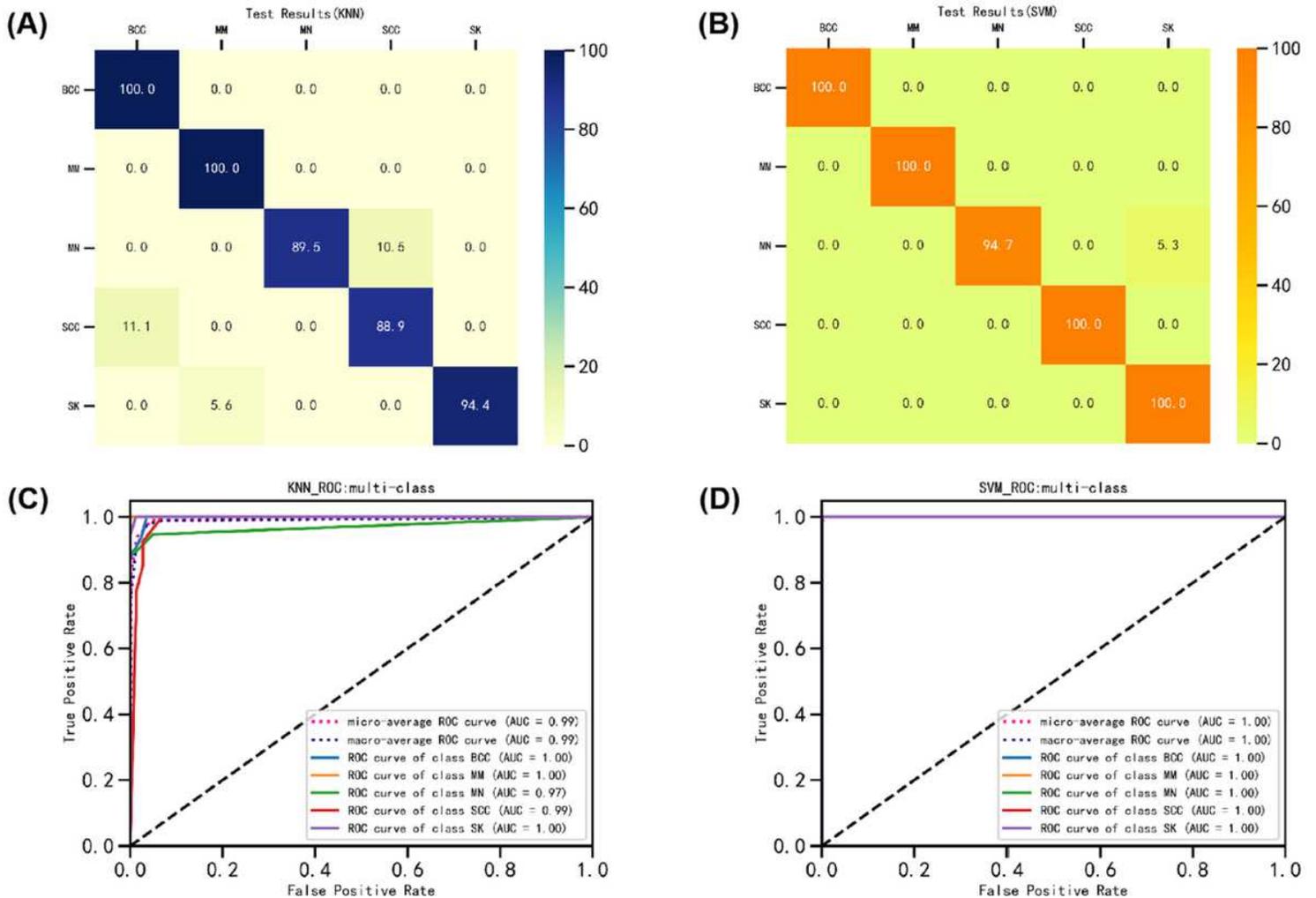
**Figure 4**

Normalized Raman intensity for bands at 720 cm<sup>-1</sup>, 752 cm<sup>-1</sup>, 853 cm<sup>-1</sup>, and 1002 cm<sup>-1</sup> of 5 skin lesions. RI, Raman Intensity. Data were presented as mean  $\pm$  SD. \* P < 0.05; \*\* P < 0.01; \*\*\* P < 0.001. (Student's t-test)



**Figure 5**

Visualized clustering results of 5 skin lesion types after dimensionality reduction in t-SNE (A) and PCA (B).



## Figure 6

Confusion matrix and ROC curves of test results. (A, B) Confusion matrixes of 20% RS data test results in KNN (A) and SVM (B). (C, D) The ROC curves for the recognition probabilities of 5 skin lesion categories in KNN (C) and SVM (D).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table.doc](#)