Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

# Raman optical identification of renal cell carcinoma via machine learning

Chang He [a,1], Xiaorong Wu [b,1], Jiale Zhou [b], Yonghui Chen [b,*], Jian Ye [a,c,*]

[a] State Key Laboratory of Oncogenes and Related Genes, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, PR China
[b] Department of Urology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, PR China
[c] Shanghai Key Laboratory of Gynecologic Oncology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, PR China

HIGHLIGHTS

- Raman spectroscopy can *ex vivo* rapidly distinguish renal tumor from normal tissues and fat.
- Raman-based SVM model can determine tumor boundary, consistent to pathological analysis.
- Raman spectrum-based SVM model can distinguish renal tumor subtypes and grades.

GRAPHICAL ABSTRACT

Early diagnosis of renal cell carcinoma and the according adjustment of surgical plan are greatly beneficial to patients. We established optical Raman spectrum-based supporting vector machine (SVM) models to accomplish diagnosis of renal cell carcinoma including identification of tumors and tumor boundary as well as classification and differentiation of tumors. This method shows high potential for rapid and accurate clinical diagnosis of renal cancers preoperatively and intraoperatively.
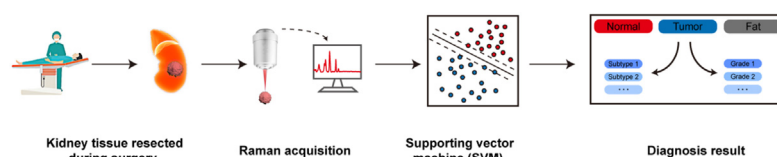
ARTICLE INFO

ABSTRACT

High pathologic tumor-node-metastasis (pTNM) stage grade or Fuhrman grade indicates poor oncological outcome in renal cell carcinoma (RCC). Early diagnosis and screening of these RCCs and adjust surgical planning accordingly are greatly beneficial to patients. Raman spectroscopy is a highly specific fingerprint spectrum on molecular level, pretty appropriate for label-free and noninvasive cancer diagnosis. In this work we established a Raman spectrum-based supporting vector machine (SVM) model to accurately *ex vivo* distinguish human renal tumor from normal tissues and fat with an accuracy of 92.89%. The model can also be used to determine tumor boundary, showing consistent results to pathological staining analysis. This method can be additionally used to accomplish the classification purposes of renal tumor subtypes and grades with an accuracy of 86.79% and 89.53%, respectively. Therefore, we prove that Raman spectroscopy has great potential in the rapid and accurate clinical diagnosis of renal cancers.

1. Introduction

With modern bioimaging techniques including magnetic resonance imaging (MRI) and computed tomography (CT), an increasing number of renal cell carcinomas (RCCs) can be diagnosed at early stage recently [1]. The bioimaging techniques also assist in the determination of the surgical plan through tumor-node-metastasis (TNM) stages of different RCCs. Clinically, nephron sparing surgery (NSS) is well recognized as the first choice for the treatment of RCCs in early stage (e.g., T1a or selective T1b RCC) to preserve renal function, and radical nephrectomy is strongly recommended for advanced RCCs (e.g., T3 ~ T4) according to the European Association of Urology (EAU) guidelines [2]. One issue, which needs to be resolved before choosing a surgical plan for patients with RCCs, is the difficulty of identification of perinephric fat invasion [3]. According to the TNM staging system, the RCCs with perinephric fat invasion was classified into a T3a stage, which is

* Corresponding authors at: Department of Urology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, No. 1630 Dongfang Road, Shanghai 200127, PR China (Y. Chen). School of Biomedical Engineering, Shanghai Jiao Tong University, No. 1954 Huashan Road, Shanghai, 200030, PR China (J. Ye).
   E-mail addresses: chenyonghui011488@renji.com (Y. Chen), yejian78@sjtu.edu.cn (J. Ye).
   [1] These authors contributed equally.

accompanied with a poor prognosis [4,5]. This issue can hardly be resolved by preoperative bioimaging methods or intraoperative examinations by naked eyes, which may lead to a serious underestimation of TNM stage and an improper surgical plan [6,7]. For example, partial nephrectomy may be adopted for localized RCCs if cancer satellitic focus or perinephric fat invasion is not correctly identified, which may cause positive surgical margin and a higher possibility of local recurrence [8–11]. Therefore, identification of perinephric fat invasion intraoperatively is of great importance for choosing a proper nephrectomy strategy for renal tumors.

Preoperative imaging techniques including either enhanced CT or MRI are hard to differentiate the subtypes of renal tumors such as papillary renal cell carcinoma (PRCC), chromophobe renal cell carcinoma (CRCC) and eosinophiloma, which have quite different prognoses. Currently, the pathology based on immunohistochemistry of paraffin specimen is considered as the gold standard for the subtype diagnosis of RCCs [12,13], but it is postoperative and time consuming. The urologists cannot obtain pathological results before finishing the surgery. The intraoperative frozen pathology is an alternative crucial approach to provide the pathological diagnosis during the surgery. However, the whole process of frozen pathology typically takes 30–60 min and the accuracy cannot be guaranteed due to the inability to perform immunohistochemistry. Meanwhile, the tumor subtype is hard to be differentiated by the intraoperative frozen pathology [14]. Thus, it is highly crucial to develop a rapid and accurate intraoperative diagnostic method to distinguish the subtype of the renal tumor so that more personalized treatment strategies can be implemented.

Raman spectrum is a highly specific fingerprint scattering signal on the molecular level and it provides the molecular structure information of the matter [15–18]. Raman spectroscopic characteristic band position, intensity and line width provide information on molecular vibration and rotation, which can reflect different chemical bonds or functional groups in molecules. Raman spectroscopy, as a fast and noninvasive testing, provides spectral information without sample pre-treatment and modification. The occurrence of cancer is mainly due to changes in the structure of the essential components of the cells, such as proteins, nucleic acids, lipids and carbohydrate, which can be possibly recognized by Raman signals [19–22]. Samples of human tissue, liquids and secretions have a large content of water, whose Raman signal is nearly negligible. This makes Raman spectroscopy a pretty appropriate detection modality. Therefore, Raman spectroscopy is quite suitable for detection and diagnosis of cancer cells and tissues *in vivo* and *ex vivo* [23–30]. Bensalah et al. were among the first to apply Raman spectroscopy to distinguish the tumor tissue and normal tissue with an accuracy of 84%. They also classified the deterioration and subtypes of the tumor tissue with an accuracy of 82% and 93%, respectively [31]. Recently, Yosef et al. combined Raman spectroscopy and the random forest algorithm to perform the label-free identification of urothelial cells, showing that Raman spectroscopy can offer sufficient biological information at the cellular level [24]. Toward clinical practice, Jermyn et al. developed a handheld Raman probe to successfully differentiate normal brain from the dense cancer and normal brain invaded by cancer cells, with a sensitivity of 93% and a specificity of 91% [32]. However, we also found that Raman spectroscopy can be further improved when applied to tissue detection. For example, clear cell renal cell carcinomas (CCRCCs) are rich in lipids, which have similar composition compared to fat at the molecular level [33]. Thus, perinephric fat invasion is very easily ignored if the discrimination is only based on the Raman spectra of normal and tumor tissue, which can affect the stage classification of kidney tumor. In addition, there are few reports about the use of Raman spectroscopy for the identification of tumor boundaries, which is particularly important for surgeons to further improve the therapeutic effect

of tumors. The determination of tumor boundaries in current studies is primarily realized by labelled optical probes such as fluorescent dyes or surface-enhanced Raman spectroscopy (SERS) nanoprobes, but this generally requires the specific marking of tumors with specific probes [34–40].

Recently, supporting vector machine (SVM), as one of the powerful machine learning approaches, has attracted great attention due to its satisfactory computation efficiency, classification capability and robustness [41]. SVM is a learning method that is trained by finding a hyperplane to separate labelled training data with the maximal margin between the training data and the hyperplane. For Raman spectroscopy, collecting a large number of samples is always time-consuming and sometimes even unpractical. SVM is rather suitable for Raman spectrum-related calculation since it performs well on a small training set.

Herein we reported a label-free and noninvasive near-infrared (NIR) Raman spectroscopy-based technique combined with the SVM classification algorithm with the aim of intraoperative diagnosis of RCCs including tumor discrimination from normal tissue and fat, tumor boundary identification and the classification of tumor subtypes and grades. In a typical workflow (Fig. 1) of SVM model establishment, surgeons removed normal renal tissue, fat and tumor tissue during an RCC operation (partial and radical nephrectomy). Each piece of resected tissue was divided into two parts, one for Raman spectroscopy acquisition and the other for pathology analysis. Raman spectra and pathology results were jointly used to train the SVM model. Once the SVM model was built, the suspected excised tissues were tested using Raman spectroscopy and predicted by the established model.

2. Materials and methods

2.1. Sample acquisition and pathological diagnosis

All tissue samples used for Raman spectrum acquisition were obtained from patients (n = 77, 38 for clinical specimens and 39 for frozen ones), who were diagnosed as RCC based on preoperative imaging examination and afterwards received either radical nephrectomy or partial nephrectomy. The tissue samples were collected from different positions on the surgical specimens resected during surgery, including tumor tissues, renal parenchyma tissues, perirenal fat tissues and tumor border samples. Each tissue sample was cut into two pieces, one for Raman spectrum acquisition and the other for pathological analysis by following the normal procedure in Ren ji Hospital, School of Medicine, Shanghai Jiao Tong University. All frozen specimens for distinguishing subtypes of RCC were taken from the −80 °C sample bank. The acquisition of tissues samples and all experiments in this study were approved by the Ethics Committee of Ren ji Hospital, School of Medicine, Shanghai Jiao Tong University. All tissues were stored in an ice box at 0 °C environment until Raman spectra were collected.

2.2. Raman spectrum acquisition and histologic section

Before the Raman spectrum acquisition, all tissues were taken out from the 0 °C environment and left at room temperature for 20 min. To avoid the spectral interference from blood, the tissues were cleaned with ultra-pure water before the Raman spectrum acquisition. Then the tissues were placed on a quartz coverslip and the Raman measurements were performed on a confocal Raman system (XploRA INV, Horiba). A 785 nm single mode diode laser (29.8 mW) was employed to each spectrum with a 5 s acquisition time. Two Raman spectra at each point were collected and their average is taken in order to increase the signal-to-noise ratio. For each tissue, 100 spectra were collected in the Raman mapping,
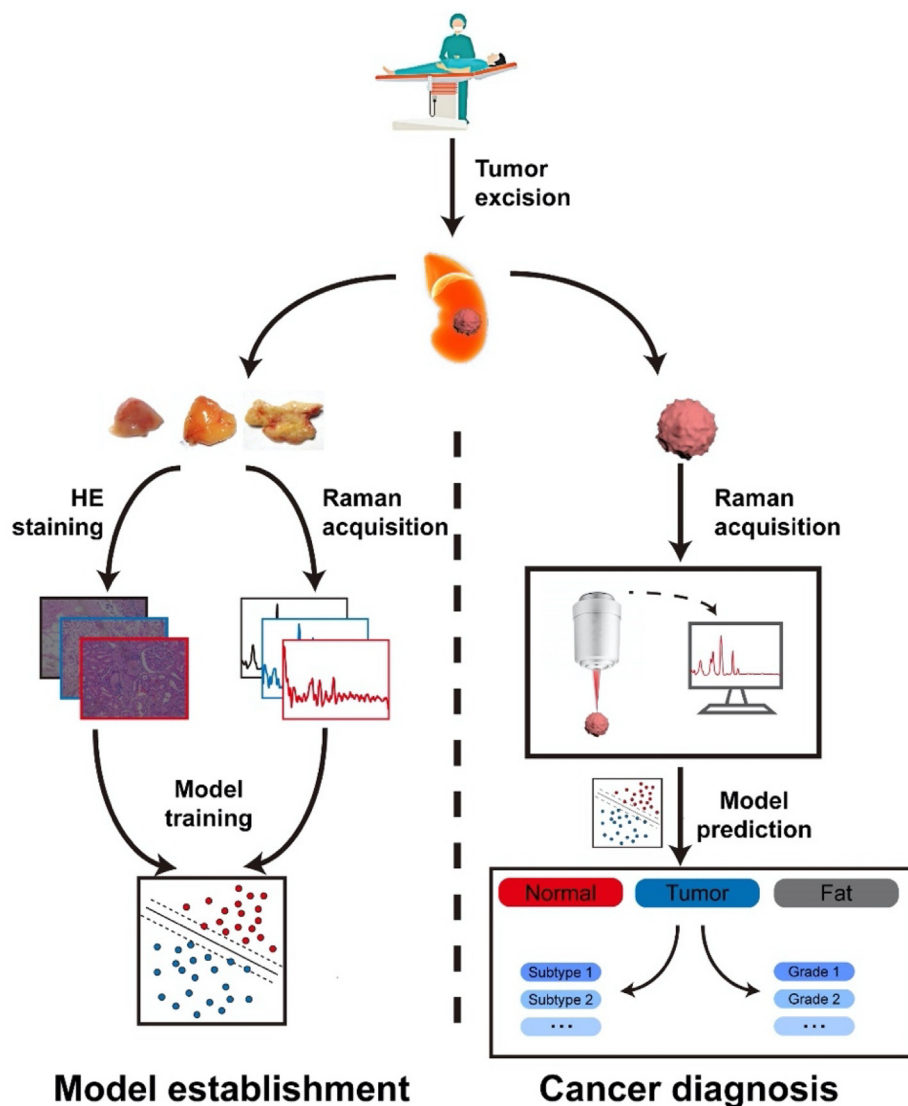
Fig. 1. Schematic workflow of the kidney cancer diagnosis by Raman spectroscopy and supporting vector machine (SVM) algorithm. In model establishment, surgeons removed normal tissue, fat and tumor tissue about 2 mm in diameter from the kidney. Each piece of tissue was divided into two parts, one for Raman spectroscopy acquisition and the other for pathology analysis. Raman spectra and histopathology results were used to train the SVM model. In the intraoperative cancer diagnosis process, only the spectra of suspected tumor tissue were taken and predicted by the established model.

but only 30 spectra were obtained as dataset to establish the classification model.

In order to accurately determine the boundary between tumor tissue and other tissue, we conducted uniform and fine mapping for each boundary tissue which collected more than 1000 Raman spectra. Each spectrum of boundary tissue was predicted by the well-trained classification model. In order to compare the predicted results with the actual results, frozen sections and pathological staining were performed on the boundary tissues after spectral collection. Due to the deformation of the tissues during the measurement, each of the boundary tissues was placed in an environment of −80 °C after the measurement and frozen sections were performed along the plane where the spectra were measured 24 h later. Each frozen section was cut into 10 μm in size and the pathological results were observed after Hematoxylin-Eosin (H&E) stained.

2.3. Data preprocessing and classification

All Raman spectra of tissues were first smoothened to reduce noises by performing Savitzky–Golay filter [42,43]. Then the algo-rithm of adaptive iteratively reweighted penalized least squares (air-PLS) (order = 3, lambda = 90) was employed for the baseline correction [44,45]. Then the Z-score normalization method was used to map the Raman spectral intensity to the data range with a mean of zero and a standard deviation of 1. SVM has extensive applications in the classification of spectral data and tends to find a hyperplane with a largest distance to the nearest data point of classes [46]. The data points that lie on the margin are support vectors (SVs). It is worth mentioning that SVs are usually fewer than the number of data points. In normal conditions, the original spectrum usually is in a thousand-dimension space and it is not linearly separable. To make the separation easier, it is supposed to map the sample space into higher dimensional space with a kernel function. In this work, a linear kernel was used as the kernel function, which was represented as dot product of all points $K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i^{\mathbf{T}} \cdot \mathbf{x}$, where $\mathbf{x}_i$ is support vector. The hyperplane in the higher-dimensional space was calculated by the points whose dot product with SVs in feature space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of support vectors $\mathbf{x}_i$. The points $\mathbf{x}$ in the feature space are mapped into higher-dimensional space and classified by the relation

$$\boldsymbol{\sigma}(\mathbf{x}) = \sum_{i=1}^{m} y_i \alpha_i K(\mathbf{x_i}, \ \mathbf{x}) - b \qquad (1)$$

where m is the number of SVs, y_i is the classification label, the number is 1 or -1, α_i is the weight of support vector $\mathbf{x}_i$ and b is the cut-off value. The relevant source codes are available from the corresponding authors upon reasonable request.

2.4. Leave-one-out cross validation

The leave-one-out cross validation (LOOCV), which is a common validation method for small sample sizes, was employed to verify the fitting degree of the training set. LOOCV is an extreme case of k-fold cross validation, where k equals to the number of patients in training set here. In LOOCV, the Raman data of one patient in original training set is used as validation set and the others are training set [47,48]. Different from the role of test set, the purpose of LOOCV is to optimize the accuracy of model cross validation by adjusting parameters, so as to ensure that the model can fit the data to the maximum extent. However, this set of parameters is only applicable to the current training set, which cannot guarantee the accuracy of prediction for unknown data. When the verification accuracy is almost equal to the test accuracy obtained from the test set, it indicates that the model has a better estimation effect.

3. Results and discussion

3.1. Differentiation of tumor and normal tissues

The fresh tissues obtained during the surgery were from 38 clinical patients, as shown in Table 1. To perform the diagnosis of kidney cancer, we took normal tissue, fat tissue and tumor tissue from the kidneys of patients who had a renal removal. For patients who had a partial excision, only tumor tissues were taken. The diameter of each piece of tissue is about 2 mm. Bright-field (BF) images, histopathology images and Raman spectra of three representative normal, fat and tumor tissues resected from a typical grade 2 sample of CCRCC are shown in Fig. 2. The collection of clinical specimens typically takes a long time. Therefore, additional 39 frozen specimens were added to the training set to avoid the unrepresentative classification performance due to a too small sample size. For each tissue, 100 Raman spectra were collected, but some spectra were excluded duo the signal saturation and typically only 30 spectra for tumor tissues were used to train and test the machine learning models after preprocessing. For the purpose of keeping the balance between three types of tissues, 90 spectra for fat and normal tissues were used to train the models. From the histopathology images (Fig. 2B), the normal tissue mostly con-

sists of the renal essence, which contains its characteristic composition such as glomerular. Compared with normal cells, the cytoplasm of tumor cells is relatively larger and transparent, which is similar to the pathological image of the fat tissue. This can be explained by the fact that CCRCC cells are rich in cholesterol and lipids [33]. Consequently, it leads to a high similarity in Raman spectrum between tumor and fat tissues, including the position and the relative intensity of main Raman bands (Fig. 2C). In contrast, the Raman spectrum of normal tissues shows much larger difference compared to the other two. Previously ones often ignored to identify the fat in either the model establishment or clinical diagnosis [28,49,50]. It should be noted that the fat tissue is most likely to be recognized as the tumor tissue if we established a distinguishing model using only Raman spectra from normal tissue and tumor tissue. This may result in a failure to identify the perinephric fat invasion in RCCs and consequently inappropriate staging evaluation and surgical plan.

Different machine learning models built with the LOOCV method using 4860 spectra from 54 kidney cancer patients (15 clinical and 39 frozen specimens) were employed to accomplished classification of different types of tissues, including SVM with a radial basic function (RBF) kernel or a linear model, artificial neural network (ANN), k-nearest neighbors (KNN), random forest and Adaboost. In order to prevent the over-fitting caused by the superabundant features of Raman spectra, the feature selection based on a variance threshold, used to remove features with low variance, was employed to reduce the number of features to 200 before training the models. Finally, the SVM model with a RBF kernel showed the best classification performance among the six models with a LOOCV accuracy rate of 93.15% (see Table 2). The accuracy represents a quite acceptable classification performance on the training set. This can be meanwhile confirmed by the receiver operating characteristic (ROC) curve and the corresponding area under curve (AUC) value (0.9534) of the classification (Figure S1). The other 2070 spectra from 23 clinical patients as the test set were employed to evaluate the classification performance of six models as unknown samples. The SVM model with the RBF kernel reaches an accuracy of 92.89% which is higher than other models, pretty close to the LOOCV accuracy. The mean specificity of three types of tissue on test set is 92.20% and the sensitivity is 93.15%. Meanwhile, the mean Matthews correlation coefficient (MCC) is 0.9076. These result proves that the model has a good generalization ability for unknown data.

3.2. Determination of tumor boundary

One of the major problems in tumor removal surgery is the difficulty to accurately determine the tumor boundary, so that surgeons have to resect more benign tissue near the tumor to ensure the complete removal of the tumor as much as possible. However, this obviously causes more damage and prolongs the recovery period. Therefore, a model that can accurately determine the location of tumor boundaries is greatly valuable. To identify the renal tumor boundary, we applied the SVM model that was previously used to distinguish between tumor tissues, normal tissues and fats. Two pieces of boundary tissue from different patients, who had a renal removal, were tested and the results are shown in Fig. 3. The BF images were recorded before Raman spectra acquisition (Fig. 3A). Due to water loss during the process of spectral acquisition, the tissues showed partial deformation. Hence, the subsequent H&E staining images do not match the BF images exactly (Fig. 3C). In contrast, the identification images based on the SVM model exhibit a better fitting with the histopathology images (Fig. 3B). Most tumor tissue and normal tissue are accurately distinguished, except that a very small portion of tumor tissue is misclassified by the model as normal tissue or fats (indicated

Table 1
Patient demographics and histological diagnosis.

		Clinical specimens	Frozen specimens
Age		N/A	N/A
	32 ~ 81 (Avg:54)		
Clear cell renal cell carcinoma		29	10
	Grade 1	5	N/A
	Grade 2	19	N/A
	Grade 3	4	N/A
	Grade 4	1	N/A
Papillary renal cell carcinoma		2	15
Chromophobe renal cell carcinoma		1	9
Collecting duct carcinoma		1	5
Total		38	39
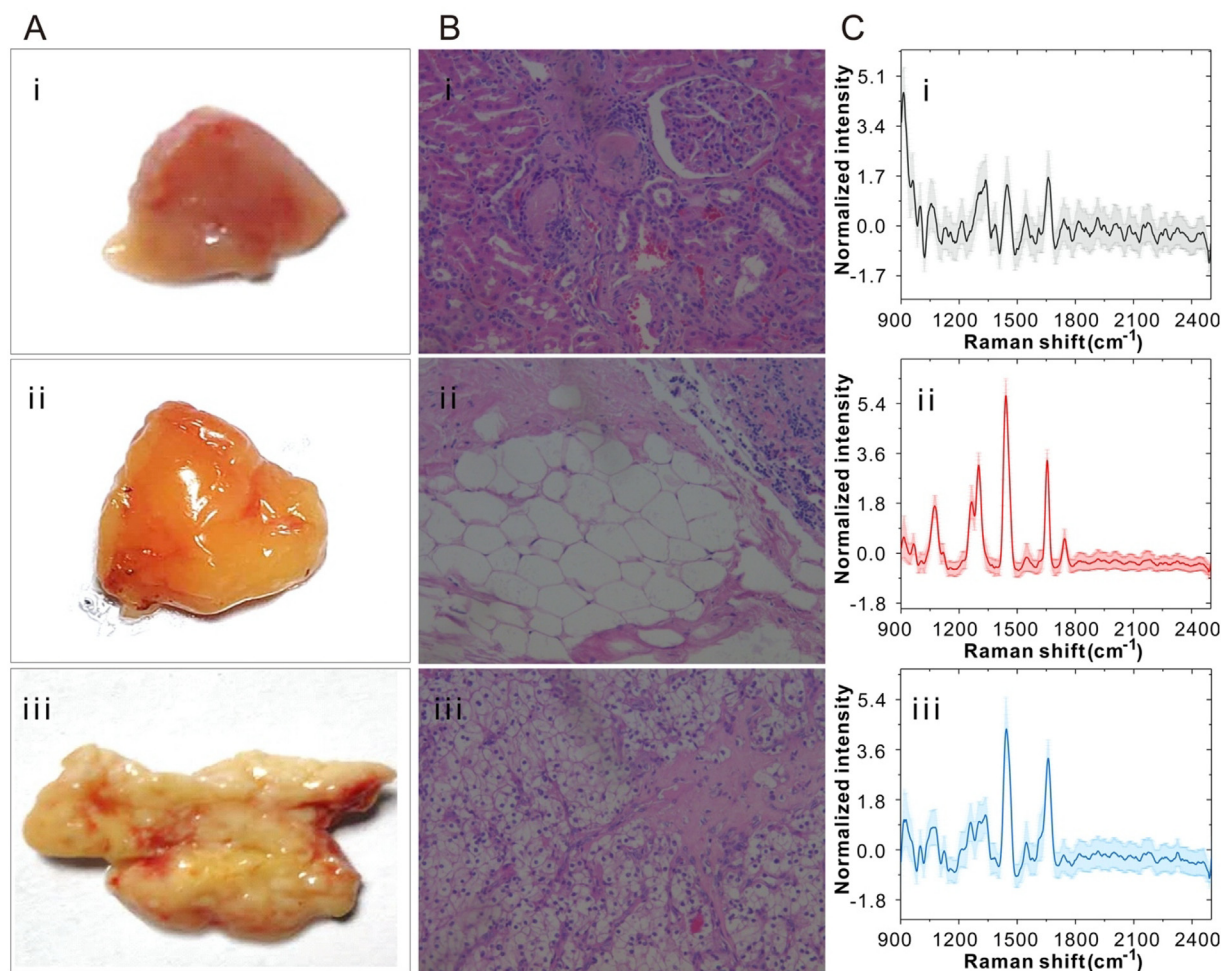
Fig. 2. Identification of kidney normal tissue, fat and tumor tissue. (A) Bright-field (BF) images, (B) histopathology images and (C) Raman spectra of (i) kidney normal tissue, (ii) fat and (iii) tumor tissue. In panel C, the solid curves represent the average Raman spectra and the shaded areas represent their standard deviations.

Table 2
The diagnosis results of the distinction between normal tissue, fat and tumor tissue.

		LOOCV	Test
Patients	Number of patients	54	23
	Number of samples	4860	2070
	Normal tissue	1620	690
	Fat	1620	690
	Tumor tissue	1620	690
Accuracy	SVM (RBF)	93.15%	92.89%
	SVM (Linear)	91.86%	91.53%
	ANN	92.22%	92.29%
	KNN	91.67%	90.90%
	Random forest	90.75%	89.90%
	Adaboost	87.93%	87.57%

by red arrows). There is one noteworthy point that kidney tumors are often surrounded with a capsule layer to isolate the tumor from the external physiological environment and create a microenvironment suitable for tumor growth. It can be clearly seen from Fig. 3C ii that there is a tumor pseudocapsule between tumor and normal tissue (the region indicated by the black dotted line) with a thickness of around 0.6 mm. Since our SVM model did not contain the spectral data of the pure tumor capsule (partially also due to the difficult to remove the capsule separately), the capsule is recognized as the tumor tissue. In fact, this does not affect the surgical plan and the prognosis since the tumor capsule should be removed

during surgery to avoid the recurrence of the cancer as much as possible.

3.3. Distinction of tumor subtypes

Accurate and real-time determination of tumor subtype is indicative to the adjustment of medical plan for doctors. However, it is difficult to identify the subtype of tumor from preoperative bioimaging techniques such as MRI (Fig. 4A). Pathological results usually require a part of tumor tissue after surgery following a series of complex encapsulation and staining processes, making it difficult to guide surgical protocols pre- or intraoperatively. Herein, we additionally aim to build an SVM model based on Raman spectra to diagnose the subtype of kidney cancer. Four subtypes of kidney cancer including CCRCC, PRCC, CRCC and collecting duct carcinoma (CDC) were collected to train and verify the SVM model. The collected fresh clinical samples include patients of 29 CCRCC, 2 PRCC, 1 CRCC and 1 CDC (see Table 1). The Raman spectra of four subtypes are shown in Fig. 4C. There is a large similarity in position of Raman bands in fingerprint region among different subtypes. In contrast, the relative intensities of Raman bands show much difference. This phenomenon can be explained by the fact that the types of biomarkers expressed by different subtypes of tumor cells are roughly the same, but the expression levels are dramatically different [51]. Comparing with CCRCC, the other three subtypes are scarce, which may result in an extremely uneven distribution of
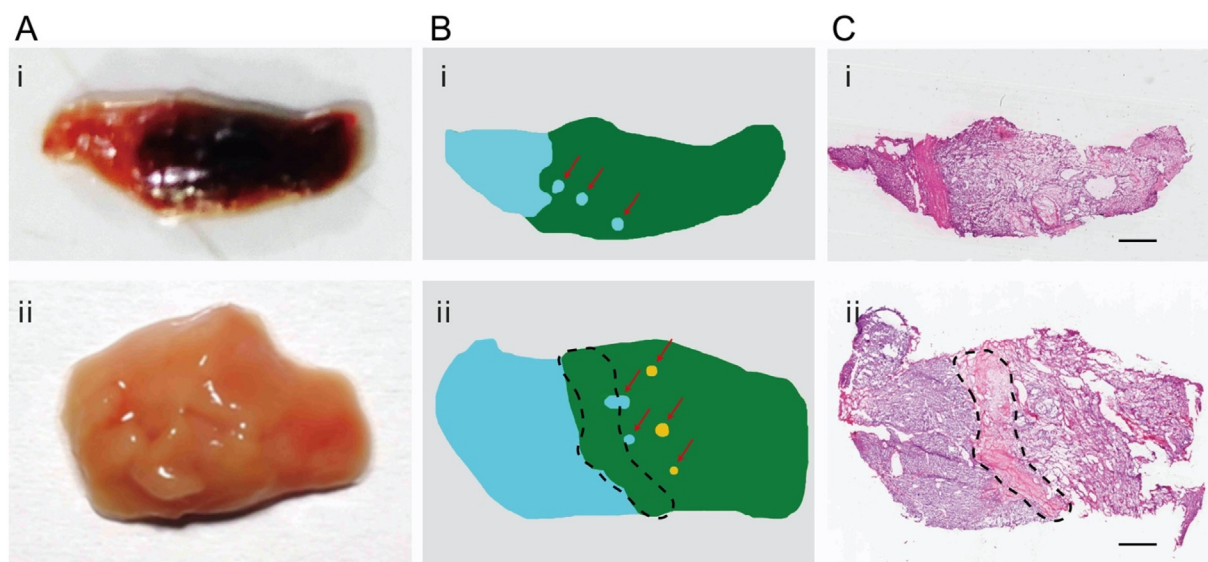
Fig. 3. Identification of the tumor boundary of kidney tissues by the SVM model. (A) BF images, (B) identification images by the SVM model, and (C) histopathology analysis of (i, ii) two intraoperative tumor tissues. Normal tissue, tumor and fat are represented by azure, turquoise and yellow color, respectively, in panel B. The scale bar is 2 mm.

data. To solve the problem of insufficient sample size, we took additional 39 frozen specimens including 10 CCRCC, 15 PRCC, 9 CRCC and 5 CDC (Table 1). The SVM model with the RBF kernel was built with 28 frozen specimens and evaluated by 20 specimens (including 11 frozen and 9 fresh clinical specimens) containing the above-mentioned four carcinoma subtypes (see details in Table 3). The 89.35% LOOCV accuracy and 88.52% test accuracy on the frozen specimens indicate that the model has great potential for distinguishing the carcinoma subtypes from frozen samples. The same SVM model was additionally employed to perform the prediction on clinical specimens. The test accuracy of 86.79%, mean specificity of 88.75% and mean sensitivity of 85.23% mean that this model is also quite suitable for the differentiation of cancer subtypes in clinical kidney specimens.

3.4. Diagnosis of tumor differentiation

One factor that has a significant impact on a doctor's decision on a treatment plan is the tumor differentiation. The clinical CCRCC specimens in this work were divided into four grades based on pathological results according to the standard of Furhman grade of renal cell carcinoma: 5 samples for grade 1, 19 for grade 2, 4 for grade 3 and 1 for grade 4 (see Table 1) [5]. From the preoperative MRI images in Fig. 5A, the tumor at grade 4 with a higher degree of differentiation shows a much larger volume than that at grade 1. However, the tumor size is no longer a good indicator for differentiation between grade 2 and 3. For example, sometimes the grade 2 tumor is even slightly larger than the grade 3 one (Fig. 5A, B). In this work, a SVM model was also developed to distinguish different grades of renal carcinoma. It has been found that the SVM model shows relative low identification accuracies with a LOOCV accuracy of 70.32% and a test accuracy of 68.84% (Table 4), which can be possibly attributed to the heterogenicity of the tumor such as the nonuniform deterioration.

Although histopathology images show that there are some differences in cell nuclei and cell density in different grades of tumors, we have noticed that not all regions of the entire tumor tissue are at the same rate of deterioration. As a result, it is possible that the acquired Raman spectra does not match the histopathology grades. On the other hand, the large difference in the sample size of different grades is also the cause of inaccurate judgement. The Raman spectra of different four grades of tumors display a con-

sistent indication that there is a large similarity in band position and intensity of Raman spectra between adjacent grades of tumors (Fig. 5C). For example, the spectral similarity is particularly pronounced between the grade 3 and 4, making it quite challenge to distinguish between these two grades. The difference in relative intensity of Raman bands between the grade 2 and 3 tumors is a characteristic feature that can be classified. To reduce the impact of the nonuniform deterioration and the small sample size, we combined the grade 1 and 2 into the early stage, and the grade 3 and 4 into the advanced stage. The judgement result shows an improvement that the LOOCV accuracy is 89.74% and test accuracy is 89.53%. The specificity is 91.23% and sensitivity of 89.20%. These improved high accuracies mean that the SVM model can accomplishes a simple two-level classification of tumor deterioration, thus may play an important role for the planning of the treatment strategy.

4. Conclusions

In summary, we successfully established SVM models after training more than 3000 Raman spectra from 77 patients (38 fresh specimens and 39 frozen specimens) to accomplish the following four diagnostic purposes for RCC: tumor identification, tumor boundary identification, tumor classification and tumor differentiation. To the best of our knowledge, we are the first to successfully to distinguish the CCRCC tumor from the fat in addition to normal renal tissue with a test accuracy of 92.89%. More importantly, the SVM model allows to accurately determine tumor boundary using the Raman mapping on the tumor tissue, which has been confirmed by the H&E staining results. Moreover, the trained SVM models can accomplish the classification of tumor subtypes and tumor grades of renal carcinoma with a test accuracy of 86.79% and 89.53%, respectively. These results suggest that the Raman spectroscopy combined with the SVM model can act as an optical noninvasive method to identify renal tumor for preoperative biopsy specimens or intraoperative fresh tissues, which is able to provide an effective guidance for renal surgical plan. The traditional pathological diagnosis requires complicated processing and always takes tens of minutes or even hours which is time-consuming. The acquisition of Raman spectrum only takes several seconds and the predication was accomplished automatically by machine learning model, so the proposed method has a great
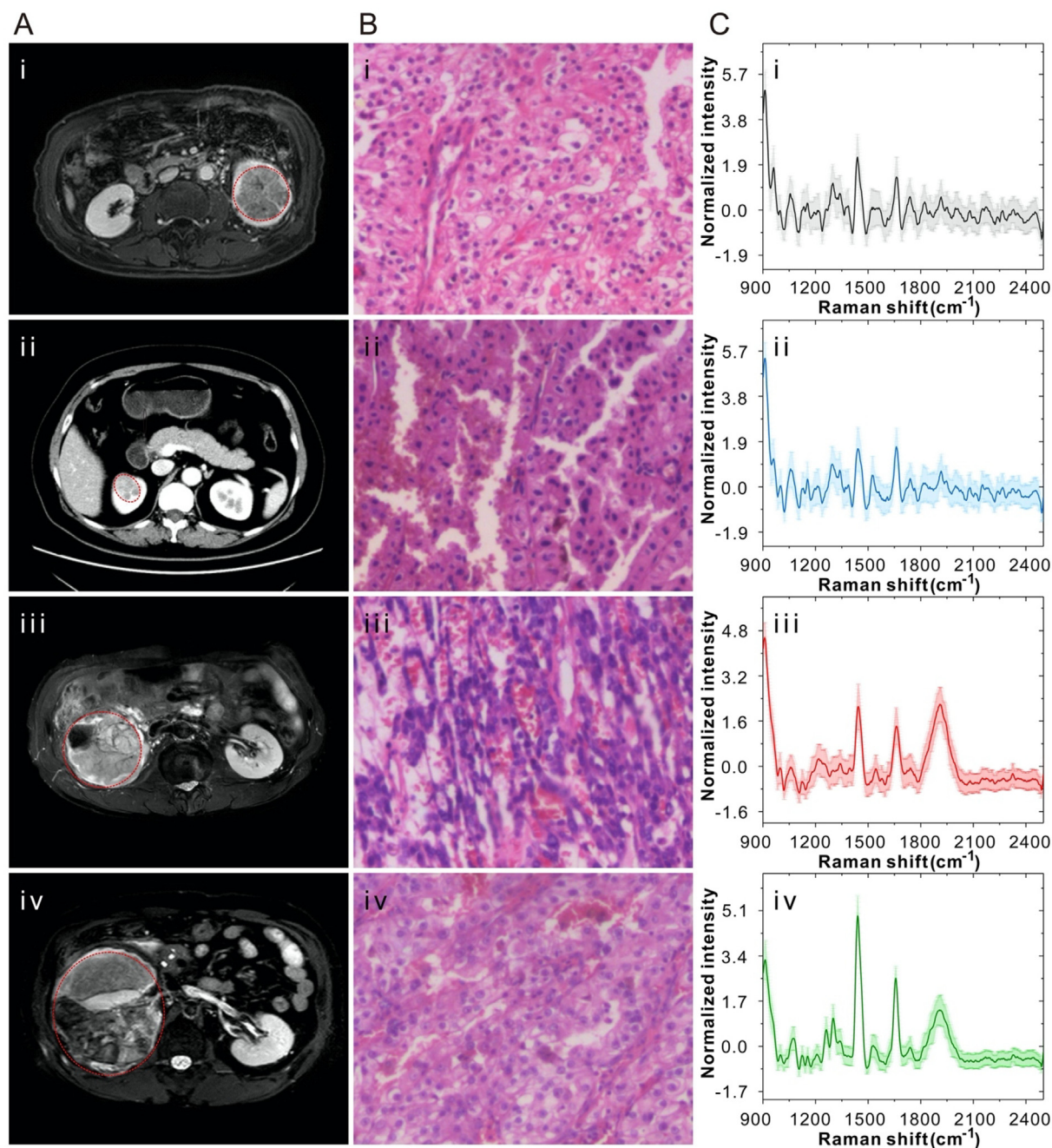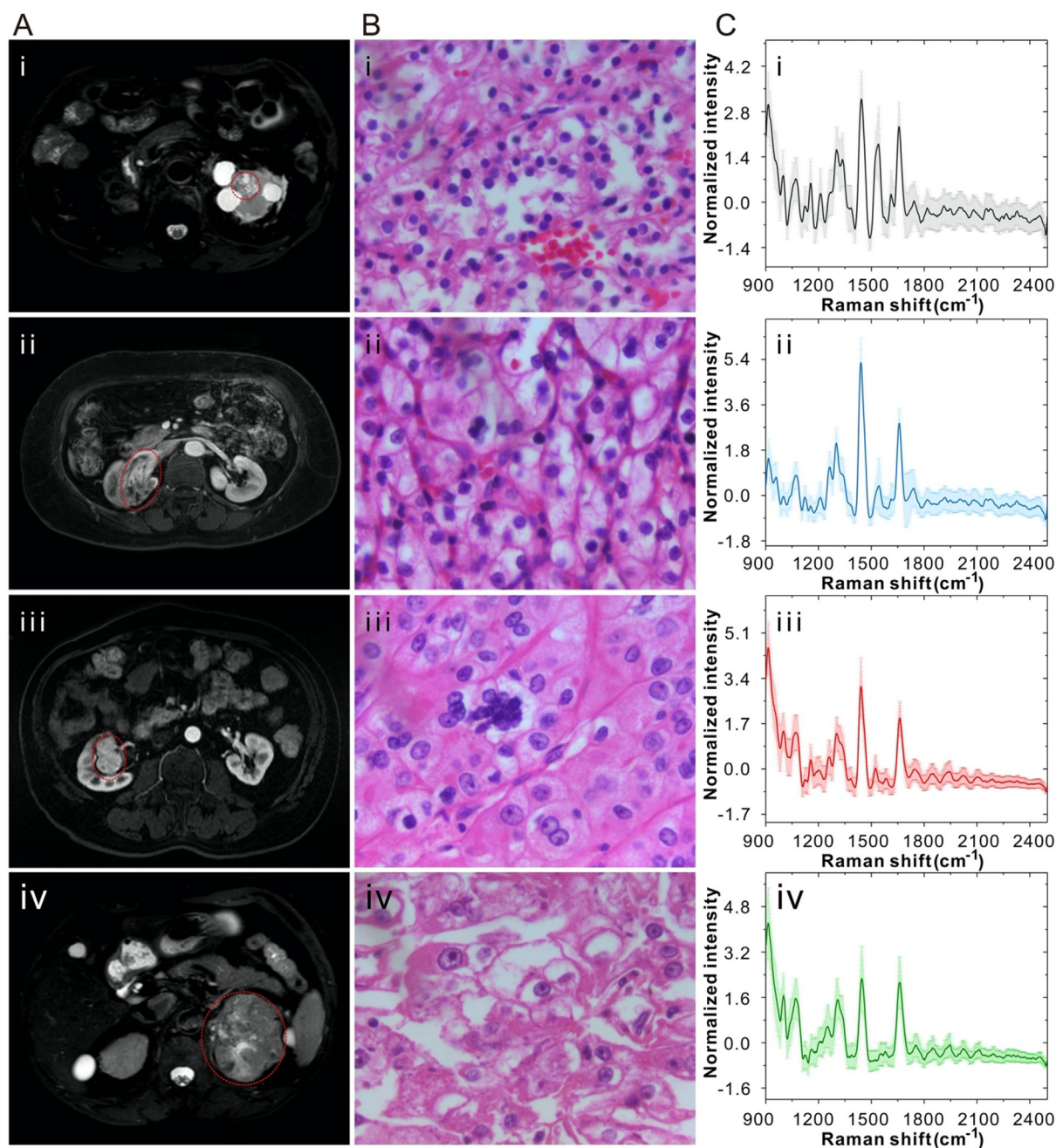
Fig. 4. Identification of four different subtypes of renal carcinoma tissues. (A) Magnetic resonance (MR) images, (B) histopathology images and (C) Raman spectra of different renal carcinoma tissues including (i) clear cell carcinoma, (ii) papillary renal cell carcinoma, (iii) chromophobe renal cell carcinoma and (iv) collecting duct carcinoma. In panel A, the tumors are indicated by red dotted circles. In panel C, the solid curves represent the average Raman spectra and the shaded areas represent their standard deviations.

Table 3
The SVM result of the distinction between four different subtypes of renal carcinoma.

	Frozen specimens		Clinical specimens
	LOOCV	Test	Test
Clear cell carcinoma	7	3	5
Papillary renal cell carcinoma	10	5	2
Chromophobe renal cell carcinoma	7	2	1
Collecting duct carcinoma	4	1	1
Number of samples	840	330	270
Accuracy	89.35%	88.52%	86.79%

Fig. 5. Identification of different grades of renal carcinoma tissues. (A) MR images, (B) histopathology images and (C) Raman spectra of renal carcinoma tissues of grade (i) 1, (ii) 2, (iii) 3 and (iv) 4. In panel A, the tumors are indicated by red dotted circles. In panel C, the solid curves represent the average Raman spectra and the shaded areas represent their standard deviations.

Table 4
The SVM result of the distinction between different stages of renal carcinoma.

		LOOCV	Test
Early grade	Grade 1	3	2
	Grade 2	13	6
Terminal grade	Grade 3	3	1
	Grade 4	1	0
Number of samples		20	9
Accuracy	4 Grades	70.32%	68.84%
	Early and advanced stage	89.74%	89.53%

potential as the real-time intraoperative tumor identification method. Moreover, this technique doesn't rely on the subjective judgement of doctors, which is different from pathological diagnosis. Additionally, the noninvasion of the proposed method indicates that the tissue after Raman acquisition can still be diagnosed by other methods, such as pathological staining and imaging diagnosis. Raman spectroscopy can be further seamlessly integrated with the optical-fiber based endoscope and medical robotics for *in vivo* real-time intraoperative diagnosis in order to achieve better cancer therapy including avoidance of tumor residues and excessive excision of normal tissues [52].

CRediT authorship contribution statement

Chang He: Methodology, Software, Formal analysis, Writing - review & editing. **Xiaorong Wu:** Funding acquisition, Methodol-

ogy, Writing - review & editing. **Jiale Zhou:** Methodology, Writing - review & editing. **Yonghui Chen:** Conceptualization, Supervision, Funding acquisition. **Jian Ye:** Conceptualization, Supervision, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2021.119520.

References

[1] P. Cairns, Renal cell carcinoma, Cancer Biomark. 9 (1–6) (2010) 461–473.
[2] B. Ljungberg, L. Albiges, Y. Abu-Ghanem, K. Bensalah, S. Dabestani, S. Fernandez-Pello, et al., European association of urology guidelines on renal cell carcinoma: the 2019 update, Eur. Urol. 75 (5) (2019) 799–810.
[3] D. Zhu, J. Cao, C. Zhi, T. Guo, Y. Li, Z. Lang, et al., Prognostic significance of the sub-classification of stage pT3a renal tumors by perinephric and sinus fat invasion, Oncol. Lett. 19 (3) (2020) 1721–1726.
[4] C.M. Russell, A.H. Lebastchi, J. Chipollini, A. Niemann, R. Mehra, T.M. Morgan, et al., Multi-institutional survival analysis of incidental pathologic T3a upstaging in clinical T1 renal cell carcinoma following partial nephrectomy, Urology 117 (2018) 95–100.
[5] S.A. Fuhrman, L.C. Lasky, C. Limas, Prognostic significance of morphologic parameters in renal cell carcinoma, Am. J. Surg. Pathol. 6 (7) (1982) 655.
[6] H.K. Sokhi, W.Y. Mok, U. Patel, Stage T3a renal cell carcinoma: staging accuracy of CT for sinus fat, perinephric fat or renal vein invasion, Br. J. Radiol. 88 (1045) (2015) 20140504.
[7] J. Landman, J.Y. Park, C. Zhao, M. Baker, M. Hofmann, M. Helmy, et al., Preoperative computed tomography assessment for perinephric fat invasion: comparison with pathological staging, J. Comput. Assist Tomogr. 41 (5) (2017) 702–707.
[8] P.H. Shah, T.D. Lyon, C.M. Lohse, J.C. Cheville, B.C. Leibovich, S.A. Boorjian, et al., Prognostic evaluation of perinephric fat, renal sinus fat, and renal vein invasion for patients with pathological stage T3a clear-cell renal cell carcinoma, BJU Int. 123 (2) (2019) 270–276.
[9] P. Mouracade, J. Dagenais, J.S. Chavali, O. Kara, R.J. Nelson, M.J. Maurice, et al., Perinephric and sinus fat invasion in stage pT3a tumors managed by partial nephrectomy, Clin. Genitourin Cancer 16 (5) (2018) e1077–e1082.
[10] J. Bedke, S. Buse, M. Pritsch, S. Machergoeppinger, P. Schirmacher, A. Haferkamp, et al., Perinephric and renal sinus fat infiltration in pT3a renal cell carcinoma: possible prognostic differences, BJUI 103 (10) (2009) 1349–1354.
[11] M. Chevinsky, M. Imnadze, A.I. Sankin, A. Winer, R. Mano, C. Jakubowski, et al., Pathological stage T3a significantly increases disease recurrence across all tumor sizes in renal cell carcinoma, J. Urol. 194 (2) (2014) 310–315.
[12] R. Montironi, L. Cheng, M. Scarpelli, A. Lopez-Beltran, Pathology and genetics: tumours of the urinary system and male genital system: clinical implications of the 4th edition of the WHO classification and beyond, Eur. Urol. 70 (1) (2016) 120–123.
[13] H. Moch, A.L. Cubilla, P.A. Humphrey, V.E. Reuter, T.M. Ulbright, The 2016 WHO classification of tumours of the urinary system and male genital organs-Part A: Renal, penile, and testicular tumours, Eur. Urol. 70 (1) (2016) 93–105.
[14] L.D. Truong, B. Krishnan, S.S. Shen, Intraoperative pathology consultation for kidney and urinary bladder specimens, Arch. Pathol. Lab. Med. 129 (12) (2005) 1585–1601.
[15] A.S. Haka, K.E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R.R. Dasari, M.S. Feld, Diagnosing breast cancer by using Raman spectroscopy, Proc. Natl. Acad. Sci. 102 (35) (2005) 12371–12376.
[16] Z. Movasaghi, S. Rehman, I.U. Rehman, Raman spectroscopy of biological tissues, Appl. Spectrosc. Rev. 42 (5) (2007) 493–541.
[17] O. Old, L. Fullwood, R. Scott, G. Lloyd, L. Almond, N. Shepherd, et al., Vibrational spectroscopy for cancer diagnostics, Anal. Methods 6 (12) (2014) 3901–3917.
[18] K.A. Antonio, Z.D. Schultz, Advances in biomedical Raman microscopy, Anal. Chem. 86 (1) (2014) 30–46.
[19] T. Fan, G. Sun, X. Sun, L. Zhao, R. Zhong, Y. Peng, Tumor energy metabolism and potential of 3-bromopyruvate as an inhibitor of aerobic glycolysis: implications in tumor treatment, Cancers (Basel) 11 (3) (2019).
[20] L.M. Phan, S.C. Yeung, M.H. Lee, Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies, Cancer Biol. Med. 11 (1) (2014) 1–19.
[21] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (5) (2011) 646–674.
[22] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell 100 (1) (2000) 57–70.
[23] Y. Zhou, C.H. Liu, Y. Pu, B. Wu, T.A. Nguyen, G. Cheng, et al., Combined spatial frequency spectroscopy analysis with visible resonance Raman for optical biopsy of human brain metastases of lung cancers, J. Innov. Optical Health Sci. 12 (2) (2019).
[24] H.K. Yosef, S.D. Krauss, T. Lechtonen, H. Jutte, A. Tannapfel, H.U. Kafferlein, et al., Noninvasive diagnosis of high-grade urothelial carcinoma in urine by Raman spectral imaging, Anal. Chem. 89 (12) (2017) 6893–6899.
[25] J. Yan, F. Shi, M. Zhao, Z. Wang, Y. Yang, S. Chen, Confocal Raman sensing based on a support vector machine for detecting lung adenocarcinoma cells, IEEE Sensors J. 19 (21) (2019) 9624–9633.
[26] T.J.E. Hubbard, A. Shore, N. Stone, Raman spectroscopy for rapid intra-operative margin analysis of surgically excised tumour specimens, Analyst 144 (22) (2019) 6479–6496.
[27] S. Holler, E. Mansley, C. Mazzeo, M.J. Donovan, M. Sobrero, B.A. Miles, Raman Spectroscopy of head and neck cancer: separation of malignant and healthy tissue using signatures outside the "fingerprint" region, Biosensors (Basel) 7 (2) (2017).
[28] M. Haifler, I. Pence, Y. Sun, A. Kutikov, R.G. Uzzo, A. Mahadevan-Jansen, et al., Discrimination of malignant and normal kidney tissue with short wave infrared dispersive Raman spectroscopy, J. Biophotonics 11 (6) (2018) e201700188.
[29] P. Piredda, M. Berning, P. Boukamp, A. Volkmer, Subcellular Raman Microspectroscopy Imaging of Nucleic Acids and Tryptophan for Distinction of Normal Human Skin Cells and Tumorigenic Keratinocytes, Anal Chem. 87 (13) (2015) 6778–6785.
[30] A. Karuna, F. Masia, M. Wiltshire, R. Errington, P. Borri, W. Langbein, Label-free volumetric quantitative imaging of the human somatic cell division by hyperspectral coherent anti-stokes RAMzAN scattering, Anal. Chem. 91 (4) (2019) 2813–2821.
[31] K. Bensalah, J. Fleureau, D. Rolland, O. Lavastre, N. Rioux-Leclercq, F.O. Guillé, et al., Raman spectroscopy: a novel experimental approach to evaluating renal tumours, Eur. Urol. 58 (4) (2010) 602–608.
[32] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, et al., Intraoperative brain cancer detection with Raman spectroscopy in humans, Sci. Transl. Med. 7 (274) (2015), 274ra19-ra19.
[33] K. Hoffmann, J. Blaudszun, C. Brunken, W.W. Hopker, R. Tauber, H. Steinhart, Lipid class distribution of fatty acids including conjugated linoleic acids in healthy and cancerous parts of human kidneys, Lipids 40 (10) (2005) 1057–1062.
[34] Y.Y. Qiu, Y.Q. Zhang, M.W. Li, C.C. Fan, K. Cui, J.B. Wan, et al., Intraoperative detection and eradication of residual microtumors with gap-enhanced Raman tags, ACS Nano 12 (8) (2018) 7974–7985.
[35] N.G. Khlebtsov, L. Lin, B.N. Khlebtsov, J. Ye, Gap-enhanced Raman tags: fabrication, optical properties, and theranostic applications, Theranostics 10 (5) (2020) 2067–2094.
[36] Y.Q. Zhang, Y.Q. Gu, J. He, B.D. Thackray, J. Ye, Ultrabright gap-enhanced Raman tags for high-speed bioimaging, Nat. Commun. 10 (1) (2019) 3905.
[37] H. Hu, Y. Sheng, M. Ye, Y. Qian, J. Tang, Y. Shen, A porphyrin-based magnetic and fluorescent dual-modal nanoprobe for tumor imaging, Polymer 88 (2016) 94–101.
[38] B. Qian, D. Kyuno, M. Schäfer, W. Gross, A. Mehrabi, E. Ryschich, Liver segment imaging using monocyte sequestration: a potential tool for fluorescence-guided liver surgery, Theranostics 8 (22) (2018) 6101.
[39] W. Duan, Q. Yue, Y. Liu, Y. Zhang, Q. Guo, C. Wang, et al., A pH ratiometrically responsive surface enhanced resonance Raman scattering probe for tumor acidic margin delineation and image-guided surgery, Chem. Sci. 11 (17) (2020) 4397–4402.
[40] S. Chen, C. Bao, C. Zhang, Y. Yang, K. Wang, B.V. Chikkaveeraiah, et al., EGFR antibody conjugated bimetallic Au@Ag nanorods for enhanced SERS-based tumor boundary identification, targeted photoacoustic imaging and photothermal therapy, Nano Biomed. Eng. 8 (4) (2016).
[41] Y.R. Zheng, Biological applications of support vector machines, Briefings Bioinform 5 (4) (2004) 328–338.
[42] S.J. Barton, T.E. Ward, B.M. Hennelly, Algorithm for optimal denoising of Raman spectra, Anal. Methods 10 (30) (2018) 3759–3769.
[43] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (11) (1964) 1627–1639.
[44] Z.M. Zhang, S. Chen, Y.Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, Analyst 135 (5) (2010) 1138–1146.
[45] H.F.M. Boelens, R.J. Dijkstra, P.H.C. Eilers, F. Fitzpatrick, J.A. Westerhuis, New background correction method for liquid chromatography with diode array detection, infrared spectroscopic detection and Raman spectroscopic detection, J. Chromatogr. A 1057 (1) (2004) 21–30.

[46] S.J. Huang, N.G. Cai, P.P. Pacheco, S. Narandes, Y. Wang, W.N. Xu, Applications of Support Vector Machine (SVM) learning in cancer genomics, Cancer Genom. Proteom. 15 (1) (2018) 41–51.

[47] T.-T. Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognit. 48 (9) (2015) 2839–2846.

[48] S. Yadav, S. Shukla (Eds.), Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification, in: 2016 IEEE 6th International conference on advanced computing (IACC), IEEE, 2016.

[49] W.C. Zuniga, V. Jones, S.M. Anderson, A. Echevarria, N.L. Miller, C. Stashko, et al., Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy, Sci. Rep. 9 (1) (2019) 14639.

[50] J.A. Caddedu, Raman spectroscopy: a novel experimental approach to evaluating renal tumours, J. Urol. 58 (4) (2011) 602–608.

[51] E. Oosterwijk, P.F.A. Mulders, Renal Cancer Including Molecular Characterization, in: C.R. Chapple, W.D. Steers, C.P. Evans (Eds.), Urologic Principles and Practice, Springer International Publishing, Cham, 2020, pp. 551–559.

[52] Z.W. Huang, S.K. Teh, W. Zheng, J.H. Mo, K. Lin, X.Z. Shao, et al., Integrated Raman spectroscopy and trimodal wide-field imaging techniques for real-time in vivo tissue Raman measurements at endoscopy, Opt. Lett. 34 (6) (2009) 758–760.