Photodiagnosis and Photodynamic Therapy 40 (2022) 103177

Contents lists available at ScienceDirect

Photodiagnosis and Photodynamic Therapy

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

Rapid and sensitive detection of esophageal cancer by FTIR spectroscopy of serum and plasma

Hongjun Chen [a], Xianchang Li [b,c,*], Shiding Zhang [b], Haijun Yang [a], Qianqian Gao [b], Fuyou Zhou [a,**]

[a] Anyang Tumor Hospital, The Affiliated Anyang Tumor Hospital of Henan University of Science and Technology, Henan key medical laboratory of precise prevention and treatment of esophageal cancer, Anyang 455001, Henan Province, China
[b] Henan Joint International Research Laboratory of Nanocomposite Sensing Materials, Anyang Institute of Technology, Anyang 455000, Henan Province, China
[c] Huzhou College, Huzhou 313000, Zhejiang Province, China

ARTICLE INFO

Keywords:
Esophageal squamous cell carcinoma
Transmission FTIR
Multivariate classification analysis
Serum and plasma
Spectroscopy

ABSTRACT

Fourier transform infrared (FTIR) spectroscopy, as a platform technology for cancer detection, must be up to the challenge of clinical transformation. To this end, detection of esophageal squamous cell carcinoma (ESCC) was hereby explored using serum and plasma scrape-coated on barium fluoride (BaF_2) disk by transmission FTIR method, and the classification model was built using six multivariate statistical analyses, including support vector machine (SVM), principal component linear discriminant analysis (PC-LDA), decision tree (DT), k-nearest neighbor (KNN) classification, ensemble algorithms (EA) and partial least squares for discriminant analysis (PLS-DA). All statistical analyses methods demonstrated that late-stage cancer could be well classified from healthy people employing either serum or plasma with different anticoagulants. Resulting PC-LDA model differentiated late-stage cancer from normal group with an accuracy of 99.26%, a sensitivity of 98.53%, and a specificity of 100%. The accuracy and sensitivity reached 97.08% and 91.43%, respectively for early-stage cancer discrimination from normal group. This pilot exploration demonstrated that transmission FTIR provided a rapid, cost effective and sensitive method for ESCC diagnosis using either serum or plasma.

1. Introduction

Given that routine therapies are rarely curative in the occurrence of distant metastasis, cancer, which can originate from different site of the body, such as lung, stomach, bladder, skin etc., is considered one of the most fatal diseases. As one of the top 6 most frequently diagnosed cancers that lead to fatal death [1], esophageal cancer has two major histological sub-types, i.e., esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC). EAC mainly occurs in the western countries, whereas ESCC is more prevalent worldwide with a high incidence in those developing countries in Asia and south-eastern Africa. Despite the epidemiological differences, both types of oesophageal cancer have a poor prognosis, and profound improvements can be seen if they are treated prior to invasion [2,3]. In this case, one major goal in esophageal cancer research is to develop a rapid, cheap and sensitive detection method for the diagnosis before the cancer metastasizes to distant sites.

Among all the current detecting approaches for the diagnosis of oesophageal cancer and preinvasive lesions, endoscopy is considered the gold-standard technique owning to its high sensitivity [4]. However, considering the high cost, the invasiveness of the procedure and the psychological burden, it is not accepted as an excellent national screening program [5]. Tissue-based tumor profiles are, generally speaking, subject to sampling bias, lack of repeatability, inconvenience of operation and invasiveness to human body [6]. For example, the esophageal balloon cytology method has been extensively used for detecting early manifestations of cancer in patients. The current blind sampling method of a large organ may miss small lesions and rare abnormal cells. Moreover, it can easily cause heartburn symptoms for patients. Therefore, current precision oncology has increasingly paid attention to liquid biopsies, since they are non-invasive and can be repeated at multiple time points [7].

Plenty of substances that have been found in the peripheral blood make blood the most used fluids in liquid biopsies. These substances are

* Corresponding author.
** Co-corresponding author.
E-mail addresses: lxc_198008@163.com (X. Li), ayzhoufuyou@163.com (F. Zhou).

https://doi.org/10.1016/j.pdpdt.2022.103177
Received 10 September 2022; Received in revised form 21 October 2022; Accepted 26 October 2022
Available online 28 October 2022

closely related to tumor stage and might serve as new biomarkers, including circulating tumor cells (CTCs), exosomes, circulating tumor DNAs (ctDNAs), and a series of cancer-related proteins. However, the complexity of blood samples and the extremely low concentration of the analytes challenge the sensitivity and reliability of current detection methods for clinical applications [8]. Fourier-transform infrared (FTIR) spectroscopy is a well-established optical technique capable of providing the biochemical information of the samples on the molecular level. It has multiple advantages, including high sensitivity, low cost, high repeatability, rapid and non-destructive, which renders it a perfect candidate for clinical translation. Up to now, using blood serum and plasma, FTIR has been employed to diagnose rheumatoid arthritis [9], HIV/AIDS [10], leukemia [11], depression [12], alzheimer's [13], malaria parasitemia [14,15], glucose [16], breast cancer [17–20], gastric cancer [21], liver cancer [22], ovarian cancer [23], lung cancer [24], endometrial cancer [25], bladder cancer [26,27], brain tumor [28,29], oral submucous fibrosis, colitis screening [30,31], etc.

As a potential clinical detecting method, FTIR technique has been used for discriminating esophageal cancer cell [32–35] and Barrett's esophagus tissues [36–39]. For biofluids detection, a recent study has applied the attenuated total reflection (ATR) FTIR to the interrogation of the oesophageal stages of transformation to EAC, employing the biofluids plasma, serum, saliva, or urine [40]. However, the transition from FTIR research to routine clinical testing remains elusive and is exposed to several major technological challenges, such as the disability to identify the disease type, lack of standardization of the technique, and validation of large cohort of samples etc. [41,42]. In addition, it is widely acknowledged that the current esophageal cancer study using FTIR mainly focus on EAC, with few on ESCC, making it of great significance of value to establish a rapid, cheap and sensitive detecting system employing FTIR technique in ESCC diagnosis for developing countries.

In this study, exploration was conducted on the detection of esophageal squamous cell carcinoma (ESCC) using serum and plasma by transmission FTIR method. Customized Barium fluoride (BaF_2) slices enable high-throughput FTIR measurement for clinical research. The changes which occur between FTIR spectra of different stage cancer samples can often be small and not easily artificial observable, therefore machine learning algorithms are frequently used to differentiate these differences in spectral data [43–47]. Decision Tree (DT), Support Vector Machine (SVM), *k*-Nearest Neighbour (KNN) classification, Ensemble Algorithms (EA), Partial Least Squares for Discriminant Analysis (PLS-DA) and Principal Component Linear Discriminant Analysis (PC-LDA) were used to discriminate the FTIR spectral profiles in this paper. Additionally, previous study has shown that the choice of anticoagulant is noticeable in the resultant infrared spectrum [14], which may influence the diagnostic performance. To this end, the diagnostic performance by serum and plasma with different anticoagulants was also assessed.

2. Materials and methods

This work was granted ethical approval by the Ethics Committee of Anyang Tumour Hospital with No. AZLL022020065200407.

2.1. Blood samples

All the samples were obtained from patients or healthy volunteers who attended Anyang Tumour Hospital with fasting for more than 8 h, and the samples were divided into those for late-stage cancer, those for early-stage cancer, or the normal group based on biopsy diagnosis. Blood samples of the late-stage cancer group were collected from 68 patients with clinically confirmed esophageal squamous cell carcinomas, confirmed to be either stage II ($n = 40$) or stage III ($n = 28$). The age was 46 to 78 years old with a mean of 66, and the ratio of male to female was about 1.3:1. The early cancer group was comprised of 35 patients with early-stage cancers (stage 0 and I). The age was 49 to 68 years old, with a mean of 62, and the ratio of male to female was about 1.3:1. The normal group was made up of 68 healthy people with no inflammation. The age was 36 to 76 years old, with a mean of 54, and the ratio of male to female was about 1.1:1. It should be noted that the patient group were confirmed using the H&E pathological diagnosis method, while the normal group were determined by no clinical symptoms and no inflammation through the blood routine test.

2.2. Serum and plasma samples preparation

Plasma samples were collected in the tube with anticoagulants of citrate (C), heparin (H) and EDTA (E) from 68 late-stage cancer patients and 68 healthy volunteers, respectively. To minimize the deviation caused by sample difference, the serum samples were obtained from the same patients or volunteers. Only plasma sample with anticoagulants of heparin were obtained from 35 patients which were diagnosed as stage 0 and I. The plasma and serum samples were centrifuged at 3000 rpm for 5 min to remove the red blood cells, and then transferred into cryogenic tubes and stored at −70 °C until taking out for FTIR measurements.

2.3. FTIR spectra measurements and data process

All the blood samples were fully thawed at room temperature before measurement. 1 μL plasma or serum sample was scrape-coated on the customized BaF_2 disk, and then dried in air at room temperature for about 10 min to form a homogeneous blood film. Considerable blood films can be produced for high-throughput FTIR measurement using this very method. Then, the BaF_2 disk was mounted on a hollow metal disk with a recessed inner edge that can be fixed on the infrared light path. All the spectra were collected using an IRtracer-100 FTIR spectrometer (Shimadzu) in the range of 400–4000 cm^{-1}, with a spectral resolution of 4 cm^{-1}. Each spectrum was collected with an integration time of 1 s, co-added over 20 scans. The IR spectra were slightly smoothed using the LabSolutions IR software, while the baseline of smoothed spectra were corrected over the 3800–400 cm^{-1} region using Origin 2022 software (OriginLab) with the 2nd derivative (zeroes) method, and then normalized with the intensity of peak 1656 cm^{-1}. The mean spectra and standard deviations were then obtained using Origin 2022 software.

2.4. Multivariate discrimination analysis

Two-, three- and four-group classification models were hereby built based on DT, SVM, KNN, EA, PLS-DA and PC-LDA, respectively. A 10-fold cross-validation was applied into the model building. To perform a 10-fold cross-validation procedure, all the FTIR data for model construction were randomly partitioned into 10 equal fold. The optimal model was established within 30 iterations of training and validation. In each cross validation, nine folds of the data were used for model training, and the remaining fold of data was left-out for validation. The models were trained from the training set and used to make predictions of the validation set. The confusion matrices of model prediction and area-under-the-ROC curve (AUC) were calculated using the correspond algorithm, and the sensitivity, specificity, and accuracy were calculated using the following the equations:

$$\text{Sensitivity}(100\%) = \frac{TP}{TP + FN} \times 100\%$$

$$\text{Specificity}(100\%) = \frac{TN}{TN + FP} \times 100\%$$

$$\text{Accuacy}(100\%) = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%$$

where TP stands for true positives; TN, true negatives; FP, false positives; and FN, false negatives. All multivariate data analyses and plots were

Table 1
Summary of the samples and the clinical characteristics of different stages.

Group	Clinical diagnosis	Clinical characteristics	Number ofsamples	Sex	Age range/ mean age
Normal	–	No diseases and no inflammation	68	32F, 36M	36–76 years/ 54 years
Early-stage cancer	Stage 0	Cancer cells present in the inner lining of the esophagus wall.	12	15F,	49–68 years/
	Stage I	Cancer cells spread into the mucosa layer, thin muscle layer, or submucosa layer of the esophagus wall.	23	20M	62 years
Late-stage cancer	Stage II	Cancer cells spread into the thick muscle layer or connective tissue layer of the esophagus wall.	40	30F,	46–78 years/
	Stage III	Cancer cells spread into the thick muscle layer or connective tissue layer of the esophagus wall, and present in the nearby lymph nodes, or spread into other organs.	28	38M	66 years

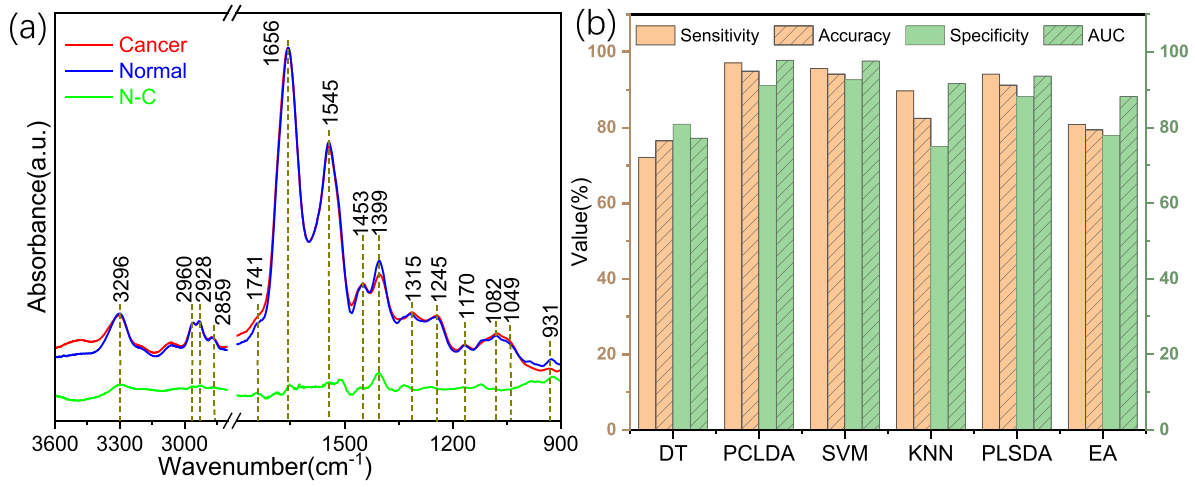

Fig. 1. (a) Mean and differential (N—C: green line) FTIR spectra of serum samples of normal (Normal-blue line) and late stage cancer (Cancer-red line) group. (b) Performance for serum spectral classification based on six different multivariate algorithms.

Table 2
The vibrational mode assignment of FTIR peaks and its primary source.

Peaks(cm^{-1})	Vibrational mode	Primary source
3296	O-H stretch	carbohydrates
2960	CH_3 asymmetrical stretch	lipids, nucleic acids and proteins
2928	CH_2 asymmetrical stretch	lipids
2859	CH_2 symmetrical stretch	lipids
1741	C=O stretch	lipids
1656	C=O stretch	Amide I of proteins
1545	C—N stretch, N—H bend	Amide II of proteins
1453	C-H bend	proteins
1399	C—H bend, COO^- stretch	lipids and proteins
1315	C-H stretch, N—H bend	Amide III of proteins
1245	PO_2^- asymmetrical stretch	nucleic acids
1170	C—O stretch	glycogen or proteins
1082	PO_2^- symmetrical stretch	nucleic acids
1049	C-O stretch, C—O bend	glycogen
931	C—C stretch	residue α-helix

implemented within MATLAB (version 2019a, Math-Works) and OriginPro (version 2022, OriginLab).

3. Results and discussions

3.1. FTIR spectra of serum

Detailed sample information is listed in Table 1. Fig. 1(a) shows the mean and differential FTIR spectra of serum samples of normal (blue line) and late-stage cancer (red line) group in the fingerprint region of 900–1800 cm^{-1} and the high wavenumber region of 2800–3600 cm^{-1}. Besides, the mean spectra and standard deviation are shown in Fig S1. Most of these labelled infrared peaks are in well conformity with those reported in previous literatures by the study of the FTIR spectra of blood and cells, and the assignment of these peaks has also been well established in the literature [27,29–31,38] and shown in Table 2. In the high wavenumber region, the peak around 3296 cm^{-1} was classified to the O—H stretching mode of hydroxyl groups, while that around 2960 cm^{-1} was assigned to the C—H stretching of CH_3 from fatty acids, lipids and proteins. The other two peaks of 2926 and 2859 cm^{-1} were attributed by C—H stretching of CH_2 from lipids. In the fingerprint region, 1741 cm^{-1} was assigned to the C=O stretching of lipids. The two strong peaks at around 1656 and 1546 cm^{-1}, assigned to the stretching vibrations of the C=O, C—N and bending vibrations of N—H, respectively, were mainly representatives of amide I and II of β-pleated sheet structured proteins. Additionally, the medium strength peaks at around 1453 and 1399 cm^{-1} were assigned to the C—H deformation of lipids methylene group. The bands in the 1000~1300 cm^{-1} region, assigned to the stretching vibrations of C—O, C—C, P-O and bending vibration of C—H, were mainly classified to phospholipids, glucose and nucleic acid, and the peak at 931 cm^{-1} was assigned to C—C residue of α-helix.

The green line in Fig. 1(a) depicts the differential FTIR spectra of serum samples of normal and late-stage cancer group, where it can be seen that the main differences of these two groups are in the region of 1500–1250 cm^{-1}. The higher intensity of 1399 cm^{-1} in normal group shows relatively high content of lipid component in healthy people rather than cancer patients. In addition, the intensity of 1545 and 931 cm^{-1} in normal group is also slightly stronger than those in cancer group. The spectral differences suggest that serum can be used for distinguishing ESCC cancer from healthy people.

Two group classification models were built based on six different multivariate algorithms, which are commonly used for spectral classification, including DT, SVM, PC-LDA, KNN, PLS-DA and EA. The performance for serum spectral classification based on six different

Table 3
Detailed classification performance of six multivariate algorithms for serum and plasma with different anticoagulants. S: serum, E: EDTA, H: heparin, C: citrate.

	Sensitivity(%)				Specificity(%)				Accuracy(%)				AUC(%)			
	S	E	H	C	S	E	H	C	S	E	H	C	S	E	H	C
DT	72.1	79.4	85.3	66.2	80.9	72.1	82.3	80.9	76.5	75.7	83.8	73.5	77.0	77.0	84.0	73.0
PC-LDA	97.1	98.5	100	98.5	91.2	95.6	97.1	92.6	94.9	97.1	98.1	95.6	97.0	97.0	98.0	96.0
SVM	95.6	97.1	94.1	92.6	92.6	83.8	94.1	89.7	94.1	90.4	94.1	91.2	97.0	95.0	97.0	95.0
PLS-DA	94.1	91.2	91.2	86.8	88.2	97.1	92.6	80.9	91.2	94.1	91.9	83.8	93.0	95.0	97.0	90.0
KNN	89.7	85.3	94.1	76.5	75.0	72.1	82.3	86.8	82.4	78.7	88.2	81.6	91.0	79.0	93.0	85.0
EA	80.8	79.4	91.2	83.8	77.9	80.9	85.3	80.9	79.4	80.1	88.2	82.3	88.0	86.0	94.0	89.0

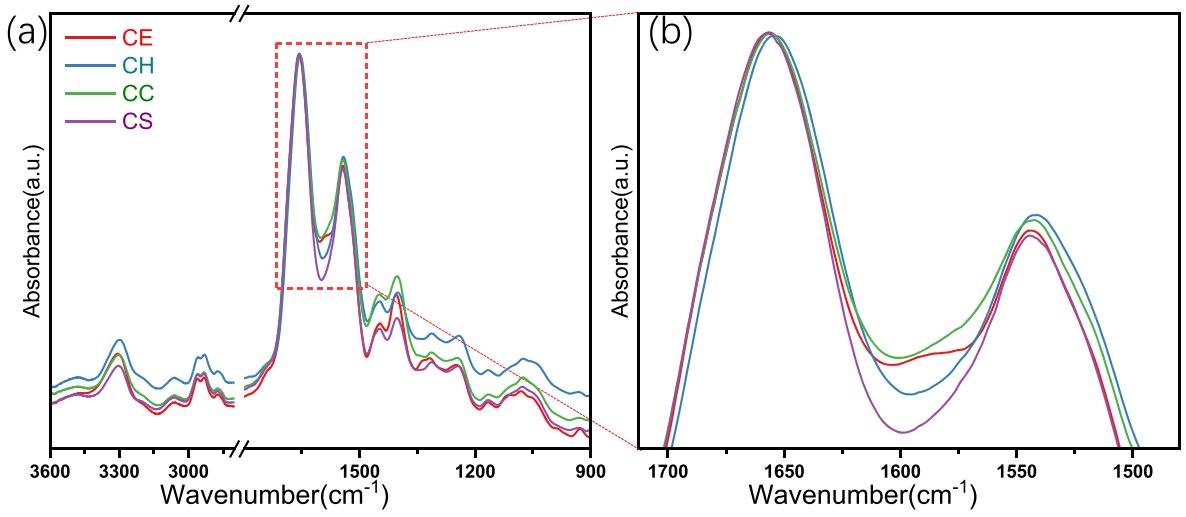

Fig. 2. Mean serum (CS-purple line) and plasma spectra with anticoagulants of citrate (CC-olive line), EDTA (CE-red line) and heparin (CH-dark cyan line) from late stage cancer groups.

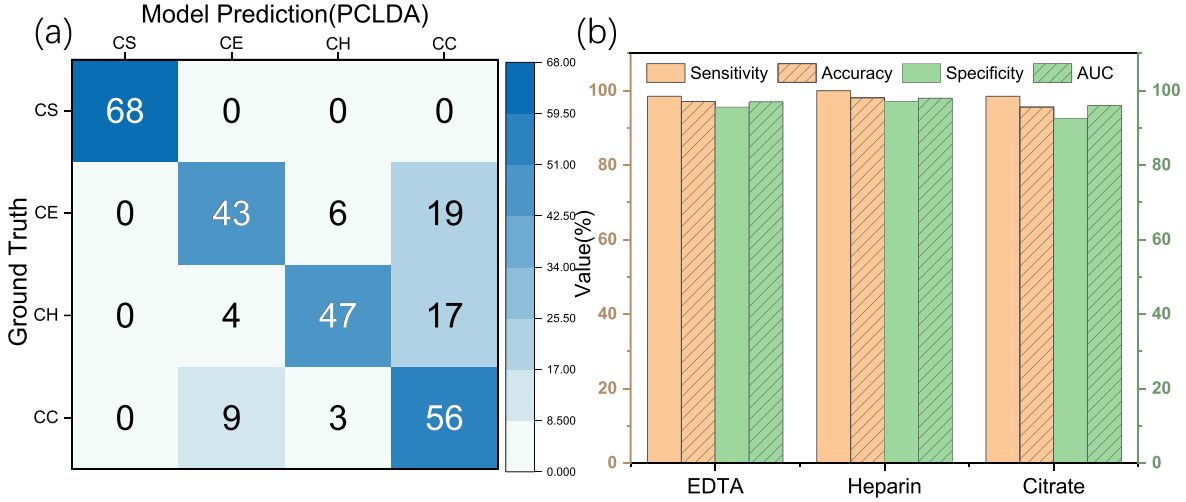

Fig. 3. (a) Confusion matrix of model prediction for spectra of serum (CS) and plasma with anticoagulants of citrate (CC), EDTA (CE) and heparin (CH) from late stage cancer groups based on PC-LDA model. (b) Classified performance based on PC-LDA algorithm for plasma spectra of late cancer group and normal group with anticoagulants of EDTA, heparin and citrate.

multivariate algorithms is shown in Fig 1(b), and the relevant values are further listed in detail in Table 3. Better classification results were achieved using SVM or PC-LDA models, which successfully classified ESCC cancer from normal group with >94.1% accuracy, >95% sensitivity, >90% specificity, and an AUCs >= 97%.

3.2. *Effect of different anticoagulants in plasma for discrimination*

Fig. 2(a) shows the mean serum (S) and plasma spectra with anticoagulants of citrate (C), EDTA (E) and heparin (H) from late-stage cancer groups. The characteristic spectra of anticoagulants were not observed in the plasma spectra, although they had been already found in the anticoagulant tubes in this study. Their infrared characteristic peaks might be submerged by plasma likely because there were fewer

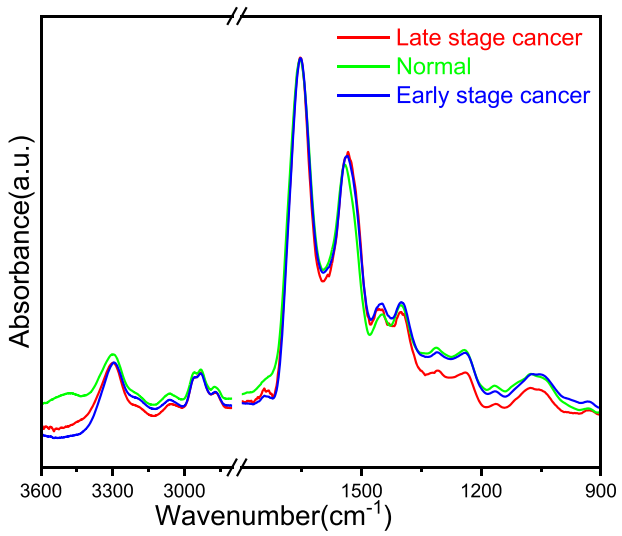

Fig. 4. Mean plasma spectra of late stage cancer group (red line), normal group (green line) and early stage cancer group (blue line) with anticoagulant of heparin.

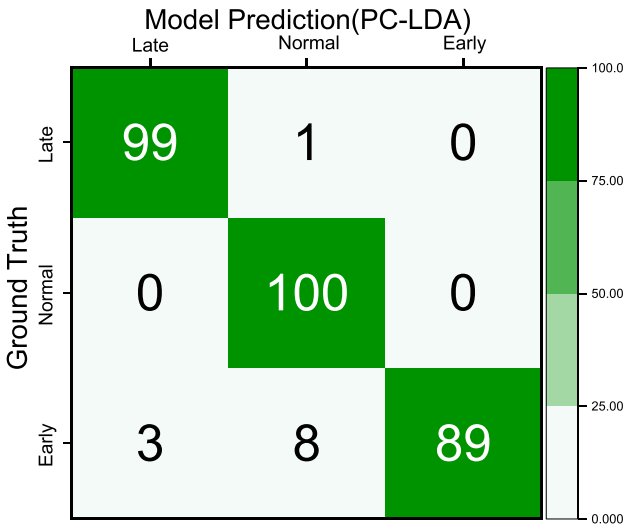

Fig. 6. Confusion matrix of the classification results showing percentages of the accurately classified stages.

anticoagulants in plasma sample after centrifugation. As can be seen from Fig. 2(a), both spectra of serum and plasma display the same waveform, which also indicates that the contribution of anticoagulants in infrared spectra of plasma is rather limited. Areas where significant differences could be found were in the trough region of 1650–1550 cm^{-1}, as shown in Fig. 2(b). Spectra of plasma with citrate and EDTA had a relatively high absorption in this trough region, which may be possibly attributed to the contribution of the strong signals from the asymmetric and symmetric carboxylate groups at 1616 and 1572 cm^{-1}, respectively [14]. However, the spectra of plasma with heparin presented a slightly higher absorption than those of serum, which may presumably result from the fact that heparin has a several signature bands around 1608 cm^{-1} [14].

Four group classification models were built based on DT, PC-LDA, SVM, KNN and EA to distinguish the spectra of serum and plasma with different anticoagulants from late-stage cancer group. The classification result was obtained using PC-LDA algorithm and the confusion matrix of model prediction based on the PC-LDA model is shown in Fig. 3 (a). The confusion matrix illustrates that spectra of serum can be rather well distinguished from spectra of plasma, and that spectra of plasma with different anticoagulants can be finely classified, but not rather well, which is also supported by the classification results using other four models shown in Fig. S2. The above results further illustrate that anticoagulants may lead to the difference of the obtained spectra. However, the effect of different anticoagulants in plasma on discriminating cancer from healthy people needs further exploration.

Plasma of healthy people was obtained using the same batch of anticoagulant tubes, and corresponding mean spectra and standard deviation are shown in Fig. S3. The performance of two group classification models built based on PCLDA algorithm is plotted in Fig. 3(b), while relevant values acquired using six different multivariate algorithms are listed in more details in Table 3. For plasma classification, all four values reached more than 95% using the PC-LDA model, and both SVM and PLS-DA models also displayed a good classification performance. The results show that late-stage cancer can be well classified from heathy people either using serum or plasma with different anticoagulants.

3.3. *Early-stage cancer diagnostics using plasma*

An initial exploratory analysis was performed with PC-LDA to classify the early-stage ESCC cancer, late-stage ESCC cancer and normal group. FTIR spectra with plasma anticoagulant of heparin were employed for constructing the PC-LDA model, and mean spectra of these three groups are shown in Fig. 4. The first model construction step was to build the separate PCA sub-model. Eigenvalues and cumulative amount of variance explained by each PC number were shown in Fig. 5 (a). The eigenvalues of PC1 and PC2 were greater than those

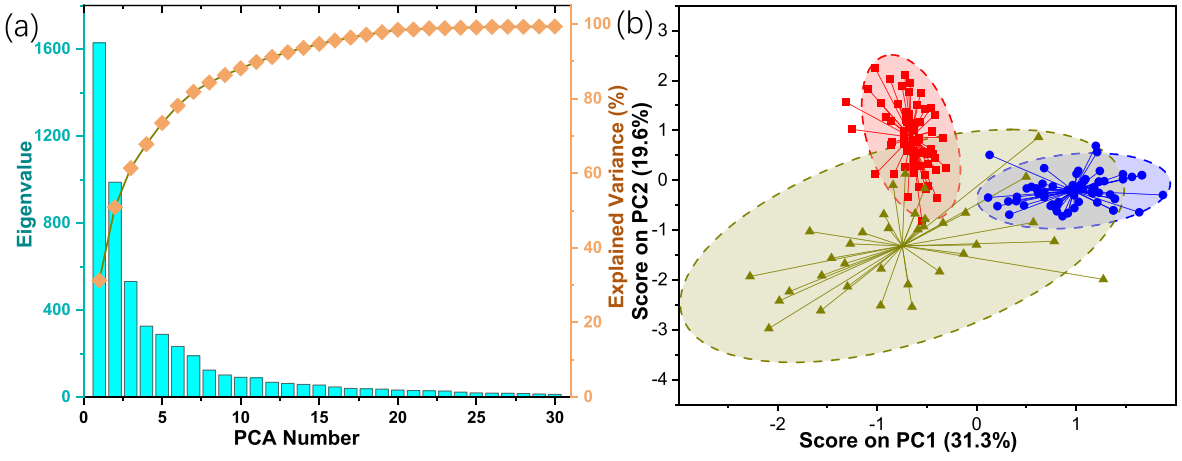

Fig. 5. (a) Eigenvalues and explained variance with PCA numbers. (b) The score values with 95% confidence ellipse for the first two PCs.

corresponding random eigenvalues of other PC numbers (histogram in Fig. 5(a)). The score values with 95% confidence ellipse for the first two PCs are shown in Fig. 5(b), and distinct clusters were observed corresponding to the three different types of FTIR data. With the increase of PCs number, the corresponding eigenvalue first decreased rapidly, and PCs = 15 can be used as an inflection point, accounting for a rather significant proportion (94.71%) of the total explained variance in the data (wireframe in Fig. 5(a)). Therefore, 15 was chosen as the optimal PC number for LDA classification combined with 10-fold cross-validation.

Confusion matrix of the classification results showing percentages of the accurately classified stages are shown in Fig. 6, illustrating that late-stage cancer can be differentiated from normal with an accuracy of 99.26% (135/136), and the sensitivity and specificity are 98.53% (67/68) and 100% (68/68), respectively. Early-stage cancer can be differentiated from normal group with an accuracy of 97.08% (100/103), and the sensitivity and specificity are 91.43% (32/35) and 100% (68/68), respectively. The combined late stage and early-stage cancer group can be differentiated from normal group with an accuracy of 97.66% (167/171), and the sensitivity is 96.12% (99/103).

4. Discussion

ESCC is provided with a high mortality rate in developing counties mainly due to its late diagnosis and lack of common screening method. Developing a rapid, low-cost, and sensitive diagnostic technology to be used in asymptomatic individuals who are exposed to risk factors can significantly improve the chances of cure and survival. In our previous study [48], resonance Raman spectroscopy was used for efficiently detecting early-stage ESCC using blood plasma. As an alternative biofluid spectroscopic technique, FTIR spectroscopy is well suited to analyze blood samples due to its flexible sampling modes and high sensitivity to subtle biological changes [49]. Using serum or plasma for the triage of cancer is considered a rather important link for transitioning FTIR spectroscopy into current clinical pathways since blood serum or plasma tests at the primary care level are already ordered as the standard. However, the potential confounding variables for the diagnostic performance on blood samples with various anticoagulants still need to be further studied.

A previous study has shown that anticoagulants contribute significantly to the spectral profile of plasma/blood when dried, but less when in the aqueous phase [14]. In this study, characteristic peaks related to anticoagulants were not found in the FTIR spectral profile of plasma, and slight waveform difference was found in the region of 1650–1550 cm^{-1} by comparing the spectra of serum and plasma (Fig. 2(b)). Further multivariate classification analysis demonstrates that spectra of serum can be rather well distinguished from spectra of plasma, while spectra of plasma with different anticoagulants cannot be well classified, which means that the waveform of FTIR spectra obtained directly from the anticoagulant tube with no other processing will be affected by different anticoagulants. Even so, the diagnostic accuracy of late-stage cancer from healthy people using the same batch of anticoagulant tubes was still more than 95% for all three anticoagulants combined with the PC-LDA model, illustrating that late-stage cancer can be well classified from healthy people using plasma with different anticoagulants.

In general, there are two common measurement modes when FTIR is applied to biofluid testing, i.e., ATR and transmission mode. It has been shown that both ATR and transmission method are endowed with their own advantages and limitations [50], and both of them have plenty of room for clinical applications. ATR is currently the most favoured method in measurement of serum and plasma samples [41,42], which can produce high reproducible and high quality spectral data, but fails to achieve high-throughput detection unless a multi-ATR crystal device or the microfabricated internal reflection element is created [51]. Maitra's study has demonstrated that excellent efficacy was achieved for discrimination in plasma from patients of oesophageal transformation to adenocarcinoma using the ATR method [40]. This study provides a more robust platform because of the use of large number of blood samples for constructing the model. It assesses the diagnostic performance by serum and plasma with different anticoagulants and enables high-throughput detection using the transmission measurements. Although the raw spectra were slightly influenced by light scattering, refraction and dispersion, more than 170 samples could be well classified with the help of proper spectral processing and the PC-LDA model.

The ultimate research goal is to establish a rapid, sensitive, and low-cost early ESCC cancer detection system based on the employment of blood. As a pilot exploration, the PC-LDA model can differentiate early-stage cancer from normal group with an accuracy of 97.08% and a sensitivity of 91.43%. Currently, the early cancer group is limited in sample size, which may affect the classification performance. It is thus planned to explore FTIR and serum or plasma for early ESCC detecting in the future study by including more early-stage ESCC patients and relatively younger healthy volunteers.

5. Conclusion

In summary, the possibility of detecting ESCC was hereby investigated using customized BaF_2 disk with transmission FTIR spectroscopy, and multivariate analysis method, including DT, SVM, PC-LDA, KNN, PLS-DA and EA, was used for classifying serum FTIR spectra of late-stage cancer and normal group. The results show that both SVM and PC-LDA models provide excellent classification performance, and the diagnostic accuracy, sensitivity, and specificity are better than 94%, 95% and 90%, respectively. Plasma FTIR spectra with anticoagulants of citrate (C), EDTA (E) and heparin (H) from late-stage cancer people and normal group were also obtained. However, the characteristic spectrum of anticoagulants was not observed in plasma FTIR spectra. Four group classification models based on DT, SVM, PC-LDA, KNN and EA illustrate that spectra of serum can be rather well distinguished from spectra of plasma, while spectra of plasma with different anticoagulants are not well classified with each other. Nonetheless, diagnostic accuracy, sensitivity, and specificity of late-stage cancer from healthy people obtained using the same batch of anticoagulant tubes were better than 95%, 98% and 92%, respectively for all three anticoagulants combined with the PC-LDA model. Additionally, a three-group classification model with PC-LDA was built to classify the early stage ESCC cancer, late stage ESCC cancer and normal group. Confusion matrix of the classification results illustrates that early-stage cancer is differentiated from normal group with an accuracy of 97.08%, and the sensitivity and specificity are 91.43% and 100%, respectively. For late-stage cancer, the established model PC-LDA can differentiate ESCC from normal group with an accuracy of 99.26%, a sensitivity of 98.5%, and a specificity of 100%. To conclude, the present study demonstrates that the transmission FTIR method can diagnose ESCC with a superior sensitivity, high specificity and accuracy using either serum or plasma.

CRediT authorship contribution statement

Hongjun Chen: Writing – original draft, Resources, Data curation. **Xianchang Li:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **Shiding Zhang:** Software, Methodology, Formal analysis. **Haijun Yang:** Writing – original draft, Validation, Project administration, Formal analysis. **Qianqian Gao:** Writing – review & editing, Supervision, Investigation. **Fuyou Zhou:** Supervision, Project administration, Methodology, Funding acquisition.

Declaration of Competing Interest

There are no conflicts of interest to declare.

Acknowledgements

This research is supported by the Scientific and Technological Development in Henan Province (No. 52102310015, HNGD2022015), the Key Scientific Research Universities Program Foundation of Henan Province (No. 16A140001), the China Postdoctoral Science Foundation (No. 2016M602262), Major Projects of Science and Technology Department in Anyang City (ZDKJJH2020006), Joint Construction Project of Medical Science and Technology Research Program in Henan Province (LHGJ20200811).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.pdpdt.2022.103177.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249.
[2] G. Wang, G. Jiao, F. Chang, W. Fang, J. Song, N. Lu, D. Lin, Y. Xie, L. Yang, Long-Term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening, Ann Thorac Surg. 77 (2004) 1740–1744.
[3] G. Prasad, T. Wu, D. Wigle, N. Buttar, L. Wongkeesong, K. Dunagan, L. Lutzke, L. Borkenhagen, K. Wang, Endoscopic and surgical treatment of mucosal (T1a) esophageal adenocarcinoma in barrett's esophagus, Gastroenterology 137 (2009) 815–823.
[4] P. Lao-Sirieix, R. Fitzgerald, Screening for oesophageal cancer, Nat. Rev. Clin. Oncol. 9 (2012) 278–287.
[5] A. Kunzmann, U. McMenamin, A. Spence, R. Gray, L. Murray, R. Turkington, H. Coleman, Blood biomarkers for early diagnosis of oesophageal cancer: a systematic review, Eur. J. Gastroen. Hepat. 30 (2018) 263–273.
[6] G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer, Nat. Rev. Clin. Oncol. 14 (2017) 531–548.
[7] E. Heitzer, I. Haque, C. Roberts, M. Speicher, Current and future perspectives of liquid biopsies in genomics-driven oncology, Nat. Rev. Genet. 20 (2019) 71–88.
[8] Y. Zhang, X. Mi, X. Tan, R. Xiang, Recent progress on liquid biopsy analysis using surface-enhanced Raman spectroscopy, Theranostics 9 (2019) 491–525.
[9] K. Durlik-Popińska, P. Żarnowiec, I. Konieczna-Kwinkowska, Ł. Lechowicz, J. Gawęda, W. Kaca, Correlations between autoantibodies and the ATR-FTIR spectra of sera from rheumatoid arthritis patients, Sci. Rep-UK 11 (2021) 17886.
[10] L.G. Silva, A.F.S. Péres, D.L.D. Freitas, C.L.M. Morais, F.L. Martin, J.C.O. Crispim, K.M.G. Lima, ATR-FTIR spectroscopy in blood plasma combined with multivariate analysis to detect HIV infection in pregnant women, Sci. Rep-UK 10 (2020) 20156.
[11] R. Chaber, A. Kowal, P. Jakubczyk, C. Arthur, K. Łach, R. Wojnarowska-Nowak, K. Kusz, I. Zawlik, S. Paszek, J. Cebulsk, A preliminary study of FTIR spectroscopy as a potential Non-Invasive screening tool for pediatric precursor B lymphoblastic leukemia, Molecules 26 (2021) 1174.
[12] J. Depciuch, M. Parlinska-Wojtan, Comparing dried and liquid blood serum samples of depressed patients: an analysis by Raman and infrared spectroscopy methods, J. Pharmaceut. Biomed. 150 (2018) 80–86.
[13] M. Paraskevaidi, C. Morais, K. Lima, J. Snowden, J. Saxon, A. Richardson, M. Jones, D. Mann, D. Allsop, P. M-Hirsch, F.L. Martin, Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood, Proc. Natl. Acad. Sci. 114 (2017), 7929-3798.
[14] M. Martin, D. Perez-Guaita, D. Andrew, J. Richards, B. Wood, P. Heraud, The effect of common anticoagulants in detection and quantification of malaria parasitemia in human red blood cells by ATR-FTIR spectroscopy, Analyst 142 (2017) 1192–1199.
[15] S. Roy, D. Perez-Guaita, D. Andrew, J. Richards, D. McNaughton, P. Heraud, B. Wood, Simultaneous ATR-FTIR based determination of malaria parasitemia glucose and urea in whole blood dried onto a glass slide, Anal. Chem. 89 (2017) 5238–5245.
[16] E. Bernardes-Oliveira, D.L. Dantas de Freitas, C.L. Medeiros de Morais, M. Cornetta, J. Camargo, K.M. Gomes de Lima, J. Crispim, Spectrochemical diferentiation in gestational diabetes mellitus based on attenuated total refection Fourier-transform infrared (ATR-FTIR) spectroscopy and multivariate analysis, Sci Rep-UK 10 (2020) 19259.
[17] F. Elmi, A. Movaghar, M. Elmi, H. Alinezhad, N. Nikbakhsh, Application of FT-IR spectroscopy on breast cancer serum analysis, Spectrochim. Acta A. 187 (2017) 87–91.
[18] R.C. Tomas, A.J. Sayat, A.N. Atienza, J.N. Danganan, M.R. Ramos, A. Fellizar, K... I.. Notarte, L.M. Angeles, R. Bangaoil, A. Santillan, P.M. Albano, Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks, PLoS ONE 17 (2022), e0262489.
[19] U. Zelig, E. Barlev, O. Bar, I. Gross, F. Flomen, S. Mordechai, J. Kapelushnik, I. Nathan, H. Kashtan, N. Wasserberg, O. Givon, Early detection of breast cancer using total biochemical analysis of peripheral blood components: a preliminary study, BMC Cancer 15 (2015) 408.
[20] K.V. Kepesidis, M. Bozic-Iven, M. Huber, N. Abdel-Aziz, S. Kullab, A. Abdelwarith, A.A. Diab, M.A. Ghamdi, M.A. Hilal, M.R.K. Bahadoor, A. Sharma, F. Dabouz, M. Arafah, A.M. Azzeer, F. Krausz, K. Alsaleh, M. Zigman, J.-.M. Nabholtz, Breast-cancer detection using blood-based infrared molecular fngerprints, BMC Cancer 21 (2021) 1287.
[21] Z. Guleken, H. Bulut, G.I. Gültekin, S. Arıkan, I. Yaylım, M.T. Hakan, D. Sonmez, N. Tarhan, J. Depciuch, Assessment of structural protein expression by FTIR and biochemical assays as biomarkers of metabolites response in gastric and colon cancer, Talanta 231 (2021), 122353.
[22] X. Yang, Q. Ou, W. Yang, Y. Shi, G. Liu, Diagnosis of liver cancer by FTIR spectra of serum, Spectrochim. Acta A 263 (2021), 120181.
[23] P. Giamougiannis, C.L.M. Morais, B. Rodriguez, N.J. Wood, P.L. Martin-Hirsch, F. L. Martin, Detection of ovarian cancer (± neo-adjuvant chemotherapy effects) via ATR-FTIR spectroscopy: comparative analysis of blood and urine biofluids in a large patient cohort, Anal. Bioanal. Chem. 413 (2021) 5095–5107.
[24] D. Yonar, M. Severcan, R. Gurbanov, A. Sandal, U. Yilmaz, S. Emri, F. Severcan, Rapid diagnosis of malignant pleural mesothelioma and its discrimination from lung cancer and benign exudative effusions using blood serum, BBA-Mol, Basis Dis. 1868 (2022), 166473.
[25] D. Mabwa, K. Gajjar, D. Furniss, R.R. Schiemer, R. Crane, C. Fallaize, P.L. Martin-Hirsch, F.L. Martin, T. Kypraios, A.B. Seddon, S. Phang, Mid-infrared spectral classification of endometrial cancer compared to benign controls in serum or plasma samples, Analyst 46 (2021) 5631–5642.
[26] J. Ollesch, S. Drees, H. Heise, T. Behrens, T. Brüning, K. Gerwert, FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification, Analyst 138 (2013) 4092–4102.
[27] J. Ollesch, M. Heinze, H. Heise, T. Behrens, T. Bruning, K. Gerwert, It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy, J. Biophotonics. 7 (2014) 210–221.
[28] J. Hands, K. Dorling, P. Abel, K. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. Jenkinson, R. Lea, C. Walker, M.J. Baker, Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples, J. Biophotonics. 7 (2014) 189–199.
[29] H.J. Butler, B.R. Smith, R. Fritzsch, P. Radhakrishnan, D.S. Palmer, M.J. Baker, Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy, Analyst 14 (2018) 6121–6134.
[30] V. Rai, R. Mukherjee, A. Routray, A. Ghosh, S. Roy, B. Ghosh, P. Mandal, S. Bose, C. Chakraborty, Serum-based diagnostic prediction of oral submucous fibrosis using FTIR spectrometry, Spectrochim. Acta A 189 (2018) 322–329.
[31] J. Titus, H. Ghimire, E. Viennois, D. Merlin, A. Unil Perera, Protein secondary structure analysis of dried blood serum using infrared spectroscopy to identify markers for colitis screening, J. Biophotonics. 11 (2018) 1–8.
[32] D. Maziak, M. Do, F. Shamji, S. Sundaresan, D. Perkins, P. Wong, Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: an exploratory study, Cancer Detect. Prev. 31 (2007) 244–253.
[33] R. Zhao, L. Quaroni, A. Casson, Fourier transform infrared (FTIR) spectromicroscopic characterization of stem-like cell populations in human esophageal normal and adenocarcinoma cell lines, Analyst 135 (2010) 53–61.
[34] J. Ingham, M. Pilling, D. Martin, C. Smith, B. Ellis, C. Whitley, M. Siggel-King, P. Harrison, T. Craig, A. Varro, D. Pritchard, A. Varga, P. Gardner, P. Weightman, S. Barrett, A novel FTIR analysis method for rapid high-confidence discrimination of esophageal cancer, Infrared Phys. Techn. 102 (2019), 103007.
[35] T. Nguyen, A. Maguire, C. Mooney, N. Jackson, N. Lynam-Lennon, V. Weldon, C. Muldoon, A. Maguire, D. O'Toole, N. Ravi, J. Reynolds, J. O'Sullivan, A. Meade, Prediction of pathological response to neo-adjuvant chemoradiotherapy for oesophageal cancer using vibrational spectroscopy, Translational Biophotonics 3 (2021), e202000014.
[36] O. Old, G. Lloyd, J. Nallala, M. Isabelle, L. Almond, N. Shepherd, C. Kendall, A. Shore, H. Barr, N. Stone, Rapid infrared mapping for highly accurate automated histology in Barrett's oesophagus, Analyst 142 (2017) 1227–1234.
[37] L. Quaroni, A. Casson, Characterization of Barrett esophagus and esophageal adenocarcinoma by Fourier-transform infrared microscopy, Analyst 134 (2009) 1240–1246.
[38] T. Wang, G. Triadafilopoulos, J. Crawford, L. Dixon, T. Bhandari, P. Sahbaie, S. Friedland, R. Soetikno, C. Contag, Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy, Proc. Natl. Acad. Sci. 104 (2007) 15864–15869.
[39] O. Old, G. Lloyd, M. Isabelle, L. Almond, C. Kendall, K. Baxter, N. Shepherd, A. Shore, N. Stone, H. Barr, Automated cytological detection of Barrett's neoplasia with infrared spectroscopy, J. Gastroenterol. 53 (2018) 227–235.
[40] I. Maitra, C. Morais, K. Lima, K.R. Ashton, F.L. Martin, Attenuated total reflection Fourier-transform infrared spectral discrimination in human bodily fluids of oesophageal transformation to adenocarcinoma, Analyst 144 (2019) 7447–7456.
[41] M. Paraskevaidi, P. Martin-Hirsch, F.L. Martin, ATR-FTIR spectroscopy tools for medical diagnosis and disease. Investigation Nanotechnology Characterization Tools For Biosensing And Medical Diagnosis, 2018 chapter 4.
[42] A. Sala, D. Anderson, P. Brennan, H. Butler, J. Cameron, M. Jenkinson, C. Rinaldi, A. Theakstone, M.J. Baker, Biofluid diagnostics by FTIR spectroscopy: a platform technology for cancer detection, Cancer Lett. 447 (2020) 122–130.
[43] R. Gao, B. Yang, C. Chen, F. Chen, C. Chen, D. Zhao, X. Lv, Recognition of chronic renal failure based on Raman spectroscopy and convolutional neural network, Photodiagn. Photodyn. 34 (2021), 102331.
[44] C. Ma, Z. Yan, J. Mo, W. Han, X. Lv, C. Chen, C. Chen, X. Nie, Rapid identification of benign and malignant pancreatic tumors using serum Raman spectroscopy combined with classification algorithms, Optik (Stuttg) 208 (2020), 164473.

[45] X. Xie, C. Chen, T. Sun, G. Mamati, X. Wan, W. Zhang, R. Gao, F. Chen, W. Wu, Y. Fan, X. Lv, G. Wu, Rapid, non-invasive screening of keratitis based on Raman spectroscopy combined with multivariate statistical analysis, Photodiagn. Photodyn. 31 (2020), 101932.
[46] C. Chen, J. Wang, C. Chen, J. Tang, X. Lv, C. Ma, Rapid and efficient screening of human papillomavirus by Raman spectroscopy based on GA-SVM, Optik (Stuttg) 210 (2020), 164514.
[47] H. Leng, C. Chen, C. Chen, F. Chen, Z. Du, J. Chen, B. Yang, E. Zuo, M. Xiao, X. Lv, P. Liu, Raman spectroscopy and FTIR spectroscopy fusion technology combined with deep learning: a novel cancer prediction method, Spectrochim. Acta A. 285 (2023), 121839.
[48] X. Li, H. Chen, S. Zhang, H. Yang, S. Gao, H. Xu, L. Wang, R. Xu, F. Zhou, J. Hu, J. Zhao, H. Zeng, Blood plasma resonance Raman spectroscopy combined with multivariate analysis for esophageal cancer detection, J. Biophotonics. 14 (2021), e202100010.
[49] A. Theakstone, C. Rinaldi, H. Butler, J. Cameron, L. Confield, S. Rutherford, A. Sala, S. Sangamnerkar, M.J. Bake, Fourier-transform infrared spectroscopy of biofluids: a practical approach, Translational Biophotonics 3 (2021), e202000025.
[50] C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N.W. Clarke, P. Gardner, Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum, J. Biophotonics. 7 (2014) 180–188.
[51] H. Butler, P. Brennan, J. Cameron, D. Finlayson, M. Hegarty, M. Jenkinson, D. Palmer, B. Smith, M.J. Baker, Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer, Nat. Commun. 10 (2019) 4501–4509.