scientific reports

Check for updates

OPEN

Multimodal separation and cross fusion network based on Raman spectroscopy and FTIR spectroscopy for diagnosis of thyroid malignant tumor metastasis

Haitao Song[1,2,6], Xuguang Zhou[3,6], Cheng Chen[3,4,5], Chao Dong[1,2], Yuyang He[3], Mingtao Wu[3], Jun Yu[3], Xiangnan Chen[3], Yanpeng Li[3] & Binlin Ma[1,2]✉

The diagnosis of cervical lymph node metastasis from thyroid cancer is an essential stage in the progression of thyroid cancer. The metastasis of cervical lymph nodes directly affects the prognosis and survival rate of patients. Therefore, timely and early diagnosis is crucial for effective treatment and can significantly improve patients' survival rate and quality of life. Traditional diagnostic methods, such as ultrasonography and radionuclide scanning, have limitations, such as complex operations and high missed diagnosis rates. Raman spectroscopy and FTIR spectroscopy can well reflect the molecular information of samples, have characteristics such as sensitivity and specificity, and are simple to operate. They have been widely used in clinical research in recent years. With the development of intelligent medical diagnosis technology, medical data shows a multi-modal trend. Compared with single-modal data, multi-modal data fusion can achieve complementary information, provide more comprehensive and valuable diagnostic information, significantly enhance the richness of data features, and improve the modeling effect of the model, helping to achieve better results. Accurate disease diagnosis. Existing research mostly uses cascade processing, ignoring the important correlations between multi-modal data, and at the same time not making full use of the intra-modal relationships that are also beneficial to prediction. We developed a new multi-modal separation cross-fusion network (MSCNet) based on deep learning technology. This network fully captures the complementary information between and within modalities through the feature separation module and feature cross-fusion module and effectively integrates Raman spectrum and FTIR spectrum data to diagnose thyroid cancer cervical lymph node metastasis accurately. The test results on the serum vibrational spectrum data set of 99 cases of cervical lymph node metastasis showed that the accuracy and AUC of a single Raman spectrum reached 63.63% and 63.78% respectively, and the accuracy and AUC of a single FTIR spectrum reached 95.84% respectively and 96%. The accuracy and AUC of Raman spectroscopy combined with FTIR spectroscopy reached 97.95% and 98% respectively, which is better than existing diagnostic technology. The omics correlation verification obtained correlation pairs of 5 Raman frequency shifts and 84 infrared spectral bands. This study provides new ideas and methods for the early diagnosis of cervical lymph node metastasis of thyroid cancer.

Keywords Raman spectroscopy, FTIR spectroscopy, Feature separation, Cross fusion, Multimodal fusion, Diagnosis of thyroid malignancy

[1]Department of Breast and Thyroid Surgery, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi 830017, Xinjiang, China. [2]The Clinical Medical Research Center of Breast and Thyroid Tumor in Xinjiang, Urumqi 830017, Xinjiang, China. [3]College of Software, Xinjiang University, Urumqi 830046, China. [4]Department of Cardiology, People's Hospital of Xinjiang Uyghur Autonomous Region, Xinjiang, China. [5]Xinjiang Key Laboratory of Cardiovascular Homeostasis and Regeneration Research, Xinjiang, China. [6]These authors contributed equally: Haitao Song and Xuguang Zhou. ✉email: mbldoctor@126.com

In recent years, the incidence of thyroid cancer has increased significantly worldwide, and the problem of cervical lymph node metastasis has attracted increasing attention from researchers. Cervical lymph node metastasis of thyroid cancer is a serious and heterogeneous stage in disease progression. Cancer cells spread to cervical lymph nodes through the blood or lymph system, which is one of the common forms of metastasis of thyroid cancer[1]. Cervical lymph node metastasis not only causes tissue damage in the neck region but may also spread to other distant organs, such as the lungs and bones, thereby significantly increasing the risk of disease recurrence and death[2]. Statistics show that the incidence of central lymph node metastasis is approximately 50%[3]. Cervical lymph node metastasis is not only a critical factor in evaluating the efficacy of thyroid cancer treatment,but also an important indicator affecting patient prognosis. Although the advancement of modern medicine has improved the treatment effect, the five-year survival rate of patients with cervical lymph node metastasis is still significantly lower than that of patients without metastasis, and their survival rate may drop to less than 70%. In comparison, the five-year survival rate of patients without metastasis is still significantly lower than that of patients without metastasis. The annual survival rate can be as high as over 95%. This highlights the importance of early diagnosis and effective management to reduce the occurrence of this serious complication.

Because cervical lymph node metastasis of thyroid cancer is usually small and hidden in the early stage, and the symptoms are often not obvious, it is difficult to locate and detect in time[4] accurately. Although traditional medical diagnostic methods (such as ultrasound, CT, MRI, etc.) play a certain role in diagnosing cervical lymph node metastasis of thyroid cancer, they have limitations in detecting micro-metastasis, and the diagnostic accuracy is usually between 60 and 80%. %, and the misdiagnosis rate may be as high as 20%-40%[5,6]. Take ultrasonic examination[7] and radionuclide scanning commonly used in clinical practice as examples. Although the ultrasonic examination is a commonly used non-invasive method, it has high operational and technical requirements. Detecting some microscopic lesions or special types of metastasis may be difficult.This leads to an increase in the rate of missed diagnosis; while radionuclide scans (such as whole-body thyroid scans) assess thyroid metastasis by ingesting and accumulating radioactive iodine or other radioactive isotopes, but are prone to interference from a variety of factors, affecting accuracy. Current clinical diagnostic methods have certain limitations, including low accuracy and high misdiagnosis rates.

The continuous development of Raman spectroscopy and FTIR spectroscopy technology has been widely used in the biomedical field to diagnose diseases such as lung cancer, glioma, and breast cancer[8,9]. They can reflect the molecular differences of samples from different angles.Compared with traditional detection methods, infrared and Raman spectroscopy do not require tissue sampling, significantly reducing the invasive operation on patients. In addition, it can detect trace molecular changes, including biomarkers in the early stages of cancer, with high sensitivity and specificity. In addition, using Raman and infrared spectroscopy technology also has the advantages of being fast and efficient and obtaining results quickly[10]. In 2021, Wang et al. proposed a diagnostic method based on serum Raman spectroscopy combined with multiple algorithms to diagnose abnormal thyroid function and chronic renal failure[11]. However, this study only used a single vibration spectrum. A single vibration spectrum only provides a data source and lacks the ability to conduct comprehensive analysis with other related data, which is susceptible to many limitations. In 2022, Song et al. proposed a method for quickly distinguishing thyroid papillary carcinoma and micropapillary carcinoma based on serum Raman spectroscopy combined with a machine learning model[12]. Although the article proposes a novel non-invasive, low-cost, and rapid diagnostic method, it also only uses a single vibration spectrum of serum Raman spectroscopy data for diagnosis. Although Raman spectroscopy technology has the advantages of being non-destructive, low-cost, and fast, compared with multimodal data fusion, a single spectral modality cannot fully utilize the complementary information between different modalities, and the information obtained is not accurate and comprehensive enough, thereby reducing the accuracy of diagnosis.

Reliable multimodal data is conducive to improving the diagnostic efficiency of diseases. Since different vibrational spectroscopy technologies or modalities have unique characteristics and information, information complementarity and enrichment can be achieved by fusing these data to obtain more comprehensive feature representations from different modalities, thereby improving the accuracy of diagnosis and prediction.In recent years, the fusion of multimodal vibrational spectra has developed rapidly in the medical field. For example, in 2020, Du et al. used multimodal fusion technology, combined with Raman spectroscopy and Fourier transform infrared spectroscopy, and adopted pattern recognition algorithms to diagnose thyroid dysfunction serum. Compared with a single spectrum, this method obtains more spectral features and significantly improves the detection accuracy[13]. In 2022, Chen et al. fused Fourier transform infrared spectroscopy, Raman spectroscopy and its first-order derivative spectra to establish a rapid diagnosis method for renal cancer to enhance the accuracy and efficiency of diagnosis. This spectral fusion method provides a new method for the early diagnosis of renal cancer[14]. Wu et al. also used Fourier transform infrared spectroscopy and deep learning algorithms to screen autoimmune diseases quickly[15]. This study also used multimodal spectral fusion to provide a rapid and effective method for the early diagnosis of autoimmune diseases. In 2023, Leng et al. proposed a new cancer diagnosis technology combining multispectral fusion and deep learning based on the limited molecular information reflected by a single spectrum and the complementarity of FTIR spectroscopy and Raman spectroscopy, and achieved certain results[16]. In 2024, Huang et al. proposed a dual-branch residual attention network to fuse SERS and original Raman spectral data.By using the information complementarity of the two spectral data entirely, the diagnostic accuracy of systemic lupus erythematosus was improved, providing new possibilities for applying vibrational spectroscopy in clinical diagnosis[17].Xin et al. combined Raman and Fourier transform infrared spectroscopy based on multimodal fusion vibrational spectroscopy technology to identify the ingredients of Uyghur medicines. This method improved the accuracy of ingredient identification, and the study highlighted the effectiveness of multimodal fusion in vibrational spectroscopy[18]. Zhang et al. also achieved good results using machine learning algorithms to estimate wheat protein and wet gluten content based on multimodal fusion of hyperspectral data and other modal data[19]. Zhou et al. collected Raman spectra, FTIR spectra and

metabolomics data of 119 samples of systemic lupus erythematosus, lupus nephritis and healthy control groups. They diagnosed systemic lupus erythematosus after fusing the three modal data. The best results were achieved in all indicators. The experimental results show that multi-omics medical data has great potential in disease diagnosis[20].

The above studies have achieved good results and further proved the feasibility of multimodal vibrational spectral fusion. By making full use of spectral techniques of different modalities, these studies have demonstrated the advantages of spectral data fusion and improved the ability to solve complex problems. However, existing studies mostly use cascade processing, ignoring the important correlation between multimodal data, and not making full use of the intra-modal relationship that is also beneficial to prediction. Therefore, this study proposes a deep learning method that can perform dynamic modeling of inter-modal and intra-modal data well, achieving excellent performance in predicting lymph node metastasis of thyroid malignant tumors.

In this paper, we explore the feasibility of using Raman spectroscopy and FTIR spectroscopy to predict lymph node metastasis of thyroid malignant tumors and propose a new multimodal model framework called multimodal separation cross fusion network (MSCNet). The framework consists of four modules: spectral fully connected neural network encoder module, multimodal feature separation module, feature cross fusion network module and malignant tumor prediction module. In the multimodal separation cross-fusion network, we first extract and reduce the original Raman spectrum and FTIR spectrum through the fully connected neural network encoder module to obtain a feature space of unified dimension. Then, through the multimodal feature separation module, we separate and interact the differences between the two spectral data in the feature space to obtain their complementary information on molecular properties. Next, the feature cross-fusion network module constructs multimodal fusion features by further modeling the separated features' intra-modal and inter-modal dynamics to improve the information's comprehensiveness and accuracy. Finally, we input the fused features into the bidirectional long short-term memory network for malignant tumor prediction, and use the sequential decision reconstruction method to obtain more sufficient decision information. The main contributions of this paper can be summarized as follows:

1. A new multimodal model framework is proposed for malignant tumor prediction. It integrates the information of Raman spectroscopy and FTIR spectroscopy to more comprehensively understand molecules' properties and characteristics. It improves the accuracy of thyroid malignant tumor prediction and brings new ideas to the field of multimodal medical information fusion.
2. Feature separation modules for different spectral modalities are designed to extract and utilize their complementary information on molecular properties effectively. The model can more accurately capture the differences and connections between the two spectral data through separation and interaction, providing a basis for subsequent information fusion.
3. A feature cross-fusion network module was developed, which realized information fusion within and between modalities through dynamic modeling and made full use of the interactive information between modalities. Thus, the model could more comprehensively understand the association between spectral data, thereby improving the accuracy and robustness of prediction.
4. A prediction module based on BiLSTM was proposed. This module uses sequential decision reconstruction to fully utilize the fused features for the final prediction of malignant tumors. This module uses time series information and sequential decision reconstruction to improve the accuracy and reliability of prediction.

Materials and methods

Sample collection

The study was approved by the Affiliated Tumor Hospital of Xinjiang Medical University to obtain the Raman spectroscopy serum and infrared spectroscopy serum samples required for the experiment. Among them, 99 patients with thyroid and cervical lymph node malignant tumors contained 49 malignant tumor metastasis and 50 malignant tumor non-metastasis samples. The importance of predicting metastasis in malignant tumor detection was further emphasized. Each subject will collect Raman spectroscopy data and FTIR spectroscopy data to ensure the consistency of the final multimodal data. The total number of experimental samples is 198. All subjects participating in the study were informed of the purpose of the experiment and signed the "Informed Consent Form for Sample Retention in the Affiliated Tumor Hospital of Xinjiang Medical University". The ethical approval of the Ethics Review Committee of the Affiliated Tumor Hospital of Xinjiang Medical University has been obtained (2024BC003). We obtained confirmation from the committee that approved the study that all studies were conducted in accordance with relevant regulations and informed consent had been obtained from all participants. All serum samples were taken from fresh blood samples without any anticoagulant. 3 mL of fresh blood was collected from each sample. Then, the blood samples were centrifuged at 4000 r/min at 4 °C, and the top transparent liquid was extracted to obtain serum. The serum was separated into centrifuge tubes and stored in a -80 °C refrigerator for sample testing. The serum was collected in the morning and the subjects were fasted for 12 h to avoid interference of diet and other factors on serum components. Specific sample information is shown in Table 1.

Raman spectrum acquisition

The serum Raman spectrum was obtained by a high-resolution confocal Raman spectrometer (LabRAM HR Evolution, gora Raman spectroscopy, ideal optics, China), with a spectral range of 500–2000 cm^{-1}, an excitation wavelength of 785 nm, an integration time of 15 s, a laser power of 160 mW, and a continuous acquisition mode. The Raman spectral data acquisition grating was 600 lines and a blaze angle of 800 nm. The laser beam was focused on the sample surface through a 10 X lens. When measuring the sample, take it out of the refrigerator and thaw it naturally. After waiting for ten minutes, shake it well, and use a 10ul pipette to draw the serum and

Dataset	Spectral type	Acquisition Instruments	Wavelength range	Data dimension	Type of disease	Number of samples
Cervical lymph node metastasis dataset of thyroid cancer	Raman spectra	LabRAM HR Evolution, gora Raman spectroscopy, ideaoptics, China	500–2000 cm^{-1}	312	Tumor metastasis	49
					Tumor not metastasis	50
	IR spectra	FTIR-850,Gangdong Scientific, China	700–4000 cm^{-1}	1654	Tumor metastasis	49
					Tumor not metastasis	50

Table 1. Details of the thyroid cancer dataset with cervical lymph node metastasis.

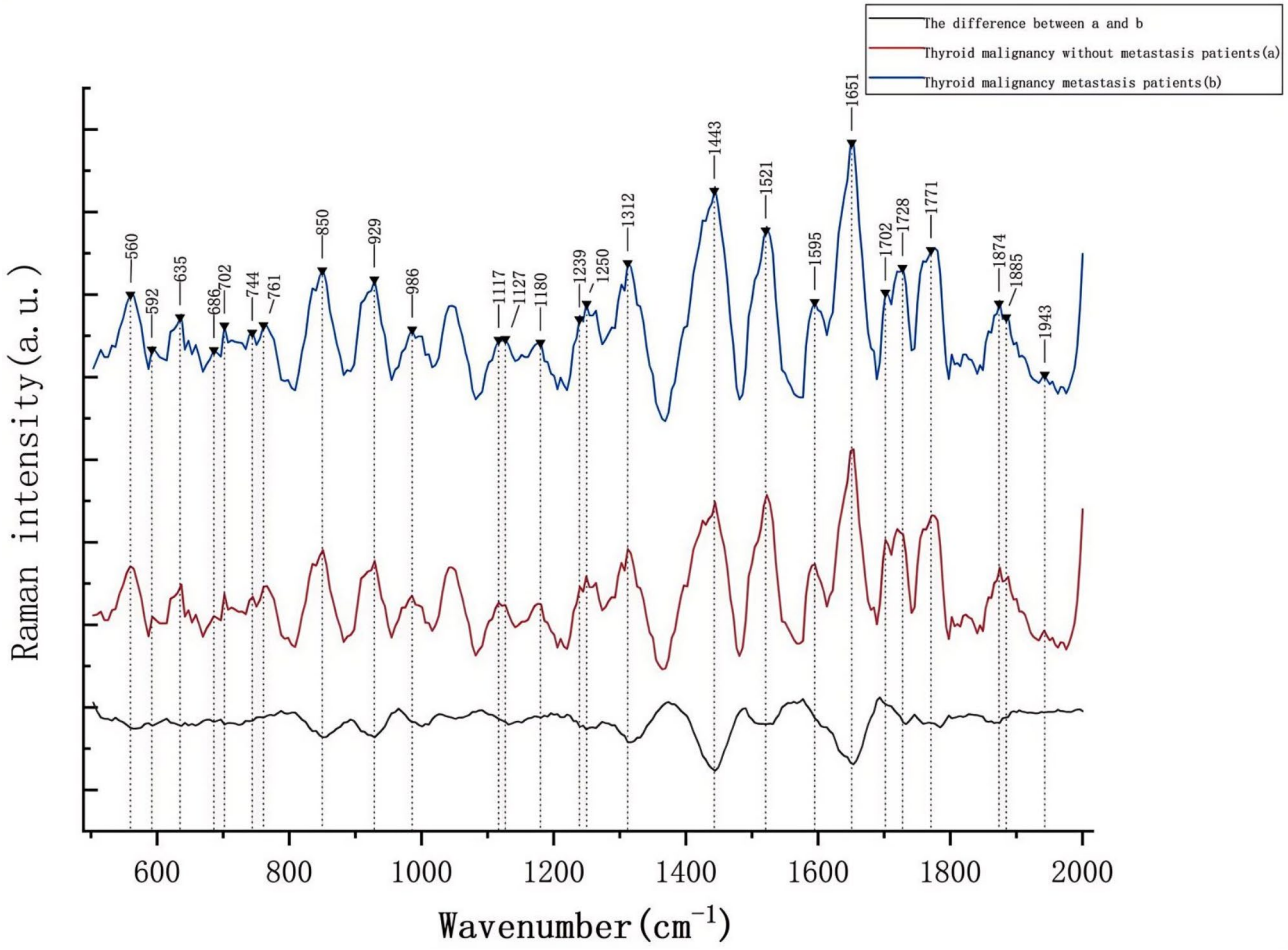


Fig. 1. Raw Raman spectra after preprocessing.

drop it on the tin foil substrate. Wait for two minutes to start the test. To avoid errors in the spectral acquisition process, record 3 spectra of each sample at different positions, and obtain the average spectrum of each sample for further analysis. Finally, a total of 99 serum Raman spectra of the two samples were collected. We used airPLS to subtract the fluorescence background and used the smoothing algorithm to remove noise. The preprocessed Raman spectrum is shown in Fig. 1, which contains the average spectra of the two classes and the difference spectra between the classes.

FTIR spectrum acquisition

The serum infrared spectrometer was a Fourier transform infrared spectrometer (FTIR-850, Gangdong Scientific, China), with a spectral wavelength range of 700–4000 cm^{-1}, a resolution of 4 cm^{-1}, and 32 scans. The spectra were obtained under constant temperature and air humidity. Before measurement, the samples were thawed at room temperature for 15 min. Then, 5 μl of the sample was dropped on the lens, placed on the spectrometer, and left to dry for 8 min.After each measurement, it was cleaned with alcohol. Although the spectrum was collected in a relatively closed environment, it was still affected by the air content. Studies have shown[21] that the band around 2350 cm^{-1} represents CO_2-related substances. In order to reduce the interference of air background on

the spectrum, the characteristics of the 2283–2397 cm^{-1} band were deleted in the experiment. In order to avoid errors in the spectrum acquisition process, 3 spectra were recorded at different positions of each sample, and the average spectrum of each sample was obtained for further analysis. Finally, a total of 99 serum infrared spectra of the two samples were collected. After stripping off the CO_2 interference peak, we subtracted the fluorescence background by using airPLS. The preprocessed infrared spectrum is shown in Fig. 2, which contains the average spectrum of the two categories and the difference spectrum between the categories.

Spectral analysis

Figure 1 shows the standardized Raman spectra of thyroid patients with and without malignant tumor metastasis from 500 to 2000 cm^{-1}. The Raman spectra of sera from patients with and without malignant tumor metastasis are different mainly at 850, 929, and 1443 cm^{-1}. There are significant differences in Raman peaks, which are related to whether the malignant tumor has metastasized. Table 2 lists some biological molecules corresponding to the Raman characteristic peaks. The band at 635 cm^{-1} was confirmed to be the C-S twisted amino acid methionine. Due to the substance methionine, tumor cells meet their needs for rapid proliferation through metabolic reprogramming, and methionine metabolism plays a key role in this process. Tumor cells promote cell growth and proliferation by increasing the uptake and utilization of methionine. Malignant tumor cells absorb and utilize a large amount of methionine from the tumor microenvironment through the methionine transporter SLC3A2, resulting in a lack of methionine utilization in T cells, which in turn affects the function and killing ability of T cells and initiates immune escape[22]. The peak at 702 cm^{-1} is attributed to cholesterol. Studies have shown that, in some cases, saturated fats may also be related to the occurrence and development of tumors. The peak at 929 cm^{-1} is mainly caused by the stretching of C–C bonds in proteins. It is attributed to proteins. In malignant tumor cells, the structure and function of proteins may change, which may lead to changes in the vibration mode of C–C bond stretching[23]. The band at 1117 cm^{-1} is confirmed to be C–C bond stretching, which is attributed to glucose. Malignant tumors usually increase energy through glycolysis during metastasis. The lower the content of this peak in thyroid malignant tumor metastasis, the more energy is consumed during metastasis. The Raman peak at 1239 cm^{-1} is generated by carbon–nitrogen stretching and nitrogen–hydrogen bond coupling, which can be mixed with the vibration of the side chain. It is attributed to amide III. Amide III usually refers to a structure in the secondary structure of proteins, also known as β-turn. It is a curved structure in the protein chain that helps maintain the spatial configuration of proteins[24]. The Raman peak at

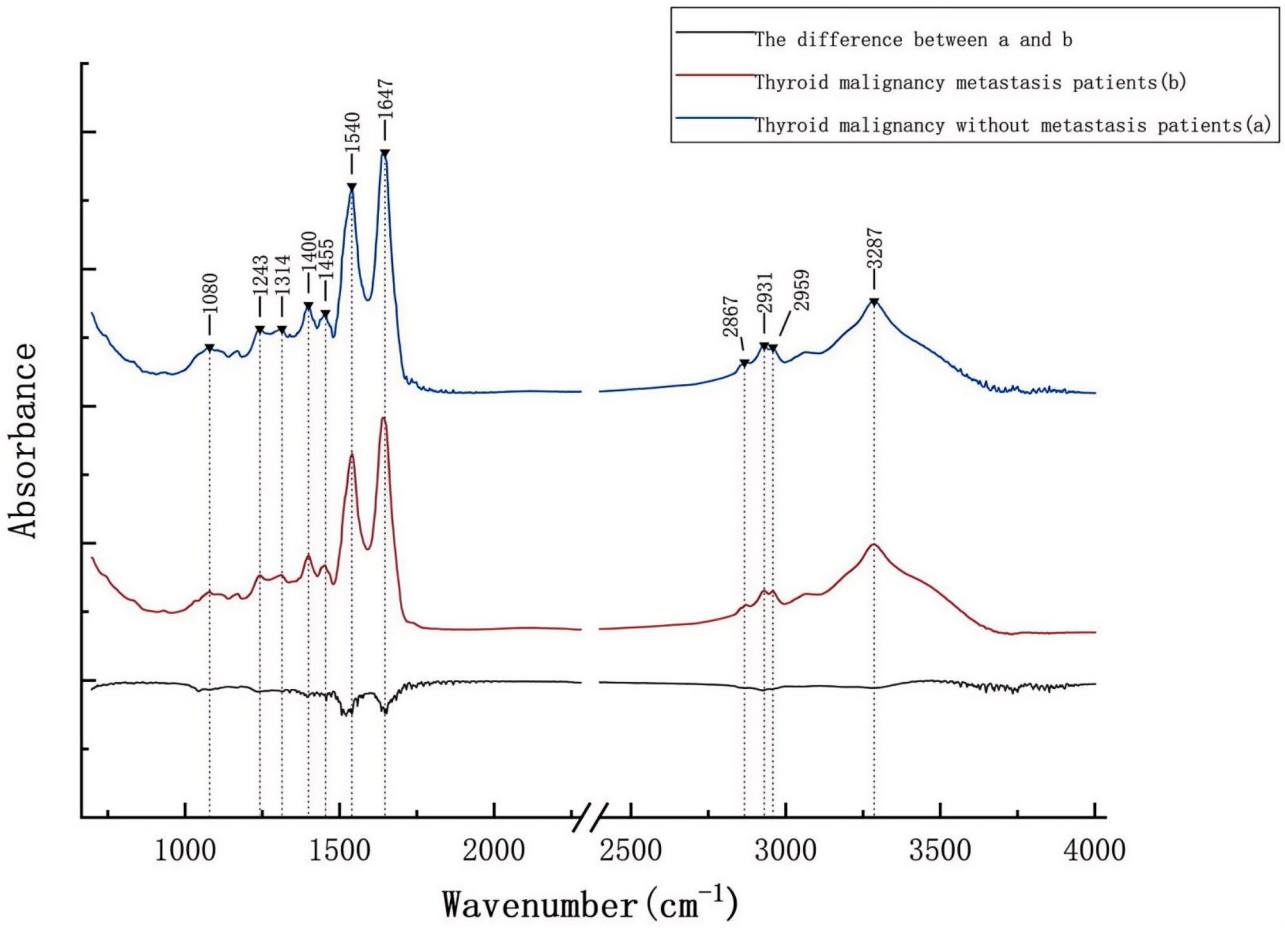


Fig. 2. Raw FTIR spectra after preprocessing.

Peak position(cm^{-1})	Initial attribution
635[25]	ν(C-S) gauche (amino acid methionine)
702[26]	Cholesterol, cholesterol ester
850[27]	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides Tyrosine (Fermi resonance of ring fundamental and overtone)
929[28]	ν(C–C), stretching; probably in amino acids proline and valine (protein band)
1117[26]	Glucose
1127[29]	ν(C-N)
1239[30]	Amide III
1443[31]	CH_2 deformation (lipids and proteins) Triglycerides (fatty acids)
1521[32]	Carotene
1702[25]	Carbonyl stretch (C-O), amide I

Table 2. Main peak positions and preliminary attribution of Raman spectra in 99 patients with and without thyroid metastasis.

Peak position(cm^{-1})	Initial attribution
1080	DNA,RNA(C-O)
1243	$\nu_{as}(PO^-_2)$ of nucleic acid
1314	Amide III
1400	ν(COO-) of amino acid
1455	CH_3
1540	(N–H) of amide II band
1647	Amide I
2867	CH_2
2931	CH_2
2959	CH_3

Table 3. The main peak positions and preliminary attribution of IR spectra in 99 patients with and without thyroid metastasis. The following information is from LibreTexts and is shared under a CC BY-NC-SA 4.0 license. From LibreTexts. (n.d.). Infrared Spectroscopy Absorption Table. Retrieved from https://chem.libretexts.org/@go/page/22645.

1443 cm^{-1} is generated by CH2 deformation, corresponding to lipids and amino acids. Studies have shown that most malignant tumor cells affect lipid metabolism, and elevated serum cholesterol levels are associated with cancer risk. The peak at 1521 cm^{-1} belongs to carotene, which has anti-cancer effects and is slightly higher in patients with thyroid malignant tumors without metastasis. The differences in amino acids, lipids, proteins and other components in the serum of thyroid patients can be reflected in the changes in spectral peak intensity. These differences provide a basis for using serum Raman spectroscopy to diagnose whether patients with thyroid malignant tumors have metastasized.

Figure 2 shows the standardized infrared spectra of thyroid patients with and without malignant tumor metastasis from 500 to 4000 cm^{-1}. There are common infrared peaks at 1080, 1243, 1314, 1400, 1455, 1540, 1647, 2867, 2931, and 2959 cm^{-1}. The infrared spectra of serum from patients with malignant tumor metastasis and those without metastasis are mainly different at 1400, 1455, 1540, and 1647 cm^{-1}, and there are significant differences in infrared peaks. Table 3 lists some biological molecules corresponding to infrared characteristic peaks. The band at 1080 cm^{-1} was identified as PO2 symmetric stretching, corresponding to nucleic acids and phospholipids, and is related to cell proliferation, differentiation, and apoptosis. In thyroid malignant tumor cells, changes in the expression levels of specific DNA and RNA are closely related to the occurrence, development, and metastasis of tumors. The DNA and RNA content in cells of patients with thyroid malignant tumor metastasis is higher than that of patients without metastasis. The band at 1314 cm^{-1} was identified as C-N bond stretching, which belongs to the amine III band. In the marker analysis of infrared spectroscopy, the amine III band is a characteristic vibration mode of the secondary structure of proteins, usually located in the 1200–1300 cm^{-1} region of the infrared spectrum. The amine III band is mainly related to the C-N stretching vibration and N–H bending vibration in proteins, and it can reflect the β-folding content of proteins[33]. For patients with thyroid cancer, changes in amine III may be related to the biochemical characteristics of malignant tumors, and the content of amine III is higher in the cells of patients with thyroid malignant tumor metastasis. The bands at 1400 and 1455 cm^{-1} were identified as S=O bond stretching, corresponding to the infrared peaks caused by protein COO-stretching vibration-CH3 asymmetric stretching vibration, respectively. Changes in protein COO-stretching vibration-CH3 asymmetric stretching vibration may be related to the metabolic activity of tumor cells, the remodeling of the extracellular matrix, and the expression of tumor-related proteins, all of which may affect the invasiveness and metastasis of tumors. The peak intensities of patients with thyroid malignant tumor

metastasis at 1400 and 1455 cm^{-1} were slightly higher than those of patients with thyroid malignant tumor without metastasis. The peak intensities of amide II band and amide I band at 1540 and 1647 cm^{-1} in patients with thyroid malignant tumor metastasis were much greater than those in patients with thyroid malignant tumor without metastasis, and the difference was obvious. The bands at 2867, 2931 and 2959 cm^{-1} were determined to be C-H bond stretching, which were attributed to substances with methylene and methyl groups, respectively. The presence of methylene and methyl groups and changes in their chemical environment may be related to the metabolic activity of thyroid malignant tumor cells, the remodeling of the extracellular matrix, and the expression of tumor-related proteins[34], providing a basis for subsequent classification and finding the most sensitive spectral band.

Methods

The framework of the MSCNet model is shown in Fig. 3. We first perform preliminary feature extraction and dimensionality reduction on the FTIR and Raman spectral data through the Sepctral FC (Fully Connected Layer) module to obtain representations of the same dimension in two dimensions. The spectral representations of the same dimension are helpful for the implementation of the subsequent multi-modal feature separation module and the cross-feature fusion network module. Secondly, we designed the Multi-modal feature separation module to perform feature separation interaction on the modal representations of the two preliminary extracted features, aiming to obtain the complementary information of the two spectra on the molecular information, and finally obtain two separated features. Next, we need to fuse the separated features, so we designed the Cross-feature module. The Cross-feature module constructs multi-modal fusion features for the two separated features obtained in the previous module through intra-modal and inter-modal dynamic modeling. To this end, we constructed the Inter-fusion network, Intra-fusion network and decoder to better fuse the rich information within and between modalities, and obtain two Cross fusion features. Finally, after the FTIR and Raman spectral data are learned through the above three modules, we use the obtained two cross fusion features as the input of Malignant tumor prediction, and finally obtain the prediction results of thyroid malignant tumors. The following subsections will introduce the details of each module of MSCNet in detail.

Spectral-fully connected neural network encoder module

In the original data of each mode, there may be a large amount of redundant information that is unfavorable to the dichotomy prediction of thyroid malignancy, so we need to feature extract the original data and map them to the feature space of the unified dimension. To extract the preliminary features for FTIR and Raman spectra and make the same input representation dimensions of Multi-modal feature separation module, we designed the spectral fully connected neural network encoders E_I and E_R extracted the preliminary features of the two modalities. The input of this module is the original dimension data sum of two modes $X_I \in R^{d_I \times N}$, $X_R \in R^{d_R \times N}$, where d_I and d_R represent the original dimension of the two spectra, and N represents the number of data samples. Specifically, the processing process of E_I and E_R is three layers of nonlinear transformation of the two original dimensions respectively (fully connected layers, plus ReLU as the activation function). The process by which the encoder processes the infrared and Raman spectral data can be summarized in the following formula:

$$V_I = E_I(X_I) \tag{1}$$

$$V_R = E_R(X_R) \tag{2}$$

where V_I and V_R indicate the tensor measured by the encoder E_I and E_R after preliminary feature extraction and dimension reduction, respectively.

Multimodal feature separation module

The purpose of the multimodal feature separation module is to decompose the initially extracted features V_I and V_R into multiple relatively independent factors or attributes. Multi-omics data usually contains common

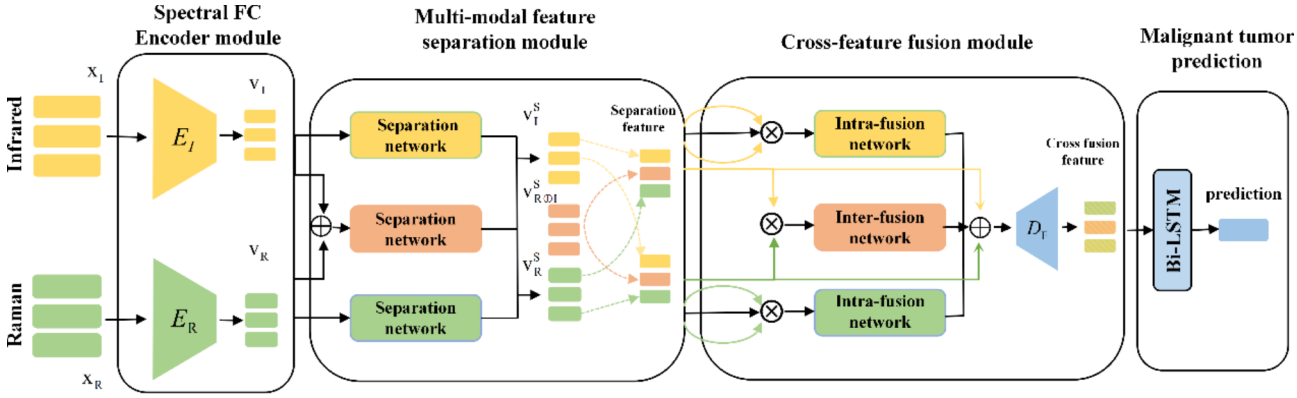


Fig. 3. MSCNet framework diagram.

information and omics-specific information (information unique to each omics). The multimodal feature separation module is able to separate these two kinds of information, allowing the model to learn better features common and unique to infrared and Raman spectroomics. There may be a complex coupling relationship between different omics. The feature separation module can effectively decouple the coupling relationship between different omics, enabling the model to be more focused on learning the features of each omics, so as to improve the effect of the model. Specifically, this module designs the separation network $f_{Sep1}()$ and $f_{Sep2}()$ performs nonlinear mapping to obtain the unique V_I^S, V_R^S and common representations $V_{I\oplus R}^C$ of infrared V_I and Raman spectra V_R. This process can be described as the following mathematical transformation:

$$V_I^S = f_{Sep1}\left(V_I\right) \tag{3}$$

$$V_R^S = f_{Sep2}\left(V_R\right) \tag{4}$$

$$V_{I\oplus R}^S = f_{Sep1}\left(V_I \oplus V_R\right) \tag{5}$$

Next, we fuse the unique representations and common representations, split the common representations before splicing, and then alternately fuse the unique representations in the two-dimensional direction, and finally get the two separated features V_I^{Sp}, V_R^{Sp}.

Feature cross-fusion network module

Using effective methods to fuse the feature information of different modalities is one of the main challenges of multimodal medical information fusion[35]. In general, according to the period of multimodal information fusion, we divide multimodal learning into early fusion, late fusion, and tensor-based fusion (Tensor-based)[36]. The method used in this paper is tensor fusion to separate features V_I^{Sp} and V_R^{Sp} information fusion. The purpose of the feature cross-fusion module is to construct the multimodal fusion features by modeling the two separated features obtained in the previous module. Specifically, the two separated features are first Kronecker-accumulated separately, and they are used as input to Intra-fusion module to obtain the fusion representation between the two modes. The two separated features were then used as Kronecker-products, which were used as input to Inter-fusion network to obtain inter-modal representations. Finally, the three representations and separation features, and five tensors, are cascade to obtain the cross-fusion features. The process can be described as the following mathematical operation:

$$V_{I\oplus IR\oplus R} = V_I^{Sp} \oplus V_R^{Sp} \oplus f_{Intra}\left(V_I^{Sp} \otimes V_I^{Sp}\right) \ \oplus f_{Inter}\left(V_I^{Sp} \otimes V_R^{Sp}\right) \oplus f_{Intra}\left(V_R^{Sp} \otimes V_R^{Sp}\right) \tag{6}$$

$$V^F = D_E\left(V_{I\oplus IR\oplus R}\right) \tag{7}$$

where $V_{I\oplus IR\oplus R} \in R^{Concat}, f_{Intra}(), f_{Inter}()$ represent the mathematical operation of Intra-fusion module and Inter-fusion network respectively, specifically, $f_{Intra}()$ and $f_{Inter}()$ is a nonlinear transformation of the input tensor. The specific operation of D_E is to perform four layers of nonlinear transformation on $V_{I\oplus IR\oplus R}$, and finally obtain the cross-feature fusion vector V^F.

The malignant tumor prediction module

The bidirectional long-and short-term memory network (Bi-LSTM) is an excellent recurrent neural network (RNN) architecture that shows a significant advantage in processing sequence prediction tasks. The core feature of Bi-LSTM is its bidirectional information processing power, which is able to consider both the temporal dependence and temporal predictivity in the sequence data, thus capturing the comprehensive context information[37] of the sequence. This unique structure makes Bi-LSTM particularly suited to understand and predict dynamic changes in sequence data, especially in the field of complex pattern recognition such as spectral sequence analysis.

In this study, we used Bi-LSTM to deeply model the timing data V_F. We constructed an excellent malignancy prediction module that is able to accurately predict whether a patient has thyroid malignancy metastasis. Our model outputs a prediction vector $\widehat{y} \in R^{2\times N}$, which contains an accurate assessment of the likelihood of metastasis from thyroid malignancy. The advantage of Bi-LSTM is that it can use long and short-term memory (LSTM) mechanism to effectively avoid the gradient disappearance or explosion of traditional RNN when dealing with long sequences. Moreover, the bidirectional learning strategy of Bi-LSTM not only enhances the understanding of the model on the time-series data, but also improves the accuracy and reliability of the predictions. In this way, our model is able to comprehensively consider the past and future information of sequence data, providing a more comprehensive and precise basis for clinical decision-making. The process can be formulated as the following mathematical operation:

$$\widehat{y} = BiLSTM\left(V^F\right) \tag{8}$$

The advantages of Bi-LSTM in sequence prediction tasks are reflected not only in its ability to capture the two-way context of sequence data, but also in its efficient processing of long sequence information and its excellent performance on prediction tasks. These properties make Bi-LSTM an ideal choice for us to solve the problem of metastasis prediction in thyroid malignancy.

Experimental setup and discussion of results

Training setup

In order to increase the robustness of the model, a fivefold cross-validation method of stratified sampling was adopted to divide the training sample into k parts, and then k cross-validation was carried out. Each time, a different part was used as the verification set, and the remaining K-1 parts were combined as the training set. Finally, the training set and the test set are divided according to 8:2. Stratified sampling can ensure that the proportion of samples in different categories in each subset is consistent with that in the original data set, and can alleviate the problem of overfitting or bias. The mean value of each evaluation index after 5 times is taken as the final evaluation standard. An Adam optimizer with an initial learning rate of 5e-4 was used in all contrast and ablation experiments, with a total number of nonlinear hidden units of 80 and epoch 200. The cross-entropy objective function is trained as a loss function:

$$L_{\log} = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \widehat{logy}_i \tag{9}$$

where is the class label of the $\widehat{y}_i$ predicted value (0: metastatic lymph node malignancy, 1: no metastatic lymph node malignancy), y_i is the true sample label value, and n is the total number of patients in the dataset. PyTorch, a high-level neural network frame-work in python, is used to implement MSCNet under Windows with GPU NVIDIA GeForce RTX 3090.

Baselines

To verify the MSCNet model's generalization ability, we selected 6 multimodal baseline models for comparative experiments. These baselines include classic neural network architectures in time series data analysis and multimodal fusion.

- EF-LSTM[38] is a deep learning model that combines the Flask backend and the extended short-term memory network (LSTM). It uses the embedding layer to convert the input data into a vector representation and captures the data's time series characteristics and long-term dependencies through the LSTM network. EF-LSTM has significant advantages in processing sequence data and time series prediction, and is an important milestone in deep learning in the time series data analysis field.
- LF-DNN[39] is a hybrid model that combines the long short-term memory network (LSTM) and the deep neural network (DNN) to process sequence data and predict time series. The model uses LSTM to capture the time series characteristics and long-term dependencies of the data and DNN to process the data's nonlinear relationship.
- AlexNet[40] is a major breakthrough in deep learning in computer vision, proposed by Alex Krizhevsky et al. in 2012. It is the first convolutional neural network to make significant breakthroughs on the large-scale image dataset ImageNet. AlexNet contains multiple convolutional layers, pooling layers, and fully connected layers and uses Dropout and ReLU activation functions, successfully opening a new chapter in deep learning in image recognition and classification.
- VGG-19[41] is the first convolutional neural network proposed by Karen Simonyan and Andrew Zisserman in 2014 that has made significant breakthroughs on the large-scale image dataset ImageNet. VGG-19 contains multiple convolutional layers, pooling layers, and fully connected layers and uses ReLU activation functions, demonstrating the great potential of deep learning in image recognition and classification.
- ResNet[42] is a deep residual network proposed by Kaiming He et al. in 2015. It solves the problem of profound network training difficulty by introducing residual blocks, allowing the network to reach deeper layers. Each residual block contains skip connections, which enables the model to be better optimized during training. The feature extraction layer of ResNet allows the network to learn feature representations with different scales, which is beneficial for processing tasks with various image sizes and complexities.
- DenseNet[43] is an efficient convolutional neural network architecture proposed by Gao Huang et al. in 2017. It introduces a dense connection method so that each layer of the network is directly connected to all subsequent layers, achieving feature reuse and reducing the number of parameters. DenseNet has achieved excellent performance in multiple image recognition tasks and has impacted deep learning.

Model evaluation index Model evaluation indexes

The model evaluation index uses six standard measures, namely: accuracy, accuracy, sensitivity, specificity, F1 score and AUC values, as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{10}$$

$$\text{Precison} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{11}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{12}$$

$$\text{Specificity} = \frac{TN}{TN+FN} \tag{13}$$

$$\text{F1 score} = \frac{2*\text{Precision}*Recall}{\text{Precision}+Recall} \tag{13}$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
EF-LSTM	57.68	53.60	57.78	57.78	53.17	57.78
LF-DNN	62.58	63.40	62.78	62.78	62.02	62.78
AlexNet	64.58	65.63	64.78	64.78	63.80	64.78
VGG-19	56.58	60.71	57.00	57.00	53.82	57.00
ResNet	59.63	59.70	59.67	59.67	59.41	59.67
DenseNet	56.63	56.46	56.56	56.56	55.63	56.56
MSCNet	63.63	66.58	63.78	63.78	62.61	63.78

Table 4. Single-mode Raman comparison experiment.

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
EF-LSTM	94.00	94.34	94.00	94.00	93.98	94.00
LF-DNN	94.00	94.52	94.00	94.00	93.97	94.00
AlexNet	85.00	89.63	85.00	85.00	84.07	85.00
VGG-19	73.89	83.43	74.00	74.00	70.89	74.00
ResNet	92.00	92.43	92.00	92.00	91.95	92.00
DenseNet	91.89	92.36	92.00	92.00	91.88	92.00
MSCNet	95.84	96.59	96.00	96.00	95.80	96.00

Table 5. Single-mode FTIR comparison experiment.

The ROC curve is a curve drawn with the true positive rate (True Positive Rate) as the ordinate and the false positive rate (False Positive Rate) as the abscissa. Usually, the closer the ROC curve is to the upper left corner means that the model has a higher true positive rate and a lower false positive rate, and therefore has better performance. The AUC represents the area under the ROC curve and is a measure of the overall performance of the model. The larger the AUC value indicates the stronger the classification ability of the model.

Model comparative analysis

In order to prove the effectiveness of the MSCNet model, we selected six multimodal baseline models for comparison, namely EF-LSTM, LF-DNN, AlexNet, VGG-19, and DenseNet, and conducted binary classification experiments on the thyroid malignant tumor dataset. It is worth noting that we used the same training standard for single-modal comparison experiments, but because the multimodal feature separation module and feature cross-fusion module are input for multimodal data, a single modality cannot operate the two fusion modules, and it will affect the actual effect of the single modality. We use the corresponding nonlinear layer for filling, and do not use the above two multimodal modules.

In the experimental results of Table 4, there are significant differences in the performance of each model on Raman spectroscopy single modal data, reflecting the differences in the adaptability of different architectures to spectral features. Traditional convolutional neural networks (CNNs) such as AlexNet perform well in capturing local features, while VGG-19 and DenseNet may overfit due to the high model complexity, resulting in performance degradation. In contrast, although ResNet alleviates the gradient vanishing problem through residual connections, its generalization ability is still limited on spectral data. In addition, although LSTM can process sequence information, it is not as effective as the CNN model in capturing spectral features, while the DNN model has achieved a relative balance in extracting deep nonlinear features. MSCNet, as the model proposed in this paper, can more effectively extract key information in the data through feature separation modules and feature cross-fusion modules, and achieves improvements in multiple indicators such as precision, sensitivity, and F1 score, demonstrating its potential and superiority in spectral analysis. The architectural advantages of MSCNet show that it helps to enhance the model's ability to interpret complex biological signals.

In the single-modal FTIR comparison experiment in Table 5, the performance differences of different models on FTIR spectral data clearly demonstrate their adaptability to spectral feature processing. Traditional convolutional neural networks such as AlexNet show some ability in capturing local features, but are still more effective than VGG-19, which indicates that deeper network structures do not significantly improve performance, and the performance of the model may be reduced due to overfitting. ResNet and DenseNet improve the depth and diversity of feature extraction by introducing residual connections and dense connections, and perform better than AlexNet and VGG-19. EF-LSTM and LF-DNN show strong learning ability when processing FTIR data, indicating that the combination of sequence and nonlinear features plays an important role in spectral analysis. In contrast, MSCNet shows the best performance, achieving an accuracy of 95.84%. The design of its feature separation module and feature cross-fusion module makes it superior to other models in multiple indicators. This shows that the innovative architecture of MSCNet can more effectively integrate and interpret complex features in FTIR spectra, improve the robustness and reliability of classification, and further verify its potential application value in biomedical spectral analysis.

Ablation study

In this study, in order to verify the generalization ability of the model, we designed a multi-modal separated cross-fusion network (MSCNet) based on Raman spectroscopy and FTIR spectroscopy. The innovative architecture of MSCNet mainly includes feature separation module (Feature Separation, FS) and cross-fusion module (Cross-fusion, CF), and the effectiveness of these modules is explored through ablation experiments. In Table 6, we conducted multi-modal ablation experiments on models such as EF-LSTM, LF-DNN, AlexNet, VGG-19, ResNet, and DenseNet, with the input being single-modal Raman spectroscopy and FTIR spectroscopy data, to verify the impact of different models and modules on performance.

From the experimental results, MSCNet has shown excellent performance without adding feature separation and cross-fusion modules, with an accuracy of 93.95% and an AUC value of 94.00%. This shows that its spectral fully connected neural network encoder module has high effectiveness in dimensionality reduction and feature extraction. At the same time, through the bidirectional long short-term memory network (Bi-LSTM), MSCNet can capture the full-band information of spectral data and further enhance the modeling ability of complex spectral features. When only the feature separation module is added, the accuracy of MSCNet is improved to 95.84% and the AUC reaches 96.00%. This shows that the feature separation module effectively analyzes the initial interaction information between Raman and FTIR spectra, and by separating common features from unique features, the expression of multimodal information is clearer, and the feature representation ability of the model is enhanced.

After introducing the feature separation and cross-fusion modules at the same time, MSCNet's performance reaches the best, with an accuracy of 97.95% and an AUC value of up to 98.00%. This result shows that the cross-fusion module realizes deep feature interaction between different spectral modalities, effectively models cross-modal information, and fully explores the potential correlation between the two spectral data. This fusion strategy not only improves the problem of insufficient feature interaction that cannot be solved by a single modality, but also greatly improves the robustness and generalization ability of the model, verifying the significant advantages of MSCNet in multimodal spectral tasks.

Among other comparison models, EF-LSTM and LF-DNN also showed good comprehensive performance, especially LF-DNN performed best, which further proved the advantage of nonlinear layers in processing spectral data features. AlexNet is good at capturing local features in the convolution layer, and although VGG-19 has weaker performance, it has improved compared with the single-modal experiment in Sect. "Model comparative analysis", verifying the value of multimodal data fusion. In addition, although the multimodal performance of ResNet and DenseNet is better than that of VGG-19, there is still a large gap compared with MSCNet, which shows that simply increasing the number of layers and connection methods in traditional convolutional networks cannot fully tap the potential of multimodal spectral data.

In summary, MSCNet has shown excellent performance and robustness in multimodal spectral fusion tasks, and its feature separation module and cross fusion module effectively solve the problem of insufficient feature interaction in spectral fusion. The successful application of this model in complex biological spectral analysis has verified the potential of multimodal spectral fusion, and it has broad application prospects in early diagnosis of clinical diseases, disease classification and prediction in the future. This result shows that optimizing feature extraction and fusion strategies is crucial to improving the performance of multimodal models, and provides new ideas and directions for future multimodal research.

Omics correlation validation

This section further explores the interpretability of Raman spectroscopy and infrared spectroscopy data fusion based on the Spearman correlation coefficient r and p-value and explores the multimodal data fusion potential of Raman spectroscopy and infrared spectroscopy[44]. In this section, a total of 5 Raman frequency shifts and 84 infrared spectral band correlation pairs were obtained. Among them, the content of glutathione is higher in patients with thyroid malignant tumor metastasis, which can help cancer cells resist oxidative damage. In addition, the content of nucleotides is also higher. Cancer cells need a large amount of nucleotides to synthesize DNA and RNA during proliferation, accelerating cell proliferation and differentiation. At the same time, the cytoskeletal structure of cancer cells in patients with thyroid malignant tumor metastasis may change, resulting in changes in cell morphology and behavior. The expression and modification status of actin and tubulin are

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
EF-LSTM	91.89	93.03	92.00	92.00	91.84	92.00
LF-DNN	92.00	92.16	92.00	92.00	91.99	92.00
AlexNet	88.95	89.41	89.00	89.00	88.90	89.00
VGG-19	74.79	82.30	74.89	74.89	71.82	74.89
ResNet	84.89	86.34	85.00	85.00	84.41	85.00
DenseNet	85.89	87.73	86.00	86.00	85.57	86.00
MSCNet(without FS& CF)	93.95	94.27	94.00	94.00	93.94	94.00
MSCNet(FS)	95.84	96.59	96.00	96.00	95.80	96.00
MSCNet(FS& CF)	97.95	98.09	98.00	98.00	97.94	98.00

Table 6. Multimodal ablation experiment.

different from those of normal cells. A heat map was drawn. Figure 4 shows the correlation diagram of Raman spectroscopy and infrared spectroscopy data.

As shown in Fig. 4a, we can find that each Raman frequency shift is correlated with at least one infrared spectral band, and the overall correlation coefficient ranges from -0.35 to 0.35. In this paper, the correlation degree of the Spearman correlation coefficient can be divided into 4 levels. When the absolute value of the Spearman correlation coefficient is less than or equal to 0.15, it indicates that the correlation is weak. The absolute value is between 0.15 and 0.30, indicating medium correlation. The absolute value is between 0.30 and 0.50, indicating a strong correlation. The absolute value is more significant than 0.5, indicating a strong correlation. In multimodal omics fusion, a high correlation coefficient between the fused omics may mean less complementary information between the modalities and a lack of fusion value. A low correlation coefficient may mean that the correlation between the two omics could be better, and the difficulty of fusion will increase accordingly.As shown in Fig. 4 (b), in order to more intuitively show the correlation between Raman spectroscopy and infrared spectroscopy, we counted the number of infrared spectral bands corresponding to Raman frequency shifts with an absolute value of the Spearman coefficient greater than or equal to 0.31 and a p value less than 0.05. We can find that the infrared bands corresponding to Raman frequency shifts such as Raman_1635 and Raman_1345 are far more than the number of other Raman frequency shifts, which indicates that some chemicals represented in the Raman frequency shifts have a strong correlation with the chemicals represented by the infrared spectral bands, and this correlation provides a certain theoretical basis for the interpretability of the fusion of Raman and infrared spectroscopy data.

Raman spectroscopy and FTIR can collect specific molecular features of all biological systems (cells, tissues, body fluids), showing the presence of basic components such as proteins, nucleic acids, lipids and carbohydrates. The information collected by Raman spectroscopy and FTIR in the spectrum is complementary. For example, FTIR is sensitive to proteins, esterified lipids, carbohydrates and nucleic acids (especially the secondary conformations of these molecules), while Raman spectroscopy is more specific to the presence of blood proteins, lipids, nucleotides and amino acid residues[45]. A huge advantage of vibrational spectroscopy

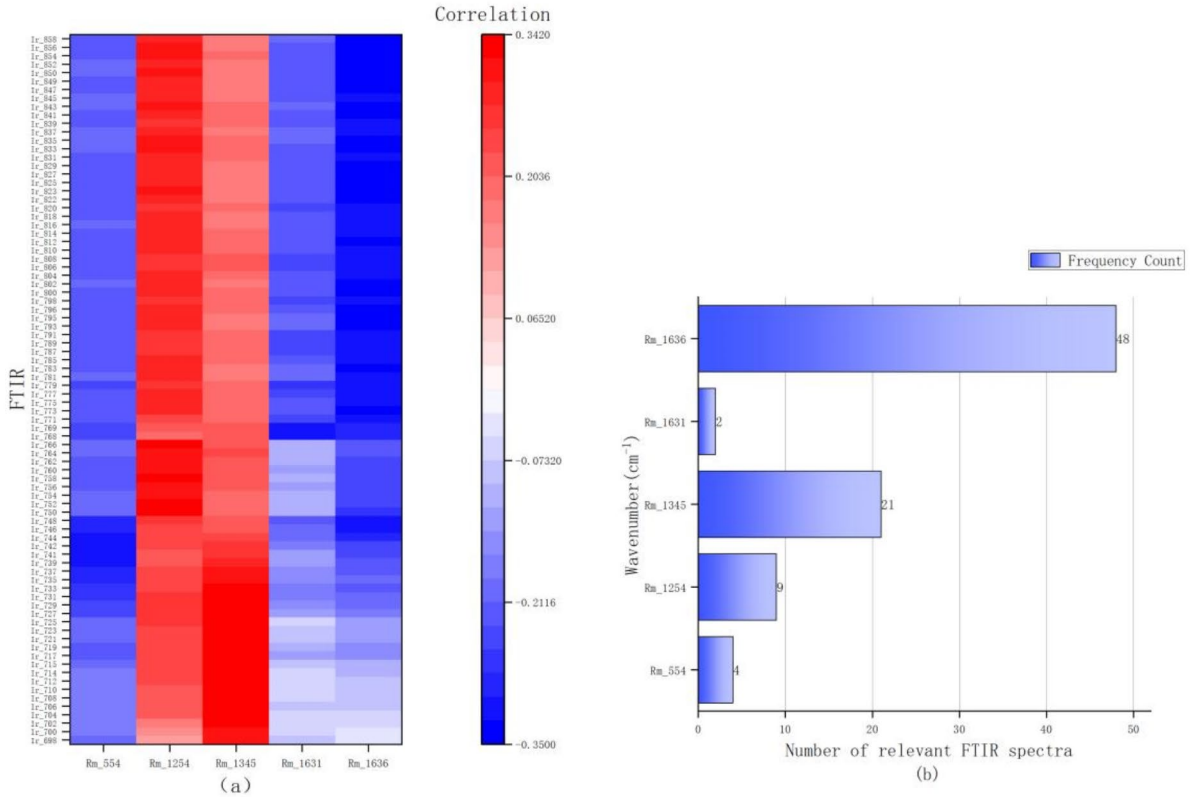


Fig. 4. Correlation diagram of Raman spectroscopy and infrared spectroscopy data (**a**) Heat map of correlation coefficients between 5 Raman frequency shifts and 84 infrared bands (**b**) Number of infrared bands corresponding to Raman frequency shifts with absolute values of correlation coefficients ≥ 0.31.

is its imaging capability, which provides label-free and non-destructive distribution of biological components in cellular compartments. Importantly, the spatial resolution of Raman spectroscopy and FTIR imaging is as high as 0.3 μm and 3 μm, respectively. It can also semi-quantitatively evaluate the content of chemical groups[46]. Therefore, this shows that the use of Raman spectroscopy and infrared spectroscopy or the fusion of two omics can be used to efficiently diagnose diseases. For example, Kujdowicz, M et al. proposed a study on the analysis of urothelial carcinoma based on Fourier transform infrared and Raman imaging[47].

In summary, the study verified that some frequency shifts of Raman spectroscopy have a certain correlation with some bands of infrared spectroscopy through the correlation analysis of Raman spectroscopy data and infrared spectroscopy data. Raman spectroscopy features are an effective supplement to infrared spectroscopy features in multimodal omics fusion classification. The fusion of the two is interpretable.

Conclusion

This paper introduces a new artificial intelligence disease classification network based on multimodal spectral data. It focuses on discussing the key issues in this field and solving the application at the level of spectral data fusion. By integrating information from different spectral types, we expect to improve the accuracy of disease diagnosis and the effect of non-invasive screening programs from a fusion-based perspective. A multimodal separation cross-fusion network is proposed. According to the encoder module, the original high-dimensional spectral data is uniformly reduced in dimension and sent to the multimodal feature separation module for feature decoupling to obtain separation features. Subsequently, the features are cross-fused through intra-modal and inter-modal interactions. The separation features and cross-features are cascaded and sent to the BiLSTM module for malignant tumor prediction, deciding whether the cervical lymph node tumor has metastasized. And through omics correlation verification, 5 Raman frequency shifts and 84 infrared spectral bands provided medically significant feature information. The results show that the model can fuse multimodal data and provide an efficient multimodal diagnosis framework for auxiliary diagnosis of medical diseases. In the future, we will consider the impact of sample size on spectral data fusion and the generalization performance under different disease types. Explore more about the complementary theories of Raman and infrared spectroscopy from a deep learning perspective.

Data availability

The datasets generated and analyzed during the current study are not publicly available due to data privacy laws, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare no competing interests.

Received: 14 June 2024; Accepted: 19 November 2024
Published online: 25 November 2024

References

1. White, M. L., Gauger, P. G. & Doherty, G. M. Central lymph node dissection in differentiated thyroid cancer. *World J. Surg.* **31**, 895–904 (2007).
2. Nixon, I. J. et al. Metastasis to the thyroid gland: A critical review. *Annals Surg. Oncol.* **24**, 1533–1539 (2017).
3. Wang, T. S. & Sosa, J. A. Thyroid surgery for differentiated thyroid cancer—recent advances and future directions. *Nat. Rev. Endocrinol.* **14**(11), 670–683 (2018).
4. Zhang, M. B. et al. Cervical lymph node metastasis prediction from papillary thyroid carcinoma US videos: A prospective multicenter study. *BMC Med.* **22**(1), 1–13 (2024).
5. Zhou, Y. et al. Added value of arterial enhancement fraction derived from dual-energy computed tomography for preoperative diagnosis of cervical lymph node metastasis in papillary thyroid cancer: Initial results. *Eur. Radiol.* **34**(2), 1292–1301 (2024).
6. Zhao, W. et al. Clinical value of dual-energy CT for predicting occult metastasis in central neck lymph nodes of papillary thyroid carcinoma. *Eur. Radiol.* **34**(1), 16–25 (2024).
7. Chen, F. et al. A nomogram based on clinicopathological and ultrasound characteristics to predict central neck lymph node metastases in papillary thyroid cancer. *Front. Endocrinol.* **14**, 1267494 (2024).
8. Qi, Y. et al. Applications of Raman spectroscopy in clinical medicine. *Food Front.* https://doi.org/10.1002/fft2.335e (2024).
9. Wang, Z. et al. Rapid biomarker screening of Alzheimer's disease by interpretable machine learning and graphene-assisted Raman spectroscopy. *ACS Nano* **16**(4), 6426–6436 (2022).
10. Zhou, X., Chen, C., Zuo, E., Chen, C. & Lv, X. Cross Branch Co-Attention Network multimodal models based on Raman and FTIR spectroscopy for diagnosis of multiple selected cancers. *Appl. Soft Comput.* **166**, 112204 (2024).
11. Wang, H. et al. Serum Raman spectroscopy combined with multiple algorithms for diagnosing thyroid dysfunction and chronic renal failure. *Photodiagn. Photodyn. Ther.* **34**, 102241 (2021).
12. Song, H. et al. Rapid identification of papillary thyroid carcinoma and papillary microcarcinoma based on serum Raman spectroscopy combined with machine learning models. *Photodiagn. Photodyn. Therapy* **37**, 102647 (2022).
13. Chen, C. et al. Exploration research on the fusion of multimodal spectrum technology to improve performance of rapid diagnosis scheme for thyroid dysfunction. *J. Biophoton.* **13**(2), e201900099 (2020).
14. Leng, H. et al. Raman spectroscopy and FTIR spectroscopy fusion technology combined with deep learning: A novel cancer prediction method. *Spectroch. Acta Part A: Mol. Biomol. Spectrosc.* **285**, 121839 (2023).
15. Chen, C. et al. A novel diagnostic method: FT-IR, Raman and derivative spectroscopy fusion technology for the rapid diagnosis of renal cell carcinoma serum. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **269**, 120684 (2022).
16. Shuai, W. et al. Rapid diagnosis of rheumatoid arthritis and ankylosing spondylitis based on Fourier transform infrared spectroscopy and deep learning. *Photodiagn. Photodyn. Therapy* **45**, 103885 (2024).
17. Huang, Y. et al. SLE diagnosis research based on SERS combined with a multi-modal fusion method. *Spectroch. Acta Part A: Mol. Biomol. Spectrosc.* **315**, 124296 (2024).
18. Xin, X. et al. A method for accurate identification of Uyghur medicinal components based on Raman spectroscopy and multi-label deep learning. *Spectroch. Acta Part A: Mol. Biomol. Spectrosc.* **315**, 124251 (2024).
19. Zhang, S. et al. Estimation of wheat protein content and wet gluten content based on fusion of hyperspectral and RGB sensors using machine learning algorithms. *Food Chem.* **448**, 139103 (2024).

20. Zhou, X. et al. CMACF: Transformer-based Cross-Modal Attention Cross-Fusion model for systemic lupus erythematosus diagnosis combining Raman spectroscopy, FTIR spectroscopy, and metabolomics. *Inform. Process. Manag.* **61**(6), 103804 (2024).
21. de Brito, E. C. et al. Paracoccidioidomycosis screening diagnosis by FTIR spectroscopy and multivariate analysis. *Photodiagn. Photodyn. Ther.* **39**, 102921 (2022).
22. Cerezo, M. & Rocchi, S. Cancer cell metabolic reprogramming: A keystone for the response to immunotherapy. *Cell Death Dis.* **11**(11), 964 (2020).
23. Rygula, A. et al. Raman spectroscopy of proteins: A review. *J. Raman Spectrosc.* **44**(8), 1061–1076 (2013).
24. Hur, J. H. et al. AC-motif: A DNA motif containing adenine and cytosine repeat plays a role in gene regulation. *Nucleic Acids Res.* **49**(17), 10150–10165 (2021).
25. Shetty, G., Kendall, C., Shepherd, N., Stone, N. & Barr, H. Raman spectroscopy: Elucidation of biochemical changes in carcinogenesis of oesophagus. *Br. J. Cancer* **94**(10), 1460–1464 (2006).
26. Krafft, C., Neudert, L., Simat, T. & Salzer, R. Near infrared RAMAN spectra of human brain lipids. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **61**(7), 1529–1535 (2005).
27. Gniadecka, M., Wulf, H. C., Nymark Mortensen, N., Faurskov Nielsen, O. & Christensen, D. H. Diagnosis of basal cell carcinoma by Raman spectroscopy. *J. Raman Spectrosc.* **28**(2–3), 125–129 (1997).
28. Larraona-Puy, M. et al. Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma. *J. Biomed. Optics* **14**(5), 054031–054031 (2009).
29. Lakshmi, R. J. et al. Tissue Raman spectroscopy for the study of radiation damage: Brain irradiation of mice. *Rad. Res.* **157**(2), 175–182 (2002).
30. Faolain, E. O. et al. A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vibrat. Spectrosc.* **38**(1–2), 121–127 (2005).
31. Stone, N., Kendall, C., Smith, J., Crow, P. & Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday discussions* **126**, 141–157 (2004).
32. Malini, R. et al. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. *Biopolym. Orig. Res. Biomol.* **81**(3), 179–193 (2006).
33. Saito, K., Xu, T. & Ishikita, H. Correlation between C═ O stretching vibrational frequency and p K a shift of carboxylic acids. *J. Phys. Chem. B* **126**(27), 4999–5006 (2022).
34. Jermain, P. R., Fischer, A. H., Joseph, L., Muzikansky, A. & Yaroslavsky, A. N. Fluorescence polarization imaging of methylene blue facilitates quantitative detection of thyroid cancer in single cells. *Cancers* **14**(5), 1339 (2022).
35. Baltrusaitis, T., Ahuja, C. & Morency, L. Multimodal machine learning: A survey and taxonomy. *RAL* **41**(2), 423–443 (2019).
36. Zhu, L., Zhu, Z., Zhang, C., Xu, Y. & Kong, X. Multimodal sentiment analysis based on fusion methods: A survey. *Inform. Fusion* **95**, 306–325 (2023).
37. Siami-Namini, S., Tavakoli, N., & Namin, A. S. The performance of LSTM and BiLSTM in forecasting time series. in 2019 *IEEE International conference on big data (Big Data)* (pp. 3285–3292). IEEE (2019, December).
38. Williams, J., et al. Recognizing emotions in video using multimodal dnn feature fusion." Proceedings of Grand Challenge and Workshop on Human Multimodal Language(Challenge-HML). 2018.
39. Han, H., Liang, Y., Bella, G., Giunchiglia, F. & Li, D. LFDNN: A novel hybrid recommendation model based on DeepFM and LightGBM. *Entropy* https://doi.org/10.3390/e25040638 (2023).
40. Krizhevsky, A., Ilya, S., Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25
41. Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2015).
42. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 770–778, (2016).
43. Huang, G, et al. Densely connected convolutional networks. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
44. Zhang, W. Y., Wei, Z. W., Wang, B. H. & Han, X. P. Measuring mixing patterns in complex networks by Spearman rank correlation coefficient. *Physica A* **451**, 440–450 (2016).
45. Diem, M. et al. Molecular pathology via IR and Raman spectral imaging. *J. Biophoton.* **6**(11–12), 855–886 (2013).
46. Wiercigroch, E. et al. FT-IR spectroscopic imaging of endothelial cells response to tumor necrosis factor-α: To follow markers of inflammation using standard and high-magnification resolution. *Anal. Chem.* **90**(6), 3727–3736 (2018).
47. Kujdowicz, M. et al. In vitro spectroscopy-based profiling of urothelial carcinoma: a Fourier transform infrared and Raman imaging study. *Cancers* **13**(1), 123 (2021).

Acknowledgements

This work was supported by Basic Scientific Research Funds of Universities in Xinjiang Uygur Autonomous Region(XJEDU2023P012).

Author contributions

H.S.and X.Z.wrote the main manuscript text. C.C.:Conceptualization, Methodology, Writing—Original Draft Preparation C.D.: Data Curation, Formal Analysis, Software Y.H.: Investigation, Validation, Visualization M.W.: Resources, Funding Acquisition, Supervision J.Y.: Writing—Review & Editing, Project Administration X.C.: Methodology, Software, Validation Y.L.: Data Curation, Formal Analysis B.M.: Investigation, Resources All authors have read and agreed to the published version of the manuscript.

Additional information

Correspondence and requests for materials should be addressed to B.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024