Thyroid Classification in Ultraound by Deep Multimodal Learning

Jiong Shao, Hanshuo Xing, Mengyin Li and Xinglong Wu
School of Computer Sciences and Engineering, Wuhan Institute of Technology, Wuhan, China

Keywords: Deep Learning, Multimodal Learning, LSTM, Thyroid Cancer.

Abstract: Purpose: Biopsy results are the gold standard for testing the benignity and malignancy of thyroid cancer, but they also brings the problems of overdiagnosis and overtreatment. This is a challenging task to avoid these two problems while ensuring diagnostic accuracy and efficiency. In this paper, we use deep learning multimodal models to assist physicians in diagnosis and improve diagnostic accuracy.

Methods: This paper presents a multimodal deep learning model to assist physicians in the diagnosis of thyroid tumor. The model uses ultrasound images of the patient, geometric features of the lesion site, and clinical information to fuse modeling, with clinical information as the first modality, geometric features of the images as the second modality, and medical images as the third modality. The results are compared with other single-modal models to analyze and validate the performance of the multimodal model.

Results: For the dataset used, the multimodal model had an accuracy of 0.884, precision of 0.865, recall of 0.859, and F1 of 0.862, the Area Under Curve (AUC) of the multimodal mode was 0.933, the AUC of the ResNet50 was 0.639, the AUC of the InceptionResnetV2 was 0.612, the AUC of the Densenet121 was 0.654, and the AUC of the EfficientNetB3 was 0.649.

Conclusion: The multimodal model has high accuracy, sensitivity, and specificity in distinguishing benign and malignant thyroid tumors, and its performance is significantly better than the four single-modal deep learning classification models used for comparison. The proposed method is therefore valuable and is expected to help clinicians diagnose thyroid cancer efficiently.

1 INTRODUCTION

Thyroid cancer is one of the common malignancies we encounter with our life (Durante et al., 2018), and its incidence is increasing worldwide. According to the latest national cancer statistics released by the National Cancer Center in February 2022 (Changfa et al., 2022), thyroid cancer ranked seventh in incidence, with 50,000 new cases of cancer in men and 170,000 in women, who remain the most prevalent group. Thyroid cancer has received widespread attention because of the youthfulness of the incidence population and the increasing incidence rate year by year.

In medicine, thyroid tumors are mainly discovered by ultrasonic diagnosis (Zhaohui et al., 2010), which is used to detect thyroid nodules. If a nodule is found to be abnormal, a biopsy by puncture of cells (CHEN & JIANG, 2017) is required. Puncture biopsy is the gold standard for diagnosing benign and malignant thyroid tumors (Qingwen et al., 2017), but because of the prevalence of thyroid masses in daily life, direct biopsy can again lead to overdiagnosis and overtreatment. In order to reduce this, the current popular diagnostic method is mainly diagnosed by the physician's observation of the patient's ultrasound images before deciding whether a biopsy is needed However, the diagnosis by the naked eye of a professional doctor is not only time-consuming but also inaccurate and subject to false detection.Thus the use of deep learning methods to assist physicians in diagnosis became gradually popular (Fujita, 2020).In recent years, deep learning techniques had been successful in several areas of medical- assisted diagnosis (Juan-Xiu et al., 2018) , and their features such as end-to-end and automatic learning (Jeong et al., 2019) also provide new solutions for assisted diagnosis (Guang-Yuan et al., 2018). At the same time deep learning for text, audio and video data analysis has also achieved a lot of results (Sun et al., 2022).we propose a joint deep learning model using data onto multiple modalities for modeling, that is, a multimodal (Ngiam et al.,

Shao, J., Xing, H., Li, M. and Wu, X.
Thyroid Classification in Ultraound by Deep Multimodal Learning.
DOI: 10.5220/0011918100003612
In *Proceedings of the 3rd International Symposium on Automation, Information and Computing (ISAIC 2022)*, pages 216-223
ISBN: 978-989-758-622-4; ISSN: 2975-9463

2011) approach to enhance the model..

In this paper, we use the data on three different modalities of patient's ultrasound images, clinical information, and image geometric features of modeling to achieve benign and malignant classification of thyroid tumors, aiming to assist physicians in diagnosis and improve the efficiency and accuracy of diagnosis.

2 RELATED WORK

2.1 Research on Medical-Assisted Diagnosis by Single-Modal Deep Learning

Zhu et al. used VGG-16T for the diagnosis of benign and malignant thyroid tumors (YiCheng et al., 2021), with additional BN and Drop-out layers in addition to the fully connected layer, and performed a 10-fold cross-validation with high sensitivity, specificity, and accuracy of the results.. Peng et al. proposed ThyNet for the classification of thyroid tumors (Peng et al., 2021), ThyNet was constructed using a combination of three networks. Comparing the results obtained by this method of professional physicians, the results are superior.Li et al. implemented a deep CNN model on thyroid tumor diagnosis based on ultrasound images (Li et al., 2019), where the team used a DCNN model for training. The method achieved better results, with diagnostic performance essentially equal in sensitivity and higher specificity compared to experienced imaging physicians.

The Classification of thyroid cancer using a single-modality deep learning model, while achieving good results, is not medically interpretable in the opinion of clinician

2.2 Research on Medical-Assisted Diagnosis by Multimodal Deep Learning

Multimodal thyroid cancer classification studies to refer to the classification of thyroid cancer using features of different modalities. The methods of multimodal feature fusion (Ramachandram & Taylor, 2017) can be generally classified into two types: early fusion and late fusion.

Gong et al. propose a late fusion approach (Gong, 2013) for benign and malignant classification of thyroid tumors based on a composite weight multi-classifier fusion method of a non-Bayesian fusion framework, which outperformed a single-modal classifier in terms of correctness, sensitivity, and specificity.Song et al. propose an image fusion method to assist in Alzheimer's disease diagnosis (AD) (Song et al., 2021) by fusing gray matter tissue regions of MRI with FDG-PET images and then using 3DMulti-Scale CNN for evaluation. This method has better overall performance and outperforms state-of-the-art AD diagnostic methods.. Zheng et al. propose a multimodal fusion lung adenocarcinoma pathology classification model based on attentional mechanisms using deep learning multimodal techniques combining clinical information, CT images, and serum tumor markers of patients (Zheng, 2021). The method outperformed existing related studies in terms of accuracy as well as AUC.Xu et al. proposed an early fusion approach from cervical dysplasia diagnosis (Xu et al., 2016), using AlexNet network to learn image features from Cervigarm images, and then combining image features with non-image features (clinical outcomes, etc.), This method significantly outperforms using only any single source of information and previous multimodal frameworks.. Wang et al made improvements to ShuffleNet and proposed a 3DShuffleNet-based model for Alzheimer's disease assisted diagnosis (Wang et al., 2021), Their proposed method obtains relatively good results from a small computational cost.

these studies demonstrated the performance of multimodal versus single-modal, respectively, reflecting the key role of multimodal learning, and Huang et al. also demonstrated the superiority over multimodal learning (Huang et al., 2021), In contrast, few studies in the field of thyroid cancer have combined ultrasound image features with clinical features and used them for classification, so we then attempted to use multimodal data for feature fusion and use it for thyroid cancer classification.

3 METHODS

In this paper, we propose a model combining Densenet (Huang et al., 2017) , Multilayer Perceptron (MLP) and Long Short Term Memory (LSTM) (Greff et al., 2016) (Figure 1),which incorporates ultrasound images, clinical information and image geometric features to classify benign and malignant thyroid tumors of patients. DenseNet is

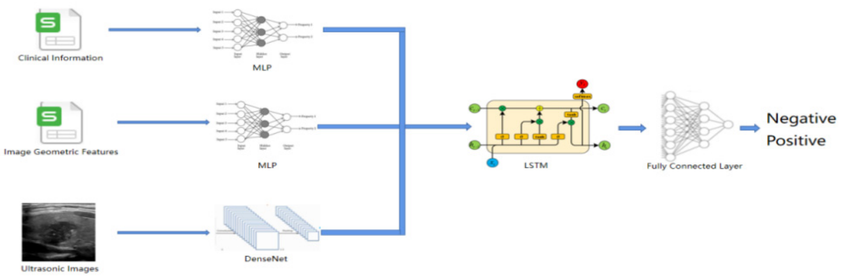

Figure 1: Diagram of thyroid cancer classification model based on multimodal deep learning

used to extract features from ultrasound images, and MLP is used to extract clinical information features and image geometry features, and the extracted feature vectors are stitched together to obtain a new feature vector to be input into the LSTM before the final classification by the classifier.

3.1 Datasets

Table 1：Baseline characteristics.

Patients		Data set
sex	Male	75 (25.8%)
	Female	215 (74.2%)
age (Years)	Age>45	175 (60.3%)
	Age<=45	115 (39.6%)

The data for this experiment were obtained from four hospitals, all of which were produced by Samsung ultrasound instruments.Because some of the patient data were missing some key information, a total of 290 patients with 558 images was retained, 170 (30%) negative images and 388 (70%) positive images, of which 25.8% were male and 74.2% were female (Table 1).

3.2 Geometric and Clinical Features of Images

Image geometric features (Zhihang, 2017): 10 features are selected from the many geometric features, which are:

(1) Number of pixels in the focal area: The number of white pixels in the focal area is calculated by MASK.

(2) Total number of pixels in the image: the number of pixels in the whole ultrasound image.

(3) Percentage of focal area: the number of pixels in the focal area.

(4) Area size of the lesion: the area of the contour formed by connecting the pixel centroids at the border of the lesion area.

(5) Perimeter of the lesion: the sum of the pixel points at the edge of the lesion area.

(6) Average gray values of the focal area: superimpose the focal area with MASK and divide the total gray value by the number of pixels.

(7) Focal area aspect ratio: the ratio of the width of the height of the outer rectangle of the focal area.

$$Aspect\ Ratio = \frac{Width}{Height} \tag{1}$$

(8) Lesion area firmness: ratio of lesion area to convex bun area.

$$Extent = \frac{Contour\ Area}{Convex\ Hull\ Area} \tag{2}$$

(9) Equivalent diameter of the lesion area: the diameter of a circle equalled to the area of the lesion area.

$$EquivalentDiameter = \frac{\sqrt{4 \times Contou\ Area}}{\pi} \tag{3}$$

(10) Focal area orientation: the angle between the short axis of the external rectangle of the focal area and the horizontal axis.

Clinical features: Han et al (Yuren et al., 2022)concluded that the benignity of thyroid cancer is related to the patient's age as well as gender, so we mainly counted and collated the patient's gender, age, true length of the lesion, true width and true area of the lesion as clinical features.

3.3 Data Preprocessing

Medical Image Preprocessing:

For the original medical images, a series of pre-processing processes are needed to remove the interference information. The preprocessing process is shown in Figure 2, where the original medical image is Gaussian blurred with the kernel size set to (9,9) and the standard deviation taken as 0. Then, it is binarized with the kernel shape of rectangle and the kernel size of (5,5) for morphological opening operation. Finally, contour detection is performed and all the found contours are sorted by area size, and the coordinates of the largest contour are recorded and cropped. For the cropped image, a mask of the lesion area (Figure 3) is created according to the annotation of the professional doctor on the image.

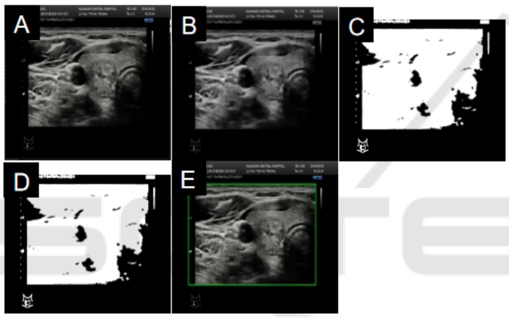

Figure 2: Ultrasound image preprocessing flow chart, original image (A), Gaussian blur (B), binarization (C), morphological open operation (D), finding maximum contour (E)

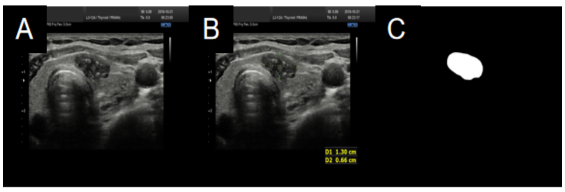

Figure 3: Original ultrasound image (A), physician labeled image (B), and lesion area labeled image (C)

Clinical data and image geometric features data preprocessing:

The gender features (male and female) in the clinical data were integrated after converting them to 0 (female) and 1 (male), and then the clinical data were normalized to 0 mean with the image geometric feature data..

$$x^* = \frac{x-\mu}{\sigma} \tag{4}$$

where is the original value of a feature, μ is the mean of the feature in all samples, σ is the standard deviation of the feature in all samples, and is the standardized feature value.

3.4 Network Design

The core of deep learning based multimodal thyroid tumor classification lies in the extraction of different modal features, fusion of multimodal features, and feature classification after fusion. In this paper, we use an early fusion approach to perform fusion at the feature layer. The proposed model in this paper has three main parts: feature extraction, feature fusion, and classification.The three parts are described in detail in the following.

Feature extraction module:

The processed clinical data and image geometry data are input to the MLP for feature extraction respectively. The MLP has two hidden layers, the first hidden layer has 32 neurons and the second hidden layer has 128 neurons, both using Relu activation function, and finally a feature vector of length 1*128 is obtained. The ultrasound image modality, on the other hand, uses Densenet121 for feature extraction, gets the output of the last convolutional layer, performs global average pooling and then reshape to 1*128.

Feature fusion module:

The three feature vectors are stitched together to obtain a 3*128 feature vector, and then the new feature vector is input into the LSTM.

Classification module:

The output of the LSTM network is fed to the fully connected layer for classification. The fully connected layer consists of two hidden layers, the first with 128 units and the second with 32 units, and both with the Relu activation function. The final output layer outputs the classification results using the Sigmod activation function.

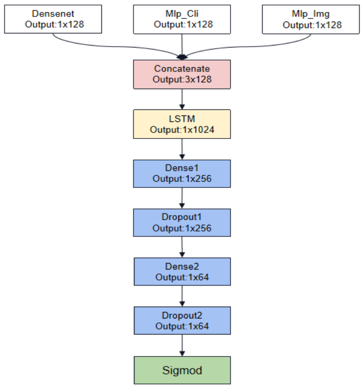

Figure 4: Overall network structure

3.5 Initialization and Training

To demonstrate the superiority in the model proposed of this paper, the model proposed to this paper is compared with some classical single- models.

For the multimodal model, Batch is set to 8, the optimizer is Adam, the learning rate is 0.0001, the loss is binary_crossentropy, and 200 rounds are trained, where the size of the ultrasonic modal input image is set to (250,250) and initialized with the pre-trained model weights.. The single-modal deep learning models used for the result comparison are ResNet50, EfficientNetB3, DenseNet121, and InceptionResnetV2, the Settings are the same as for multimodal models.

3.6 Evaluating Performance Metrics

In medical diagnosis, sensitivity and specificity are important indicators, and we prepare several indicators to evaluate the model performance.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (6)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (9)$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (10)$$

TP: Positive samples predicted by the model as positive class. TN: negative sample predicted by the model as negative class. FP: negative samples predicted by the model as positive class. FN: positive samples predicted by the model to be negative.

ROC (Koote et al., 2017)curve:Also known as subject operating characteristic curves, the ROC plot provides a quick visualization of the relationship between sensitivity and specificity.

In order to make the classification results more interpretable, Wu et al (Wu et al., 2022)proposed to use the modal contribution index (MCI) to find the contribution of each modality of the model. Considering the different degree of contribution to different modalities to the classification accuracy we proposed the WMCI, adding the accuracy of each modality as a weight.

$$\text{MCI}_{\text{m}} = \frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} \frac{\sum_{\text{j}=1}^{\text{FN}} \text{F}_{\text{i,m,j}}}{\sum_{\text{m}=1}^{\text{FM}} \sum_{\text{j}=1}^{\text{FN}} \text{F}_{\text{i,m,j}}} \quad (11)$$

$$WMCI_m = \frac{W*MCI_m}{\sum_{m=1}^{FM} W*MCI_m} \quad (12)$$

By calculating the output of the feature extraction module, where denotes the Jth element of the mth modal vector of the ith instance, FN is the number of features, FM is the modality,N is the number of instances, and W is the accuracy rate when each modality is predicted individually.

4 RESULTS

With 588 images obtained from a total of 290 cases from four hospitals, 334 (60%) images were used for the training set, 112 (20%) for the validation set, and 112 (20%) for the test set.

Figure 5 shows the ROC curves on different models. From the results in the figure, Resnet50 [28] has an AUC of 0.693, EfficientnetB3 [29] has an AUC of 0.649, Dencenet121 has an AUC of 0.654, InceptionResnetV2 has an AUC of 0.612, MLP_Cli has an AUC of 0.837, and MLP_Img has an AUC of 0.882, while the best results are achieved using our proposed multimodal model, with an AUC of 0.933. Table 2 shows the prediction results of each model on the test set, and the thyroid cancer classification model based on multimodal deep learning proposed in this paper achieves an accuracy of 0.884, precision of 0.865, recall of 0.859, and F1 of 0.862, which is better than several other comparative methods in all evaluation metrics. It indicates that the introduction of data from different modalities is beneficial for the classification of benign and malignant thyroid cancer.

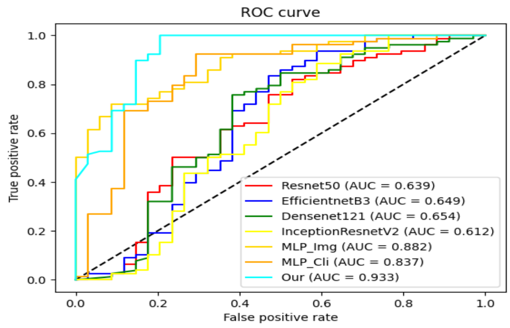

Figure 5: ROC plots for the multimodal model and each single-modal model, Mlp_cli is the Mlp model with clinical data as input and Mlp_img is the Mlp model with image geometric features as input

Table 2：Diagnostic performance of Multimodal and Single-modal models.

Model	ACC	Precision	Recall	F1	AUC
ResNet50	0.705	0.644	0.631	0.636	0.639
InceptionResnetV2	0.732	0.680	0.625	0.634	0.612
Densenet121	0.705	0.639	0.615	0.620	0.654
EfficientNetB3	0.723	0.675	0.634	0.658	0.649
Mlp_cli	0.803	0.797	0.803	0.792	0.837
Mlp_img	0.821	0.816	0.821	0.816	0.882
Our	**0.884**	**0.865**	**0.859**	**0.862**	**0.933**

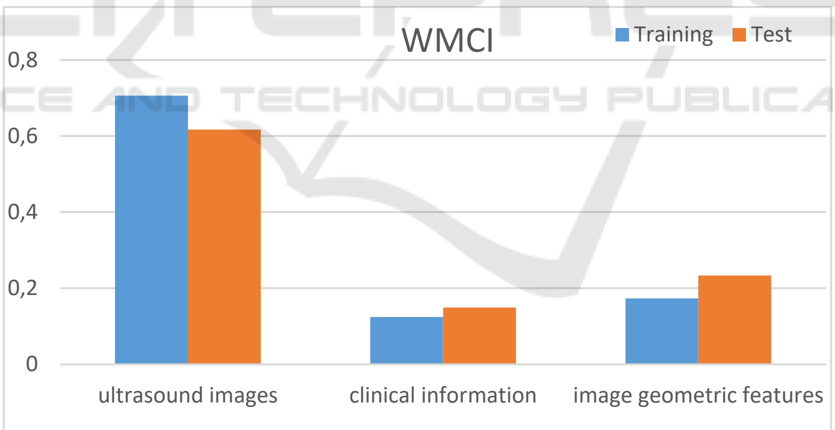

Figure 6：WMCI for Training Set and Test Set

Figure 6 shows the contribution of the model for each modality in the training set and the test set. In both datasets, the ultrasound image modality has the largest contribution, with 0.706 for the ultrasound image modality, 0.124 for the clinical information modality, and 0.170 for the image geometry feature modality in the training set, and 0.617 for the ultrasound image modality, 0.149 for the clinical This also indicates that in most cases, the multimodal deep learning model relies more on the ultrasound image modality for prediction.

5 DISCUSSION AND CONCLUSION

This paper proposes a multimodal thyroid cancer classification model based on deep learning method.

According to our results, our model outperforms the four single-modal networks used for comparison in terms of accuracy, sensitivity,,and specificity.The clinical information about patients and geometric features of images play a role in improving the classification of thyroid tumors and also validate the superiority of the model.

Our study also has some limitations:

(1) The current collected and collated multimodal dataset is relatively small, and the performance of the model may be better if more samples are available in the future.

(2) In the feature fusion part, we use early fusion, which directly splices the three output feature vectors, and can try other different feature fusion methods.

(3) In this study, the objective is to classify the benign and malignant thyroid tumors. The images input to the model are the parts of the ultrasound images that contain only the lesions, and it is still necessary to segment the images according to the doctor's labeled images when collecting and organizing the data in the preliminary stage.

To the best of our knowledge, previous studies have shown that deep learning algorithms outperform medical professionals in certain clinical outcomes, however, the use of deep learning approaches alone is not applicable in clinical settings (Ko et al., 2019), therefore, the main objective of this study is to assist physicians in diagnosis and reduce overdiagnosis and overtreatment. In future studies the multimodal model will be further improved by expanding the dataset used in the experiment and adding more different clinical data as features in the clinical information. In the feature fusion part, different fusion strategies are used to compare the effects of different fusion strategies on the model performance so as to improve the performance.

REFERENCES

Changfa, X., Xuesi, D., He, L., Maomao, C., Dianqin, S., Siyi, H., . . . Wanqing, C. (2022). Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chinese Medical Journal*.

CHEN, J., & JIANG, L. (2017). Accurate pathological diagnosis of thyroid cancer in the era of precision medicine. *Chinese Journal of Clinical Oncology*, *44* (04), 181-185.

Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S. J., & Cooper, D. S. (2018). The diagnosis and management of thyroid nodules: a review. *Jama*, *319* (9), 914-924.

Fujita, H. (2020). AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiological physics and technology*, *13* (1), 6-19.

Gong, R. (2013). *THYROID TUMOR CLASSIFICATION BASED ON MUTLI-MODE ULTRASOUND IMAGE* Harbin Institute of Technology].

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28* (10), 2222-2232.

Guang-Yuan, Z., Xia-Bi, L., & Guang-Hui, H. (2018). Survey on Medical Image Computer Aided Detection and Diagnosis Systems. *Journal of software*, *29* (05), 1471-1514.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H., & Huang, L. (2021). What makes multi-modal learning better than single (provably). *Advances in Neural Information Processing Systems*, *34*, 10944-10956.

Jeong, E. Y., Kim, H. L., Ha, E. J., Park, S. Y., Cho, Y. J., & Han, M. (2019). Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *European radiology*, *29* (4), 1978-1985.

Juan-Xiu, T., Guo-Cai, L., Shan-Shan, G., Zhong-Jian, J., Jin-Guang, L., & Dong-Dong, G. (2018). Deep Learning in Medical Image Analysis and Its Challenges. *Acta automatica Sinica*, *44* (03), 401-424.

Ko, S. Y., Lee, J. H., Yoon, J. H., Na, H., Hong, E., Han, K., . . . Park, V. Y. (2019). Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. *Head & neck*, *41* (4), 885-891.

Koote, R. S., Levin, E., Salojärvi, J., Smits, L. P., Hartstra, A. V., Udayappan, S. D., . . . Holst, J. J. (2017). Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell metabolism*, *26* (4), 611-619. e616.

Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., . . . Li, J. (2019). Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *The Lancet Oncology*, *20* (2), 193-201.

Peng, S., Liu, Y., Lv, W., Liu, L., Zhou, Q., Yang, H., . . . Zhang, X. (2021). Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: A multicentre diagnostic study. *The Lancet Digital Health*, *3* (4), e250-e259.

Qingwen, H., Yi, C., Huagui, L., & Bin, L. (2017). Comparative analysis of CT and B ultrasound diagnosis and pathological diagnosis of thyroid carcinoma. *China Foreign Medical Treatment*, *36* (28), 179-180+183.

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE signal processing magazine*, *34* (6), 96-108.

Song, J., Zheng, J., Li, P., Lu, X., Zhu, G., & Shen, P. (2021). An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Frontiers in Digital Health*, *3*, 637386.

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Wang, Y., Liu, X., & Yu, C. (2021). Assisted diagnosis of alzheimer's disease based on deep learning and multimodal feature fusion. *Complexity*, *2021*.

Wu, X., Li, M., Cui, X.-w., & Xu, G. (2022). Deep multimodal learning for lymph node metastasis prediction of primary thyroid cancer. *Physics in Medicine & Biology*, *67* (3), 035008.

Xu, T., Zhang, H., Huang, X., Zhang, S., & Metaxas, D. N. (2016). Multimodal deep learning for cervical dysplasia diagnosis. International conference on medical image computing and computer-assisted intervention,

YiCheng, Z., PengFei, J., Jie, B., Quan, J., & Ximing, W. (2021). Thyroid ultrasound image classification using a convolutional neural network. *Annals of translational medicine*, *9* (20).

Yuren, H., Limei, L., & Rui, W. (2022). Clinical Characteristics and Prognostic Factors of 'Thyroid Cancer Patients. *The Practical Journal of Cancer*, *37* (06), 1000-1002.

Zhaohui, L., Haiqing, Z., Jingtao, D., Yukun, L., Qinglong, K., Guoqing, Y., . . . Juring, L. (2010). Diagnostic value of ultrasonographic features in thyroid nodule properties. *Chinese Medical Journal* (46), 3272-3275.

Zheng, D. (2021). *Research on Intelligent Lung Tumor Analysis Technology Based on Multimodal Data Fusion* [Shanghai Institute of Technical Physics, Chinese Academy of Science].

Zhihang, Z. (2017). *Feature extraction and visualization of ultrasonic image based on TI-RADS image based on TI-RADS* [Southwest Jiaotong University].