BMC Cancer

RESEARCH Open Access

Pathology diagnosis of intraoperative frozen thyroid lesions assisted by deep learning

Tingting He[1], Shanshan Shi[1], Yiqing Liu[1], Lianghui Zhu[1], Yani Wei[2], Fenfen Zhang[2], Huijuan Shi[2*], Yonghong He[1*] and Anjia Han[2*]

Abstract

Background Thyroid cancer is a common thyroid malignancy. The majority of thyroid lesion needs intraoperative frozen pathology diagnosis, which provides important information for precision operation. As digital whole slide images (WSIs) develop, deep learning methods for histopathological classification of the thyroid gland (paraffin sections) have achieved outstanding results. Our current study is to clarify whether deep learning assists pathology diagnosis for intraoperative frozen thyroid lesions or not.

Methods We propose an artificial intelligence-assisted diagnostic system for frozen thyroid lesions that applies prior knowledge in tandem with a dichotomous judgment of whether the lesion is cancerous or not and a quadratic judgment of the type of cancerous lesion to categorize the frozen thyroid lesions into five categories: papillary thyroid carcinoma, medullary thyroid carcinoma, anaplastic thyroid carcinoma, follicular thyroid tumor, and non-cancerous lesion. We obtained 4409 frozen digital pathology sections (WSI) of thyroid from the First Affiliated Hospital of Sun Yat-sen University (SYSUFH) to train and test the model, and the performance was validated by a six-fold cross validation, 101 papillary microcarcinoma sections of thyroid were used to validate the system's sensitivity, and 1388 WSIs of thyroid were used for the evaluation of the external dataset. The deep learning models were compared in terms of several metrics such as accuracy, F1 score, recall, precision and AUC (Area Under Curve).

Results We developed the first deep learning-based frozen thyroid diagnostic classifier for histopathological WSI classification of papillary carcinoma, medullary carcinoma, follicular tumor, anaplastic carcinoma, and non-carcinoma lesion. On test slides, the system had an accuracy of 0.9459, a precision of 0.9475, and an AUC of 0.9955. In the papillary carcinoma test slides, the system was able to accurately predict even lesions as small as 2 mm in diameter. Tested with the acceleration component, the cut processing can be performed in 346.12 s and the visual inference prediction results can be obtained in 98.61 s, thus meeting the time requirements for intraoperative diagnosis. Our study employs a deep learning approach for high-precision classification of intraoperative frozen thyroid lesion distribution in the clinical setting, which has potential clinical implications for assisting pathologists and precision surgery of thyroid lesions.

Keywords Deep learning, Thyroid cancer, Frozen section, Whole slide image, Histopathological classification, Acceleration

*Correspondence:
Huijuan Shi
shihj@mail.sysu.edu.cn
Yonghong He
heyh@sz.tsinghua.edu.cn
Anjia Han
hananjia@mail.sysu.edu.cn
Full list of author information is available at the end of the article

Background

Pathological diagnosis provides a theoretical and practical basis for the diagnosis, treatment and prevention of diseases through the study of the causes and pathogenesis of diseases and the morphological structure of the diseased organism during the disease process. The production of hematoxylin and eosin (H&E) sections usually takes about three days to get the diagnosis result from the initial taking of specimen [1]. The number of pathologists is limited globally [2], and the lack of pathologists in hospitals has resulted in senior pathology doctors completing residents' work. Nevertheless, the surgeon needs to know the information about the lesion as soon as possible to adjust the surgical plan. This requires rapid sectioning to obtain efficient diagnostic results from pathologists [3], and intraoperative frozen sectioning is the most widely used method [4]. It is a procedure in which the tissue cut during surgery is frozen in a frozen sectioning machine to allow for rapid cooling of the tissue through low-temperature conditions and then for production. Frozen sections are obtained during the process and are a significant guide to surgical treatment such as benign or malignancy for the patient's lesion which might not match the original diagnostic surgical plan, and then affect the operation procedure [5]. On the contrary, many primary care hospitals have limited clinical experience in issuing frozen pathology reports because of a severe shortage of pathologists, which often makes it difficult to meet clinical needs.

Over recent years, the application of deep learning artificial intelligence approaches in digital pathology has shown excellent results, both in the broad application of different tissue diseases such as liver cancer [6, 7], breast cancer [8–10] and esophageal cancer [11, 12] and in the optimization and updating of cellular tissues in detection [13], segmentation [14] and classification [15] methods, all of which show great potential. For example, deep learning methods enable networks for Digital whole slide images (WSIs) lymphocyte measurement and segmentation tasks simultaneously [14]; artificial intelligence approaches for lung cancer histopathology classification by supervised or weakly supervised strategies [16]; deep learning for prediction from H&E images of follicular thyroid tumors [17]; and differential diagnosis of follicular thyroid tumors based on histopathological images using deep learning techniques [18]. However, existing deep learning methods for histopathological classification of thyroid cancer have been studied with paraffin-embedded H&E sections WSIs, but ignore the significance of intraoperative frozen sections. In this study, we attempt to develop a deep learning-based five-class classifier for identifying a broader and more detailed range of thyroid lesions. We refined the frozen sections' analysis into five classification problem classes through differences in the histological characteristics of different cancers, including papillary thyroid carcinoma (PTC), follicular thyroid tumor (FTT), medullary thyroid carcinoma (MTC), anaplastic thyroid carcinoma (ATC), and non-cancerous thyroid lesion (NTC). ResNet [19], multi-processing, TensorRT [20], and graphics processing units (GPUs) were used to improve efficiency. To validate the performance, efficiency, and sensitivity of clinical applications, comparative model experiments, prior knowledge cascade experiments, time-cost acceleration analyses, and visualization of the effects of applications were carried out.

Due to the high resolution of digitized frozen slices, which can reach $100{,}000 \times 100{,}000$ pixels, direct complete WSIs processing with large-scale down-sampling loses many detailed features, which is not conducive to a classification approach regarding morphological features. Also training a convolutional neural network (CNN) on the whole slide tissue image (WSI) is computationally tricky to achieve. Cropping WSIs into tiles for analysis is a standard solution for studying histopathological WSIs by training a tiles-level classifier and later performing tiles fusion in an Expectation-Maximization (EM) manner to summarize the prediction results and obtain an effect map of the whole slice prediction [21]. Wei et al. [22] used sliding windows over the entire slide to generate small tiles and classified each tile by a neural network. They used heuristics to determine the primary and secondary histological patterns across the slide so that labels for different regions could be obtained.

In this study, the high accuracy performance and interpretability of our proposed system served as our primary goal. Against this backdrop, we ensured the accuracy of the tissue regions of candidate training tiles by double manual annotation of regions of interest (ROIs) by pathologists and pathologists. Then we segmented the localized tissue regions into small tiles and simulate real clinical diagnostic scenarios by a cascade diagnostic system to diagnose the presence of a cancerous lesion and subdivide the type of the cancerous lesion to predict the category of each tidbit separately. The final classification criterion was a fusion recovery of the entire tiles of the pathology slices, using a heuristic to identify the significant and minor histological patterns across the WSI. In addition, we confirmed the performance sensitivity of our system with test results of papillary thyroid microcarcinoma and different scale thermograms with accelerated plug-ins to speed up the diagnostic inference of the system. Our system not only allows for a more detailed classification of thyroid lesions but also allows for the system's sensitivity to localize suspicious lesions, saving diagnostic time and

cost and assisting the pathologist in making a diagnosis in copious ways. The main contributions of this paper can be summarized as follows:

- A tile-supervised WSI-label learning-based classification and diagnosis system for frozen thyroid lesions is designed. We propose a Two-step cascade diagnostic system (TSCD), which utilizes a priori knowledge cascading into a five-classification diagnostic system that fits the clinical diagnostic scenarios, and for the first time achieves fine-grained classification of lesions, realizes a high-accuracy classification task and demonstrating advanced performance.
- We conducted a systematic working test on papillary thyroid microcarcinoma alone. To ensure the accuracy of clinical application, we strictly selected cases with lesion sizes within 1 cm, and the results showed that a lesion size of 2 mm could be accurately localized.
- A diagnostic acceleration component is designed, this acceleration component accelerates the processing from multi-processing of data to TensorRT acceleration of model inference, reducing consumption time to 729.37% of the original speedup. In addition, this component can be embedded in all tiles-based classification inference models for broad applicability.
- A multi-scale thermographic visualization of the system is proposed to test the sensitivity, which is reflected by the selection of different scales of lesion areas from different cases to visualize the system's identification results due to the high resolution of the WSI.
- For the consideration of the clinical application of the diagnostic system, a heuristic strategy of tumor prioritization based on thresholds is proposed to clarify the diagnostic process by thresholding the WSI primary and secondary tissues in a heuristic manner to promote the development of deep learning methods for the clinical application of frozen histopathological images.
- A database of frozen thyroid digital pathology sections was established, and the developability of deep learning in frozen sections was confirmed by designing a deep learning model to analyze data and assist in diagnosing pathology library data.

Methods

In this section, we will describe the preparation of data, the selection of deep learning models, composition of acceleration components, and the generation of the system for the final frozen diagnosis of thyroid lesions from several perspectives.

Slides collection and categorization

We collected frozen thyroid sections from the First Affiliated Hospital of Sun Yat-sen University to form a library of frozen thyroid digital pathology sections Four certified pathologists independently reviewed the organized frozen thyroid sections. The diagnosis was determined with the aid of the results of paraffin-embedded H&E sections and immunohistochemistry staining sections. In order to ensure the accuracy of the data, three rounds of screening were conducted sequentially through diagnostic pathology reports, four pathologists, and two pathologists in the process of selecting the data, and finally the data that were all approved were selected. Figure 1 shows the morphology of five types of typical thyroid sections, Fig. 2 shows their high magnification histopathology, and Table 5 lists the clinical management of different thyroid lesions. It can be seen that differentiating the types of thyroid cancer can facilitate efficient and targeted clinical diagnosis and help patients heal better. This study was approved by the Ethics Committee of the First Affiliated Hospital, Sun Yat-sen University.

Thyroid cancer is divided into papillary thyroid cancer, medullary thyroid cancer, anaplastic thyroid cancer and follicular thyroid cancer. However, if a follicular thyroid tumor is determined to be benign and less than 1 cm in diameter, it can be treated with an observation instead of follow-up surgery [23]. Follicular thyroid cancer which is malignant follicular thyroid tumor requires demonstration of capsular and/or vascular invasion. Ultrasound, puncture, intraoperative frozen pathology, and even postoperative paraffin-embedded pathology may only partially guarantee accuracy [24]. So there exists a tendency for various findings to be benign but may be malignant, so it is not easy to fully determine whether it is benign or malignant. Therefore, our classification tends to be divided into five parts: papillary carcinoma, follicular tumor, medullary carcinoma, anaplastic carcinoma, and non-cancer lesion, where the noncancer frozen section consists of data from nodular goiter, chronic lymphocytic thyroiditis (CLT), and subacute thyroiditis (ST) together.

WSI datasets

The database of frozen thyroid WSIs we created consisted of 4409 frozen thyroid sections from the First Hospital of Sun Yat-sen University (SYSUFH) and the First Affiliated Hospital of Sun Yat-sen University (FAHSYSU), covering 3873 cases diagnosed as PTC, FTT, ATC, MTC or NTC, and the inclusion criteria for the dataset were that each frozen section was able to be included in the above categories. Before WSI annotation, all sections were observed by two senior pathologists at SYSUFH through microscopy, including immunohistochemically

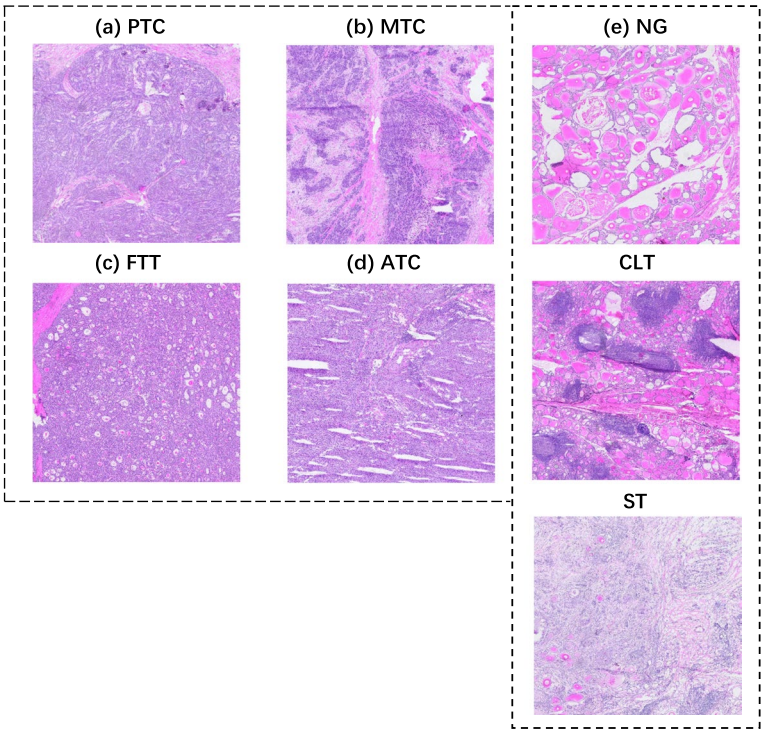

Fig. 1 Frozen section examples WSI, **A** for papillary thyroid carcinoma nodule, **B** for medullary thyroid carcinoma nodule, **C** for follicular tumor nodule, **D** for anaplastic carcinoma nodule, and **E** for all non-cancerous frozen section slides WSI, from top to bottom, for nodular goiter, chronic lymphocytic thyroiditis, and subacute thyroiditis

stained Sects. [25, 26] used as an aid to diagnosis. The diagnostic reports of all cases were also checked as a way to determine the accuracy of the collected data. Slides were then scanned with the Slide Scan Imaging System SQS-600P scanner at 40x magnification and digitized into SDPC format. For an unbiased evaluation, classification was performed according to the 5th edition of WHO thyroid tumor classification criteria. For the scanned digitized pathology slides, four pathologists first annotated WSI for regional ROI, and two pathologists then checked the annotated slides with at least seven years of clinical experience in the pathology department of SYSUFH to ensure the accuracy of the annotated information.

We collected three batches of frozen thyroid slices for database development, and all the data in the library were used in our system's training validation and testing. It contained 4409 thyroid frozen sections (PTC, 2300; FTT, 225; MTC, 49; ATC, 7; NTC, 1904) for model construction, randomly divided into training and testing 11 groups in a 10:1 ratio. Of these, 2894 (PTC, 1479; FTT, 224; MTC, 49; ATC, 7; NTC, 1135) were used to construct the model, 127 (PTC, 127) were used to evaluate the sensitivity of the model to small cancers, and 1388 (PTC, 694; FTT, 1; NTC, 769) were used in the external test set to evaluate the system's generalization ability. The number of follicular variant of papillary thyroid carcinoma (FVPTC) cases was 135, containing 163 WSIs, or 3.49% of 3873 cases, and 3.70% of 4409 WSIs, and non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) The number of cases was 11, containing 14 WSIs, representing 0.28% of 3873 cases and 0.32% of 4409 WSIs.

Based on the patch cropping method, 105622 papillary carcinoma patches, 42644 follicular tumor patches, 27491 medullary carcinoma patches, 8507 anaplastic carcinoma patches, and 215418 non-cancerous lesion patches were cropped from the training slides used for system construction. Of these, 78343 papillary carcinoma patches, 42407 follicular tumor patches, 27491 medullary carcinoma patches, 8507 anaplastic carcinoma patches, and 118895 non-cancerous lesion patches were cropped from the training slides used for model construction and the rest of the patches used for testing, as shown in Table 1.

Data pre-processing

The pathology slide scanner used to obtain the WSIs was the SQS-600P slide scanning imaging system from Shenzhen Shengqiang Technology Co. (Shenzhen, China). The labeling software is also the digital pathology reading

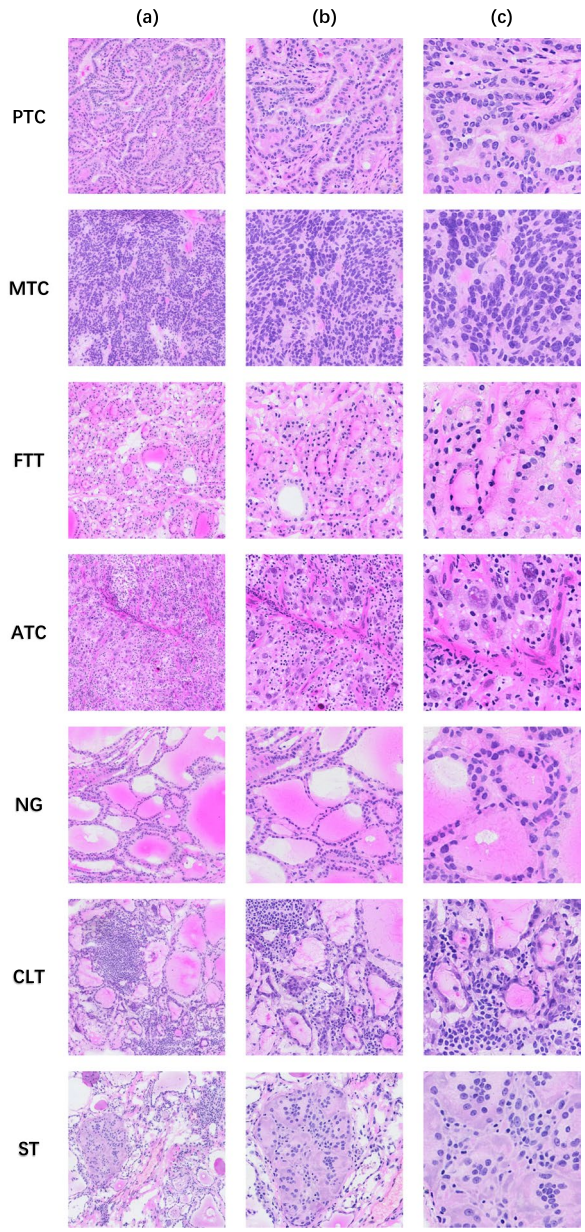

Fig. 2 High-magnification histopathology, **a** listed as WSI at 20x magnification, **b** listed as WSI at 40x magnification, and **c** listed as WSI at 80x magnification, from top to bottom, PTC, MTC, FTT, ATC, NG, CLT, ST

software ImageViewer provided by Shenzhen Sangqiang Technology Co., Ltd, version DPVIEW V2.0.1.0927. We use a supervised approach to train the model, which requires us to annotate the ROI region carefully.

The pathologists manually annotated the specific histopathological thyroid tissue types in each WSI by using colored irregular shapes, and the annotated WSIs were reviewed by two pathologists to ensure the accuracy of the annotation. The colored irregularly annotated regions were cut at the tile level after ROI region annotation to provide the tissue details of the cut patches, as shown in Fig. 3. Different categories of tiles obtained by cropping are shown in Fig. 4. All processing was performed at a magnification of 40x, based on a pixel scale of 0.2065 μm. We traversed all manual annotations by reading the SDPL file (the file generated by manual annotation). Use the multiplicity of the rectangle marked in the blank area as the reference coordinate system, and calculate the ratio of the reference coordinate system to the absolute coordinate system using the screen position of the colored outline, and then deflating the annotated points to the reference coordinate system, taking the center point of the annotation as the distribution of tiles is described in Table 1.

Deep neural networks

According to the research content and its applicability, what we need is a neural network framework with high accuracy while ensuring low cost. The more commonly used neural network backbones include ResNet19, DenseNet [27], EfficientNet [28], ShuffleNet [29], etc. In addition, with the extension of the transformer to the computer vision (CV) field, the generic backbone network Swin-Transformer [30] is also showing superior performance in vision tasks such as image generation, medical image segmentation, etc. performance. We choose ResNet101 as the backbone network for histopathology classification tasks. The residual structure of ResNet is proposed to solve the problem of network degradation, and the gradient disappearance or explosion has been translated by normalized initialization, etc. The model can make learning less and more manageable, thus ensuring sure accuracy while reducing the cost. And Resnet101 is used for our histopathology classification task by replacing the last fully connected layer with five outputs representing papillary thyroid carcinoma, medullary carcinoma, follicular tumor, anaplastic carcinoma, and non-carcinoma lesion, respectively. For training, testing and optimization, we divided all the datasets into a training set, a validation set, and a test set. Although we used more than 399682 patches of data for system development, we still trained the thyroid patch model by transfer learning from models pre-trained on the ImageNet recognition task. By initializing the pre-trained model, the patch classifier can achieve better performance. At the same time, we also fine-tune models such as Swin Transformer, EfficientNet-B5, and GoogLeNet [31] using the same data and settings as Resnet101 and also test the models using the same test set to make qualitative comparisons among different kinds of model results.

Table 1 Statistical information of the dataset used for model development testing. The dataset was categorized into three subsets, dataset A, dataset B and dataset C, depending on the purpose

Dataset	PTC			FTT			MTC			ATC			NTC			SUM		
	C	W	P	C	W	P	C	W	P	C	W	P	C	W	P	C	W	P
DA	1266	1479	78343	185	224	42407	43	49	27491	7	7	8507	986	1135	118895	2487	2894	275643
DB-Small	101	127	3220	-	-	-	-	-	-	-	-	-	-	-	-	101	127	3220
DC-East	619	694	24059	-	-	-	-	-	-	-	-	-	591	769	96757	1285	1388	120816
SUM	1986	2300	105622	186	225	42644	43	49	27491	7	7	8507	1577	1904	215418	3873	4409	399682

Dataset A was used for model development and evaluation. Dataset B is used to evaluate the sensitivity of the model on tiny cancers. Dataset C is used as an external test set to evaluate the system generalization capability. Datasets A, B were obtained from the First Affiliated Hospital of Sun Yat-sen University (FAH-SYSU), and dataset C was obtained from the East Hospital of the First Affiliated Hospital of Sun Yat-sen University

DA Dataset A, *DB* Dataset B, *DC* Dataset C, *C* Case, *W* WSI, *P* Patch

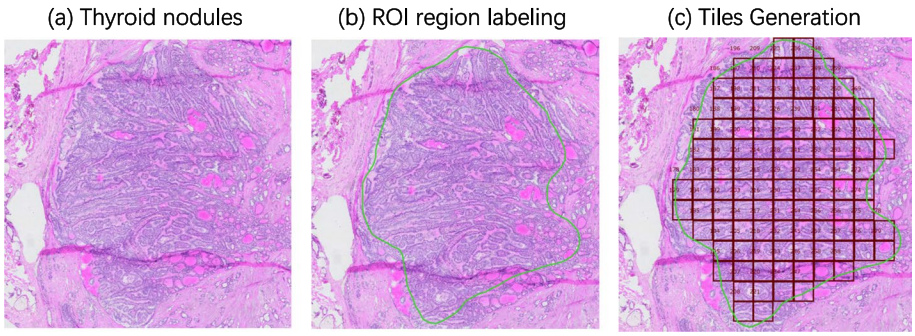

Fig. 3 Diagram of the process of cropping WSI into tiles. **a** shows the WSI of the original frozen thyroid nodule, and **b** shows the annotation of the contours by the pathologist, the green contour is the ROI area, i.e., the area with lesions and non-cancerous areas can also be present in the slides with lesions. figure (**c**) shows the tiles obtained by cropping from the annotated section, the number of patches obtained by cropping is proportional to the area of the annotated area

A priori knowledge and classification cascade

In image classification tasks, a priori knowledge can be obtained by pre-training on large-scale image datasets [32]. Through pre-training, the model can learn the basic features and patterns of the images, and the introduction of a priori knowledge as pre-training weights can effectively improve the accuracy and efficiency of image classification. Modeled after the actual diagnostic process of clinical thyroid freezing, the TSCD system was designed as a categorization cascade of two classification cascades for the presence of cancerous lesions and a categorization cascade of four classification cascades for subdivided lesion types. However, the pre-training dataset ImageNet [33] used now has a feature extraction method that does not fit the pathology image, and the training data of the two parts of the cascade model is not as much as that of the five-classification model, which is insufficiently fine-grained for the classification task. In order to make the cascade model have similar feature extraction ability as the five-classification model and fit the pathological images of thyroid lesions, the five-classification baseline model's is used as the pretrain model of the cascade model.

Network training and implementation details

Our training, validation and testing data are not publicly available due to strict controls on medical management and strict policies on sample privacy. Even though we have acquired several 399062 tiles for the model training process, we still apply transfer learning to our network model training to obtain better performance. By initializing the network with the default weights transferred from the ImageNet dataset, the entire network is fine-tuned to fit our data target better. The initial learning rate is set to 0.00003 and the optimizer is Adam. Momentum and attenuation are both set to 0.9. dynamic data enhancement, including horizontal flip or in the vertical direction, color dithering includes automatic adjustment of luminance in the range of (0.65, 1.35), automatic adjustment of contrast in the field of (0.5, 1.5), automatic adjustment of automatic saturation adjustment, and 0.3 size adjustment for hue to increase

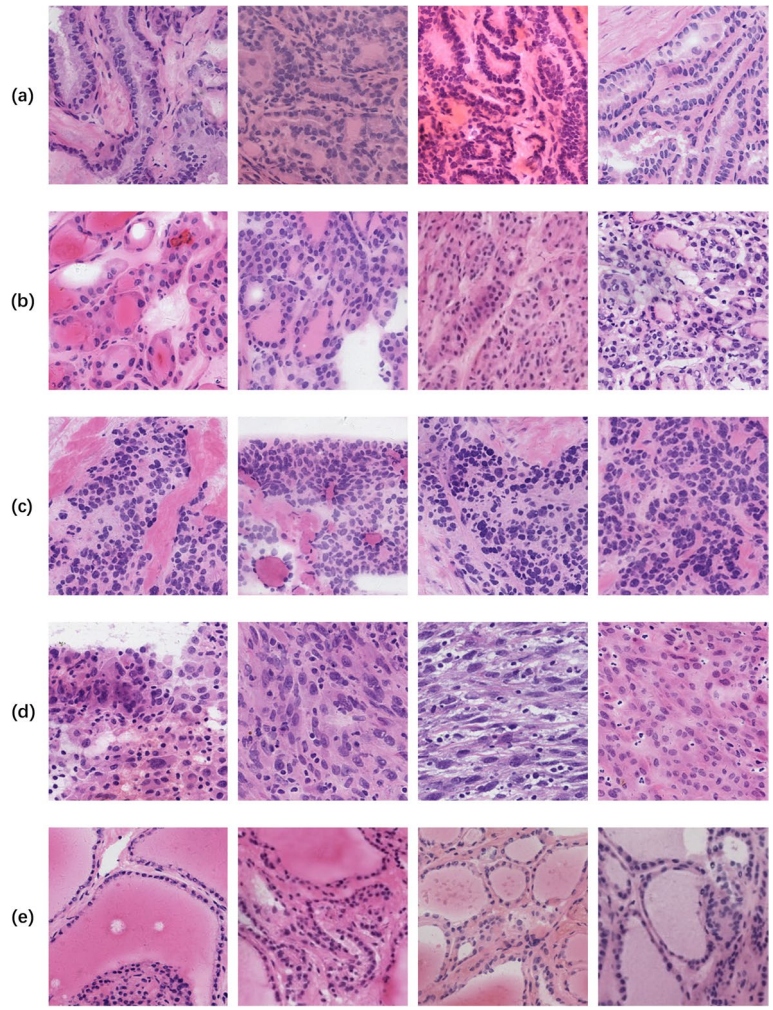

Fig. 4 Different categories of tiles obtained by cropping, **A** for PTC, **B** for FTT, **C** for MTC, **D** for ATC, and **E** for NTC, with a pixel size of 224 × 224

the data variation. To improve the learning characteristics of convergence, pixels were rescaled from 0 to 255 to 0–1 by dividing by 255, and the Z-score was normalized using the mean (0.485, 0.456, 0.406) and STD (0.229, 0.224, 0.225). The training process lasts 40 epochs, and the optimization model with the most negligible loss is saved and used. We used the deep learning framework PyTorch 1.10.2 to implement all CNN models, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory was used for CNN model training and evaluation. Training ResNet18, ResNet34 models took about 32 h, and Resnet50 model took about 44 h, and ResNet101 model took about 48 h. GoogLeNet model took about 28 h. Swin Transformer model took about 67 h. EfficienNet-B5 took about 47 h. Vision-Transformer took about 88 h. AlexNet, DenseNet121, VGGNet-16, ShuffleNet-v2 models took about 16 h, 36 h, 43 h and 25 h.

Tile splitting and prediction

For the sake of achieving the integrity of tissue details and clarity of morphological features, we cut the labeled area into non-overlapping tiles of 224*224 pixels in size. Tiles were filtered with the criterion of 50% or more of class-labeled components to obtain a dataset with a high degree of lesion fit from both WSI and tile-level perspectives. We use our deep learning system to perform classification prediction for these tiles from manually labeled regions. After obtaining the initial tile-level class labels, we perform WSI fusion according to our tiles fusion strategy to get the WSI-level label prediction, which is the auxiliary lesion diagnosis result. Figure 5 illustrates the diagnostic flow chart of a whole frozen thyroid section.

Whole-slide label inferencing with tiles fusion

Since the network outputs are tile-level predictions, the final diagnosis must be integrated into WSI-level

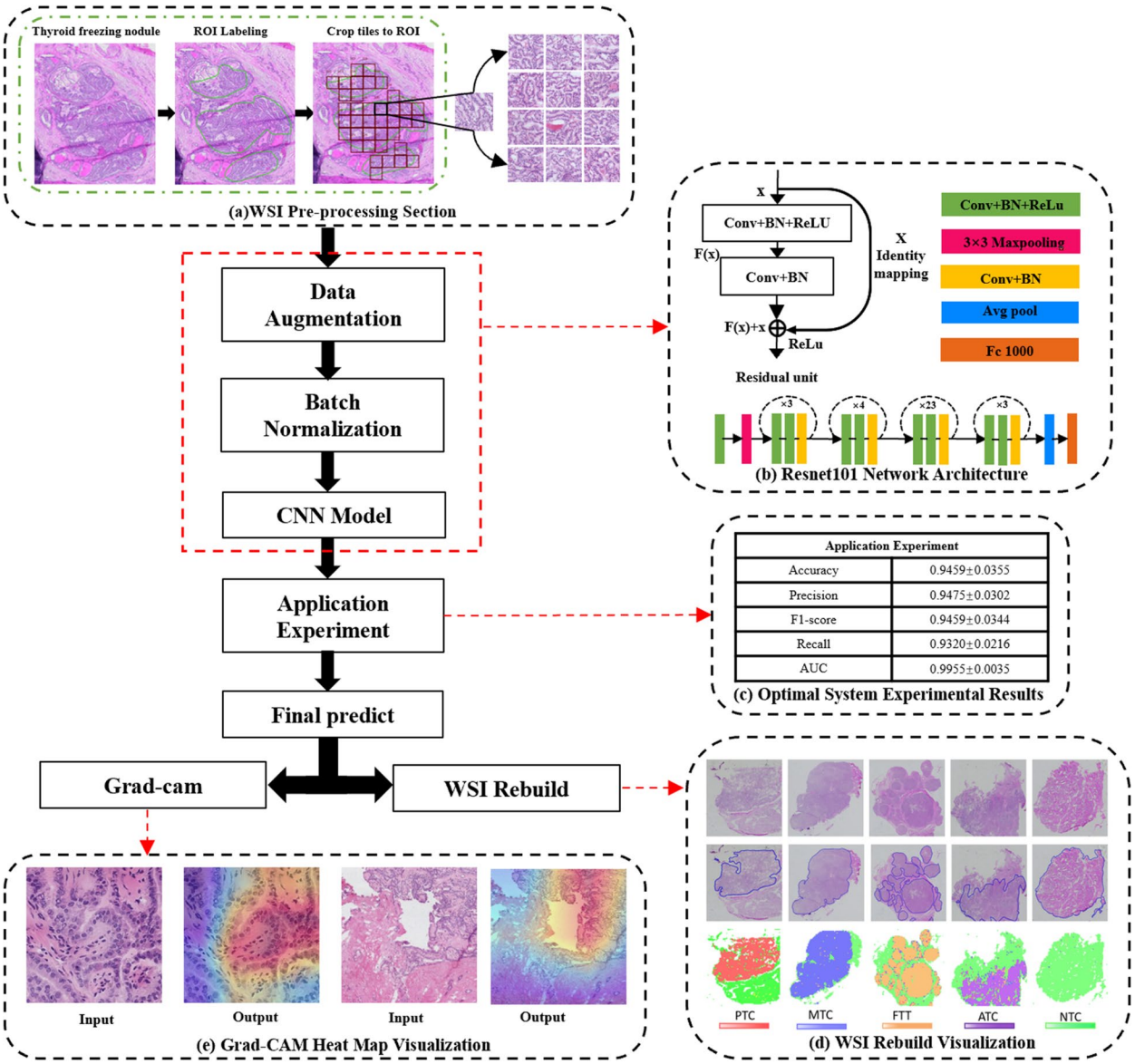

Fig. 5 Deep learning-based diagnostic workflow for frozen thyroid lesions. The pipeline contains five modules: **a** WSI pre-processing section, **b** Resnet101 network structure, **c** optimal system experimental results, **d** WSI reconstruction visualization, and **e** Grad-CAM (Gradient-weighted Class Activation Mapping) heat map visualization

forecasts. In general, the prediction results of slice-level WSI are aggregated to decide the classification based on tiles, which depend on the category to which the maximum probability belongs. Classical integration methods include averaging [34] and voting [35] procedures, which use more straightforward strategies to combine the predictions of individual learners. The majority voting method [36] is based on the classification results of a single classification model. It uses the principle of minority rule to determine the category labels predicted by the model, that is, by the number of tiles in each category and assigning the entire WSI to the one label with the highest number of corresponding types. The other is the average pooling method, where the probabilities of each category are summed and the slide labels are derived from the maximum average class probabilities [37].

In our data, tissue components of non-cancerous categories and other tumors may coexist in one WSI. For example, papillary thyroid cancer and non-cancerous parts are distributed in different regions of the same WSI. Still, the final label of one WSI is attributed to only one category. Inspired by the work of Li et al. [38], we

propose a threshold-based approach to compound the results based on the majority voting method incorporating a threshold division of pathological tissue priorities. When we encounter multiple tiles class labels co-located in a single WSI, equal treatment of tumor-containing and non-cancerous WSIs will result in the neglect of microscopic cancerous lesions and serious misclassification of the patient's diagnosis, such that the results we predict will not fit the requirements for putting into practical clinical use. Therefore, to improve the priority level of tumor-containing WSIs and enhance the sensitivity of the system to tumor regions, we propose a heuristic strategy based on threshold tumor priority, as shown in Fig. 6. Firstly, we set the threshold value according to the severity of lesions, because PTC, ATC, FTT, and MTC are all tumor types except the NTC category, so the threshold value of NTC is set to 95%. In contrast, all the rest of tumors are not differentiated by a threshold value. In addition, the pixel size of the tiles we selected was 224×224. Such a size is small even when compared with the lesion diameter of papillary thyroid microcarcinoma. As shown in Fig. 6, in our work for the prediction of WSI of papillary thyroid microcarcinoma with a lesion diameter ≤ 10 mm, the size of the lesion in the visualization result can be cropped to 20 tiles with continuity. Therefore, the category of this WSI was determined as NTC only when the weight of NTC exceeded a threshold value of 95% and the number of regional continuity tiles (RCT) was less than 20. In addition to the above, if there was a coexistence of NTC and other categories, all different categories were used as the main tissue component, and the same was valid for label selection. The task classification we implemented in this study is based on the visual features observable at the cellular or partial tissue level on the image tiles scale, so this strategy of fusing tiles is used.

Acceleration components

Due to the time-specific nature of intraoperative diagnosis, efficient diagnostic results need to be given in a short period of time, so we designed an acceleration component to enhance the diagnostic efficiency, and the component includes multiprocessing for cutting tiles and TensorRT to accelerate the inference.

Multiprocessing for cutting tiles

In the program of cutting tiles, multiple processes [39] can be run at the same time through the multi-process module, each process has its own independent execution space and resources, and can handle multiple tasks at the same time, here we use 9 processes, compared to the original time to get a single tile can now be obtained 9 tiles, in order to improve the concurrency and response speed of the system.

Diagnostic reasoning acceleration

PyTorch [40] is an open-source deep learning framework that can be used to build and train neural network models. ONNX (Open Neural Network Exchange) [41] is an open deep learning model exchange format that can be used to share and convert models between different deep learning frameworks. TensorRT [42] is an NVIDIA provided high-performance deep learning inference engine that can be used to accelerate the inference process of deep learning models. Our initial system was trained to obtain PyTorch models, and in model inference, to speed up the inference process, we converted Pytorch to ONNX files via the torch.onnx.

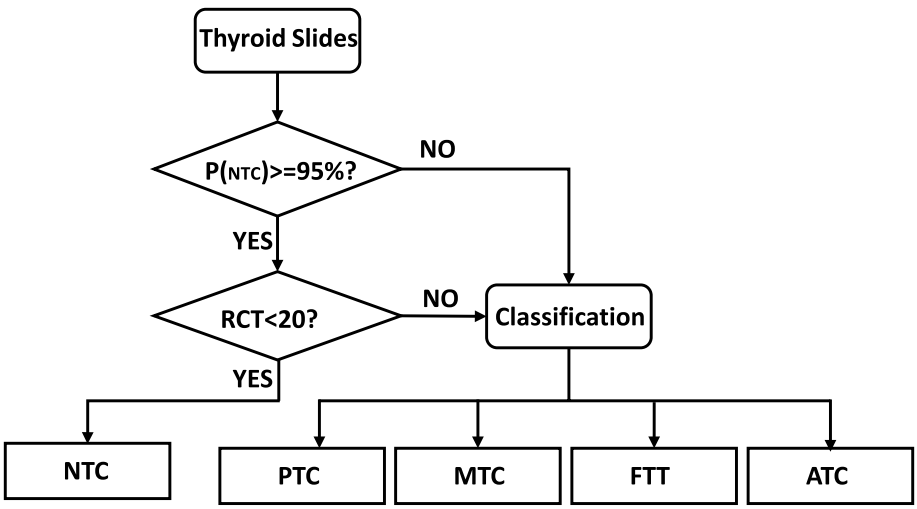

Fig. 6 Athreshold-based tumor-first frozen section diagnostic method for the thyroid is proposed

export() function, after which we used the TensorRT Python API to build the TensorRT engine with ONNX models, as depicted in Fig. 7 conversion flow so that high performance inference can be achieved in TensorRT using PyTorch-trained models.

Evaluation Metrics

In our experiments, the accuracy, precision, recall, and F1 score were used to evaluate the performance of our proposed method and the state-of-the-art techniques [43]. In the assay method, the actual category value and the prediction were the same and both were positive, then TP (True Positive); if both were negative, then TN (True Negative); the actual category value did not agree with the prediction and the prediction was positive, then false positive (FP) if the forecast was negative, then false negative (FN). Based on these basic definitions, additional evaluation metrics (Accuracy [A], Precision [P], F1-Score [F1], Recall [R], Specificity[S], TPR [True Positive Rate], FPR [False Positive Rate]) of the frozen thyroid tissue diagnostic test can be introduced as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R} \tag{3}$$

$$TPR = Sensitivity = Recall = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{TN + FP} \tag{6}$$

Results

Patch-based classification

In the twelve experimental ones of the Convolutional Neural Networks (CNN) model, the model's performance is optimized by fine-tuning the model. Among them, Resnet34, Resnet50 and Resnet101 models perform close to each other in the fine-tuned training mode, but Resnet101 shows a slight lead. Table 2; Fig. 8 offer a qualitative comparison between the system backbone network Resnet and other classical deep learning models. These results compare and validate the superior results of selecting the Resnet34, Resnet50 and Resnet101 models for the classification of frozen thyroid lesions by transfer learning, and the final visualization of the whole slice prediction is in excellent agreement with the ground. The absolute consistency of the entire prophecy is also extremely high.

Three backbone models were selected for the cascade test since the performance differences on the five-category baseline of Resnet34, Resnet50 and Resnet101model were small. Figure 9 compares the evaluation of Accuracy [A], Precision [P], F1-score [F], and Recall [R] of the direct cascade with ImageNet as the pre-training model and the TSCD system with the five base classifications of thyroid lesions as the pre-training model under the three models metrics. The experimental results show that whether it is Resnet34, Resnet50 or Resnet101 model, the evaluation indexes from BASE classification to direct cascade to pre-train cascade system are incremental effect, as this can prove the validity of our proposed cascade system modeled on clinical diagnosis. Since the cascade of the

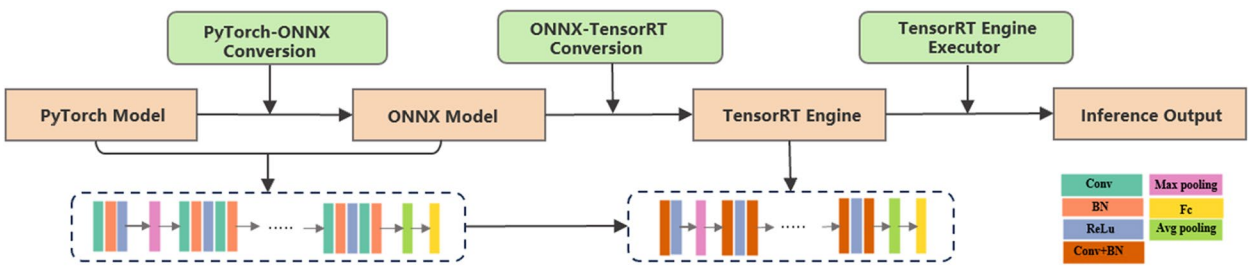

Fig. 7 TensorRT accelerates the PyTorch model inference process, which consists of three steps:(1) export the PyTorch model to an ONNX file, (2) build the TensorRT Engine, and (3) deploy the TensorRT Engine

Table 2 Average test ACC, Precision, F1-Score, recall with standard deviation on test dataset

Model	Accuracy	Precision	F1-score	Recall	AUC
ResNet18	0.9257 ± 0.0352	0.9277 ± 0.0285	0.9254 ± 0.0342	0.9007 ± 0.0200	0.9920 ± 0.0040
ResNet34	**0.9365 ± 0.0244**	**0.9389 ± 0.0206**	**0.9367 ± 0.0232**	**0.9178 ± 0.0278**	**0.9938 ± 0.0038**
ResNet50	**0.9348 ± 0.0319**	**0.9373 ± 0.0271**	**0.9344 ± 0.0298**	**0.9140 ± 0.0139**	**0.9933 ± 0.0053**
ResNet101	**0.9370 ± 0.0317**	**0.9392 ± 0.0264**	**0.9369 ± 0.0303**	**0.9170 ± 0.0366**	**0.9938 ± 0.0048**
GoogleNet	0.9086 ± 0.0409	0.9111 ± 0.021	0.9083 ± 0.0257	0.8760 ± 0.0217	0.9860 ± 0.0050
EfficientNet-B5	0.9089 ± 0.0344	0.9134 ± 0.0252	0.9085 ± 0.0305	0.8750 ± 0.0177	0.9873 ± 0.0063
Swin-Transformer	0.9085 ± 0.0314	0.9109 ± 0.0284	0.9065 ± 0.0238	0.8697 ± 0.0217	0.9877 ± 0.0057
Vision-Transformer	0.9055 ± 0.0284	0.9080 ± 0.0219	0.9030 ± 0.0284	0.8650 ± 0.0237	0.9773 ± 0.0077
AlexNet	0.8819 ± 0.0441	0.8856 ± 0.0245	0.8801 ± 0.0380	0.8357 ± 0.0230	0.9795 ± 0.0115
DenseNet121	0.9128 ± 0.0277	0.9150 ± 0.0176	0.9119 ± 0.0242	0.8757 ± 0.0226	0.9882 ± 0.0052
VGGNet-16	0.9302 ± 0.0352	0.9315 ± 0.0312	0.9298 ± 0.0340	0.9067 ± 0.0173	0.9917 ± 0.0057
ShuffleNet-v2	0.8631 ± 0.0428	0.8688 ± 0.0282	0.8631 ± 0.0407	0.8041 ± 0.0197	0.9723 ± 0.0133

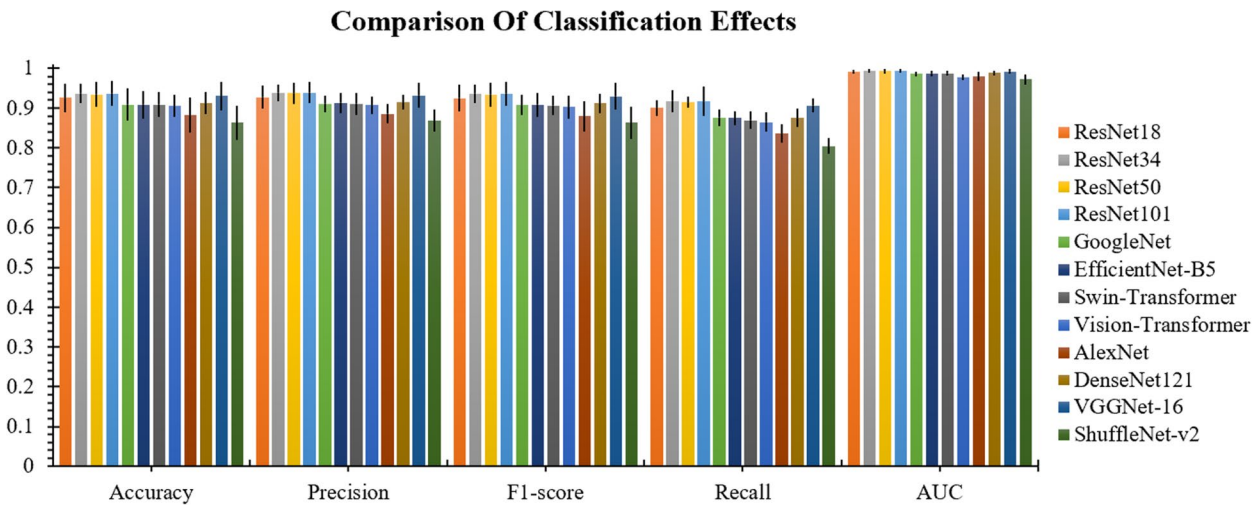

Fig. 8 Comparison of classification effects. It can be seen from the figure that Resnet34, Resnet50, Resnet101 are higher than other models in all four evaluation indexes

five classifications of thyroid lesions as a pretrained model yielded more accurate results than the direct cascade using ImageNet as a pretrained model, and the overall performance of the cascade of Resnet101 outperformed that of the other models in the system, we ultimately chose Resnet101 as the model that constitutes the backbone of our TSCD system for diagnosing frozen thyroid lesions. The final test results of our TSCD system were 94.59% ± 3.55% in Accuracy, 94.75% ± 3.02% in Precision, 94.59% ± 3.44% in F1-score, 93.20% ± 2.16% in Recall, and the value of AUC was 99.55% ± 0.35% (Table 3).

Internal and external test sets

Internal and external test sets were selected to evaluate the generalization ability of the diagnostic system. The external test set mainly focuses on the two categories of papillary thyroid cancer and non-cancer, which is due to the imbalance in the distribution of thyroid lesions themselves, with papillary cancer accounting for a much larger proportion than other lesions. Therefore, we selected the binary classification model for the presence of cancer in the first part of the TSCD system to be used to test the effectiveness of the diagnostic system on an external test set. At the same time, we divided the internal data originally used to test the five-classification diagnostic system

Table 3 Average test ACC, Precision, F1-Score, recall with standard deviation on cascade system

Model	Accuracy	Precision	F1-score	Recall	AUC
ResNet34	0.9365 ± 0.0244	0.9389 ± 0.0206	0.9367 ± 0.0232	0.9178 ± 0.0278	0.9938 ± 0.0038
ResNet50	0.9348 ± 0.0319	0.9373 ± 0.0271	0.9344 ± 0.0298	0.9140 ± 0.0139	0.9933 ± 0.0053
ResNet101	0.9370 ± 0.0317	0.9392 ± 0.0264	0.9369 ± 0.0303	0.9170 ± 0.0366	0.9938 ± 0.0048
ResNet34_direct _cascade	0.9381 ± 0.0268	0.9404 ± 0.0253	0.9384 ± 0.0267	0.9216 ± 0.0149	0.9947 ± 0.0023
ResNet50_direct _cascade	0.9429 ± 0.0350	0.9446 ± 0.0312	0.9430 ± 0.0343	0.9227 ± 0.0361	0.9952 ± 0.0038
ResNet101_direct _cascade	0.9419 ± 0.0307	0.9447 ± 0.0276	0.9422 ± 0.0304	0.9273 ± 0.0315	0.9955 ± 0.0035
ResNet34_pretrain_cascade	0.9385 ± 0.0321	0.9412 ± 0.0289	0.9389 ± 0.0314	0.9233 ± 0.0290	0.9945 ± 0.0035
ResNet50_pretrain_cascade	0.9448 ± 0.0409	0.9465 ± 0.0351	0.9448 ± 0.0398	0.9309 ± 0.0188	0.9953 ± 0.0053
ResNet101_pretrain_cascade	**0.9459 ± 0.0355**	**0.9475 ± 0.0302**	**0.9459 ± 0.0344**	**0.9320 ± 0.0216**	**0.9955 ± 0.0035**

into two categories for the comparison test between the internal and external test sets. The test results are shown in Fig. 10. The accuracy of our model on the external test set reaches 99.37% ± 0.15%, and the accuracy of the internal test data is slightly lower than the external data set because the complexity of the internal data is greater than that of the external test set, but the overall effect is still as high as 97.62% ± 0.58%. The accuracy of our diagnostic system is 97.62%±0.58% for the binary classification of cancer or not, and 99.37%±0.15% for the external test set, which shows the strong generalization ability of our system.

Visualizing predictions with heatmaps

Grad-CAM (Gradient-weighted Class Activation Mapping) can assist in analyzing the regions of interest of the system for a given class, and we also verify whether we have learned the correct features or information through the regions of interest of the network in turn. We use Grad-CAM to draw the heat map of Fig. 11. From these heat maps, we can find that our system achieves accurate localization and recognition of the regions that we classify for attention.

Slide-level performance

To obtain the predictions on the whole slide image, we mapped the system predicted tile-level results onto the original frozen sections, with different color blocks representing different categories. Figure 12 selected slices with different labels, the original WSI, the pathologist annotated WSI and the system slice-level prediction results are plotted, which visually shows the high agreement between our system classification prediction area and the pathologist annotated area.

Test on papillary thyroid microcarcinoma

To test the sensitivity of the system and for more accurate results in clinical use, we developed a sensitivity testing strategy for papillary thyroid microcarcinoma. we screened papillary thyroid microcarcinoma by lesion diameter and used them to test the system's sensitivity. We screened 101 frozen WSIs of papillary thyroid microcarcinoma, whose lesion diameters were all within the range of 1 cm, and obtained more than 3000 tiles by cropping them and inputting them into the system for testing. And obtained visualization results displayed in Fig. 13, which can accurately identify the cancerous regions, verifying the accuracy of the system.

As can be seen from Slide-level diagnosis Fig. 13, our system can accurately identify cancerous lesions even in the face of lesions as small as 2 mm in diameter when testing papillary thyroid microcarcinoma cases.

Multi-scale Thermographic visualization

We proposed a multi-scale heat map visualization, which, due to the high resolution of WSIs, the lesion areas of different cases and different scales were selected, as in Fig. 14, we chose the WSIs of four papillary carcinoma cases and cropped the four WSIs at different scales, and the pixel sizes obtained were 224×224, 512×512, 1120×1120, 2240×2240 in order, and the results of the system in different scale tiles for grad-cam heat map

(See figure on next page.)

Fig. 9 Better performing comparison in cascade system. The top three comprehensive performances of the above models are selected for a more visual comparison. **a** Comparison of the metrics performance of the three selected models on base classification and TSCD system. **b** Confusion matrix at the tile level predicted by the basis of the three models. **c** ROC curves and their AUC values for the three models on the base categorization and the TSCD system under cross-validation. It can be seen that Resnet101_pretrain_cascade is superior to other models in every index

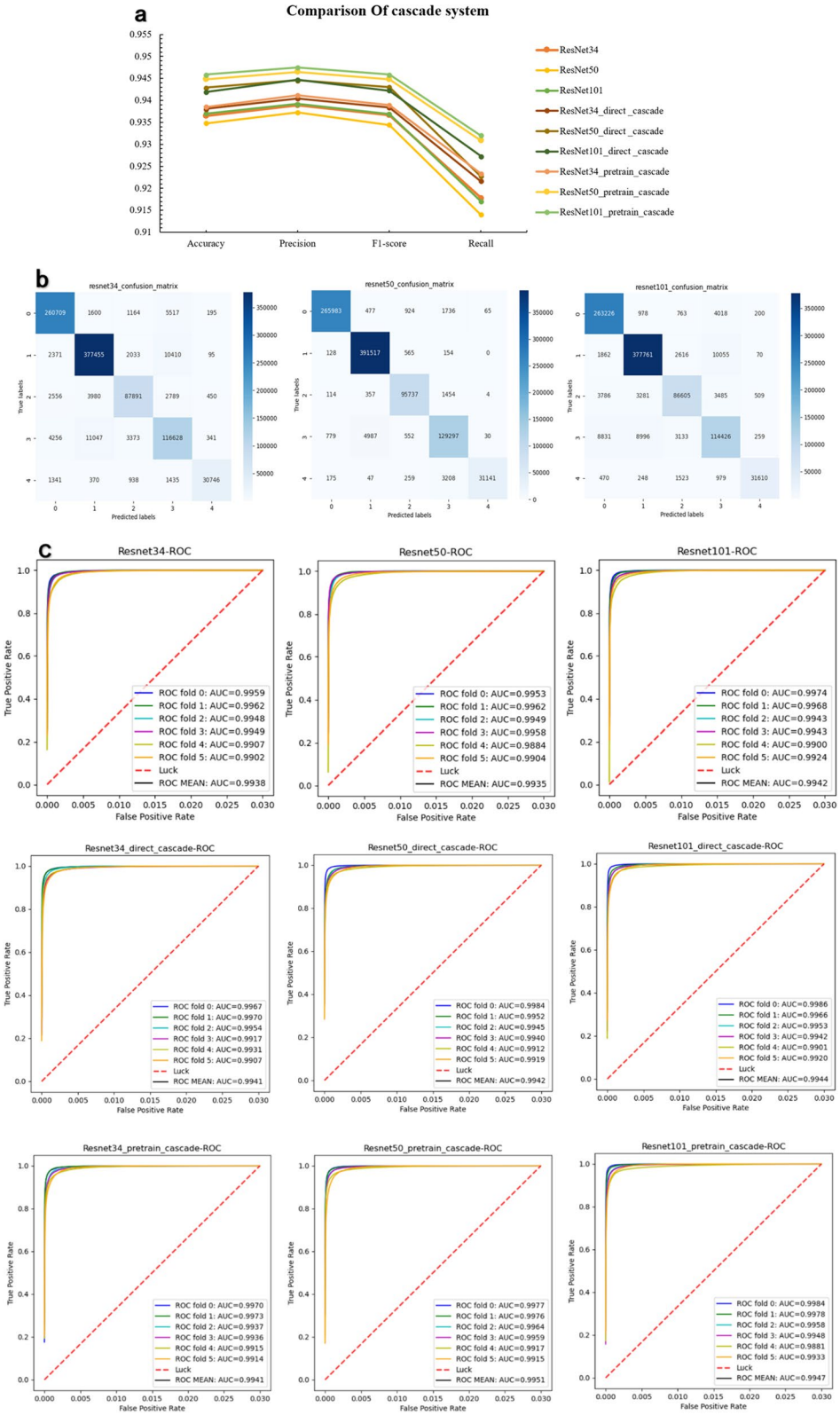

Fig. 9 (See legend on previous page.)

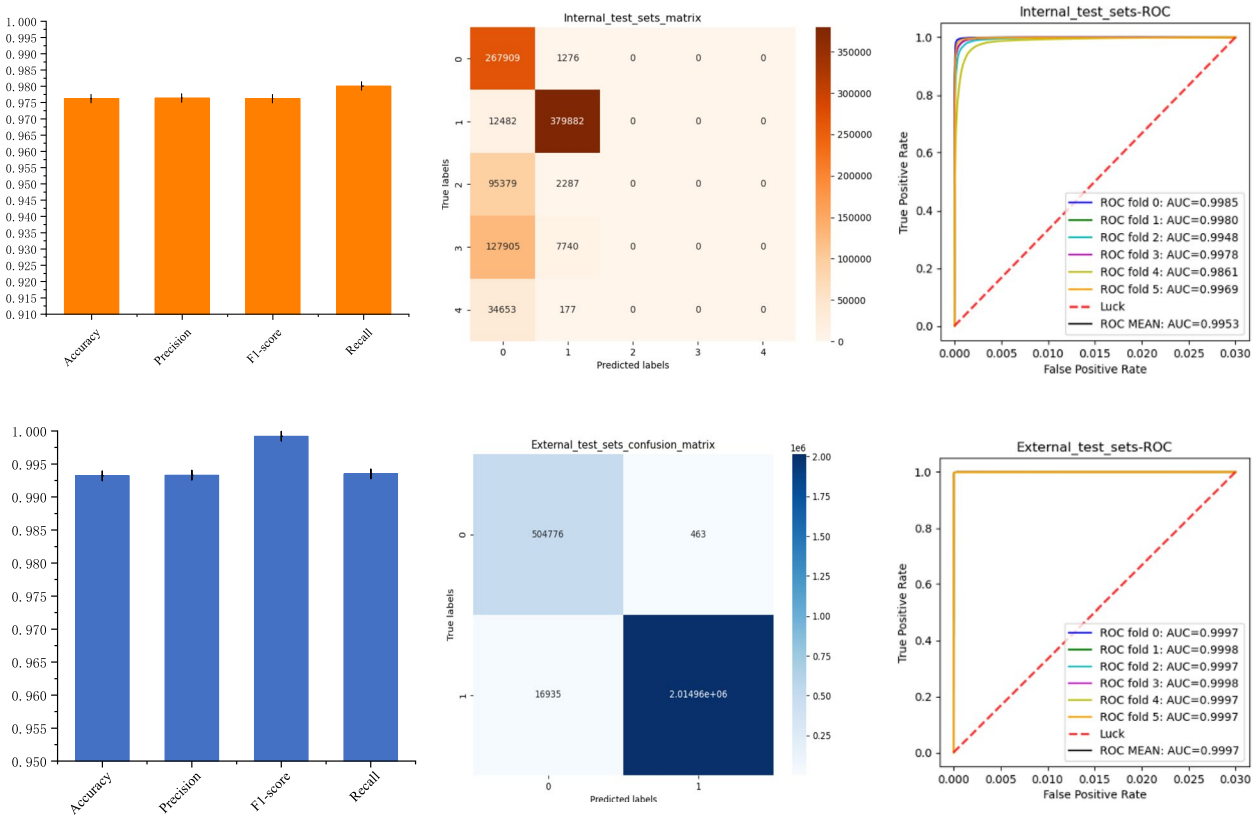

Fig. 10 The first part of the TSCD system, i.e., the dichotomous classification of the presence or absence of cancer, was selected for the test. The first row shows the effects on the internal test set originally used to construct the diagnostic system, and the second row shows the effects on the external test set obtained from the East Hospital of the First Affiliated Hospital of Sun Yat-sen University. The effects on the four metrics Accuracy [A], Precision [P], F1-score [F], Recall [R], the confusion matrix, and the AUC performance are shown from left to right

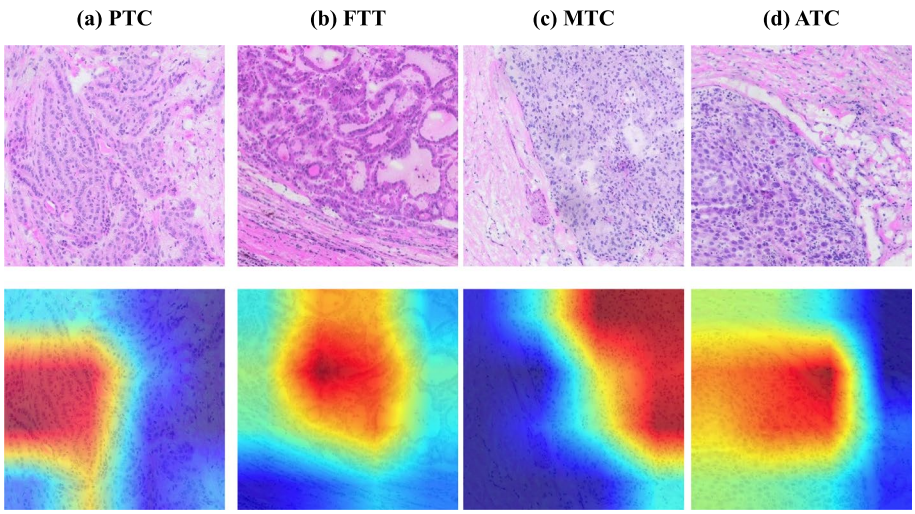

Fig. 11 Grad-cam heat map visualization of tumor classification, **a** tile level visualization for papillary carcinoma, **b** tile level visualization for follicular tumor, **c** tile level visualization for medullary carcinoma, and **d** tile level visualization for anaplastic carcinoma

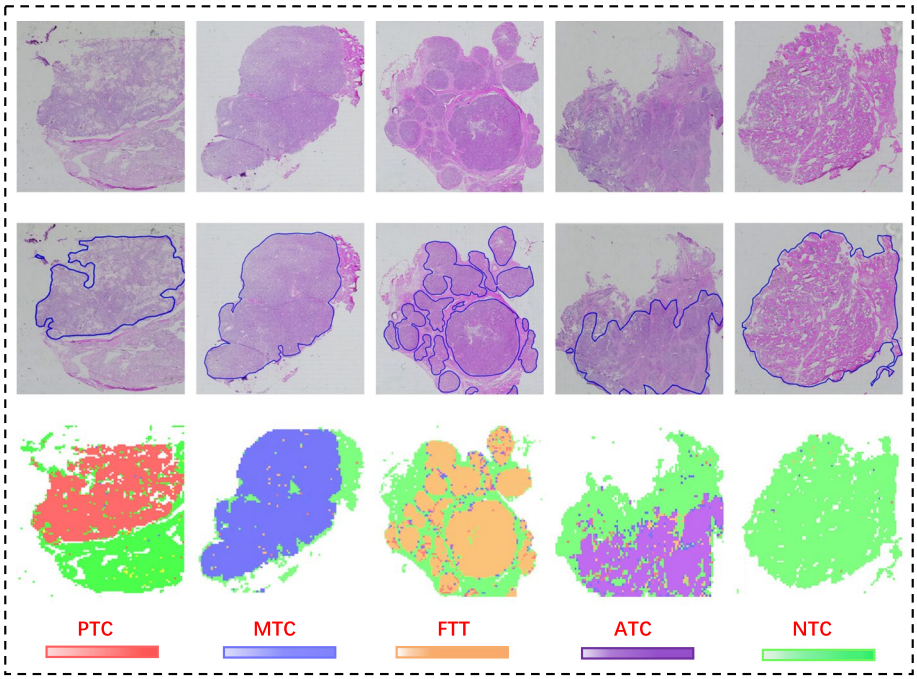

Fig. 12 From left to right, the slide-level visualization (performance) of PTC, MTC, FTT, ATC, and NTC tissue predictions are shown. The first row shows the original slides, the second row shows the annotated WSIs with closed blue curves depicting the ROIs annotated by the pathologist, and the last row shows the corresponding WSI prediction visualization

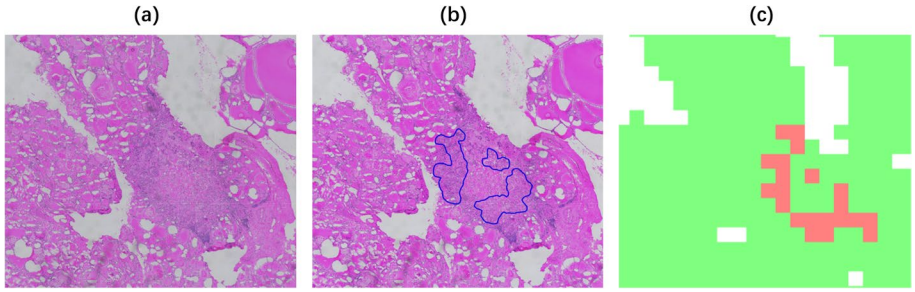

Fig. 13 Test of papillary thyroid microcarcinoma, the example of papillary thyroid microcarcinoma with lesion size of 2 mm in diameter, **A** original WSI, **B** manually labeled ROI, **C** Slide-level visualization display, the result can accurately identify the cancerous area

visualization [44], reflecting the sensitivity of the system. Figure 14 shows the classification effect of our deep neural network system in the form of a Grad-CAM visual thermal map. Even when faced with tiles of different scales, the diagnostic system was able to pinpoint the area where the tumor was located.

Accelerated component results in diagnostic time cost

After the optimization, excluding the necessary time to read and store the image brought by the hardware, the average time to cut the whole WSI is reduced from 2899.89s to 346.12s, which is an 837.83% speed improvement. The average inference time of the model's individual WSIs is reduced from 343.84s to 98.61s, which is a 348.69% speed improvement. The system's overall inference and diagnosis time is reduced from 3243.73 s to 444.73s, an improvement of 729.37%, as shown in the details in Table 4. As intraoperative freezing has very strict time requirements, doctors need to accurately grasp the freezing time to ensure the safety and effectiveness of the procedure. The shortened time helps the effective intraoperative diagnosis, thus further improving the efficiency and success rate of the surgery.

The primary time cost of WSI analysis lies in the data processing and prediction results of the WSI and its visualization generation. the level down-sampling

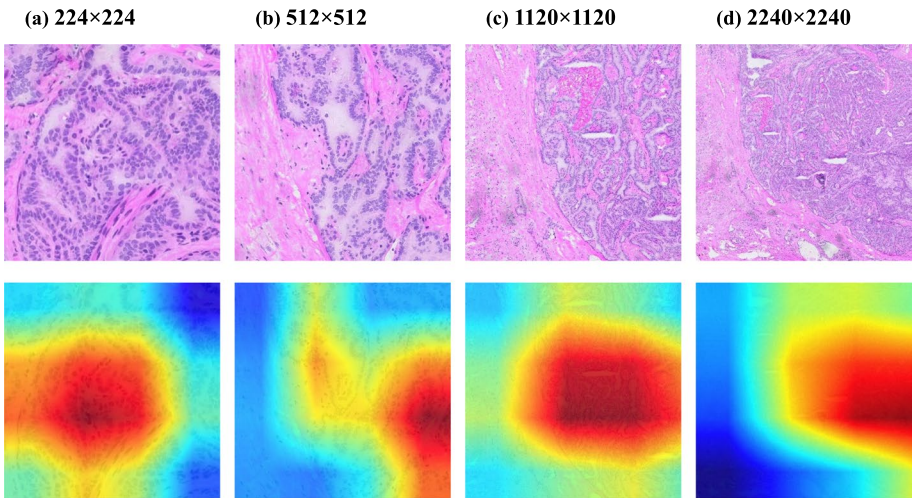

Fig. 14 Grad-cam heat map visualization of different scales of tumor classification, illustration of papillary thyroid carcinoma. The pixel size of figure (**a**) is 224×224, figure (**b**) is 512×512, figure (**c**) is 1120×1120, and figure (**d**) is 2240×2240

Table 4 Time consumed by the diagnostic system under the acceleration components

	1	2	3	4	5	6	7	8	9	10	Average	Reduction factor
A1	723.01s	3791.32s	1556.64s	6710.05s	1210.97s	1644.73s	2179.73s	3791.19s	2474.71s	4916.52s	2899.89s	-
A2	86.52s	434.15s	194.45s	764.17s	150.5s	205.9s	278.42s	464.65s	307.37s	575.1s	346.12s	837.83%
B1	86.61s	306.98s	123.46s	1258.66s	96.52s	133.79s	194.68s	376.91s	303.8s	556.98s	343.84s	-
B2	33.41s	106.67s	48.27s	388.03s	43.12s	42.77s	54.32s	127.82s	76.34s	159.96s	108.07s	318.16%
B3	30.17s	85.14s	58.03s	388.97s	31.08s	40.16s	49.89s	100.88s	66.35s	135.46s	98.61s	348.69%
A1 + B1	809.62s	4098.3s	1680.1s	7968.71s	1307.49s	1778.52s	2374.41s	4168.1s	2778.51s	5473.5s	3243.73s	-
A1 + B2	756.42s	3897.99s	1604.91s	7098.08s	1254.09s	1687.5s	2234.05s	3919.01s	2551.05s	5076.48s	3007.96s	107.84%
A1 + B3	753.18s	3876.46s	1614.67s	7099.02s	1242.05s	1684.89s	2229.62s	3892.07s	2541.06s	5051.98s	2998.5s	108.18%
A2 + B1	173.13s	741.13s	317.91s	2022.83s	247.02s	339.69s	473.1s	841.56s	611.17s	1132.08s	689.96s	470.13%
A2 + B2	119.93s	540.82s	242.72s	1152.2s	193.62s	248.67s	332.74s	592.47s	383.71s	735.06s	454.19s	714.18%
A2 + B3	116.69s	519.29s	252.48s	1153.14s	181.58s	246.06s	328.31s	565.53s	373.72s	710.56s	444.73s	729.37%

A1: Original cutting time, A2: Multi-process cutting time, B1: Pytorch Model, B2: ONNX Model, B3: TensorRT Model. B1, B2, B3 are the system inference times under different models. Numbers 1–10 are randomly selected WSIs for different cases, and the first item in each category in the Reduction factor is the original comparison object

factor in the WSI pyramidal storage structure is 2, and the annotated ROI size is one-fourth of the slide height and width, Then the size of the ROI region is 10,000 × 10,000 pixels in the level 2 slide image, which corresponds to 40,000 × 40,000 in level 0 of the WSI, and the size of the tiles extracted by the CNN model is 224 × 224.A frozen slide was scanned using a slide scanning imaging system SQS-600P scanner, and the time from getting the frozen slide to scanning to obtain the WSI was within two minutes. The scanned WSI can be directly processed by our system first by cropping the whole WSI into tiles of 224 × 224 size, and the cropped tiles are directly input for the entire slide prediction module in the system for visualizing the effect, the whole process is strictly controlled within 10 min, which can well assist the pathologist's diagnosis.

Discussion

Up to now, histopathology is still the gold standard for pathology doctors to diagnose diseases, and its development from slides under the light microscope to digital pathology images shows the significance of artificial intelligence in medical treatment. As the incidence of thyroid cancer is increasing year by year worldwide, the advancement of AI-assisted approaches to update it has received widespread attention.

Reasons for classification task selection

In recent years, there has been good development in the field of research on the artificial intelligence-assisted diagnosis of the thyroid, but almost the vast majority of the findings have focused on paraffin Sects. [45–48], and few studies have been performed on frozen sections

of thyroid cancer in clinical surgery [38, 49]. The only articles that are available only address the benignity and malignancy of nodules in thyroid sections, and the management of different thyroid cancers in clinical surgery is exceptionally variable [50, 51], as shown in Table 5.

Due to time constraints in surgery, protocols for different thyroid cancers must be adjusted and determined as soon as possible. Therefore, to solve the practical clinical problem, we designed a five-class classifier for frozen thyroid lesions, through which we can cover a broader range of thyroid lesions and assist pathologists in obtaining efficient results. Since thyroid cancers are morphologically distinct, such as papillary thyroid carcinoma nuclei showing variations in size and shape, such as enlargement, lengthening, overlapping and crowding, our deep learning model centers around morphological features for the classification task.

Development of a database of frozen thyroid WSIs

We built a Frozen Thyroid WSIs Database with 3873 cases, 4409 frozen thyroid slices containing WSIs for papillary thyroid cancer, medullary carcinoma, anaplastic carcinoma, follicular tumor, nodular goiter, chronic lymphocytic thyroiditis, and subacute thyroiditis. All the data in the database were used in our diagnostic system, of which 2894 were used for system building. The creation of the Frozen Thyroid WSIs Database lays the foundation for research into rapid intraoperative diagnosis, introducing deep learning into the surgical process. In addition, the database broadens the idea of integrating artificial intelligence into clinical care and confirms the potential of deep learning approaches in the freezing field. We will also continue to collect samples by enriching the database in terms of sample types and numbers in the future.

Distinction between FVPTC and NIFTP

In the classification diagnosis, distinguishing between NIFTP and FVPTC is crucial for the improvement of model accuracy. FVPTC and NIFTP have partial similarity in pathological features, but still have obvious distinguishing features. Nuclei in FVPTC are usually larger, heteromorphic, and irregular, whereas nuclei in NIFTP are more regular, smaller, and relatively homogeneous in shape. The follicular structure in FVPTC may appear disrupted, misshapen, or atypically arranged and have more microstructural variations. The follicular structure in NIFTP usually remains relatively regular and organized and lacks significant disruption or heterogeneity [52–55]. These pathological feature differences can be effectively learned and utilized by deep learning models to achieve better results in classification tasks.

Patch size selection

Our proposed system is a neural network model based on tiles classification so the selection of tiles has a particular influence on the model results to a large extent. We use tiles of 224×224 pixels in size, which firstly ensure that there is no significant loss of image information, and secondly, our model is mainly based on the extraction of morphological features for classification learning, and the selected tiles are also as detailed as possible to show each morphological feature to assist in making a reliable diagnosis. By comparing the performance of the benign and malignant classification work done by Li et al. [38], the size of the tiles selected by them is 2392×2392, and the average accuracy of our results presents better results under the same model action, which is also reflected in the tiles-based classification model, where our tiles size selection can better capture the morphological features of the lesion. In addition, the chosen tiles size allows the model to diagnose a single thyroid slide within 5 min in most cases, satisfying the time requirement. Therefore, considering the perspective of clinical application, the size of this tile can fulfill both the model feature extraction needs and the diagnostic time limit, thus better diagnosing from the standpoint of simulated pathologists.

Table 5 Diagnostic modalities of different cancerous lesions of the thyroid gland

Subsets	Clinical Method
PTC	If the cancer is still confined to the gland and there is no metastasis in the cervical lymph nodes, all of the affected glands together with the isthmus can be excised, and most of the contralateral gland can be excised. Additional cervical lymph node dissection is not necessary. If there is metastasis in the cervical lymph nodes, the lymph nodes on the affected side should be removed simultaneously.
FTC	In follicular carcinoma, even if the cancer is confined to one side of the gland, both glands should be excised together with the isthmus. If no cervical lymph node metastasis exists, cervical lymph node dissection is not required.
MTC	The scope of surgery for medullary carcinoma is to remove all the glands on both sides with the same isthmus. Since medullary carcinoma appears early with cervical lymph node metastasis, the affected side or both cervical lymph nodes should be removed simultaneously.
ATC	Anaplastic carcinoma overgrows and has high malignancy, usually infiltrative growth, and the possibility of surgical resection is slight. To prevent respiratory difficulties caused by cancer development, tracheotomy can be performed and the comprehensive treatment of surgery, chemotherapy and radiotherapy can be used.

As the number of cases of different lesions is relatively unbalanced, this may have some impact on the generalizability of the study's findings and the limitations of its practical application. One of the goals of the study is to generalize to a wider population or context to strengthen the validity and reliability of its practical application. We will continue to collect samples, firstly, to contain a relatively small number of instances, such as the part of anaplastic carcinoma; secondly, to collect some external data sets, and combine the two aspects with improving the robustness and generalization ability of the model.

Assistant to pathologists

Our work demonstrates that deep learning can have a fine-grained aid to the intraoperative frozen section classification diagnostic work, which helps pathologists to develop surgical strategies efficiently. On the one hand, our system can analyze digital pathology sections while the pathologist is studying frozen sections under a light microscope, visualizing the distribution of lesions and areas of cancer by a grad-cam heat map or tiles fusion. The two complement each other and work together to improve the accuracy of the pathologist's judgmental findings. On the other hand, our system can also complement the pathologist's diagnosis with specific details, for example, in areas easily missed, thus complementing each other. However, the deep learning approach is used only to aid the pathologist's diagnosis and cannot directly replace the pathologist's independent work.

Conclusion

In this paper, we propose a deep learning model classification system based on tiles to build a Frozen Thyroid WSIs Database, which is the first work on the fine classification of frozen thyroid carcinoma slices and the first digital database for frozen slices. We finally used the Resnet101 model as the final development model by comparing multiple models. We tested our data through the diagnostic system and obtained a high accuracy result of 94.59%±3.55%. Meanwhile, we proposed sensitivity tests for both systems to reflect the system's accurate control of cancer diagnosis by pooling multiple scales. In addition, to consider the clinical application, we developed a threshold-based tumor-first heuristic strategy for dividing WSI primary and secondary tissues by double thresholds, which meets the actual clinical needs. Finally, our study can diagnose typical WSIs images within 10 min, confirming the efficient performance of our system. Our work is consistent with the application to clinical scenarios in terms of the breadth of classification coverage of frozen thyroid sections, the accuracy of model results, the efficiency of diagnosis and the consistency with experienced pathologists, and it has significant implications for clinical diagnosis. In future work, we will also enrich the number of frozen digital databases and extend them to other tissue-frozen sections to build a multi-tissue system for intraoperative rapid frozen sections.

Abbreviations

AI	Artificial intelligence
CNN	Convolutional neural network
GPU	Graphics processing unit
H&E	Hematoxylin and eosin
PTC	Papillary thyroid carcinoma
FTC	Follicular thyroid carcinoma
FTT	Thyroid follicles tumors MTC: medullary thyroid carcinoma
ATC	Anaplastic thyroid carcinoma
NT	Non-carcinoma lesion
PTMC	Papillary thyroid microcarcinoma
ROI	Region of interest
SYSU	Sun Yat-sen University dataset
SYSUFH	Sun Yat-sen University First Affiliated Hospital
WSI	whole slide image
CV	Computer Vision
TP	True Positive
FN	False Negative
FP	False Positive
TN	True Negative
WHO	World Health Organization
CLT	Chronic lymphocytic thyroiditis
ST	Subacute thyroiditis
NG	Nodular goiter
RCT	Regional continuity tiles
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
TPR	True Positive Rate
FPR	False Positive Rate
Grad-CAM	Gradient-weighted Class Activation Mapping
TSCD	Two-step cascade diagnostic
NIFTP	Non-invasive follicular thyroid neoplasm with papillary-like nuclear features
FVPTC	Follicular variant of papillary thyroid carcinoma

Authors' contributions

A.J.H., T.T.H., Y.H.H., and H.J.S. conceived the study and designed the experiments. T.T.H., S.S.S completed all codes of data processing, model training and testing. T.T.H., Y.N.W., Y.Q.L., L.H.Z. and F.F.Z. performed the experiments analysis. T.T.H. prepared the manuscript. Y.H.H., A.J.H., revised the manuscript. All authors have read and agreed to publish the paper.

Funding

National Science Foundation of China (61875102 and 61975089), Natural Science Foundation of Guangdong province (2021A1515012379 and 2022A1515012550), Natural Science Foundation of Shenzhen city (JCY120200109110606054 and WDZC20202008211413490001).

Availability of data and materials

The pathological images used in this paper is not publicly available. However, they can be obtained by contacting the corresponding author (hananjia@mail.sysu.edu.cn) for scientific research purposes.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of First Affiliated Hospital of Sun Yat-sen University, approval number [2023] C-061. The requirement for informed consent was waived by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, China. [2]Department of Pathology, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China.

Received: 27 February 2024 Accepted: 26 August 2024
Published online: 29 August 2024

References

1. Rodig SJ. Preparing Paraffin Tissue Sections for Staining. Cold Spring Harb Protoc. 2021;2021(3). https://doi.org/10.1101/PDB.prot099663. PMID: 33649119.
2. Märkl B, Füzesi L, Huss R, Bauer S, Schaller T. Number of pathologists in Germany: comparison with European countries, USA, and Canada. Virchows Arch. 2021;478(2):335–41. https://doi.org/10.1007/s00428-020-02894-6. Epub 2020 Jul 27. PMID: 32719890; PMCID: PMC7969551.
3. Kraemer BB. Frozen section diagnosis and the thyroid. Semin Diagn Pathol. 1987;4(2):169–89. PMID: 3313602.
4. Novis DA, Gephardt GN, Zarbo RJ. Interinstitutional comparison of frozen section consultation in small hospitals: a college of American pathologists Q-probes study of 18532 frozen section consultation diagnoses in 233 small hospitals. Arch Pathol Lab Med. 1996;120(12):1087.
5. Albores-Saavedra J, Wu J. The many faces and mimics of papillary thyroid carcinoma. Endocr Pathol. 2006;17(1):1–18.
6. Kiani A, Uyumazturk B, Rajpurkar P, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer[J]. NPJ Digit Med. 2020;3(1):1–8.
7. Chaudhary K, Poirion OB, Lu L, et al. Deep learning–based Multi-omics Integration robustly predicts Survival in Liver CancerUsing Deep Learning to Predict Liver Cancer Prognosis[J]. Clin Cancer Res. 2018;24(6):1248–59.
8. Han Z, Wei B, Zheng Y, et al. Breast cancer multi-classification from histopathological images with structured deep learning model[J]. Sci Rep. 2017;7(1):1–10.
9. Wang D, Khosla A, Gargeya R, et al. Deep learning for identifying metastatic breast cancer[J]. arXiv preprint arXiv:1606.05718. 2016.
10. Bejnordi BE, Veta M, Van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer[J]. JAMA. 2017;318(22):2199–210.
11. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning[J]. Nat Med. 2018;24(10):1559–67.
12. Sun W, Zheng B, Qian W. Computer aided lung cancer diagnosis with deep learning algorithms[C]//Medical imaging 2016: computer-aided diagnosis. SPIE. 2016;9785:241–8.
13. Liu KL, Wu T, Chen PT, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation[J]. Lancet Digit Health. 2020;2(6):e303–13.
14. Zhang X, Zhu X, Tang K, Zhao Y, Lu Z, Feng Q. DDTNet: a dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer, Medical Image Analysis, 78,2022, 102415, ISSN 1361–8415, https://doi.org/10.1016/j.media.2022.102415
15. Wang S, Yang DM, Rong R, et al. Pathology image analysis using segmentation deep learning algorithms[J]. Am J Pathol. 2019;189(9):1686–98.
16. Yang H, Chen L, Cheng Z, Yang M, Wang J, Lin C, Wang Y, Huang L, Chen Y, Peng S, Ke Z, Li W. Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. BMC Med. 2021;19(1):80. https://doi.org/10.1186/s12916-021-01953-2. PMID: 33775248; PMCID: PMC8006383.
17. Dolezal JM, Trzcinska A, Liao CY, et al. Deep learning prediction of BRAF-RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features. Mod Pathol. 2021;34:862–74. https://doi.org/10.1038/s41379-020-00724-3.
18. Nojima S, Kadoi T, Suzuki A, Kato C, Ishida S, Kido K, Fujita K, Okuno Y, Hirokawa M, Terayama K, Morii E. Deep learning-based Differential diagnosis of follicular thyroid tumors using histopathological images. Mod Pathol. 2023;36(11):100296. Epub 2023 Jul 31. PMID: 37532181.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conf Comput Vis Pattern Recognit (CVPR). 2016;770–8. https://doi.org/10.1109/CVPR.2016.90.
20. NVIDIA TensorRT, Accessed. Mar. 23, 2021. Available: https://developer.nvidia.com/tensorrt/
21. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. Proceedings of the IEEE conference on computer vision and pattern recognition 2016:2424–33. https://doi.org/10.1109/CVPR.2016.266
22. Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Sci Rep. 2019;9(1):3358. https://doi.org/10.1038/s41598-019-40041-7. PMID: 30833650; PMCID: PMC6399447.
23. Rau JV, Fosca M, Graziani V, et al. Proof-of-concept Raman spectroscopy study aimed to differentiate thyroid follicular patterned lesions. Sci Rep. 2017;7:14970. https://doi.org/10.1038/s41598-017-14872-1.
24. Ito Y, Hirokawa M, Hayashi T, Kihara M, Onoda N, Miya A, et al. Clinical outcomes of follicular tumor of uncertain malignant potential of the thyroid: real-world data. Endocr J. 2022;69(7):757–61. https://doi.org/10.1507/endocrj.EJ21-0723. Epub 2022 Jan 26. PMID: 35082189.
25. Cheung C, Ezzat S, Freeman J, et al. Immunohistochemical diagnosis of papillary thyroid carcinoma. Mod Pathol. 2001;14:338–42. https://doi.org/10.1038/modpathol.3880312.
26. Haines DM, Chelack BJ. Technical considerations for developing enzyme immunohistochemical staining procedures on Formalin-fixed paraffin-embedded tissues for Diagnostic Pathology. J Vet Diagn Invest. 1991;3(1):101–12. https://doi.org/10.1177/104063879100300128.
27. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4700–8.
28. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR; 2019. p. 6105–14.
29. Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE; 2018. p. 6848–56. https://doi.org/10.1109/CVPR.2018.00716.
30. Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 10012–22.
31. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE; 2015. p. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
32. Kieffer B, Babaie M, Kalra S, Tizhoosh HR. Convolutional neural networks for histopathology image classification: Training vs. Using pre-trained networks, 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 2017, pp. 1–6, https://doi.org/10.1109/IPTA.2017.8310149
33. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012. p. 1097–105.
34. Jazar RN. Averaging Method. Perturbation methods in Science and Engineering. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-73462-6_8.
35. Kangas A, Kurttila M, Hujala T, Eyvindson K, Kangas J. Voting methods. Decision support for Forest Management. Managing Forest ecosystems. Volume 30. Cham: Springer; 2015. https://doi.org/10.1007/978-3-319-23522-6_10.
36. Mahdavi S, Rahnamayan S, Mahdavi A. Majority voting for discrete population-based optimization algorithms. Soft Comput. 2019;23:1–18. https://doi.org/10.1007/s00500-018-3530-1.
37. Xiaomeng Wu G, Irie K, Hiramatsu, Kashino K, Weighted Generalized Mean Pooling for Deep Image, Retrieval., IEEE Signal Processing Society SigPort, 2018. Available: https://sigport.org/documents/weighted-generalized-mean-pooling-deep-image-retrieval. Accessed: Dec. 11, 2022.

38. Li Y, Chen P, Li Z, Su H, Yang L, Zhong D. Rule-based automatic diagnosis of thyroid nodules from intraoperative frozen sections using deep learning, Artificial Intelligence in Medicine, 108,2020,101918, ISSN 0933–3657, https://doi.org/10.1016/j.artmed.2020.101918
39. Aziz A, Naseradeen Abdulqader Z, Sallow D, A. B., Khalid Omer H. Python parallel Processing and Multiprocessing: a Rivew. Acad J Nawroz Univ. 2021;10(3):345–54. https://doi.org/10.25007/ajnu.v10n3a1145.
40. Meta. From research to production. Online at https://pytorch.org/.
41. The Linux Foundation. Open neural network exchange. Online at https://onnx.ai/.
42. Zhou Y, Yang K. Exploring TensorRT to Improve Real-Time Inference for Deep Learning, 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), Hainan, China, 2022, pp. 2011–2018, https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00299
43. Ting KM. Confusion Matrix. In: Sammut C, Webb GI, editors. Encyclopedia of machine learning. Boston, MA: Springer; 2011. https://doi.org/10.1007/978-0-387-30164-8_157.
44. Selvaraju RR, Cogswell M, Das A et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE; 2017: 618–26. https://doi.org/10.1109/ICCV.2017.74.
45. Gupta N, Sarkar C, Singh R, et al. Evaluation of diagnostic efficiency of computerized image analysis based quantitative nuclear parameters in papillary and follicular thyroid tumors using paraffin-embedded tissue sections. Pathol Oncol Res. 2001;7:46–55. https://doi.org/10.1007/BF03032605.
46. Nafe R, Fritsch RS, Soudan B, Hammann A, Choritz H. Histomorphometryin paraffin sections of thyroid tumors, Pathology - Research and Practice, Volume 188, Issue 8, 1992, Pages 1042–1048, ISSN 0344 – 0338, https://doi.org/10.1016/S0344-0338(11)81249-5
47. Sui Peng Y, Liu W, Lv L, Liu Q, Zhou H, Yang et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study, The Lancet Digital Health,Volume 3, Issue 4, 2021, Pages e250-e259, ISSN 2589–7500, https://doi.org/10.1016/S2589- 7500(21)00041 – 8.
48. Buddhavarapu VG, Jothi AA. An experimental study on classification of thyroid histopathology images using transfer learning. Pattern Recognit Lett. 2020;140. https://doi.org/10.1016/j.patrec.2020.09.020. Pages 1–9, ISSN 0167–8655.
49. Chen P, Shi X, Liang Y, Li Y, Yang L, Gader PD. Interactive thyroid whole slide image diagnostic system using deep representation, computer methods and programs in Biomedicine, 195, 2020, 105630, ISSN 0169–2607, https://doi.org/10.1016/j.cmpb.2020.105630
50. National Cancer Institute. Physician Data Query (PDQ). Thyroid Cancer Treatment. 02/06/2019. Accessed at https://www.cancer.gov/types/thyroid/hp/thyroid-treatmentpdq#_313_toc. on February 20, 2019. 42.
51. Bible KC, Kebebew E, Brierley J, Brito JP, Cabanillas ME, Clark TJ Jr et al. 2021 American Thyroid Association Guidelines for Management of Patients with Anaplastic Thyroid Cancer. Thyroid. 2021;31(3):337–386. DOI: 10.1089/thy.2020.0944. Erratum in: Thyroid. 2021;31(10):1606–1607. PMID: 33728999; PMCID: PMC8349723.
52. Turan G, Sevgiye Kaçar Ôzkara. Pathological findings of the retrospective diagnosis of NIFTP (non-invasive follicular thyroid neoplasm with papillary-like nuclear features) in 84 cases from Turkey and systematic review. Annals Diagn Pathol. 2021;53:151764.
53. Tunca F, et al. Comparison of histopathological features and prognosis of classical and follicular variant papillary thyroid carcinoma. J Endocrinol Investig. 2015;38:1327–34.
54. Jug R, Jiang X. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: an evidence-based nomenclature change. Patholog Res Int. 2017;2017(1):1057252.
55. Maletta F, et al. Cytological features of noninvasive follicular thyroid neoplasm with papillary-like nuclear features and their correlation with tumor histology. Hum Pathol. 2016;54:134–42.

Publisher's note