## RESEARCH





# Annotation-free genetic mutation estimation of thyroid cancer using cytological slides from multi-centers

Siping Xiong<sup>1†</sup>, Shuguang Liu<sup>1†</sup>, Wei Zhang<sup>1†</sup>, Chao Zeng<sup>1</sup>, Degui Liao<sup>2</sup>, Tian Tang<sup>2\*</sup>, Shimin Wang<sup>1\*</sup> and Yimin Guo<sup>1\*</sup>

## Abstract

Thyroid cancer is the most common form of endocrine malignancy and fine needle aspiration (FNA) cytology is a reliable method for clinical diagnosis. Identification of genetic mutation status has been proved efficient for accurate diagnosis and prognostic risk stratification. In this study, a dataset with thyroid cytological images of 310 indeterminate (TBS3 or 4) and 392 PTC (TBS5 or 6) was collected. We introduced a multimodal cascaded network framework to estimate BARF V600E and RAS mutations directly from thyroid cytological slides. The area under the curve in the external testing set achieved  $0.902 \pm 0.063$  and  $0.801 \pm 0.137$  AUCs for BRAF, and RAS, respectively. The results demonstrated that deep neural networks have the potential in cytologically predicting valuable diagnosis and comprehensive genetic status.

**Keywords** Deep convolutional network, Thyroid cancer, Fine needle aspiration cytology, Annotation-free, Gene mutation estimation

## Introduction

Thyroid cancer is one of the most prevalent malignancies of the endocrine system, and the incidence rate of thyroid nodules has increased significantly due to the improved capabilities of ultrasound detection [1]. The histopathology of thyroid cancer can be mainly classified as papillary thyroid carcinoma (PTC,70–90%), follicular thyroid carcinoma (FTC, 5–10%), anaplastic thyroid carcinoma

<sup>†</sup>Siping Xiong, Shuguang Liu and Wei Zhang contributed equally to this work.

\*Correspondence: Tian Tang tt9772888@126.com Shimin Wang wangshm35@mail.sysu.edu.cn Yimin Guo guoym35@mail.sysu.edu.cn

<sup>1</sup> Department of Pathology, The Eighth Affiliated Hospital of Sun Yat-Sen University, Shenzhen 518000, Guangdong, China (ATC, 2%) and medullary thyroid carcinoma (MTC, 2% <) [2]. The majority of PTC exhibit a favorable long-term prognosis with a 10 year survival rate of 96% after standard treatment, 5-10% of cases manifest as advanced disease with poor prognosis [3]. Hence, early precise diagnosis information holds extremely significance for patients in avoiding over- or under- treatment.

Currently, ultrasound-guided fine-needle aspiration (FNA) combined with liquid-based thin-layer cytology (TCT) is a reliable method for clinical identification of benign and malignant thyroid nodules [4]. It is suitable for the diagnosis of thyroid nodules with typical cytologic characteristic such as papillary, flat honeycomb sheet-like architecture, and pseudoinclusions [5]. However, using The Bethesda System (TBS), pathologists often feel confused when dealing with nodules exhibiting undistinctive cytologic features which are classified as indeterminate (TBS3: Atypia of undetermined significance or TBS4: Follicular neoplasmFollicular lesion of undetermined



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

<sup>&</sup>lt;sup>2</sup> Department of Pathology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510000, Guangdong, China

significance) [6]. With the constant innovation of diagnostic technology, molecular detection has integrated in the entire thyroid cancer diagnostic and treatment process. Genetic analysis of FNA samples can not only enhance the accuracy of preoperative cytological diagnosis, but also be used to predict the risk of invasiveness and provide decision-making information for thyroid nodule patients [7].

Genetic alternations in the two signaling pathways of mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) have been identified responsible for the incidence and progression of thyroid cancer [8]. Notably, BRAF V600E mutation is observed in over 80% of PTC patients, followed by mutations of RAS (including NRAS, KRAS, HRAS) present in 10-15% of PTC. The mutations in BRAF or RAS genes activate the MAPK signaling pathway, facilitating the advancement and metastasis of thyroid tumors [9, 10]. In addition, studies have shown that BRAF V600E mutation is closely related to the reduced sensitivity of radioactive iodine treatment [11, 12]. RAS gene mutations are commonly observed in thyroid adenomas, underscoring their significance in the carcinogenesis of thyroid follicular cells [13].

The general procedure employed by genetic analysis including ARMS PCR, Sanger sequencing, next-generation sequencing, and high-performance liquid chromatography. ARMS PCR is relatively commonly used in clinical practice owing to its high sensitivity and specificity. However, its clinical application is limited by various factors such as strict requirements on laboratory and sample quality, time-consuming experimental steps, and scarcity of qualified staff [14].

With the rapid development of deep learning, many aritificial intelligence(AI)-assisted diagnostic systems have been developed for various tasks in tumor pathology, encompassing tumor classification, prognosis prediction, as well as genetic mutation estimation [15, 16]. In 2019, Guan et al. trained a deep convolutional neural network on cytological images for classification of PTC and Non-PTC [17]. Later in 2020, malignancy estimation of thyroid cytological lesions was reported [18, 19]. Furthermore, Anand et al. [20] and Xi et al. [21] introduced various AI-based diagnostic system for predicting BRAF mutation on H&E-stained slides and ultrasound images, respectively.

These methods have significantly improved the application extends of AI in thyroid pathology diagnosis. However, genetic mutation prediction from FNA slices of thyroid nodules, which can augment malignancy diagnosis in indeterminate group and imply prognostic risk in PTC group, is still very challenging. To enhance the clinical utility of identifying genetic status, we proposed Page 2 of 10

a multimodal cascaded deep convolutional framework to sequentially classify informative cellular regions and estimate genetic mutations (i.e., BRAF and RAS) on the WSIs. The proposed framework consists of a few-shot region classifier and a multimodal mutation classifier, which significantly reduces the need for manual annotations while maintaining a high level of classification generalization capability. The main contributions of this study are as follows:

- We proposed a multimodal cascaded deep convolutional framework to to sequentially detect informative cellular regions and estimate genetic mutations using FNA slices.
- We introduced a few-shot learning strategy which significantly reduces the number of annotations required for training region classifier while maintaining high classification accuracy.
- We further analyzed the effectiveness of the number of features selected for ensembling to understand its effects on the performance and stability of deep convolutional networks.

The rest of the paper is organized as follows: Firstly, we presented the datasets and methods used for this research in "Materials and methods" section. Then, we illustrated the quantitative and qualitative results in "Results" section. Finally, discussion and conclusion were presented in the "Discussion" section.

## **Materials and methods**

## Data

The cytology slides (Thin Prep, Papanicolaou stained) of 702 distinct thyroid nodules were obtained from the Department of Pathology, the Eighth Affiliated Hospital of Sun Yat-sen University (596/702) and the Department of Pathology, the Second Affiliated Hospital of Guangzhou Medical University (106/702) (The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board with protocol code 2024d013). Among the 310 indeterminate and 392 PTC nodules collected in this study, 58(18.7%) and 351(89.5%) BRAF V600E mutations were found respectively. Besides, we detected 43 (13.9%) and 9 (2.3%) RAS mutations in 310 indeterminate and 392 PTC nodules respectively. All cytology slides were diagnosed by four expert pathologists. Each cytologic sample was separated after mixing, with one half used for the creation of liquid-based cytology slides for microscopic diagnosis, and the other half employed for the generation of cell wax blocks as specimens for subsequent genetic testing. The Amplification Refractory Mutation System-polymerase chain reaction

method was used to detect the mutation status of BRAF V600E, KRAS G12C/G12V/Q61R, NRAS Q61R, and HRAS Q61R, following the protocols provided by the manufacturer (Mole Bioscience Co. Ltd., Jiangsu, China). All the 702 slides were scanned with 20x magnification ratio by PANNORAMIC 1000, 3DHISTECH (https://www.3dhistech.com/research/pannoramicdigital-slide-scanners/pannoramic-1000/). To facilitate cell cluster detection, binary masks of cell areas of the

 Table 1
 Distribution of patient nodules and corresponding clinical information

Training set	Validation set	External testing set
485	111	106
42.1 ± 12.5	43.6 ± 12.0	46.5 ± 13.8
117 (24.1%)	33 (29.7%)	23 (21.7%)
368 (75.9%)	78 (70.3%)	83 (78.3%)
IS		
294 (60.6%)	64 (57.7%)	51 (48.1%)
191 (39.4%)	47 (42.3%)	55 (51.9%)
36 (7.4%)	10 (9.0%)	6 (5.7%)
449 (92.6%)	101 (91.0%)	100 (94.3%)
	Training set 485 42.1 ± 12.5 117 (24.1%) 368 (75.9%) 15 294 (60.6%) 191 (39.4%) 36 (7.4%) 449 (92.6%)	Training set         Validation set           485         111           42.1 ± 12.5         43.6 ± 12.0           117 (24.1%)         33 (29.7%)           368 (75.9%)         78 (70.3%)           1s         294 (60.6%)         64 (57.7%)           191 (39.4%)         47 (42.3%)           36 (7.4%)         10 (9.0%)           449 (92.6%)         101 (91.0%)

WSIs were carefully annotated by experienced pathologists using QGIS (v3.22.7 LTR, https://qgis.org/).

As shown in Table 1, 702 WSIs of patient nodules were splitted into training, validation, and external testing sets. In our experiment, the size of each tile was fixed to  $512 \times 512$  pixels.

## Methodology

In this study, we proposed an annotation-free cascaded deep learning pipeline to sequentially detect informative cellular regions and then estimate the somatic gene mutations from the thyroid cytology whole-side images.

As shown in Fig. 1, the experimental workflow is consisted of three parts: (a) whole-slide image preprocessing to cut it into image patches and filter out white noise or background (b) few-shot region classification to determine informative and non-informative cellular regions (c) multimodal genetic mutation classification to estimation mutation type for each tile. Following preprocessing, a limited subset of image patches exhibiting varying abundances of follicular cells-encompassing both informative and non-informative categories-will undergo few-shot region classification. Subsequently, the well-trained region classifier will serve to exclude image patches that lack follicular cells. Only those informative patches will be retained for the training and evaluation of

#### (a) Whole-slide image preprocessing



(c) Multimodal individual classification







Fig. 1 Experimental workflow. a Whole-slide image preprocessing b Few-shot image patch classification for estimating informative and non-informative cellular regions c Multimodal genetic mutation classifier for tile-level somatic gene mutation estimation

the multimodal genetic mutation classification, ensuring that the model is trained on relevant, cellularly informative data.

#### WSI preprocessing

At data preprocessing, the 702 pairs of whole-side images (WSIs) and their corresponding clinical records were distributed to three groups: training (485), validating (111), and testing (106). A  $512 \times 512$  square window was applied to the whole-side image to generate image tiles. To avoid interfering of blank backgrounds and dark noises, a simple threshing strategy was applied to filter all tiles with average pixel values  $\leq 10$  or  $\geq 240$ . As shown in Table 1, there were 1 245 999, 422 517, and 450 574 image tiles within the corresponding training, validating, and testing set.

## Few-shot classification

In ultrasound-guided FNA, only a small portion of cells will be taken from patient nodules which usually leads to a relatively sparse distribution of cellular clusters within a whole-slide image. Before any diagnostic classification or gene mutation estimation, it's vital to detect and classify the informative cellular regions. Other than existing researches [16] which manually annotated all region of interests (ROIs), we introduced a few-shot training strategy, which learn from limited annotation data while maintaining good generalization capability, that significantly reduced human labor and speed-up the entire experimental workflow.

As shown in Fig. 2, a small subset (15 / 516) WSIs within internal set were annotated by experienced pathologist and then applied for training (i.e., 10 WSIs) and validating (i.e., 5 WSIs) the few-shot cellular region classifier.

For better generalization capability, the region classifier was composed of a fixed CLIP (Contrastive Language-Image Pretraining) [22] image encoder, a learnable EfficientNet [23] image encoder, and a fully-connected (FC) layer. The CLIP model introduced a contrastive pre-training scheme which bridges image and text description of a specified scene. After training with large-scale paired image-text dataset, the CLIP can serve as a robust zeroshot image encoder. Other than CLIP, we chose Efficient-Net, a classic model architecture designed to produce high accuracy under limited computational operations, as the learnable image encoder which can gradually adjust its parameters during every iterations. As shown in Table 2, we chose an ImageNet-1K [24] pretrained EfficientNet B0 (https://pytorch.org/vision/master/models/



Fig. 2 Sampled Whole-slide image from SYSU8H-Thyroid dataset. **a** The whole-side-image (WSI) **b** Non-annotated negative patches within WSI **c** Annotated positive patches within WSI

#### Page 4 of 10

**Table 2** The backbone network of EfficientNet. Each rowdescribes the stage, operation, input resolution, output channel,and the number of layers

Stage	Operator	Resolution	Channels	Layers
0		512x512	3	0
1	Conv3x3	512x512	32	1
2	MBConv1, k3x3	256x256	16	1
3	MBConv6, k3x3	256x256	24	2
4	MBConv6, k5x5	128x128	40	2
5	MBConv6, k3x3	64x64	80	3
6	MBConv6, k5x5	32x32	112	3
7	MBConv6, k5x5	32x32	192	4
8	MBConv6, k3x3	16x16	320	1
9	Conv1x1 &Pooling	16x16	1280	1

generated/torchvision.models.efficientnet\_b0.html) as the network backbone.

The final FC layer taken two encoded features of CLIP and EfficientNet as input to generate output  $(x_i)$ . As a binary classification task, we adopted sigmoid function to generate final prediction  $(p_i)$  and L1 loss as object function.

$$p_{i} = \frac{1}{1 + e^{-x_{i}}}$$

$$Loss_{L1} = \frac{1}{n} \sum_{i=1}^{n} |y_{i} - p_{i}|$$
(1)

Where  $p_i$  and  $y_i$  was the *i*<sup>th</sup> prediction and corresponding ground truth. The ranges of prediction  $p_i$  and  $y_i$  were limited to [0, 1].

With all of the above layers being trained by mini-batch stochastic gradient descent (SGD) [25] to minimize the L1 loss, the classifier learned to make category prediction on every 512x512 image patches of the SYSU8H-Thyroid dataset. The image tiles with high probability of being a informative region of training and validation set were subsequentially used for training and validating the multimodal classifier to discriminate wild type (i.e., W.T.) vs. mutant type (i.e., M.T.) of target somatic genes (i.e., BRAF and RAS), respectively. To ensure the high fidelity of selected tiles, a high threshold (i.e., 0.75) was used to filter out tiles with a low probability of being an informative cellular region.

#### Multimodal classifier

Similar to few-shot region classifier, our proprosed multimodal classifier consisted of three encoders, a multihead attention module, and two FC layers. In addition to image encoders, we adopted a tokenizer (https://github. com/mlfoundations/open\_clip) to convert text clinical records (e.g., gender of female, age of 44) to a 1024  $\times$  1 vector. The multi-head attention module, which was firstly proposed by Ashish et al. [26] in 2017, is able to capture complex relationships in data and adapt to varying context. The equation for multi-head attention can be formulated as follows:

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h) \cdot W_O$$
$$head_i = Attention(Q \cdot W_{Qi}, K \cdot W_{Ki}, V \cdot W_{Vi})$$
$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V$$
(2)

Here, Q, K, and V represented the query, key, and value matrices, respectively.  $W_{Qi}$ ,  $W_{Ki}$ , and  $W_{Vi}$  were learnable linear transformation matrices for each attention head.  $W_O$  was the output transformation matrix.  $d_k$  was the dimension of the key vectors.

The output of muli-head attention module was passed to two FC layers to generate multi-class predictions (X).

$$p_i = \frac{e^{x_i}}{\sum\limits_{j=1}^{N} e^{x_j}} \tag{3}$$

where  $x_i$  and N was an element and number of elements in the input vector  $X_i$  respectively.

Instead of L1 loss, we adopted categorical crossentropy loss [27] as our object function to address multiclass classification. The equation can be formulated as:

$$Loss(y, p) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \cdot \log(p_{ij})$$
(4)

Where p and y was the predicted probability distribution and corresponding ground truth. The i, j, and C represented sample, class indice and the number of classes, respectively.

To make a decisive conclusion on the whole-slide image (WSI) using the separated predictions of tiles, we adopted a frequency histogram to convert tile-level predictions to WSI instance-level probability. For each WSI, the probability ( $p_{wsi}$ ) can be calculated by the following equations.

$$Histogram = [f1, f2, f3, ..., fn]$$

$$p_{wsi} = ARGMAX(Histogram)/n + MAX(Histogram) * 0.1/n$$
(5)

The n represented the number of bins within range[0, 1]. Finally, the  $p_{wsi}$  and corresponding ground truth  $(y_{wsi})$  were used to compute the area under the receiver operator characteristic (ROC) curve [28] and its confidence interval (CI) [29] for performance estimation.

$$ROC-AUC = \int_0^1 TPR(FPR^{-1}(t)) dt$$
 (6)

The TPR and FPR represent true positive rate (sensitivity) and false positive rate (1-specificity) at various threshold (t) settings.

After repeating several times of the training and validation procedures, the hyperparameters, including batch size, the number of epochs, and learning rate, are optimized with the Adam stochastic optimizer [30]. Subsequently, the predictions generated by the optimized models were evaluated using the WSIs of the testing set (see details in Table 1).

## Results

## Few-shot region classification

The probability maps generated by the few-shot region classification models using tiles of the WSIs are depicted in Fig. 3. Within each sample (i.e., a and b), as the predicted probability decreases (from top to bottom), the presence of informative cell clusters in each row of tiles gradually diminishes. As the predicted probability ranges between 0.75 and 1.00, the majority of tiles contain valid diagnostic cells (i.e., 1st row). However, the predicted probability falls between 0.50 and 0.75, only a few tiles contain partial valid diagnostic information (i.e., 2nd row). When the predicted probabilities are < 0.50, the tiles essentially lack of diagnostic information (i.e., 3rd and 4th rows). To ensure the reliability of the selected tiles, only those predicted with high probability (i.e., 0.75 to 1.00) are chosen for further training and validating the following genetic mutation estimation models.

## Multimodal genetic mutation estimation

After model ensembling, the proposed method generated probability of gene mutations (i.e., BRAF and RAS) of every WSI.

Zoom-in tiles with various probability



Fig. 3 Probability maps of informatic region classification using tiles of the whole slide images (WSIs). From left to right: the whole-side image; the predicted probability map over WSI; the zoom-in tiles with various probability ranges. **a**, **b**, and **c** are randomly sampled cases



Fig. 4 The receiver operator characteristic (ROC) curve and corresponding area under the curve (AUC) of genetic mutation estimation of the whole slide images (WSIs). **a** The ROC curves of BRAF mutation estimation **b** The ROC curves of RAS mutation estimation

As shown in Fig. 4a, in BRAF mutation estimation task, the proposed method reached 0.938 (95% CI of 0.917–0.960), 0.905 (95% CI of 0.849–0.962), and 0.902 (95% CI of 0.839–0.965) AUCs in training, validation, and external testing set, respectively. In Fig. 4b, our method also displayed high accuracy in RAS mutation estimation task. The AUCs of training, validation, and external

(a) WSI

testing sets reached 0.881 (95% CI of 0.812–0.951), 0.802 (95% CI of 0.621–0.984), and 0.801 (95% CI of 0.663–0.938). Compared with RAS mutation estimation, the proposed method present higher accuracy and stability in BRAF mutation estimation.

In order to investigate the relationship between image features and gene mutations, we selected two

(b) Histogram of WSI



Fig. 5 Representative tiles of BRAF mutation **a** The whole slide image. **b** The histogram of tile-level predictions within the WSI. **c** The zoom-in tiles with various probability ranges

representative cases of BRAF and RAS mutations, respectively.

As shown in Fig. 5, the whole slide image (WSI) and the corresponding tiles of different BRAF mutation probabilities were displayed. The tiles with high probabilities of being BRAF mutant type were listed on top of Fig. 5c.

As shown in Fig. 6, the whole slide image (WSI) and the corresponding tiles of different RAS mutation probabilites were displayed. The tiles with high probabilities of being RAS mutant type were listed on top of Fig. 6c.

## Discussion

(a) WSI

### Regarding the proposed framework

Previous studies have shown the capacity of deep learning models not only for thyroid cancer diagnosis but also for predicting genetic mutations based on diverse medical images. Most current researches focus on classifying thyroid nudules as benign or malignant according to the histological diagnosis by CNNs [17–19]. Anand et al. developed a deep neural network that predict BRAF V600E mutational status from thyroid cancer H&E slides [20]. Wang et al. implemented a deep learning based on a dataset of 118 PTC cytologic WSIs to predict BRAF V600E mutation [31]. Considering the significance of providing as much valuable clinical information as possible to patients with thyroid nodules before surgery, in the present study, we firstly employed a DNN model to complete comprehensive genetic prediction within different diagnostic cytology slides. Our method employed a few-shot learning strategy, significantly reducing the need for manual annotations while maintaining a high level of classification generalization. The utilization of deep convolutional networks for estimating gene mutations (BRAF and RAS) offers pathologists a more convenient means to assess risk and guide treatment.

#### Accuracies, uncertainties, and limitations

Qualitative and quantitative assessments on the internal validation set, our model achieved  $0.905 \pm 0.056$  and  $0.802 \pm 0.181$  AUCs for BRAF, and RAS, respectively. Similarly, our method kept high accuracy in the external testing set, showing  $0.902 \pm 0.063$  AUC of BRAF and  $0.801 \pm 0.137$  AUC of RAS. Compared with BRAF mutation estimation, our method has encountered challenges in distinguishing between wild type and mutant types of RAS due to a relatively biased data distribution (i.e., 5.7%–9.0% W.T. in the training, validation and testing set). Due to the data-driven nature of deep learning models, a substantial quantity of positive samples is typically necessary for effective model training. Consequently, it



(b) Histogram of WSI

Fig. 6 Representative tiles of RAS mutation. b The histogram of tile-level predictions within the WSI. c The zoom-in tiles with various probability ranges

0.00 - 0.25

is imperative to consider the reliability of the numerical values associated with the limited positive sample size of RAS mutants.

The occurrence and progression of PTC is a complex, multifactorial process. Among the various genetic alterations, the BRAF V600E and RAS Q61R mutations are two prevalent mutations associated with PTC. Notably, the BRAF V600E mutation exhibits an almost 100% specificity for the diagnosis of PTC, making it a critical marker in clinical settings. Consequently, this study focuses exclusively on the detection and prediction of these two hotspot mutations. It is crucial to highlight that even in the absence of BRAF V600E and RAS Q61R mutations, other genetic abnormalities, such as RAS Q61K, RET and NTRK fusions, although less common, can also play a key role in the pathogenesis of PTC [32]. In this study, cases with undetected non-hot-spot mutations or fusions may have been incorrectly classified as wild-type, potentially skewing the true prevalence of BRAF-like and RAS-like phenotypes in the dataset. The heterogeneity of the wild type group, which may include cases with undetected mutations, could introduce variability into the results and reduce the model's overall accuracy. We propose incorporating comprehensive genomic profiling to capture a wider range of mutations and fusions in future studies. This would enable a more accurate classification of cases and improve the robustness of the findings.

Within our cascaded classification pipeline, the models were trained to generate tile-to-label predictions using multimodal feature encoders. However, the absence of internal connectivity with adjacent tiles within the same WSI may result in inconsistent predictions among tiles (e.g., predicting PTC for tile A and AUS for tile B). Additionally, as the region classifier and multimodal models were trained and optimized separately, the proposed framework necessitates extra computational time and storage for training, saving checkpoints, and interferring. To enhance computational efficiency, further work should explore a unified model with shared parameters and objective functions.

To thoroughly assess the effectiveness and generalization of the proposed method, further evaluation using a larger dataset that encompasses more clinical centers, racial diversity, and genetic alterations is essential.

#### **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s13000-025-01618-1.

Supplementary Material 1.

#### Acknowledgements Not applicable.

#### Code availability

All the codes and guidance for this study can be found on GitHub: https:// github.com/huster-wgm/MultimodalLearning.

#### Authors' contributions

Conceptualization, Tian Tang, Shimin Wang and Yimin Guo; Data curation, Chao Zeng and Degui Liao; Investigation, Siping Xiong, Shuguang Liu and Wei Zhang; Methodology, Siping Xiong, Shuguang Liu and Wei Zhang; Supervision, Tian Tang, Shimin Wang and Yimin Guo; Writing – original draft, Siping Xiong, Shuguang Liu, Wei Zhang, Chao Zeng and Degui Liao; Writing – review & editing, Tian Tang, Shimin Wang and Yimin Guo. All authors reviewed the manuscript.

#### Funding

The study is supported by Futian Healthcare Research Project (No. FTWS035).

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

#### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of The Eighth Affiliated Hospital of Sun Yat-Sen University (protocol code D2024R054).

#### **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 1 September 2024 Accepted: 14 February 2025 Published online: 21 February 2025

#### References

- Miranda-Filho A, Lortet-Tieulent J, Bray F, Cao B, Franceschi S, Vaccarella S, et al. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. Lancet Diabetes Endocrinol. 2021;9(4):225–34.
- Pacini F, Castagna M, Brilli L, Pentheroudakis G. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2012;23:110–19.
- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. Ca Cancer J Clin. 2023;73(1):17–48.
- Fadda G, Rossi ED. Liquid-based cytology in fine-needle aspiration biopsies of the thyroid gland. Acta Cytol. 2011;55(5):389–400.
- Feldkamp J, Führer D, Luster M, Musholt TJ, Spitzweg C, Schott M. Fine needle aspiration in the investigation of thyroid nodules: indications, procedures and interpretation. Deut Ärzteblatt Int. 2016;113(20):353.
- Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda System for reporting thyroid cytopathology. J Am Soc Cytopathol. 2023;12(5):319–25.
- Laha D, Nilubol N, Boufraqech M. New therapies for advanced thyroid cancer. Front Endocrinol. 2020;11:82.
- Fagin J. How thyroid tumors start and why it matters: kinase mutants as targets for solid cancer pharmacotherapy. J Endocrinol. 2004;183(2):249–56.
- Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. Nat Rev Endocrinol. 2011;7(10):569–80.
- Hsiao SJ, Nikiforov YE. Molecular approaches to thyroid cancer diagnosis. Endocr-Relat Cancer. 2014;21(5):T301–13.
- Dunn LA, Sherman EJ, Baxi SS, Tchekmedyian V, Grewal RK, Larson SM, et al. Vemurafenib redifferentiation of BRAF mutant, RAI-refractory thyroid cancers. J Clin Endocrinol Metab. 2019;104(5):1417–28.

- Rothenberg SM, McFadden DG, Palmer EL, Daniels GH, Wirth LJ. Redifferentiation of iodine-refractory BRAF V600E-mutant metastatic papillary thyroid cancer with dabrafenib. Clin Cancer Res. 2015;21(5):1028–35.
- Zaballos MA, Santisteban P. Key signaling pathways in thyroid cancer. J Endocrinol. 2017;235(2):R43–61.
- Sciacchitano S, Lavra L, Ulivieri A, Magi F, De Francesco GP, Bellotti C, et al. Comparative analysis of diagnostic performance, feasibility and cost of different test-methods for thyroid nodules with indeterminate cytology. Oncotarget. 2017;8(30):49421.
- Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci. 2018;115(13):E2970–9.
- Guo Y, Lyu T, Liu S, Zhang W, Zhou Y, Zeng C, et al. Learn to Estimate Genetic Mutation and Microsatellite Instability with Histopathology H &E Slides in Colon Carcinoma. Cancers. 2022;14(17):4144.
- Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer. 2019;10(20):4876.
- Fragopoulos C, Pouliakis A, Meristoudis C, Mastorakis E, Margari N, Chroniaris N, Koufopoulos N, Delides AG, Machairas N, Ntomi V, Nastos K. Radial basis function artificial neural network for the investigation of thyroid cytological lesions. J Thyroid Res. 2020;2020(1):5464787.
- Elliott Range DD, Dov D, Kovalsky SZ, Henao R, Carin L, Cohen J. Application of a machine learning algorithm to predict malignancy in thyroid cytopathology. Cancer Cytopathol. 2020;128(4):287–95.
- Anand D, Yashashwi K, Kumar N, Rane S, Gann PH, Sethi A. Weakly supervised learning on unannotated H &E-stained slides predicts BRAF mutation in thyroid cancer with high accuracy. J Pathol. 2021;255(3):232–42.
- Xi C, Du R, Wang R, Wang Y, Hou L, Luan M, et al. Al-BRAFV600E: a deep convolutional neural network for BRAFV600E mutation status prediction of thyroid nodules using ultrasound images. View. 2023;4(2):20220057.
- 22. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: International conference on machine learning. PMLR; 2021. pp. 8748–8763.
- Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. pp. 6105–6114.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6):84–90.
- Hinton G, Srivastava N, Swersky K. Overview of mini-batch gradient descent. Neural Netw Mach Learn. 2012;575(8).
- 26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
- 27. Shore J, Johnson R. Properties of cross-entropy minimization. IEEE Trans Inf Theory. 1981;27(4):472–82.
- Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30(7):1145–59.
- Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Proc Lett. 2014;21(11):1389–93.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
- Wang CW, Muzakky H, Lee YC, Lin YJ, Chao TK. Annotation-free deep learning-based prediction of thyroid molecular cancer biomarker BRAF (V600E) from cytological slides. Int J Mol Sci. 2023;24(3):2521.
- Ju G, Sun Y, Wang H, Zhang X, Mu Z, Sun D, et al. Fusion oncogenes in patients with locally advanced or distant metastatic differentiated thyroid cancer. J Clin Endocrinol Metab. 2024;109(2):505–15.

## **Publisher's Note**

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