Technology in Cancer Research & Treatment
Volume 24: 1-11

DOI: 10.1177/15330338251323168
journals.sagepub.com/home/tct
Sage

# An Early Thyroid Screening Model Based on Transformer and Secondary Transfer Learning for Chest and Thyroid CT Images

**Na Han, PhD[1,2], Rui Miao, PhD[3], Dongwei Chen, PhD[4], Jinrui Fan, MD[4], Lin Chen[3], Siyao Yue[3], Tao Tan, PhD[1], Bowen Yang, MD[5,*], and Yapeng Wang, PhD[1,*]**

**Abstract**

**Introduction:** Thyroid cancer is a common malignant tumor, and early diagnosis and timely treatment are crucial to improve patient prognosis. With the increasing use of enhanced CT scans, a new opportunity for early thyroid cancer screening has emerged. However, existing CT-based models face challenges due to limited datasets, small sample sizes, and high noise. **Methods:** To address these challenges, we collected enhanced CT scan image data from 240 patients in Guangdong and Xinjiang, China, and established a CT dataset for early thyroid cancer screening. We propose a deep learning model, the DVT model, which combines transformer DNN and transfer learning techniques to integrate time series data and address small sample sizes and high noise. **Results:** The experimental results show that the DVT model achieves a prediction accuracy of 0.96, AUROC of 0.97, specificity of 1, and sensitivity of 0.94. These results indicate that the DVT model is a highly effective tool for early thyroid cancer screening. **Conclusion:** The DVT model has the potential to assist clinicians in identifying potential thyroid cancer patients and reducing patient expenses. Our study provides a new approach to thyroid cancer screening using enhanced CT scans and demonstrates the effectiveness of deep learning techniques in addressing the challenges associated with CT-based models.

**Keywords**

thyroid cancer, vision transformer, transformer, DNN, chest CT, secondary transfer learning

## Abbreviation

In this article, all abbreviations and their full forms are listed in Table 1.

## Introduction

Thyroid cancer is a kind of malignant tumor, which can occur at any age, and its incidence is on the rise worldwide. Its early diagnosis and timely treatment are significant to improve patients' prognosis[1-3] Through standardized thyroid nodule management strategies, clinicians can more effectively identify the nature of thyroid nodules, improve the early diagnosis rate of thyroid cancer, and thus improve the treatment effect and quality of life of patients.[4,5] At present, many thyroid cancer detection models based on computed tomography (CT) image data have been proposed by researchers. For example, Arepalli et al used the deep learning algorithm of convolutional neural network to analyze, detect and classify thyroid cancer in CT thyroid images. In 2021, Zhao Hongbo et al analyzed the CT images of 880 patients with thyroid nodules, selected 5 convolutional neural networks and generated an ensemble model to identify benign and malignant thyroid nodules below.[6] In 2023,

[1] Faculty of Applied Sciences, Macao Polytechnic University, Macao, P. R. China
[2] Business School, Beijing Institute of Technology, Zhuhai, P. R. China
[3] Basic Teaching Department, Zhuhai Campus of Zunyi Medical University, Zhu Hai, P. R. China
[4] Zhuhai People's Hospital (Zhuhai Clinical Medical College of Jinan University), Zhuhai, P. R. China
[5] Image Center, The First People's Hospital of Kashi, Xinjiang, P. R. China

[*] Co-corresponding authors.

**Corresponding Authors:**
Bowen Yang, Image center, the First People's Hospital of Kashi, Xinjiang 844000, P. R. China.
Email: y434972074@163.com

Yapeng Wang, Faculty of Applied Sciences, Macao Polytechnic University, Macao, 999078, P. R. China.
Email: yapengwang@mpu.edu.mo

**Table 1.** Abbreviation.

| Abbreviation | Full Term and Explanation |
|---|---|
| AUC | Area Under the Curve: A metric used to evaluate the performance of binary classification models, specifically in the context of the Receiver Operating Characteristic (ROC) curve. |
| CNN_LSTM | Convolutional Neural Networks - Long Short-Term Memory: A combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, often used for sequence prediction tasks involving spatial and temporal data. |
| CLS Token | Classification Token: A special token used in models like BERT for classification tasks. It is added at the beginning of the input sequence to represent the entire sequence. |
| DNN | Deep Neural Network: A neural network with multiple hidden layers, enabling it to learn complex patterns in data. |
| DVT | Deep Learning-Based Transformer and Transfer Learning Thyroid Cancer Early Screening Model: A proposed model utilizing transformers and transfer learning for the early screening of thyroid cancer. |
| DenseNet_121 | Dense Convolutional Network-121: A type of Dense Convolutional Network (DenseNet) with 121 layers, known for its dense connectivity between layers. |
| GoogleNet | GoogleNet (Inception v1): A deep convolutional neural network architecture designed by Google, featuring multiple inception modules. |
| ICLR 2021 | International Conference on Learning Representations 2021: Refers to the 2021 edition of the International Conference on Learning Representations. |
| MLP | Multilayer Perceptron: A type of feedforward neural network consisting of multiple layers of neurons. |
| MHSA | Multi-Head Self-Attention: A key component of transformer models that allows the model to focus on different parts of the input sequence simultaneously. |
| ResNet | Residual Network: A type of deep neural network that uses residual connections to improve training performance. |
| ViT | Vision Transformer: A model that applies the Transformer architecture to computer vision tasks, enabling the processing of image data in a similar manner to text data. |
| VGG16 | Visual Geometry Group-16: A convolutional neural network architecture with 16 layers, developed by the Visual Geometry Group (VGG) at the University of Oxford. |
| AlexNet | AlexNet: A deep convolutional neural network architecture named after Alex Krizhevsky, one of the authors of the paper that introduced it. |

Wang Chujun et al collected 676 thyroid ultrasound images from 338 thyroid cancer patients and used ResNet-18 as the basic network to build a deep learning model to predict whether cervical lymph nodes had metastasis.[7]

However, in daily life, most patients do not initiate a specialized thyroid scan. Therefore, it is difficult for such a model to be directly used in the early screening of thyroid cancer. At present, with the COVID-19 pandemic in 2019, enhanced CT scanning for the chest has gradually become a routine test option for ordinary patients in Chinese hospitals[8,9] We have observed that many hospitals scan part or most of the thyroid area when performing chest scan, which provides a new idea for us to establish early thyroid screening based on routine detection data: In order to more effectively assist doctors in early detection of thyroid cancer, we proposed a CT image examination model combining routine examination and artificial intelligence.[10-12] When performing routine tests on ordinary patients in hospitals, the model can initially analyze enhanced CT scans containing the thyroid to determine whether the thyroid is normal.[13,14] Once the model identifies a potential cancer risk, the patient is quickly referred to a specialist for further evaluation and testing. This integrated approach is expected to significantly improve the efficiency and accuracy of thyroid cancer detection.

At present, early screening of thyroid cancer based on enhanced CT data faces four major challenges. Firstly, it is the lack of available CT data sets which includes both the thoracic and thyroid areas. Second, the current data has the problem of small sample size and loud noise in the thyroid area. Due to routine test data, the CT scan images of each patient included both the thyroid gland and the chest cavity, of which the thyroid region accounted for a small proportion. This results in fewer thyroid regions available and may contain more noise, which makes it difficult for us to build an effective thyroid region feature extraction model. Thirdly, how to extract and integrate thyroid region features of time series.[5,15-17] The thyroid image data based on enhanced CT is typical continuous time series data, and its adjacent images have significant correlation. How to extract single thyroid image features and integrate time series features is the third major challenge facing the model. Fourthly, The field of medical imaging commonly faces the issue of scarce sample availability, particularly in specific tasks such as thyroid cancer screening, where insufficient data can lead to inadequate model training, thereby affecting its generalization capability and predictive accuracy. Based on the above problems, we first collected enhanced CT scan image data of 240 patients in Zhuhai People's Hospital, Guangdong, China & The First People's Hospital of Kashi, Xinjiang, China, and established CT data set for early screening of thyroid cancer. Then, we proposed a deep learning based on transformer and transfer learning thyroid cancer early screening model (DVT model), as shown in Figure 1. The DVT model consists of two main parts. The first is the separation model of thyroid region and thoracic region. We construct the model based on vision transformer (ViT) and carry out the first transfer learning based on ImageNet dataset and retain the learning results. The second is the early screening model of thyroid cancer. We splice two adjacent thyroid images together and input the ViT to integrate

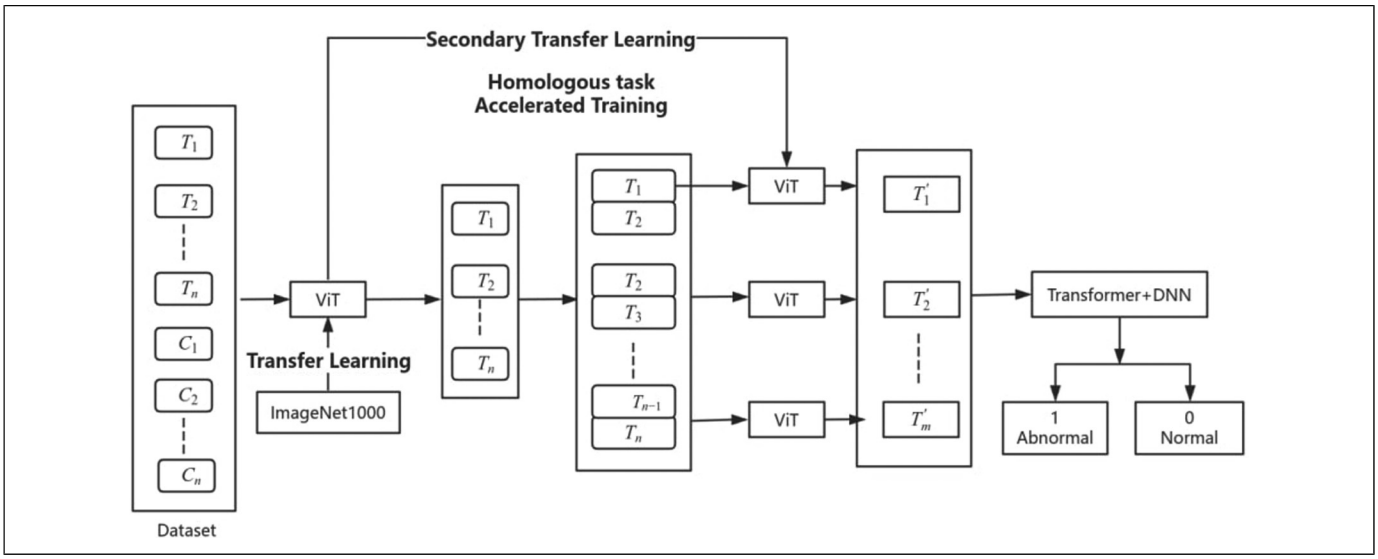


**Figure 1.** The framework of DVT model (T is thyroid, C is chest).

**Table 2.** Describes the Device Collection Parameters.

| | Zhuhai People's Hospital, Guangdong, China | The First People's Hospital of Kashi, Xinjiang, China |
|---|---|---|
| Mode | GE Revolution CT GSI | SIEMENS SOMATOM Definition Flash |
| Scanning range | C4~L2 | C4~L2 |
| Scanning direction | Head side to foot side | Head side to foot side |
| Parameter Setting | 80~120 kVp Fast switching voltage | 100 kV. Ref. kV: 120 |
| noise index, NI | 5.5 | 5.0 |
| ASIR | 50% | SAFIRE level 3 |
| Speed | 0.5 s/r | 0.5 s/r |
| Pitch | 0.992 : 1 | 0.8 |
| The collimation width | 64 × 0.625 mm | 64 × 0.625 mm |
| Layer thickness | 1.25 mm | 1.0/0.75 mm |
| The layers apart | 1 mm | 0.6~1.0 mm |
| Reconstruct the matrix | 512 × 512 | 512 × 512 |

the time series data. At this stage, the ViT model employs the homologous data from the thyroid and thoracic region segmentation model as an auxiliary domain for secondary transfer learning. There are two advantages to this approach. First, the stitching of adjacent thyroid images can better integrate the features of adjacent areas and reduce the calculation amount of the transformer model.[18-23] Second, because the task objectives of the thyroid separation model and the thyroid cancer early screening model are similar, the model can adapt to the thyroid cancer recognition task more quickly, which can solve the problem of small samples and noise in the data set.[24-26] Finally, we design a transformer-DNN model for classification output.[27-32]

We conducted three independent experiments, namely, the Guangdong dataset experiment, the Xinjiang dataset experiment, and the ablation experiment. The experimental results are shown in the second half of the paper. The results show that the DVT model is superior to the other reference model, and the results of the ablation experiment also verify the effectiveness of the proposed secondary transfer learning.

## Dataset and Model

### *Data set Construction Process*

First, the research team collected combined chest and thyroid CT scans of 120 patients at Zhuhai People's Hospital, Guangdong, China, detailed information was de-identified. The dataset consisted of 60 normal subjects and 60 thyroid cancer subjects, all of which contained thoracic region and no less than 20 thyroid region images. The dataset of the First People's Hospital of Kashi, Xinjiang, China, detailed information was de-identified. Like the People's Hospital dataset, contained 120 images of the thoracic region and no less than 20 images of the thyroid region, including 60 normal subjects and 60 thyroid cancers. Table 2 describes the device collection parameters. Table 3 summary of two batches of patient data.

**Table 3.** Summary of Two Batches of Patient Data.

| Batch | Total Patients | Number of Male Patients | Number of Female Patients | Average Age (years) | Age Range (years) | Number of Cancer Patients | Number of Non-Cancer Patients |
|---|---|---|---|---|---|---|---|
| Zhuhai | 120 | 40 | 80 | 48 | 18-75 | 60 | 60 |
| Xinjiang | 120 | 50 | 70 | 47 | 20-70 | 60 | 60 |
| Total | 240 | 90 | 150 | 47 | 18-75 | 120 | 120 |

Note: The data is presented in a summarized form without including any personal identification information.

## Model

*Thyroid Region Separation Model.* In this paper, vision transformer (ViT) is used to separate thyroid regions. The traditional Transformer structure is made up of the Encoder-decoder framework, while for ViT, only the Encoder part is used. The input of standard Transformer is one-dimensional sequence data, so the image needs to be converted into serial data. The idea of ViT model is to slice an image into patches of fixed size without overlapping and then convert each patch into a one-dimensional vector through stretching operation. Finally, the input patches are converted to a fixed-length vector called patch embedding through a linear transformation layer. Since the final output should be a label for classification tasks, the author has adjusted the input to Transformer Encoder by adding a CLS Token at the beginning of the input sequence. In 2020, Google proposed the paper An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale,[18] which has been included in ICLR 2021. Vision Transformer (ViT) is proposed for the first time to apply Transformer structure to image classification in CV field. The paper shows that compared with the current best convolutional neural network structure, ViT still achieves satisfactory results and requires fewer computing resources.

First, we assume that each patient contains n consecutive thyroid images, the size of each CT image input represented as X , as shown in equation (1)

$$X = [n,\ 224,\ 224] \tag{1}$$

Next, the picture is cut into 16 pieces of equal size, and each cut module is converted to $K$, ie,

$$K = [16,\ 3136] \tag{2}$$

That is, we can finally get the input tensor $L$ of the model.

$$L = [n, 16, 3136] \tag{3}$$

We input the tensor $L$ into a standard Transformer encoder layer consisting of Multi-Head Self-Attention (MHSA) and Multi-Layer Perceptron (MLP), as shown in equation (4). Thus, we get the calculated $H$, whose dimensions are the same as $[n, 16, 3136]$.

$$H = transformerencoder(L) \tag{4}$$

Next, we construct a Deep Neural Network (DNN) model for thyroid region separation. To meet the input requirements of the DNN model, we reshape $H$.

$$H = [n, 50176] \tag{5}$$

Next, we put the converted tensor $H$ into a three-layer DNN model, where the number of hidden layer neurons is 512 and the output number is 2, and then we get a new output tensor $G$, that is

$$G = DNN(H) \tag{6}$$

*Thyroid Region Feature Fusion and Classification Model.* Since the thyroid pictures of each patient are multiple and continuous, considering the spatial relationship between these adjacent pictures, to better extract features and reduce the complexity of the model, we splice the two adjacent pictures into one picture horizontally as the data input of the thyroid cancer recognition module. For example, 20 pictures in 3.1 are merged into 10 pictures. Currently, we resize the picture, that is, we get formula (8) from formula (7), which is convenient for transfer learning.

$$J = [n, 224, 448] \tag{7}$$

$$J = [n, 224, 224] \tag{8}$$

The same cutting operation is performed with formula (2) (3) in 3.1 to obtain formula (9), ie

$$M = [n, 16, 3136] \tag{9}$$

Similarly, we input the tensor $M$ into a standard Transformer encoder layer, as shown in equation (10). Thus, we get the calculated $M$, whose dimensions are the same as $[n, 16, 3136]$.

$$M_1 = transformerencoder(M) \tag{10}$$

To meet the input requirements of DNN model, we reshape $M_1$

$$M_1 = [n, 50176] \tag{11}$$

Next, the converted tensor $M_1$ is put into a two-layer DNN model, and the output formula (13) is obtained.

$$M_2 = DNN(M1) \tag{12}$$

$$M_2 = [n, 512] \tag{13}$$

$M$ is the number of patients, and $n$ is the combined CT images of each patient.

$$W = [m, n, 512] \tag{14}$$

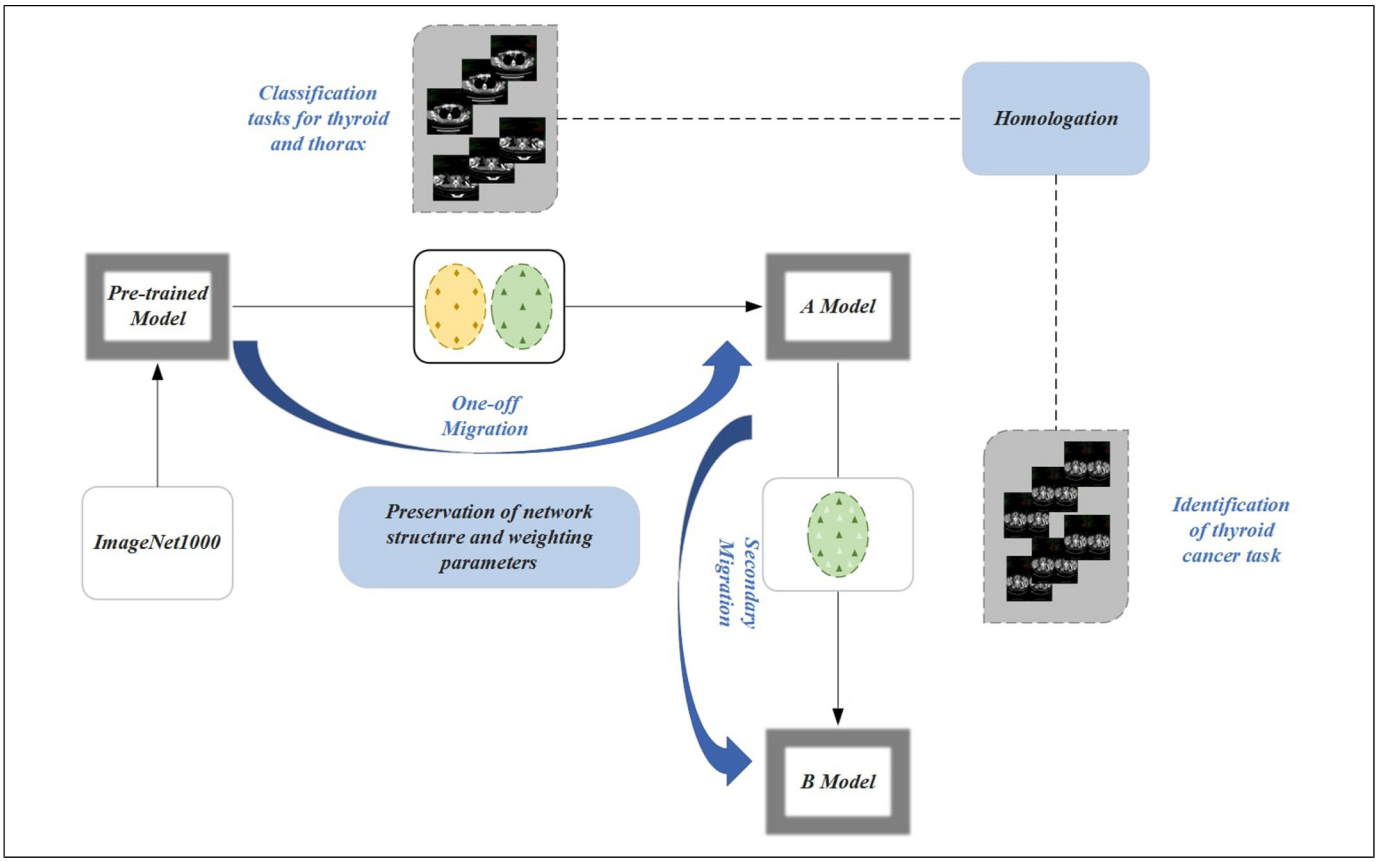


**Figure 2.** Structure of secondary migration of DVT model.

Again, we input the tensor W into a standard Transformer encoder layer, as shown in equation (9). Thus, we get the calculated $W$, whose dimensions are the same as [$m$, $n$, 512].

$$W = [m, \text{n}, 512] \tag{15}$$

$$W_1 = transformer\ encoder(W) \tag{16}$$

Next, the tensor $W_1$ is put into a three-layer DNN model, where the number of hidden layer neurons is 512 and the number of outputs is 2.

$$W_2 = DNN(W_1) \tag{17}$$

There are two kinds of output results, where [1,0] determines that the picture is diseased thyroid, and [0,1] determines that the picture is normal thyroid.

*Secondary Transfer Learning.* First, our pre-trained model uses ImageNet1000 as a data set for training, then we reuse the trained network structure and connection parameters,[33] input the patient's original CT image set into the model, and train the model to recognize the thyroid and chest cavity, as shown in formula (4). Model, this is a transfer learning; Similarly, we will re-use the model after secondary training and input the patient's thyroid picture data set into it to distinguish diseased thyroid from normal thyroid, that is, formula (10), which is secondary migration,[34,35] as shown in Figure 2.

In this way, the training of our model is a gradual process, and after the first migration, the model can better extract the biometric characteristics of the thyroid. In the secondary transfer process, the training efficiency of the model will be improved, and the recognition effect will be better.

*Statistical Analyses.* In this study, a variety of indicators were used to evaluate the performance of the model comprehensively, including Accuracy, Recall, F1-score, Area Under the Curve (AUC) and Specificity. The evaluation indicators are explained in detail as follows:

Accuracy is one of the most used evaluation indexes in classification models, which represents the proportion of samples that the model predicts correctly in the total number of samples. It is suitable for cases where the distribution of categories is more balanced, but can be misleading when the categories are unbalanced The calculation formula is as follows:

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

Where, TP (True Positive): The number of samples that the model correctly predicts to be positive.TN (True Negative): The number of samples that the model correctly predicts as a negative class. FP (False Positive): indicates the number of

samples that the model incorrectly predicts as positive. FN (False Negative): The number of samples that the model incorrectly predicts as a negative class.

Recall rate, also known as Sensitivity, measures the ability of a model to identify positive samples. Recall rates are key indicators in scenarios where missing positive samples need to be minimized, such as disease screening. The calculation formula is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The F1-score is a harmonic average of accuracy and recall and is used as a comprehensive measure of the model's performance. It is particularly important when the categories are unbalanced and need to balance considerations of accuracy and recall. The calculation formula is as follows:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where, Precision is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

AUC refers to the Area Under the Curve, which is used to evaluate the overall performance of a classification model at different thresholds. The Receiver Operating Characteristic (ROC) curve plots the relationship between the true rate (TPR) and the false positive rate (FPR). The AUC ranges from 0.5 to 1, with higher values indicating better model performance. AUC is suitable for the need to assess the consistency of the performance of models at different thresholds, especially in medical diagnosis. The calculation formula is as follows:

$$\text{AUC} = \int_0^1 \text{TPR}(\boldsymbol{FPR})\, \boldsymbol{dFPR}$$

Specificity measures the model's ability to correctly identify negative class samples. It is a key indicator in scenarios where there is a need to reduce false positives for negative samples, such as reducing misdiagnosis in cancer screening. The calculation formula is as follows:

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

Statistical analysis method:

To verify the performance of the DVT model, we first calculate the accuracy and AUROC and plot the ROC curve to compare the performance of the model. We also calculated precision, sensitivity, specificity, and f1 score indicators, considering the sample's imbalance. The experiment consists of three parts: We first build a model based on 120 data from Zhuhai People's Hospital. We chose AlexNet, DenseNet_121, Resnet, GoogleNet, VGG16 and CNN_LSTM for comparison.[36-39] The purpose of this part is to prove that the DVT model performs better than the reference model; In the second part, to prove the universality of the DVT model, we selected 120 cases from a hospital in Xinjiang for verification. In the third part, we conducted the ablation experiment, that is, we compared the results under the same conditions without transfer learning.

Through multi-dimensional evaluation indicators and rigorous statistical analysis methods, this study comprehensively evaluates the performance of the classification model to ensure the reliability and scientific results. The comprehensive application of these evaluation indexes provides a solid foundation for the optimization and practical application of the model.

## Results

### *Comparative Experiment*

The experimental results show that the DVT model is superior to the six classical AI methods mentioned above. As seen from Table 4 and Figure 3, DVT model has the best performance in accuracy, AUC and sensitivity, whose values are 0.88, 0.92 and 0.91, respectively. The AUC curve of DVT model also shows that DVT model is better than other models in the experimental group. The results show that the DVT model has su-perior performance. The specificity and sensitivity of DVT model can reach 0.84 and 0.91, respectively. These results show that the predictive performance of DVT models is still ideal, which has important implications for thyroid prediction models. The accuracy, AUROC, specificity, sensitivity, precision and F1-SCORE were 0.88, 0.92, 0.84, 0.91 and 0.87, respectively. In terms of precision, the AlexNet, DenseNet-121 and VGG16 models have high values of 0.71, 0.85 and 0.76, respectively. As for specificity, VGG16 had the best performance of specificity, reaching 1, indicating that the model had high accuracy in predicting negative classes. The Convolutional Neural Networks - Long Short-Term Memory (CNN-LSTM) model performs well in the F1-score, reaching 0.79, which indicates that the model strikes a balance between accuracy and recall.

**Table 4.** The Running Results of the First Peo ple's Hospital of Kashi Dataset in Different AI Model Groups.

| Model | AlexNet | DenseNet-121 | GoogLeNet | VGG16 | CNN-LSTM | Resnet | DTV |
|---|---|---|---|---|---|---|---|
| Precision | 0.71 | 0.85 | 0.76 | 0.76 | 0.81 | 0.79 | 0.88 |
| Recall | 0.71 | 0.79 | 0.75 | 0.76 | 0.84 | 0.79 | 0.88 |
| F1-score | 0.71 | 0.78 | 0.75 | 0.75 | 0.79 | 0.79 | 0.87 |
| AUC | 0.72 | 0.84 | 0.71 | 0.78 | 0.79 | 0.81 | 0.92 |
| accuracy | 0.71 | 0.79 | 0.75 | 0.75 | 0.79 | 0.79 | 0.88 |
| Specificity | 0.75 | 0.53 | 0.83 | 1 | 0.83 | 0.69 | 0.84 |

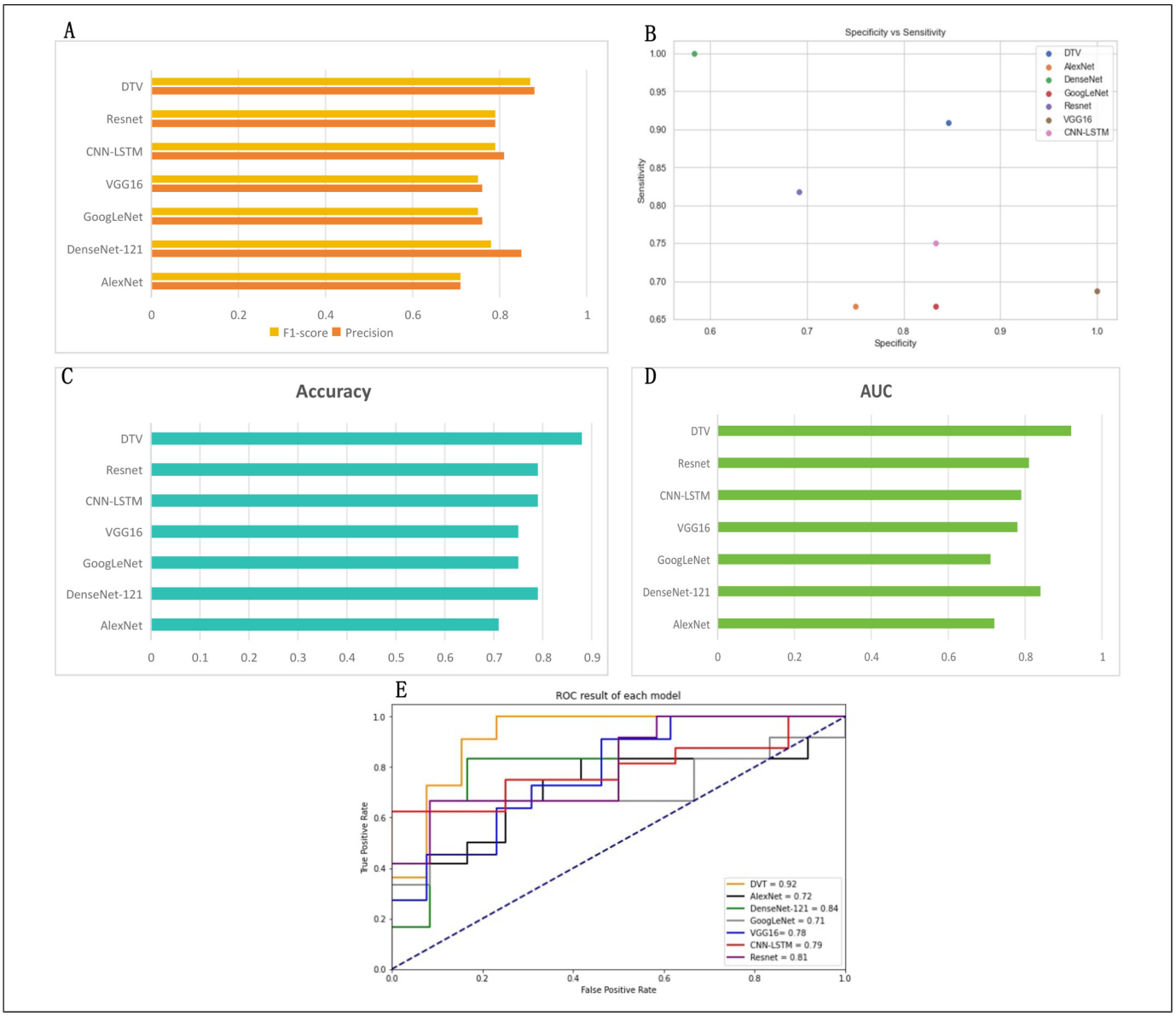


**Figure 3.** The results of chest CT dataset of Zhuhai People's Hospital: (A) F1-score and precision result of each model, (B) sensitivity and specificity result of each model, (C) accuracy result of each model, (D) AUC-ROC result of each model, (E) ROC curve of each model.

As mentioned above, the experimental results show that the DVT model has achieved satisfactory results in thyroid cancer recurrence and is superior to the existing single omics AI model on all experimental indicators.

## *Universality Experiment*

To prove the universality of the DVT model, we also selected mixed CT images of chest and thyroid of 120 patients from a hospital in Xinjiang for verification. The results are shown in Table 5 and Figure 4. The DVT model performs best in accuracy and AUC (0.88 and 0.91, respectively). This means that the model has high accuracy and good classification ability in the overall prediction. In terms of precision, both the DenseNet-121 and DVT models performed well, with 0.84 and 0.81, respectively. This shows that these two models have high accuracy in predicting positive cases. VGG16 was an outstanding model for specificity, reaching a value of 1, which meant that the model had extremely high accuracy in predicting negative cases, with almost no misdiagnosis of negative cases. CNN-LSTM model has a superior performance in terms of sensitivity, reaching 0.94, which indicates that the model has a strong ability to recognize positive examples and rarely misses positive examples. Overall, each model has its strengths and weaknesses, and choosing the model that best suits the specific task needs depends on the data characteristics and task objectives, but in this set of data, the DVT model performs well on multiple metrics and may be an excellent choice.

**Table 5.** The Running Results of the First People's Hospital of Kashi, Dataset in Different AI Model Groups.

| Model | AlexNet | DenseNet-121 | GoogLeNet | VGG16 | CNN-LSTM | Resnet | DTV |
|---|---|---|---|---|---|---|---|
| Precision | 0.81 | 0.84 | 0.81 | 0.75 | 0.78 | 0.75 | 0.81 |
| Recall | 0.79 | 0.83 | 0.78 | 0.77 | 0.83 | 0.75 | 0.88 |
| F1-score | 0.75 | 0.83 | 0.78 | 0.76 | 0.8 | 0.75 | 0.84 |
| AUC | 0.82 | 0.84 | 0.73 | 0.74 | 0.8 | 0.84 | 0.91 |
| accuracy | 0.75 | 0.83 | 0.79 | 0.79 | 0.83 | 0.79 | 0.88 |
| Specificity | 1 | 0.8 | 0.88 | 0.63 | 0.88 | 0.63 | 0.625 |

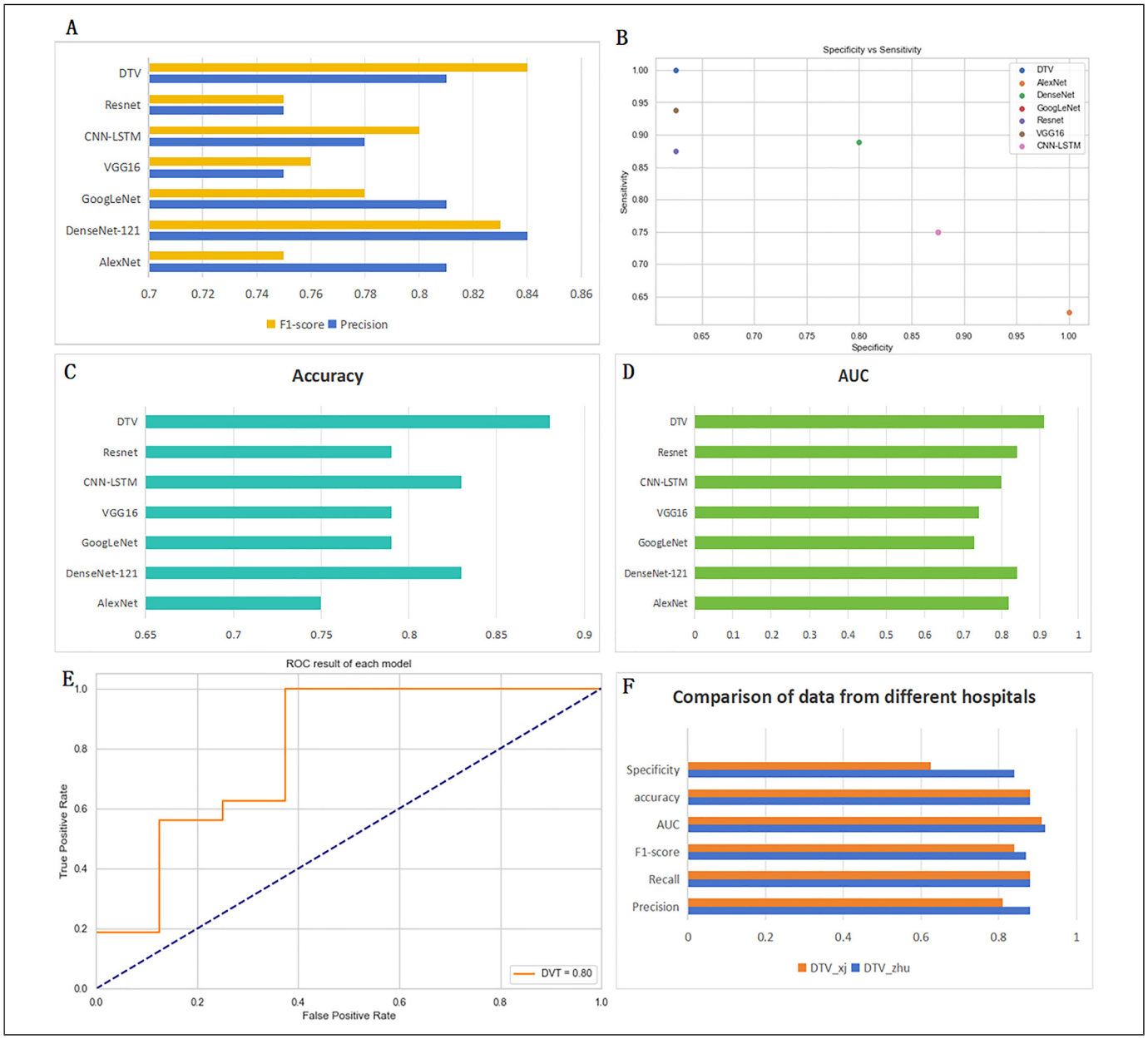


**Figure 4.** The results of chest CT dataset of a hospital in Xinjiang: (A) F1-score and precision result of each model, (B) sensitivity and specificity result of each model, (C) accuracy result of each model, (D) AUROC result of each model, (E) AUC curve of each model, (F) comparison of data from dif-ferent hospitals(DVT_xj is results of the First People's Hospital of Kashi, DVT_zhu is results of the Zhuhai People's Hospital).

**Table 6.** The Results of Data Set of Zhuhai People's Hospital Before and After Ablation Experiment (DVT_zhu d Results Before the Ablation Experiment, DVT_xr_zhu Results After the Ablation Experiment).

| | Precision | Recall | F1-score | AUC | accuracy | Specificity |
|---|---|---|---|---|---|---|
| DVT_zhu | 0.88 | 0.88 | 0.87 | 0.92 | 0.88 | 0.84 |
| DVT_xr_zhu | 0.69 | 0.68 | 0.71 | 0.8 | 0.79 | 0.375 |

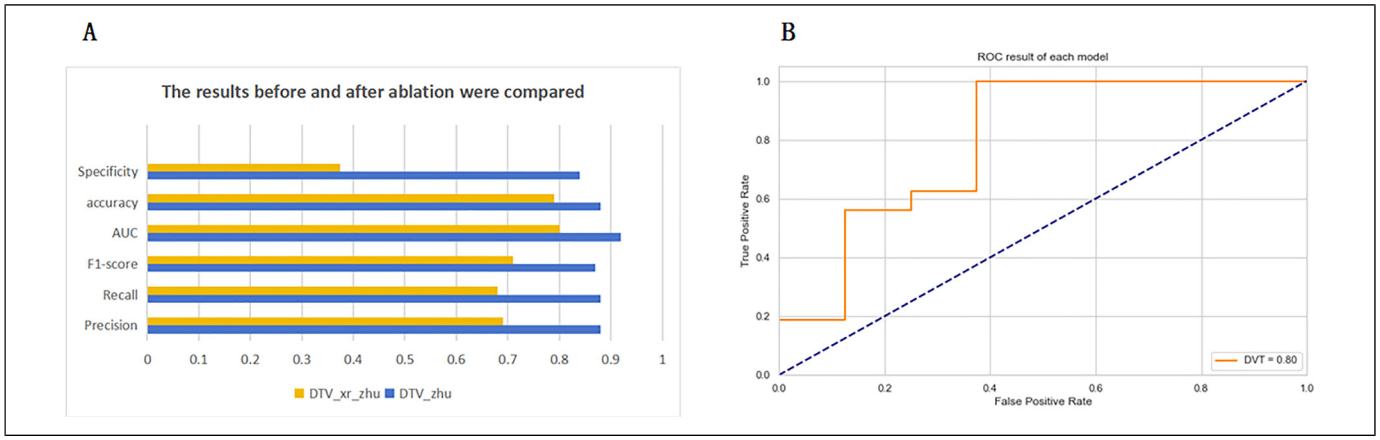


**Figure 5.** The results of chest CT dataset of Zhuhai People's Hospital: (A) comparison of model performance before and after ablation study (DTV_xr_zhu is results of DTV model ablation experiment, DTV_zhu is results of DTV model experiment), (B) ROC curve of DTV model ablation experiment.

### *Ablation Experiment*

In order to verify the selectivity of this method, we conducted ablation experiments,[40,41] that is, under the same conditions, no transfer learning was used. From Table 6 and Figure 5, it can be clearly seen that without transfer learning, the first is that the result is significantly worse, and the second is that the prediction effect of malignant patients is significantly worse. In addition, the model tends to bias the prediction results to 1 more often, which also indicates that the model may have potential overfitting problems.

## Discussion

Chinese tertiary hospitals and above in different regions, built an additional multi-center data set to solve the data challenges; Secondly, for the data, there are problems such as less available area, small sample size and large noise. In this paper, a secondary transfer learning method is proposed. The idea of this method is to use homologous tasks for multiple transfer learning, so that the model can quickly adapt to small samples and high noise thyroid cancer prediction tasks. It is suggested that chest CT containing thyroid gland can be used for early screening of thyroid cancer. Finally, the time series of thyroid continuous CT scan data. In this paper, transformer model framework is used for feature fusion. In this paper, adjacent images are spliced together for feature calculation, which can effectively reduce the calculation load of transformer and better calculate the features of adjacent images. The experimental results of data sets in Guangdong and Xinjiang both show that the results of DVT model are better than the benchmark model. Ablation experiments show that transfer learning can effectively improve the learning effect of the model. We have reason to believe that the DVT model can effectively help clinicians identify potential patients in the early screening of thyroid cancer, intervene in advance, and reduce the mortality of patients.

Although the experiment has yielded satisfactory results, there are still four main limitations in this study. First, the sample size is relatively small, and the data from 240 patients may not be sufficient to train a model with adequate robustness. Second, the training speed remains relatively slow. Compared to traditional convolutional neural networks (CNNs), transformer models (Transformers) significantly increase computational load, which limits their efficiency in practical applications. Third, the experiment was conducted only on enhanced chest CT images with relatively high image quality and has not been validated on standard chest CT images. Finally, this study did not perform more detailed classification of thyroid cancer, such as the analysis of potential metastatic lymph nodes. Future work will involve collecting additional data and exploring the combination of traditional convolutional neural networks with efficient transformer algorithms to optimize both the model's predictive accuracy and training speed. Furthermore, through image enhancement techniques, we plan to extend the application of the model (DVT) to standard chest CT images, aiming to achieve similar high accuracy on these images. At the same time, the research will continue to

focus on data collection and model optimization to enable the model to accurately classify additional categories, such as potential metastatic lymph nodes.

## Conclusions

This study evaluated the performance of the DVT model, demonstrating its superiority over six classical artificial intelligence models in thyroid cancer prediction tasks. Based on the results for accuracy, AUC, sensitivity, specificity, precision, and F1-score, the DVT model outperformed all other models across all key metrics, including accuracy (0.88), AUC (0.92), sensitivity (0.91), and precision (0.88). The AUC curve further validated the DVT model's exceptional predictive capability, indicating its greater robustness and reliability compared to other models. We believe that the DVT model can effectively assist clinicians in identifying potential thyroid cancer patients during early screening, enabling timely interventions and reducing patient mortality rates.

### Acknowledgments

This work is funded by Macao Polytechnic University under grant no.: RP/FCA-04/2022, and under submission control code fca.3ca7.92b0.8. Guangdong Provincial Department of Education youth innovative talent project (No. 2023KQNCX155), Postdoctoral training project of Zunyi Medical University (No.2023F-ZH-019).

### Author Contributions

Conceptualization, N.H. and R.M.; methodology N.H., R.M., and D.W. C.; Software, M.R. and N.H.; Validation, L.C. and Y.S.Y.; Formal Analysis, Y.P.W.; Survey, D.W. C.; Resources, G.R.F., D.W.C., and B.Y.; Data Management, L.C.; Writing - Original manuscript preparation, N.H. and R.M. ; Writing - Reviewing and editing, N.H. and R.M.; Visualization, R.M., D.W.C.; Supervised Y.P.W. and B.Y.; Project Management, R.M.; All authors have read and agreed to the published version of the manuscript.

### Declaration of Competing Interest

All authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Ethical Compliance

Research experiments conducted in this article with animals or humans were approved by the Ethical Committee Zhuhai People's Hospital, approval number is 2024.109. and the Ethical Committee of the First People's Hospital of Kashgar Region, approval number is 2024.93.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Industry-University-Research cooperation and basic and applied basic research project cooperation, Zhuhai, grant number 2220004002437. Guangdong Provincial Department of Education youth innovative talent project (No. 2023KQNCX155), Postdoctoral training project of Zunyi Medical University (No.2023F-ZH-019). This work is supported by Science and Technology Development Fund of Macao (0021/2022/AGJ).

### ORCID iD

Na Han https://orcid.org/0009-0003-9740-2689
Rui Miao https://orcid.org/0000-0001-5292-4377
Dongwei Chen https://orcid.org/0000-0001-6005-1748
Jinrui Fan https://orcid.org/0009-0001-5707-7223
Lin Chen https://orcid.org/0009-0007-6342-2736
Siyao Yue https://orcid.org/0009-0006-5767-6599
Tao Tan https://orcid.org/0000-0001-5403-0887
Bowen Yang https://orcid.org/0009-0004-5060-8417
Yapeng Wang https://orcid.org/0000-0002-1085-5091

### References

1. Zhou Y, Walter FM. Understanding cancer risk in patients at lower risk to improve early cancer diagnosis. *Lancet Oncol.* 2023;24:1166-1167. doi:10.1016/S1470-2045(23)00514-4
2. Yan S, Yansong L, Yiming L, et al. National guidelines for diagnosis and treatment of thyroid cancer 2022 in China (English version). *Cancer Res China: Engl*. 2021;34:131-150. doi:10.21147/j.issn.1000-9604.2022.03.01
3. Zhang Q, Zhang S, Li J, et al. Improved diagnosis of thyroid cancer aided with deep learning applied to sonographic text reports: a retrospective, multi-cohort, diagnostic study. *Cancer Biol Med*. 2022;5:733-741.
4. Sung H, Ferlay J, Siegel RL, et al. Lobal cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209-249.
5. Zhu YC, Du H, Jiang Q, et al. Machine learning assisted Doppler features for enhancing thyroid cancer diagnosis. *J Ultrasound Med.* 2022;41(8):1961-1974. doi:10.1002/jum.15873
6. Zhang XY, Lee VCS, Lee JC, Liu F. Deep convolutional neural networks in thyroid disease detection: A multi-classification comparison by ultrasonography and computed tomography. *Comput Methods Programs Biomed.* 2022;220:106823. doi:10.1016/j.cmpb.2022.106823
7. Wang CJ. Preliminary study of deep learning model for diagnosing thyroid nodules and predicting lymph node metastasis in thyroid cancer based on ultrasound images. Master's thesis. Nanchang University; 2023.
8. Yan Q, Wang B, Gong D, et al. COVID-19 chest CT image segmentation – a deep convolutional neural network solution. arXiv e-prints. 2020. doi:10.48550/arXiv.2004.10987
9. Yan Q, Wang B, Gong D, et al. COVID-19 chest CT image segmentation network by multi-scale fusion and enhancement operations. *IEEE Trans Big Data*. 2021;7(1):13-24. doi:10.1109/TBDATA.2021.3056564
10. Han N, Fan J, Chen DW, Wang YP. Application of artificial intelligence in the diagnosis of thyroid cancer with enhanced computed

tomography. *J Mech Med Biol.* 2024;24:2440017.1-2440017.12. doi:10.1142/S0219519424400177
11. Cao K, Xi Y, Yao J, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat Med.* 2023;29(12):3033-3043. doi:10.1038/s41591-023-02640-w
12. Rorat M, Jurek T, Simon K, Guziński M. Value of quantitative analysis in lung computed tomography in patients severely ill with COVID-19. *PLoS One*. 2023;16(5):34-45. doi:10.1371/journal.pone.0251946
13. Leelavathi A, Venkata RK, Madhavi D. Deep learning approach for analyzing and predicting thyroid cancer in computed tomography images: Review. *SN Comput Sci*. 2023;4(5):622. doi:10.1007/s42979-023-02010-w
14. Wang H, Wang J, Wang X, et al. Comments on national guidelines for diagnosis and treatment of thyroid cancer 2022 in China (English version). *Chin J Cancer Res*. 2023;34(5): 447-450.
15. Lopez B, Fligor SC, Randolph GW, James BC. Inequities in thyroid cancer care: populations most at risk for delays in diagnosis and treatment. *Thyroid*. 2023;33(6):724-731. doi:10.1089/thy.2022.0723
16. Nagendra L, Pappachan JM, Fernandez CJ. Artificial intelligence in the diagnosis of thyroid cancer: recent advances and future directions. *Artif Intell Cancer*. 2023;4(1):1-10.
17. Zhang Y, Lu YY, Li W, et al. Lymphatic contrast-enhanced US to improve the diagnosis of cervical lymph node metastasis from thyroid cancer. *Radiology*. 2023;307(4):e221265-e221273. doi:10.1148/radiol.221265
18. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. arXiv. doi:10.48550/arXiv.2010.11929
19. Liu Z, Lin Y., Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. arXiv. doi:10.48550/arXiv.2103.14030
20. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training data-efficient image transformers & distillation through attention. arXiv. 2020. doi:10.48550/arXiv.2012.12877
21. Touvron H, Cord M, Sablayrolles A, et al. Going deeper with image transformers. arXiv.2103.17239. 2021. doi:10.48550/arXiv.2103.17239
22. Xiao T, Singh M, Mintun E, Darrell T, Dollár P, Girshick R. Early convolutions help transformers see better. arXiv.2106.14881. 2021. doi:10.48550/arXiv.2106.14881
23. Chong CF, Wang Y, Ng BK, Yang X. Projective transformation rectification for camera-captured chest x-ray photograph interpretation with synthetic data. arXiv. 2022. doi:10.48550/arXiv.2210.05954
24. Fang X, Chong CF, Wong KL, Wang Y, Ke W, Zhang T. Analysis of the two-step heterogeneous transfer learning for laryngeal blood vessel classification: issue and improvement. arXiv. 2024. doi:10.48550/arXiv:2402.19001
25. Meng J, Tan Z, Yu Y, Wang P, Liu S. TL-med: a two-stage transfer learning recognition model for medical images of COVID-19. *Biocybern Biomed Eng*. 2024;42(3):842-855. doi:10.1016/j.bbe.2022.04.005
26. Yin L, Liuy Y, Pei M, Li J, Wu M, Jia Y. Laryngoscope8: laryngeal image dataset and classification of laryngeal disease based on attention mechanism. *Pattern Recognit Lett.* 2021;150:207-213. doi:10.1016/j.patrec.2021.06.034
27. Song Y, Tang H, Meng F, et al. A transformer-based low-resolution face recognition method via on-and-offline knowledge distillation. *Neurocomputing*. 2024;509:193-205. doi:10.1016/j.neucom.2022.08.058
28. Narayanan AL, Wu H, Ma J, Huang M. RADFormers for detection and classification of hand gestures. *IEEE Sensors J.* 2024;24(6):9093-9103. doi:10.1109/JSEN.2024.3352492
29. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv. 2017. doi:10.48550/arXiv.1706.03762
30. Bogoychev N. Not all parameters are born equal: attention is mostly what you need. arXiv.11859. 2020. doi:10.48550/arXiv.2010.11859
31. Chen C, Zha Y, Zhu D, Ning K, Cui X. Attention is all you need for general-purpose protein structure embedding. Cold Spring Harbor Laboratory. 2021. doi:10.1101/2021.01.31.428935
32. Li C. Data & channel efficient vision transformer. 2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET), 2023:1-5. doi:10.1109/CCET59170.2023.10335121
33. Hamida S, Gannour OE, Cherradi B, Ouajji H, Raihani A. Handwritten computer science words vocabulary recognition using concatenated convolutional neural networks. *Multimed Tools Appl.* 2023;82(15):23091-23117. doi:10.1007/s11042-022-14105-2
34. Tang H, Xie Y. Deep transfer learning for connection defect identification in prefabricated structures. *Struct Health Monit*. 2023;22:2128-2146. doi:10.1177/14759217221 119537
35. Loc CV, Viet TX, Viet TH, et al. Pre-trained language model-based deep learning for sentiment classification of Vietnamese feedback. *Int J Comput Intell Appl.* 2023;22(03):2350016. doi:10.1142/s1469026823500165
36. Gehlot M, Saini ML. Analysis of different CNN architectures for tomato leaf disease classification. 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2021:1-6. doi:10.1109/ICRAIE51050.2020.9358279
37. Víctor V, Daniel RF. Synaptic metaplasticity for image processing enhancement in convolutional neural networks. *Neurocomputing*. 2021;462:534-543. doi:10.1016/j.neucom.2021.08.021
38. Ozyurt F. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *J Supercomput.* 2020;76(11):8413-8431. doi:10.1007/s11227-019-03106-y
39. Gautam S, Kumar A. Automatic traffic light detection for self-driving cars using transfer learning. *Intell Sustain Syst*. 2022;333:597-606. doi:10.1007/978-981-16-6309-3_56
40. Xuebao W, Ao H, Min Z, Li J, Li S, Li X. Preparation and ablation properties study of a single component ceramifiable RTV silicone rubber. *J Rubber Res*. 2023;26(5):391-405. doi:10.1007/s42464-023-00222-6
41. Li M, Cheng K, Wen J, Tian R. Study on diffusion and migration characteristics of impurities in secondary side of simplified u-tube steam generator. *Nucl Eng Des.* 2023;409:112368.