



RESEARCH ARTICLE

Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network

Olfa Moussa¹  | Hajer Khachnaoui¹ | Ramzi Guetari²  | Nawres Khelifa¹

¹Laboratoire de Biophysique et Technologies Médicales, Université de Tunis El Manar, Tunis, Tunisia

²Institut Supérieur d'Informatique, Université de Tunis El Manar, Tunis, Tunisia

Correspondence

Olfa Moussa, Laboratoire de Biophysique et Technologies Médicales, Université de Tunis El Manar, 1006 Tunis, Tunisia.
Email: olfa.moussa@gmail.com

Abstract

Ultrasonography AKA diagnostic sonography is a noninvasive imaging technique that allows the analysis of an organic structure, thanks to the ultrasonic waves. It is a valuable diagnosis method and is also seen as the evidence-based diagnostic method for thyroid nodules. The diagnosis, however, is visually made by the practitioner. The automatic discrimination of benign and malignant nodules would be very useful to report Thyroid Imaging Reporting. In this paper, we propose a fine-tuning approach based on deep learning using a Convolutional Neural Network model named resNet-50. This approach allows improving the effectiveness of the classification of thyroid nodules in ultrasound images. Experiments have been conducted on 814 ultrasound images and the results show that our proposed approach dramatically improves the accuracy of the classification of thyroid nodules and outperforms The VGG-19 model.

KEYWORDS

computer-aided diagnosis (CAD) system, deep convolutional neural network, deep learning, fine-tuning, thyroid nodule, ultrasound image

1 | INTRODUCTION

The thyroid is an endocrine gland that secretes thyroid hormones. These hormones mainly influence the metabolic rate and protein synthesis. They also have many other effects especially those having an importance regarding the development of the human body.

Various disorders can affect the good functioning of the thyroid. Among the problems that can affect the thyroid there is the appearance of a thyroid nodule. This latter is a small mass formed at the level of the thyroid gland and requires a medical assessment in order to make an accurate diagnosis of the cause of the nodule and determine the possible treatment.

There are several types of thyroid nodules. The colloidal nodule is the most common one and consists of normal cells. The cysts, mostly benign, are formations, up to several centimeters in diameter, filled with fluid. An inflammatory nodule can occur most often to people with thyroiditis. The nodule can also be linked to an adenoma, benign tumor,

which evolves slowly. In about 5% to 10% of the cases, it can be a cancer of the thyroid which concerns the women in 75% of the cases and which is usually perfectly treated.

The thyroid assessment is systematically requested by doctors when they note the presence of nodules. It consists in a set of biological analyzes performed from a blood test. It usually does not prejudge the malignant or benign nature of the nodule, but it allows to know if the operation of the thyroid is normal or not. The examination of the neck region by ultrasound is systematically performed also. This examination allows doctors to measure the size of the thyroid, the number, the size as well as the shape of the nodules and to detect the possible presence of other abnormalities.

A thyroid ultrasound examination provides information about the structure of the nodules and their characteristics. This helps the diagnosis of various thyroid disorders, including echogenicity, texture, composition (cystic, solid, or mixed), existence or absence of microcalcification, halo, and intranodular blood flow.

Based on these clinical features, Horvath et al¹ proposed a system called TI-RADS, for Thyroid Imaging Reporting and Data System that allows the categorization and the classification of thyroid nodules according to their probability of malignancy. The categories of TI-RADS score are designated by TI-RADS 2 (Benign), TI-RADS 3 (Probably benign), TI-RADS 4A (Weakly suspicious for malignancy), TI-RADS 4B (Suspicious for malignancy), TI-RADS 4C (Highly suspicious for malignancy), and TI-RADS 5 (Probably malignant). However, the evaluation of thyroid nodules using TI-RADS is rarely used in daily clinical practice.

In clinical radiology, final diagnoses of radiological examinations are made by human beings. Therefore, they are not exempt from misinterpretation or even diagnostic errors. For instance, the majorities of the thyroid nodules are heterogeneous and provide different internal echogenicity. As a consequence, this reduces the judgment capability of specialists and radiologists and also makes them confused about the variability of the echo patterns of the thyroid nodules. Moreover, the accuracy of the identification of malignant or benignant nodules depends on the radiologists' personal experiences.

The double reading (review) of medical image examinations by radiologists is very helpful and may reduce errors and enhance radiologists' performance. Unfortunately, the increasing number of radiological examinations and the time required for reading make this option impractical. Therefore, an automatic or a semiautomatic classification system is certainly a significant contribution that improves the accuracy of the classification in the case of the thyroid nodules.

The computer-aided diagnosis (CAD) is a computerized technique that may act as a second objective and accurate opinion to assist radiologists and physicians in medical image interpretation and diagnosis by enhancing the accuracy and consistency of radiological diagnosis and by reducing radiological examinations reading time.

Our objective is to design a deep convolutional neural network (DCNN)-based CAD system for thyroid ultrasound images classification using a fine tuning approach of a pretrained deep residual network named ResNet-50. In this work, we pretrain the ResNet-50 on a large collection of natural images which composed of 1.2 million images and 1000 categories. Besides, we update the pretrained weights of the ResNet-50 model in order to adapt it to our data set.

The remainder of the paper is organized as follows: Section 2 provides a profound analysis of the recent studies of the ultrasound image classification approaches. Section 3 introduces the fundamental materials and methods by describing the deep residual learning and the proposed model. Experimental results are discussed in section 4. The conclusion is outlined in section 5.

1.1 | Comparative analysis of classification ultrasound thyroid nodule studies

In this section, we introduce a comparative state of the art of the recent ultrasound CAD system for thyroid nodule classification methods.

Generally, the ultrasound CAD system could be divided into two categories. The first one is the traditional learning ultrasound CAD system, which are based on machine learning algorithms and the second one is the deep-learning (DL) CAD system, which are an advanced form of machine learning techniques.

1.1.1 | Traditional ultrasound CAD system

The conventional ultrasound CAD systems consist of four basic steps including the image preprocessing, the extraction of the region of interest (ROI), the features extraction, and classification. Figure 1 illustrates the flowchart of a general machine learning ultrasound CAD system.

Due to the considerable speckle noise and artifacts of ultrasound images, the medical image resolution and contrast are usually reduced and degraded. The speckle is a multiplicative noise. It is an inherent characteristic of ultrasound imaging that could make some image processing procedures complicated. Thus, the image preprocessing is the basic requirement to provide a better "input" for the next CAD system steps. An important number of research works²⁻⁴ have been developed to achieve this objective. These steps are ordered as follows: the extraction of the ROI and the extraction of relevant features from the ROIs for the classification. The extracted features are used to perform a supervised classification through the usual machine learning classifiers. An effective features extraction step is considered as one of the most important and complicated tasks, which could potentially produce an accurate nodule differentiation and classification.

In the literature, the most conventional machine learning CAD system uses features extraction methods that are based on textural, statistical, shape, or/and gradient based approaches.

In 2011, Ding et al⁵ selected the optimal subset of statistical and textural features, which are extracted from color thyroid elastograms, in order to classify the nodules as malignant or benign using support vector machine (SVM) classifier. The purpose of this study was to find a quantitative metric that can discriminate the benign from the malignant nodules more efficiently and accurately on the thyroid elastograms. The lesion regions were delineated by radiologists, and statistical and textural features were extracted by using a minimum redundancy-maximum relevance algorithm. Thereby, the selected features were fed to a support vector machine classifier. The classification accuracy was 93.6% using 125 thyroid elastograms. Despite the encouraging results achieved by this work, there were some gaps. For

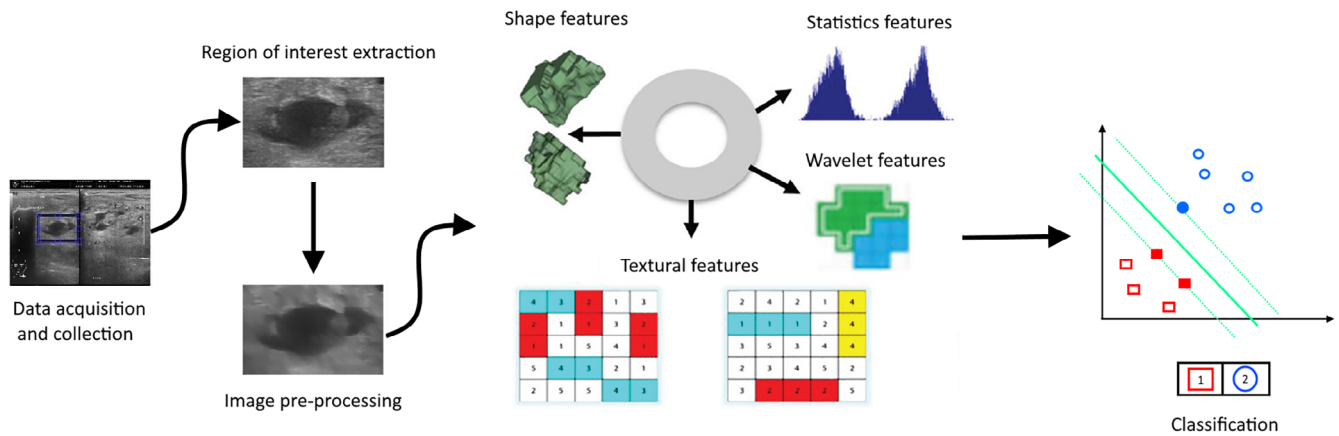


FIGURE 1 Block diagram of a general computer-aided diagnosis system for ultrasound imaging [Color figure can be viewed at wileyonlinelibrary.com]

example, the lesion regions were selected manually and only solid nodules were considered in this work.

In 2012, Acharya et al⁶ presented a CAD technique for the thyroid nodules classification. They extracted the textural features from the 3D contrast-enhanced ultrasound (CEUS) and the 3D high-resolution ultrasound (HRUS) images. These features are the fractal dimension, the local binary pattern (LBP), the Fourier spectrum descriptor, and the Laws texture energy. The results show that the combination of the textural features coupled with the SVM and the Fuzzy classifiers presented a classification accuracy of 100% for the HRUS data set. The Gaussian mixture model classifier yielded an accuracy of 98.1% for the CEUS data set. However, these results are only evaluated on a small database. Thus, these techniques need to be evaluated on larger databases to further used as diagnostic tools in daily clinical practice.

In 2013, Gopinath and Shanthi⁷ developed an automated CAD system for thyroid cancer pattern in fine-needle aspiration cytology (FNAC) microscopic images using statistical textural features and SVM classifier. The ROI of thyroid cell is extracted by mathematical morphology segmentation method. Statistical textural features of the segmented images are subsequently extracted using Gabor filter-based method. A total of 110 FNAC images were used to perform the diagnosis of thyroid cancer. The SVM achieved a diagnostic accuracy of 96.7% with sensitivity and specificity of 95% and 100%, respectively.

Recently, an automated system for the classification of the benign and the malignant thyroid nodules using HRUS was presented in Reference 9. HRUS images of the thyroid nodules are preprocessed using Gabor transform to extract its texture information in terms of its frequency and orientation. Various entropy features are also extracted from these transformed images including Fuzzy, Shannon, Renyi, HOS, Kapur, and Vajda. These features are further subjected to locality sensitive discriminant analysis (LSDA), in order to

reduce insignificant features, and then ranked by Relief-F method. Cross-validation technique with Wilcoxon signed rank test, Friedmans test, and Iman-Davenport test are used for the comparison of classifier performance and classification of the data set using SVM, MLP, kNN, and C4.5 classifier. The C4.5 decision tree classifier yielded an accuracy of 94.3% using 242 thyroid HRUS images.

It was observed from the above-mentioned ultrasound CAD systems that the methods based on the combination of texture-based features with a SVM classifier are commonly used for the identification and classification of nodules according to their malignancy risk.

Despite the encouraging results obtained by the most of the traditional machine learning-based ultrasound CAD systems, the selection of the most significant and useful features is still a complicated CAD task that requires several passes of trial and error design to check which features and classifiers are most suitable for each specific problem.

1.1.2 | DL CAD systems

The DL CAD systems have recently been introduced to the medical image fields by providing a reliable solution to the shortcomings of the traditional CAD (T-CAD) systems. With the DL models, there is no need to confuse about which the most informative features that should be extracted for a particular task, because this is decided by the deep neural network in the training process. CAD techniques based on deep neural networks can automatically learn features for medical images classification and diagnosis. Moreover, DL models have shown promising results on various medical image diagnosis problems including the breast lesion diagnosis, the lung nodule diagnosis, and the thyroid nodule diagnosis.

In breast ultrasonography, Zhang et al⁹ developed a CAD system using a DL architecture for automatic classification

of breast tumors from the shear-wave elastography (SWE) images. It is the first study that attempt to evaluate learned-from-data image features in SWE image analysis. The proposed method was validated using 5-fold cross validation on a set of 227 SWE images composed 135 of benign tumors and 92 of malignant tumors. It achieved a diagnostic accuracy of 93.4%, with a sensitivity of 88.6% and specificity of 97.1%. Recently in 2017,¹⁰ the authors developed a hybrid breast CAD methodology for classification of benign and malignant breast lesions by combining the high and low-level features extracted from a pretrained VGG19 network with handcrafted features computed using conventional CAD methods. The proposed methodology was tested on three different clinical imaging modalities including mammography, ultrasound, and MRI. It achieved an area under curve (AUC) = 0.90 for ultrasound modality.

For lung nodule diagnosis problem, Raunak et al¹¹ proposed in 2018 the use of four 3D DCNN to classify the 3D lung nodules in CT images into benign or malignant categories. These four networks consist of a basic 3D CNN, a novel multi-output network, a 3D DenseNet, and an augmented 3D DenseNet with multi-outputs. The networks are evaluated on the public LIDC-IDRI data set and a private lung nodule data set. For the LIDC-IDRI data set, the best performance is achieved by the 3D DenseNet, resulting in 90.40% accuracy in a cross-validation setting.

For the thyroid nodule diagnosis problem, Ma et al^{12,13} proposed a cascade deep convolutional neural network (CNNs)-based model for ultrasound thyroid nodule diagnosis. The proposed model consists of the fusion of two different CNNs and a new splitting method. Particularly, a DCNN is used to learn the segmentation probability maps from the ground true data. Hence, the splitting method is employed to split all the segmentation probability maps into different connected regions. Finally, another DCNN is used to automatically detect the thyroid nodules from the ultrasound thyroid images. The proposed model achieves an AUC = 98.51%.

In addition, Chen et al¹⁴ proposed in 2017 a DL-based ultrasound text classifier for the prediction of the benign and the malignant thyroid nodules. The proposed ultrasound text classifier is trained by the labeled ultrasonic text with benign or malignant label of pathology. Experimental results show that this method has a high accuracy rate of 93% and 95% both on a real medical data set and an UCI (standard data set).

Generally, training a DCNN from scratch (or full training) is often complicated and required extensive computational and memory resources. A promising alternative to the full training strategy is to use the transfer learning. Transfer learning is the reuse of a pretrained model on a different data set. It has shown promising result in thyroid nodule diagnosis problem. The fine-tuning strategy is an advanced technique of transfer learning. It is to not only retrain the pretrained model

on the new data set, but to also fine-tune the weights of the pretrained network by continuing the backpropagation.

An example of this is the study of Liu et al¹⁵ in 2017, which presented a feature extraction method for ultrasound images to classify the thyroid nodules into benign and malignant. They transfer the CNNs model learned from ImageNet as a pretrained feature extractor to the new ultrasound image data set. This method combines the traditional low-level features that extracted from histogram of oriented gradient (HOG) and LBPs, and high-level deep features extracted from CNN model together, to form a hybrid feature space. Experimental results on 1037 images show that the accuracy of the proposed method is 93.10%.

Another example is the study of Reference 16 in 2017, which presented a DL CAD system for classifying thyroid nodules in ultrasound images using features extracted from a fine-tuned pretrained GoogLeNet. The extracted features of the thyroid ultrasound images are fed to a cost-sensitive random forest classifier to classify the images into malignant and benign classes. The experimental results show that the proposed fine-tuned GoogLeNet model achieves an accuracy of 98.29%, with a sensitivity of 99.10% and specificity of 93.90% for the images in an open access database which contains 2754 samples and 549 samples, respectively.

In 2016,¹⁷ the authors introduced an ensemble of fine-tuned of different CNN architectures for classifying medical images. The ensemble fused the fine-tuned CNN models to derive a more powerful image classification scheme. The proposed method was evaluated on the image CLEF 2016 medical image public data set that contains 30 modalities; 6776 training images and 4166 test images.

From these examples, we see that deeply fine-tuned CNN performed better than other CNN baselines.

Answers to the following central question “Can the use of pre-trained deep CNNs, with sufficient fine-tuning, eliminate the need for training a deep CNN from scratch?” is given in this work.¹⁸ The authors have demonstrated using extensive experiments that deeply fine-tuned CNNs are useful for medical image analysis, performing as well as fully trained CNNs and even outperforming the latter when limited training data are available. Based on four distinct medical imaging applications from three different imaging modality systems, the authors have also proven that the required level of fine-tuning differed from one application to another.

We believe that DL-based CAD will play an important role in radiology and in thyroid ultrasound diagnosis.

Both deeply fine-tuned CNNs and fully trained CNNs proved that they could be a feasible solution to overcome the drawbacks of T-CAD systems. Table 1 shows the comparison of the T-CAD methods with DL-based CAD methods.

TABLE 1 Summary of performance of different thyroid ultrasound computer-aided diagnosis (CAD) systems

	Techniques	Features	Classifiers	Data set	Accuracy
T-CAD	Acharya et al ⁸	Entropy and textural features	C4.5	242 thyroid HRUS images	94.3%
	Gopinath et al ⁷	Statistical textural features	SVM	110 FNAC images	96.7%
	Acharya et al ⁶	Textural features	SVM and fuzzy	small CEUS, HRUS data set	98.1%
	Ding et al ⁵	Statistical and textural features	SVM	125 thyroid elastograms	93.6%
DL	Chen et al ¹⁴	Deep neural network features	-	Real thyroid US	93.0%
	Liu et al ¹⁵	Deep-learning and low level features	Positive-sample	1037 thyroid US	93.1%
	Maa et al ¹²	Deep-learning features	Softmax classifier	15.000 thyroid US	83.02%
	Mei et al ¹⁹	Deep features, HOG and LBP	SVM	427 thyroid US	53.0%
	Chi et al ¹⁶	Deep-learning features	Cost-sensitive random forest	3303 thyroid US	98.29%

Abbreviations: CEUS, contrast-enhanced ultrasound; FNAC, fine -needle aspiration cytology; HOG, histogram of oriented gradient; HRUS, high-resolution ultrasound; LBP, local binary pattern; SVM, support vector machine.

2 | MATERIALS AND METHODS

In this section, we explain the fundamental concepts of the deep residual learning and the architecture of the ResNet-50. Then, we describe the basic idea of the proposed approach.

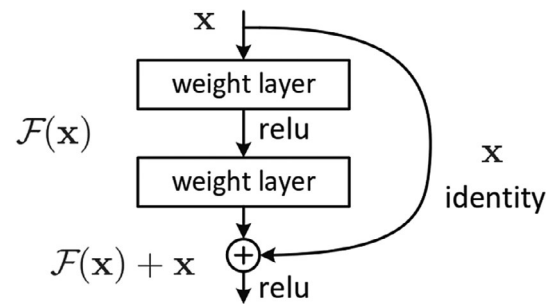
2.1 | Deep residual learning

Deep residual learning network is a very intriguing network that was developed by researchers from Microsoft Research. These networks led to 1st-place winning entries in all five main tracks of the ImageNet and COCO 2015 competitions, which covered image classification, object detection, and semantic segmentation. The robustness of ResNets has since been proven by various visual recognition tasks and by non-visual tasks involving speech and language.

This CNN was introduced by Reference 20 and allows the training of very deep networks (more than 150 layers). The difficulty in training such deep networks is related to the backpropagation of the gradient. The deeper the network is, the lower the gradient updating the low levels (first layers) is. Therefore, the deep architecture does not really allow updating these layers. The idea developed in ResNet is the use of residual connections allowing a better optimization of very deep networks. A residual connection makes it possible to pass the input through two convolution filters but also to pass this input directly to the following layers. This is done by summing the result of the two convolutional layers and the input as shown in Figure 2. With this architecture, the authors demonstrate the interest of training very deep networks through their performance and offer a way to learn them effectively.

2.2 | ResNet-50 model

ResNet is a deep residual network that has recently been introduced by He et al.²⁰ It is a very deep CNN that has a

**FIGURE 2** A residual block—the fundamental building block of residual networks²⁰

high accuracy and won both ILSVRC²¹ and COCO^{22,23} challenges. In this work, we use ResNet-50. ResNet-50, as its name would suggest, includes 50 layers. Figures 3 and 4 illustrate the ResNet architectures.

ResNet-50 holds 16 blocks, each of them includes a convolutional layer, batch normalization and ReLU. The output of each block is merged with its own input. He et al²⁰ use bottleneck architecture for each residual block. It means that the residual block consists of three layers in this order: 1×1 convolution – 3×3 convolution – 1×1 convolution.

At the first layer, ResNet uses 7×7 convolution with stride 2 to downsample the input by the order of 2 as well as the pooling layer. Then, it is followed by three identity blocks before downsampling again by 2. The last layer is the average pooling which creates 1000 feature maps (for ImageNet data). The result would be 1000 dimensional vector which then fed into softmax layer directly, so it is fully convolutional.

2.3 | Proposed method

In the present work, we present a thyroid nodule classification system using a fine-tuning strategy of the deep residual network ResNet-50.

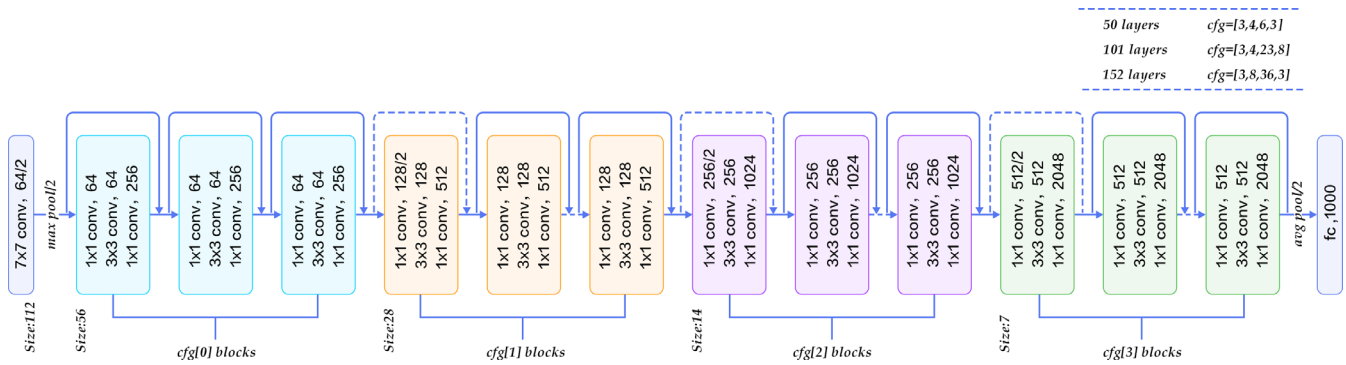


FIGURE 3 ResNet architecture²⁰ [Color figure can be viewed at wileyonlinelibrary.com]

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10 ⁹	3.6×10 ⁹	3.8×10 ⁹	7.6×10 ⁹	11.3×10 ⁹

FIGURE 4 Detailed ResNet architecture for ImageNet data set²⁰

The fine-tuning is the process of adapting the pretrained weights of a CNN to different data sets through the use of the backpropagation. Fine-tuning a pretrained network is typically much faster and easier than training a CNN or a neural network from scratch. Using pretrained deep networks allows to quickly learn new tasks without defining and training a new network, having millions of images, or long training times. The advantage of this technique is that the pretrained network has already learned a rich set of features that can be applied to a wide range of other similar tasks.

As illustrated in Figure 5, we start by pre-training the ResNet-50 on the ImageNet data set that contains 1000 classes of natural images. Thus, the ResNet-50 is initialized with a powerful set of parameters that can learn a variety of universal image features from general domains with a low error rate. Figure 6 illustrates an example of the features extraction from different layers. Then, we transfer the learned features and parameters from the general domain to our specific classification problem.

The “softmax” classifier of the original DCNN computes the probability of the last fully connected layer that contains

1000 classes of the ImageNet data set. We fine-tune the ResNet-50 by modifying this softmax classifier and initializing a new one with random values.

The new softmax classifier is trained from scratch through the use of backpropagation algorithm with the new thyroid ultrasound data, which have only two classes instead of 1000 in order to learn only two outputs, the “benign” and “probably malignant” output classes. Therefore, the weights of the last fully connected layers are fine-tuned through the use of backpropagation technique which minimizes the error function in weight using the gradient descent optimization algorithm. This algorithm has been repeatedly rediscovered at each iteration step.

To summarize, the proposed method consists of the following basic steps:

1. Pretrain the DCNN ResNet-50 architecture on the ImageNet data set.
2. Assign a label to each ultrasound thyroid image of the data set.
3. Define the ROIs of thyroid nodule in each image as delineated by radiologist.

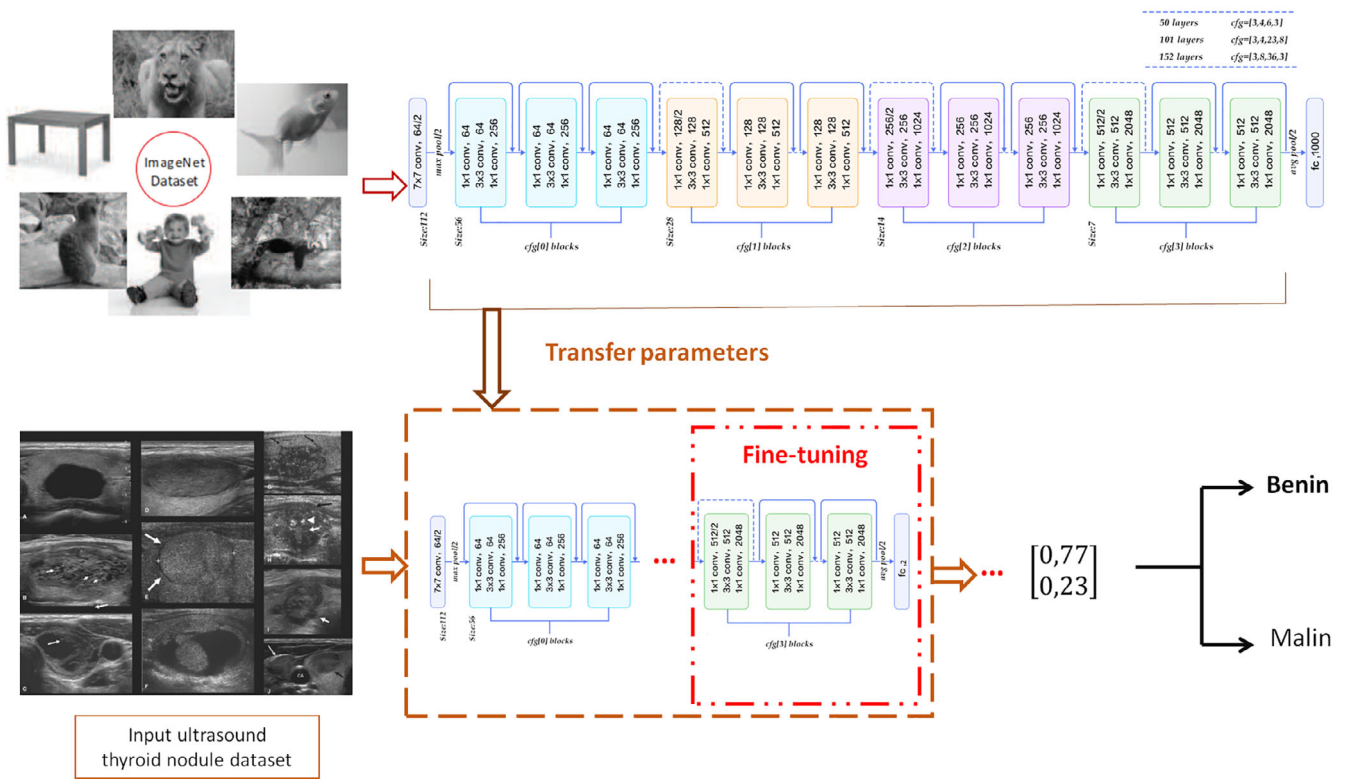


FIGURE 5 Flowchart of the proposed deep-learning-based model for prediction of thyroid nodules [Color figure can be viewed at wileyonlinelibrary.com]

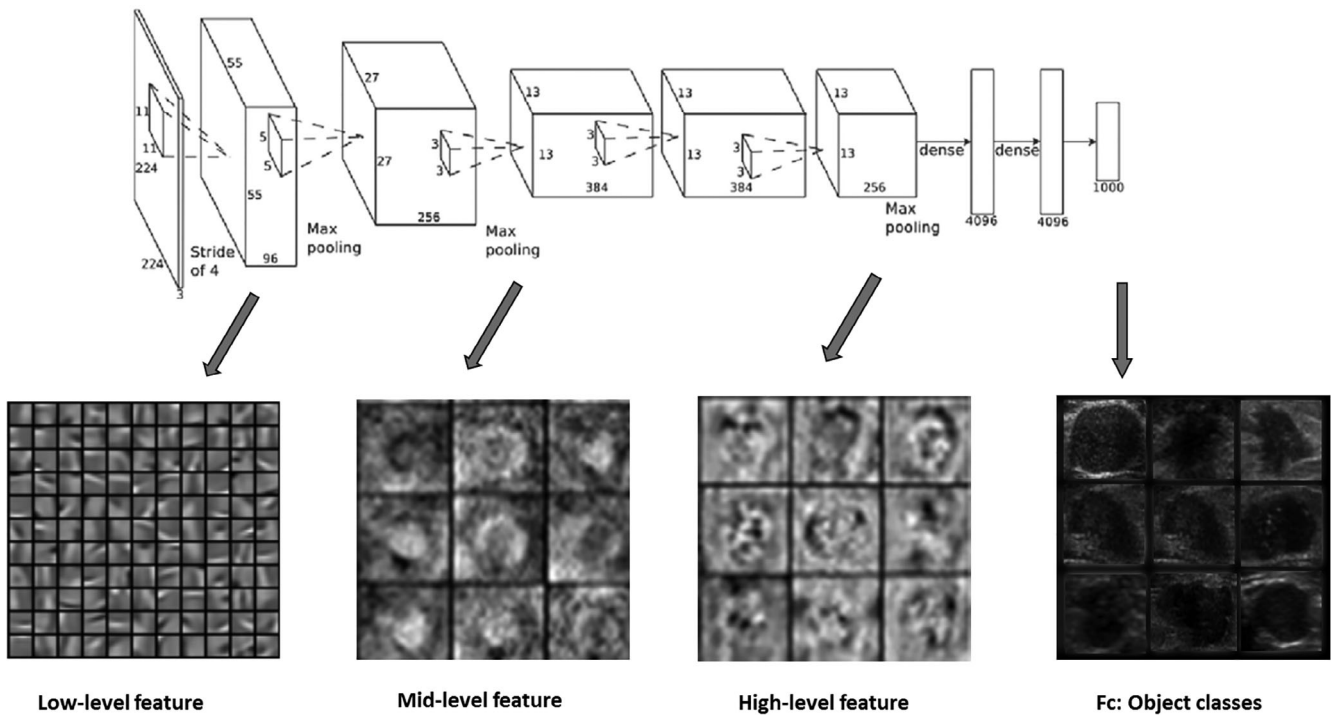


FIGURE 6 Features extraction

4. Rescale each thyroid nodule to 224×224 dimensions.
5. Fine-tune the pretrained ResNet-50 network by modifying the weights of the last fully connected layer that contains 1000 classes with layer by a two fully-connected softmax layer using the backpropagation technique.
6. Set the appropriate values of the parameters of backpropagation algorithm to achieve the optimum learning performance of the proposed network.

3 | FINE-TUNING PARAMETER SELECTION

A pretrained model has been previously trained on a data set and contains the weights and biases that represent the features of data set it was trained on. Hence, the pretrained model will already have learned the universal features including curves and edges, which are also pertinent to our classification problem. In this work, we adapted the ResNet-50 to our data set by transferring the acquired knowledge from the pretrained model to another model and by modifying the last fully connected layer of the pretrained DCNN (ResNet-50).

In order to achieve the optimum learning performance, it is important to set the appropriate values of the parameters of backpropagation algorithm. Learning rate, momentum rate, and activation functions are the most influential parameters of backpropagation algorithm which affect the learning performance of the DCNN. However, there are no clear guidelines available about the selection of the optimal values of these parameters. The only way to check which parameters perform better is through several passes of trial and error.

Higher values of learning rates frequently lead to overfitting, while lower values of learning rates lead to limited change in error across epochs. In our experiments, the learning rate $= 10^{-6}$ was empirically selected by controlling the validation error during the fine-tuning process in order to adapt smoothly to the new task, without aggressively crushing the knowledge already acquired. The momentum rate affects the stabilization of the error rates in the weight space during training and validation. Momentum rate of 0.9 has been found to be suitable for our classification problem. The softmax activation function is found to be the most appropriate when we are trying to handle just two classes in the output layer of the classifier. We run the backpropagation algorithm for 20 epochs, which optimizes the network parameters using stochastic gradient descent.

4 | RESULTS

4.1 | Image collection

The proposed network was evaluated on two public database, DDTI* (Digital Database Thyroid Image) and Ultrasoundcases.info.[†] DDTI is a thyroid ultrasound image database performed by two experts in 299 patients with thyroid disorders, containing 451 thyroid images with the size 560×360 , including 376 malignant images and 75 benign images. The images were extracted from thyroid ultrasound video sequences captured with TOSHIBA Nemio 30 and TOSHIBA Nemio MX Ultrasound devices.

The DDTI contains a complete annotation and diagnostic description of suspicious thyroid lesions, using the BETHESDA (System for Reporting Thyroid Cytopathology) report and the TI-RADS description performed by at least two expert radiologists. The TI-RADS system has been validated to correlate the image features with the pathological diagnosis. The experts independently evaluated the patient individually and described the specific features filling the TI-RADS requirements. This classification pretends avoiding the unnecessary fine-needle aspiration (biopsy) and performs an accuracy diagnostic related to the thyroid disorder. The diagnostic description of malignant lesions was confirmed by a biopsy procedure. This database is designed to allow segmented nodules and annotated labels to be instantly shared via the web and to grow over the time. Figure 7 shows a snapshot of the online annotation tool. The latter provides a simple drawing interface that allows users to outline the silhouettes of the nodule present in each image. First, the ROIs of thyroid nodule in each image are manually defined by radiologist. Then, all the images were resized to have exactly the same dimensions. Figure 8 illustrates manual nodule segmentation by the radiologist.

The second database is a publicly available ultrasound images database that contains a large number of general ultrasound cases from the Gelderse Vallei Hospital in Ede, the Netherlands. The cases have been collected over the years by the radiologists and ultrasound technicians of the hospital in an image database. The latter contains 363 thyroid ultrasound images including 231 malignant images and 132 benign images. The diagnostic description of malignant and benign lesions was confirmed by a biopsy procedure.

The two databases have been merged to guarantee that there were enough training samples for each class in the database. Figure 1 shows some examples of ultrasound thyroid nodules images with their categories of TI-RADS score.

4.2 | Results and discussion

Since databases of the most state-of-the-art models are not available and not similar (we have not the same number of

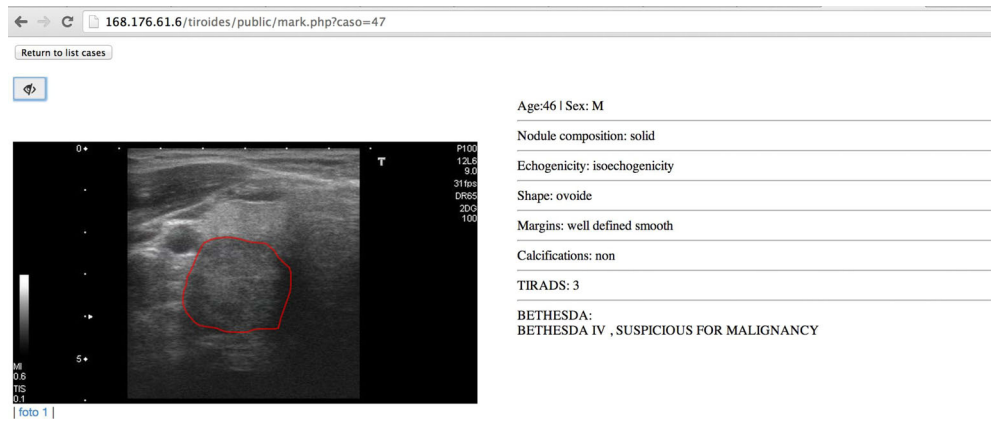


FIGURE 7 Screenshot of the online annotation tool [Color figure can be viewed at wileyonlinelibrary.com]

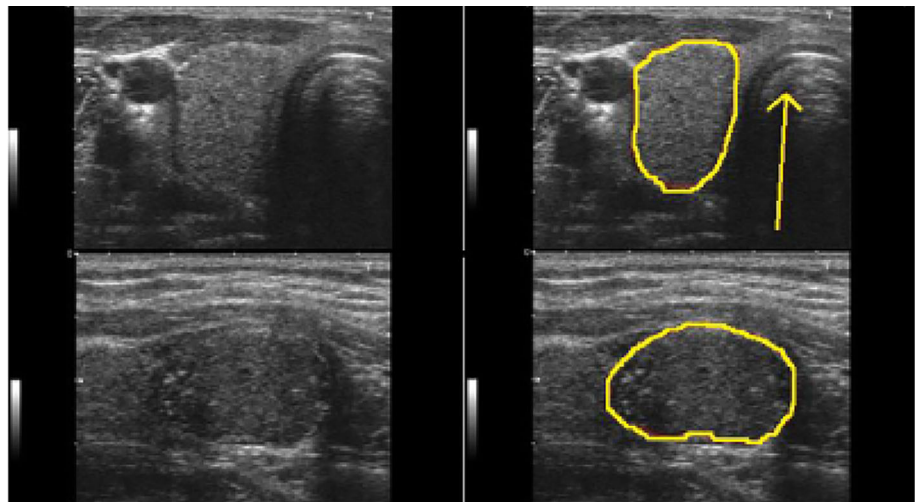


FIGURE 8 Example of thyroid nodule area delineated by radiologist [Color figure can be viewed at wileyonlinelibrary.com]

thyroid ultrasound images, from the same modality), it is unfortunately not feasible to directly compare the proposed DCNN with these models on the same database.

Taking this reason into account, we have not found in the same conditions and a comparison of the proposed method with those of the state of the art is not necessarily valid. Consequently, we decided to compare the proposed architecture to the VGG-19 architecture using the same fine-tuning technique.

4.2.1 | Performance measurement

The two models of classification were evaluated by classification accuracy, sensitivity, specificity, and area under receiver operating characteristic (ROC) curve (AUC). Sensitivity indicates the percentage of benign nodules correctly identified as benign. Specificity measures the percentage of malignant nodules correctly classified as malignant. Accuracy evaluates the percentage of correctly diagnosing thyroid nodules. ROC curve is a tool to measure the balance

between finding true positives and avoiding false positives. The AUC values that belong to the interval $[0.5, 1]$ indicate the probability of correctly prediction the malignant and benign thyroid nodules. A good classifier possesses an AUC close to 1. This value represents a larger AUC.

4.2.2 | Discussion

Table 2 shows the test accuracy, sensitivity, specificity, and AUC of each model of classification. These results clearly indicate that the proposed ResNet-50 method is the one with the highest score, in terms of test accuracy, which reached

TABLE 2 Results of the classification of thyroid nodules by the two models VGG-19 and ResNet-50

	Accuracy	Sensitivity	Specificity	AUC
VGG-19	81.83%	98.51%	20.89%	0.74
Proposed model	97.33%	80.69%	64.17%	0.78

Abbreviation: AUC, area under the curve.

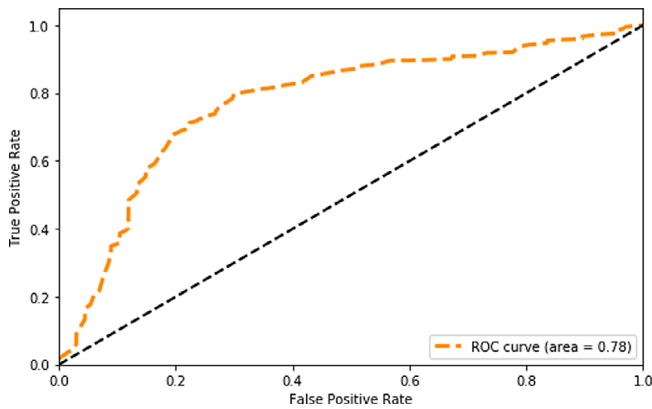


FIGURE 9 The receiver operating characteristic (ROC) curves and the area under the curve (AUC) of the proposed deep model [Color figure can be viewed at wileyonlinelibrary.com]

97.33%, which means that almost all nodules can be identified correctly, in terms of specificity which reached 64.17% and in terms of AUC which reached 0.78. However, for the sensitivity metric, the best score is obtained by the VGG-19 with a value of 98.51%.

From the two curves in Figures 9 and 10, it is clear that the area under the curve of the proposed method represents the largest area, which is the largest value of AUC. The AUC value provided in Table 2 confirms the results of Figures 9 and 10. The curve of the proposed method evolves asymptotically along the y-axis, which proves the best compromise between sensitivity and specificity.

The same hardware configuration was used for all the conducted tests, namely a computer with Intel Core i5-3210M CPU of 2.80 GHz, 8 GB memory, and Windows 7 as operating system. The experimental evaluations were conducted on 814 thyroid nodule ultrasound images, including 207 benign images and 607 malignant images and the type and location of all nodules are specified by doctors as shown in Figure 8.

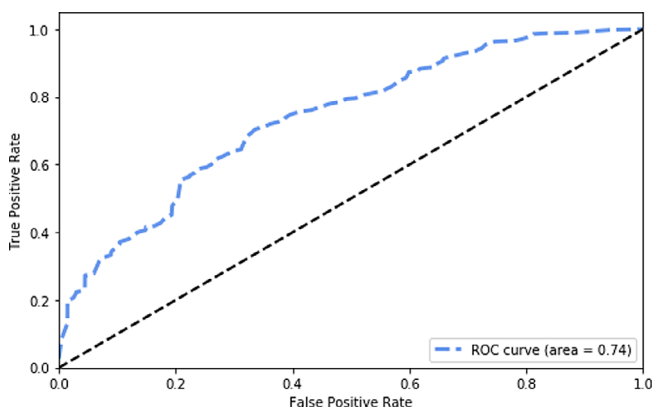


FIGURE 10 The receiver operating characteristic (ROC) curves and the area under the curve (AUC) of the VGG-19 model [Color figure can be viewed at wileyonlinelibrary.com]

The above experiments prove the benefits of using the fine-tuning of the pretrained DCNN model in the improvement of the training process by learning the most relevant image features that were adaptable for thyroid ultrasound images classification.

The main contribution of the present work is the exploration of the fine-tuning ResNet 50 in the prediction of benign and malignant thyroid nodules in ultrasound images. According to the limitation of the number of samples, the present study was not able to classify the thyroid nodules into all classes of the different TI-RADS scores and it was restricted in only two classes (benign and probably malignant).

One important constraint that exists in DL algorithms is the need to resize images to a unified dimension. However, the resizing technique might cause losing some beneficial information, such as nodule size and shape. It is also important to note that, although the TI-RADS has demonstrated good performance²⁴ in differentiating malignant thyroid nodules, it does not consider nodular size in classification systems.

In the other hand, various studies^{25,26} have reported that nodule size and nodule shape were not significant risk factors and did not affect the risk of malignancy.

5 | CONCLUSION

In this paper, we propose a thyroid nodule classification system using a fine-tuning strategy of the deep residual network ResNet-50. The fine-tuning strategy proved that the knowledge and parameters acquired from the pretrained model can be successfully transferred to thyroid nodule classification task. Using pretrained deep networks allow to quickly learning new tasks without defining and training a new network, having millions of images, or long training times. The advantage of this technique is that the pretrained network has already learned a rich set of features that can be applied to a wide range of other tasks. Moreover, the experimental results on 814 ultrasound images show that our proposed fine-tuned ResNet-50 model achieves higher thyroid nodule classification performance. In our future work, we aim to continue further tuning on the proposed DCNN in order to improve accuracy specifically for the benign nodules. Furthermore, we would like to consider all the six classes of different TI-RADS scores in our future CAD system.

ENDNOTES

*<http://cimalab.intec.co/?lang=en&mod=project&id=31>

†The website is a cooperation between Hitachi Medical Systems Europe and the Radiology Department of the Gelderse Vallei Hospital in Ede in the Netherlands. <http://www.ultrasoundcases.info/Default.aspx>

ORCID

Olfa Moussa  <https://orcid.org/0000-0002-0521-9381>

Ramzi Guetari  <https://orcid.org/0000-0002-0608-3498>

REFERENCES

- Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metabol.* 2009;94(5):1748-1751.
- Olfa M, Nawres K. Ultrasound image denoising using a combination of bilateral filtering and stationary wavelet transform. Paper presented in: International Image Processing, Applications and Systems Conference (IEEE); Sfax, Tunisia; 2014: 1-5.
- Olfa M, Nawres K. A new framework based on the trilateral filter for despeckling ultrasound images. 2016 International Conference on Control, Decision and Information Technologies (CoDIT) (IEEE); St. Julian's, Malta; 2016: 712-717.
- Olfa M, Nawres K, Nouredine BA. Video despeckling using Shearlet tensor-based anisotropic diffusion. *Computer Aided Geometric Design*, 2018;67:34-46.
- Ding J, Cheng H, Ning C, Huang J, Zhang Y. Quantitative measurement for thyroid cancer characterization based on elastography. *J Ultrasound Med.* 2011;30(9):1259-1266.
- Rajendra Acharya U, Vinitha Sree S, Muthu Rama Krishnan M, Molinari F, Garberoglio R, Suri JS. Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems. *Ultrasonics.* 2012;52(4):508-520.
- Gopinath B, Shanthi N. Support vector machine based diagnostic system for thyroid cancer using statistical texture features. *Asian Pac J Cancer Prev.* 2013;14(1):97-102.
- Acharya RU, Chowriappa P, Fujita H, et al. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. *Knowl-Based Syst.* 2016; 2213:1-11.
- Zhang Q, Xiao Y, Dai W, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics.* 2016; 72:150-157.
- Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys.* 2017;44(10):5162-5171.
- Dey R, Lu Z, Hong Y. Diagnostic classification of lung nodules using 3D neural networks. Paper presented at: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); Washington, DC. 2018.
- Ma J, Wu F, Jiang T'a, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Med Phys.* 2017;44(5):1678-1691.
- Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int J Comput Assist Radiol Surg.* 2017;12(11):1895-1910.
- Chen D, Niu J, Pan Q, Li Y, Wang M. A deep-learning based ultrasound text classifier for predicting benign and malignant thyroid nodules. Paper presented at: 2017 International Conference on Green Informatics (ICGI); Fuzhou, China; 2017, 199-204.
- Liu T, Xie S, Yu J, Niu L, Sun W. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. Paper presented in: 2017 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) (IEEE); New Orleans, LA; 2017: 919-923.
- Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging.* 2017;30(4):477-486.
- Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform.* 2017;21(1):31-40.
- Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging.* 2016;35(5):1299-1312.
- Mei X, Dong X, Deyer T, Zeng J, Trafalis T, Fang Y. Thyroid nodule benignity prediction by deep feature extraction. Paper presented in: 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE) (IEEE); Washington, DC; 2017: 241-245.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV; 2015.
- Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comp Vis.* 2015;115(3):211-252.
- Lin TY, Maire M, Belongie S, et al. (2014) Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. Computer Vision – ECCV 2014. *Lecture Notes in Computer Science*, vol 8693. Springer, Cham.
- Luo S, Kim E-H, Dighe M, Kim Y. Thyroid nodule classification using ultrasound elastography via linear discriminant analysis. *Ultrasonics.* 2011;51(4):425-431.
- Macedo BM d, Izquierdo RF, Golbert L, ELS M. Reliability of thyroid imaging reporting and data system (TI-RADS), and ultrasonographic classification of the American Thyroid Association (ATA) in differentiating benign from malignant thyroid nodules. *Arch Endocrinol Metab.* 2018;62(2):131-138. <https://dx.doi.org/10.20945/2359-3997000000018>.
- Uyar O, Cetin B, Aksel B, et al. Malignancy in solitary thyroid nodules: evaluation of risk factors. *Oncol Res Treat.* 2017;40:360-363. <https://doi.org/10.1159/000464409>.
- Jabiev AA, Ikeda MH, Reis IM, Solorzano CC, Lew JI. Surgeon-performed ultrasound can predict differentiated thyroid cancer in patients with solitary thyroid nodules. *Ann Surg Oncol.* 2009;16: 3140-3145.

How to cite this article: Moussa O, Khachnaoui H, Guetari R, Khelifa N. Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network. *Int J Imaging Syst Technol.* 2019;1–11. <https://doi.org/10.1002/ima.22363>