IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE IN PARALLEL AND DISTRIBUTED COMPUTING

Received February 15, 2020, accepted February 28, 2020, date of publication March 12, 2020, date of current version March 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980290

Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images

YONGFENG WANG[1], WENWEN YUE[2], XIAOLONG LI[2], SHUYU LIU[3], LEHANG GUO[2], HUIXIONG XU[2], HEYE ZHANG[4], (Member, IEEE), AND GUANG YANG[5], (Member, IEEE)

[1]School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510006, China
[2]Department of Medical Ultrasound, Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Tongji University Cancer Center, Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, School of Medicine, Tongji University, Shanghai 200072, China
[3]School of Pharmacy, Sun Yat-sen University, Guangzhou 510006, China
[4]School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510006, China
[5]National Heart and Lung Institute, Imperial College London, London SW7 2AZ, U.K.

Corresponding authors: Shuyu Liu (lsy@mail.sysu.edu.cn) and Lehang Guo (gopp1314@hotmail.com)

This work was supported in part by the Project of Shenzhen International Cooperation Foundation under Grant GJHZ20180926165402083, in part by the Shanghai Municipal Health Commission under Grant SHLCZDZK, in part by the Shanghai Hospital Development Center under Grant 16CR3061B, in part by the Science and Technology Commission of Shanghai Municipality under Grant 19441903200, and in part by the Major Project of Guangzhou Science and Technology of Collaborative Innovation and Industry: intelligent video detection and recognition technology in cloud computing environment, under Grant 201605122151511.

ABSTRACT Thyroid nodules have a high prevalence and a small percentage is malignant. Many non-invasive methods have been developed with the help of the Internet of Things to improve the detection rate of malignant nodules. These methods can be roughly categorized into two classes: radiomics based and deep learning based approaches. In general, convolutional neural networks based deep learning methods have achieved promising performance in many medical image analysis and classification applications; however, no existing comparison has been done between radiomics based and deep learning based approaches. Therefore, in this paper, we aim to compare the performance of radiomics and deep learning based methods for the classification of thyroid nodules from ultrasound images. On one hand, we developed a radiomics based method, which consists of extracting high throughput 302-dimensional statistical features from pre-processed images. Then dimension reduction was performed using mutual information and linear discriminant analysis respectively to achieve the final classification. On the other hand, a deep learning based method was also developed and tested by pre-training a VGG16 model with fine-tuning. Ultrasound images including 3120 images (1841 benign nodules and 1393 malignant nodules) from 1040 cases were retrospectively collected. The dataset was divided into 80% training and 20% testing data. The highest accuracies yielded on the testing data for radiomics and deep learning based methods were 66.81% and 74.69%, respectively. A comparison result demonstrated that the deep learning based method can achieve a better performance than using radiomics.

INDEX TERMS Ultrasound images, thyroid nodule, thyroid cancer, nodule classification, convolutional neural network, radiomics.

I. INTRODUCTION

Thyroid nodules are a common thyroid disease. According to previous studies, the incidence of thyroid nodules is increasing [1], [2]. Thyroid nodules can be roughly divided into benign and malignant, and about 10% of patients who present thyroid nodules were diagnosed as malignant [3]. Ultrasonography has been recommended by the American Thyroid Association (ATA) [4] and is a preferred [5] method for early detection and diagnosis of thyroid nodules due to its economy, effectivity and no radiation. However, the resemblance of the manifested pattern which exists in ultrasound images of both benign and malignant thyroid nodules may cause difficulties for radiologists in the process of interpreting

The associate editor coordinating the review of this manuscript and approving it for publication was Songwen Pei.

and evaluating thyroid disease. Therefore, the need for an objective and effective method that can reduce the misdiagnosis rate is becoming more and more critical in the matter of analysis and estimation of ultrasound images for thyroid nodules.

Various machine learning based methods to automatically detect and assess thyroid nodules in ultrasound images were proposed [6]–[8]. These methods can be broadly classified into two categories: radiomics and deep learning based methods. The radiomics based method [9] was proposed to extract the quantitatively high throughput features of images like Gray-Level Co-occurrence Matrix (GLCM) [10], Histograms of Oriented Gradients [11] and Gray level run-length matrix [12] from medical images. Then after feature selection and dimension reduction, the extracted features were finally fed into a classifier to achieve the classification. Wang *et al.* [13] used a radiomics based method to build a nomogram from multiparametric MRI to predict glioma grade in brain tumor patients. Nugroho *et al.* [14] applied radiomics to classify cystic or solid thyroid in ultrasound images with an accuracy of 89.74%, a sensitivity of 88.89%, a specificity of 91.67%, the positive predictive value of 96.00% and negative predictive value of 78.57%.

Recently, due to the development of deep learning, especially convolutional neural network (CNN) [15], the performance of the Computer-aided Diagnosis system in diagnosing diseases from medical images is comparable or better than the performance of radiologists [16], [17]. For the training of deep learning based models, the amount of datasets plays an extremely important role. If a large number of datasets with different instances (million level) are fed into the model, its performance can be impressive. In addition, with the rapid development of the Internet of Thing, large-scale streaming data are collected and distributed [18]. Many deep learning models can be efficiently distributed and trained using large-scale datasets [19]–[21] in variety of fields like nature language processing, face recognition and machine translation. The VGG16 architecture of convolutional neural networks trained by the well-known ImageNet is one of them. In the domain of clinical medicine, deep learning models also have obtained good performance. For example, Rajpurkar *et al.* [16] demonstrated that the performance of detecting heart arrhythmias has exceeded board-certified cardiologists using the deep learning based method with sufficient datasets. Liu *et al.* [22] integrated domain knowledge in training deep learning models for breast cancer ultrasonography in order to classify benign tumors from malignant lumps and achieved an accuracy of 83.3% with a sensitivity of 96.7%. Yang and Zhao [8] developed a semi-supervised deep learning model, which combined generative adversarial networks and semi-supervised support vector machine to classify thyroid nodules and obtained satisfied results.

In general, the results of both radiomics and deep learning based methods are promising. However, there are still limitations of these methods. The radiomics based methods need to extract statistical features, which may be affected by factors such as image resolution, low signal to noise ratio (SNR) and other artefacts in ultrasound images. In contrast, deep learning based methods may be unstable due to insufficient datasets [8]. Yang and Zhao [8] developed a semi-supervised model to overcome insufficient datasets. There is a question also that deep learning models are susceptible to data sources. This study aims to investigate which method is more effective in detecting and assessing thyroid nodules using our ultrasound images. The pre-trained convolutional neural network (VGG16) learned from ImageNet [15] is developed as our deep learning model. The process of radiomics method in this study involves extracting commonly used high throughput features, applying mutual information (MI) and linear discrimination analysis (LDA) to realize the goals of feature selection and dimension reduction. Finally, support vector machine (SVM) is applied to classify the thyroid nodules into benign and malignant.

II. MATERIALS AND METHODS

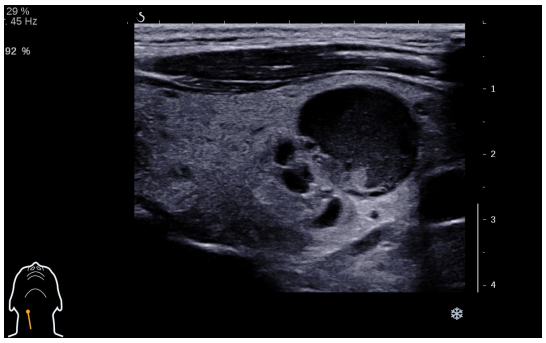
This section is organized into four sections, i.e., materials, image preprocessing, our deep learning and radiomics based methods, respectively. The first section mainly introduces how ultrasound images were collected and acquired. In the second section, the step of image preprocessing is presented. The third section describes what types of deep learning model was developed, including the variation of the VGG16 model. Finally, the last section elucidates the workflow of our radiomics based method in details, including computing and dealing with statistical features and classifying thyroid nodules using SVM.

A. MATERIALS

In this work, 3120 thyroid nodules from 1040 patients, with different ages, were retrospectively collected in the collaborated hospital. The benign and malignant nodules accounted for 59% and 45% of the total, respectively (i.e., 1841 benign and 1393 malignant nodules). In addition, transverse and longitudinal thyroid nodules images were scanned for each patient (as seen in Fig.1). According to the guidance of radiologists, because the background of the longitudinal ultrasound images with thyroid nodules is larger than that of the transverse ones, we only chose longitudinal ultrasound images for our further investigation. In addition, in order to avoid bias, all thyroid nodules were performed with the same ultrasonic system.

B. IMAGE PREPROCESSING

The collected raw ultrasound images contained a lot of irrelevant information, e.g., patient's ID and name, acquisition parameters, and screen background. Because these were useless information and/or even could confuse the network training, an automated or a manual procedure was needed to remove them. We proposed the following semi-automated process to deal with these irrelevant information. Since the boundary of the thyroid nodules in the ultrasound image was


(a)

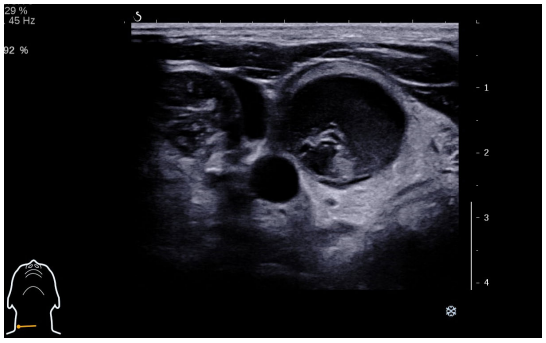
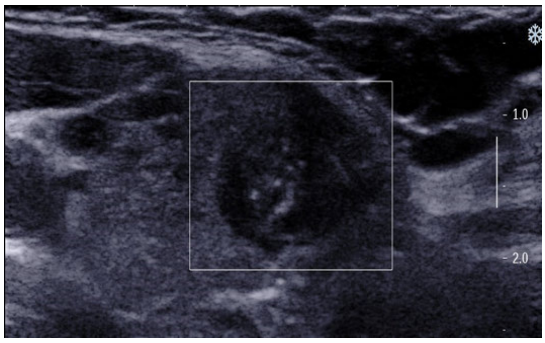

(b)

FIGURE 1. Examples of the ultrasound images with thyroid nodule. (a) The longitudinal thyroid nodules. (b) The transverse thyroid nodules.


(a)


(b)

FIGURE 2. Example of a cropped image for the ultrasound thyroid nodule localization. (a) delineated by radiologist (white rectangle). (b) cropped by automatic method.

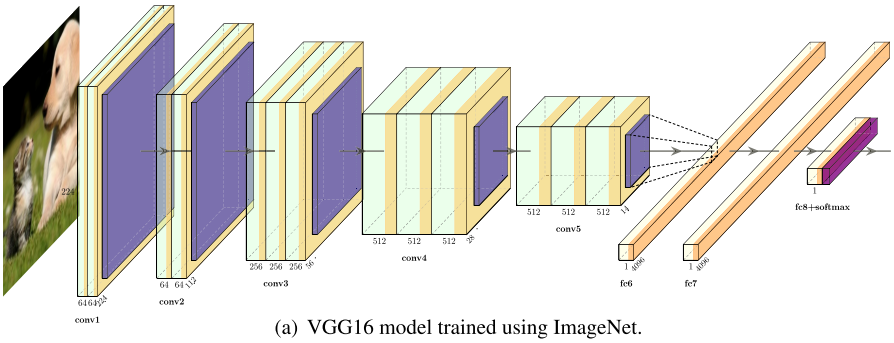
approximately delineated by experienced radiologists using a square shape with colored or white-or-black pixels, we first located that shape and executed erosion and expansion operations to get the corresponding masks. Finally, the mask multiplied with the corresponding raw image that could produce a cropped square image. If the automated procedure failed, we performed the manual cropping instead. Although this step might lose some useful information, we ensured that each cropped image contained a complete lesion area that was crucial for prediction. One cropped example is illustrated in Fig 2.

C. DEEP LEARNING BASED CLASSIFICATION METHOD

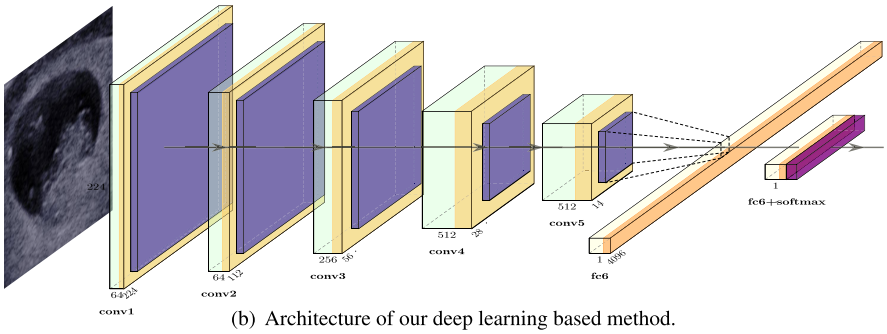
The deep learning models, specifically CNNs, are widely used and have performed extremely well in various computer vision tasks. The CNNs can extract features automatically by computer, and realize regression and classification tasks by multi-layer perceptron. In this paper, the CNNs model learned from ImageNet [15] was transferred to our ultrasound dataset as a pre-trained deep learning model to complete a specific classification task. Among of CNNs, the VGG16 model (see Fig. 3a),, which was previously trained on 1.2 million high-resolution images with 1000 different classes, was used in this paper. It can extract specific features and complete classification for our thyroid ultrasound images. The initial VGG16 model included 13 convolution layers, 5 max-pooling layers, and 3 fully-connected layers. The output of two fully-connected layers was 4096-dimensional features, and the other one aimed to classify images. Additionally, considering that our ultrasound datasets with thyroid nodules were relatively small, which would increase the complexity of training the deep learning model; therefore, reducing the number of layers of the VGG16 model was a good strategy to avoid this problem. Finally, our model included 5 convolution layers, 5 max-pooling layers, 2 fully-connected layers. The output of one fully-connected layer was 4096-dimensional, and the last one was the classified layer, i.e. benign and malignant. Fig. 3b shows the variation of the initial VGG16 model that we developed for our deep learning based method. The input ultrasound images were resized to 224*224*3. And the size of convolutional kernels was 3*3. But penultimate one was 1*1 *4096 and the last layer had a convolution depth of 2. Therefore, the complexity of the initial VGG16 model was reduced. Additionally, the fine-tuning method was used in our work. That is to say, the parameters of VGG16 model trained by ImageNet dataset were considered as the initial parameters for our developed deep learning model. For the training of our developed deep learning model, we applied dropout to prevent overfitting, and the cross-entropy loss was used to optimize the learning rate and dropout probability.

D. RADIOMICS BASED CLASSIFICATION METHOD

The radiomics based methods consist a process that converting digital medical images into mineable high-dimensional data. Therefore, the key of radiomics based methods is to extract high-dimensional feature data in order to meet the clinical needs. In this paper, 302 features were extracted from region of interest (ROI) for each ultrasound image with thyroid nodules. ROI was delineated by radiologists (see Fig. 2a). The extracted features included gray-scale histograms, intensity difference, GLCM, Gabor and wavelet based statistical



(a) VGG16 model trained using ImageNet.



(b) Architecture of our deep learning based method.

FIGURE 3. The deep learning model, where fc denotes a fully connected layer, conv denotes a convolution layer, softmax denotes a classified layer, pool denotes the max-pooling layer. In the figure, the color of pale green refers to the conv, golden refers to the relu, purple refers to the pool and violet refers to the softmax.

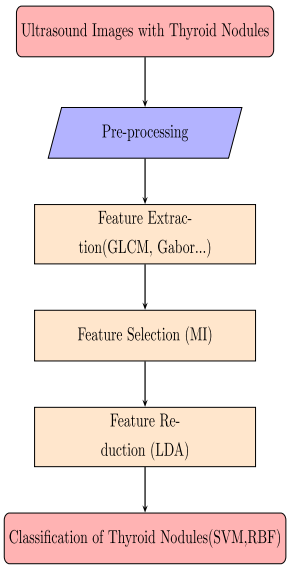


FIGURE 4. Framework of our proposed radiomics based method.

feature. After feature extraction, MI [23], [24], LDA [25], [26] and SVM [27]–[29] were applied to achieve the goal of feature selection, dimension reduction and classification task, respectively. Fig 4 displays the workflow of the radiomics based method developed in our study.

1) FEATURE EXTRACTION

Before feature extraction, the median filter, which has a proved filtering effect on salt-pepper noise and edge preserving characteristics, was applied to the ultrasound images with thyroid nodules. In total, 302 features (including 28 gray-scale histograms features, 12 intensity difference features, 96 GLCMs, 12 wavelet features and 36 gabor features) were extracted from the ultrasound images. The first-order statistical features, including mean, standard deviation, skewness, kurtosis, energy, and entropy, are the most widely used ones for gray-scale histograms, which can describe the pixel intensity distribution in digital images [27]. The mean is a measure of median intensity of gray levels of the image. The purpose of variance (Var) and standard deviation (Sd) is to express the degree of dispersion of data points in datasets. The skewness (Skew) measures data symmetry. The kurtosis (Kurt) measures the flatness of data distribution. The energy indicates how the gray levels are distributed. The entropy specifies the uncertainty / randomness in the image value, which measures the average amount of information needed to encode the image value. They are mathematically defined as follows:

$$Mean_1 = \frac{1}{M}\sum_{i=0}^{L-1} i \cdot h(i)$$

$$Var = \frac{1}{M}\sum_{i=0}^{L-1} (i - Mean)^2 \cdot h(i)$$

$$Sd = (Var)^2$$

$$Skew = \frac{1}{M \cdot \sigma^{-3}}\sum_{i=0}^{L-1} (i - \mu)^3 \cdot h(i)$$

$$Kurt = \frac{1}{M \cdot \sigma^{-4}}\sum_{i=0}^{L-1} (i - \mu)^4 \cdot h(i)$$

$$energy_1 = \frac{1}{M^2}\sum_{i=0}^{L-1}[h(i)]^2$$
$$entropy_1 = -\frac{1}{M}\sum_{i=0}^{L-1}h(i)\cdot log_2[\frac{h(i)}{M}] \quad (1)$$

Among the 7 mathematical formulas, let M be the image resolution, L be the number of gray levels, i be the current gray level, and h(i) be the number of image pixels having the gray intensity value corresponding to the ''i'' level. The features of gray-scale histogram were directly computed by the 7 mathematical formulas.

The gray difference between a pixel in an image and a pixel with only a small distance is called intensity difference. $P_\triangle(i)$ is the probability corresponding to the gray difference value i, and i is the normalized difference value. When the $P_\triangle(i)$ value is closer to the coordinate origin, the image texture is thicker. In contrast, if its distribution is more uniform, the texture is relatively thin. In general application, the intensity difference statistical method uses mean, contrast (Con), second-order moment of angle (ASM) to represent an image texture features.

$$Mean_2 = \frac{1}{M}\sum_i i\cdot P_\triangle(i). \quad (2)$$
$$Con = \sum_i i^2\cdot P_\triangle(i). \quad (3)$$
$$ASM = \sum_i [P_\triangle(i)]^2 \quad (4)$$

Gray level co-occurrence matrix (GLCM) was originally proposed by Robert and Shanmugam [10]. It is a comprehensive texture analysis method based on the assumption that the spatial distribution among pixels contains texture information of an image. It not only reflects the distribution of image brightness, but also reflects the location characteristics between pixels with the same two points or close to brightness. Additionally, it is a second-order statistical feature. There are 24 types of statistic features based on GLCM, including energy, entropy, evenness (Even), correlation (Cor). If f(i, j) is a two-dimensional digital image and S represents the pixel pairs with special spatial relation in the target region R, then the gray level co-occurrence matrix P satisfying certain spatial relation can be obtained P(i,j):

$$Energy_2 = \sum_i\sum_j P(i,j)^2. \quad (5)$$
$$Entropy_2 = -\sum_i\sum_j p(i,j)logP(i,j). \quad (6)$$
$$Even = \sum_i\sum j\frac{1}{1+(i-j)^2}\cdot P(i,j). \quad (7)$$
$$Cor = \frac{\sum\sum(i-\bar{x})(j-\bar{y})\cdot p(i,j)}{\sigma_x\cdot\sigma_y} \quad (8)$$

In addition, there are two important parameters of the GLCM: distance (d) and angle (a). The different combinations of d and a will produce abundant image features. The angle was set to 0°, 45°, 90° and 135° in this study. Furthermore, to preserve the complexity of the spatial relationship, the distance was set to 1.

The wavelet transform is a local transform of time and frequency, which has the characteristics of multi-resolution analysis and can represent the local characteristics of signals in both time and frequency domains. The wavelet transform can concentrate the energy of original image to a small number of wavelet coefficients in the direction of three local correlations, a high degree of detail components that provides a powerful condition for feature extraction. Additionally, using different scales of the wavelet band-pass filter, an image can be decomposed to a series of frequency bands. The low-pass filter and high-pass filters were used in this study, and the image was decomposed into low-frequency sub-bands and high-frequency sub-bands. In doing so, four sub-bands could be produced, LL, LH, HL, HH, and the mean and standard deviation could be used as the features of a sub-band image. Finally, the extracted 12 features of the wavelet were LL mean, LE mean, EL mean, SL mean, LS mean, EE mean, LL standard deviation, LE standard deviation, EL standard deviation, SL standard deviation, EE standard deviation, and LS standard deviation.

The Gabor transform can decompose an image into a series of channels, make full use of the accurate descriptive information of each decomposition level, and form an effective feature vector. Textural feature extraction can be achieved by filtering the image with a set of Gabor filters with selected size and direction parameters. In our study, the size was set to 3*3, 2*3 and 2*4, and the direction was set to 0°, 45°, 60° and 90°. The extracted features were mean, contrast and entropy.

It is worth mentioning that before extracting gray histogram, gray difference statistics and GLCM based features, the method of Laplacian of Gaussian was used. The Laplacian of Gaussian can be approximated by the Gaussian difference, which is obtained by the convolution of two Gaussian filters with different variables. In addition, the Laplace operator is sensitive to discrete points and noise when it is used to implement edge detection through image operation. In doing so, firstly, the image was denoised by Gaussian filter, and then the edge was detected by the Laplace operator, which could improve the robustness of noise and discrete points. In this study, the coefficient of filter was set to 0.4, the size was set to 3, 7 and 10.

2) FEATURE SELECTION, REDUCTION AND CLASSIFICATION

Uncorrelated features and even interdependence among features could lead to a prolonged feature analysis and classification model training. Moreover, too many features could easily cause ''high-dimension disaster'' or overfitting problem and the model would become very complex. The method of feature selection can eliminate irrelevant or redundant features [30]. MI and LDA were used to select and reduce extracted features, respectively. In computer vision

or medicine research studies, many classifiers, including k-nearest neighbor classifier, naive Bayes classifier, and SVM classifier, have been widely used. Among them, the SVM is one of the most robust and effective classifiers. Its computational complexity depends on the number of support vectors rather than the dimension of the sample space, which avoids the problem of "high-dimension disaster" in a sense. In addition, it is determined by a few support vectors and insensitive to outliers, which makes the model have better robustness and generalization ability. Therefore, SVM with the RBF kernel function was used as our classifier in this study.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. EVALUATION INDEXES

To evaluate the performance of the classification using both radiomics and deep learning based methods, the indexes, including the accuracy (ACC), the sensitivity (SEN), the specificity (SPC), the receiver operating characteristic curve (ROC) and the area under curve (AUC), were computed. SEN and SPC express the possibility of predicting the malignant and benign thyroid nodules, respectively. The vertical axis of the ROC is the true positive rate (TPR), and the horizontal axis stands for the false positive rate (FPR). AUC is a performance index of the classifiers.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$SEN = \frac{TP}{TP + FN} \tag{10}$$

$$SPC = \frac{TN}{TN + FP} \tag{11}$$

B. QUANTITATIVE RESULTS

In our experiments, the dataset was divided into 80% for training and 20% for an independent testing. The classification performance and the ROC curves for radiomics and deep learning based methods are presented in Table 1 and Fig 5, respecyively. The classification accuracy, sensitivity, and specificity of applying radiomics based method are 66.81%, 51.19% and 75.77%, respectively, while these evaluation indexes for the deep learning based method trained to the testing samples are 74.69%, 63.10% and 80.20%, respectively. For all these indexes, the trained deep learning model demonstrates higher value compared to radiomics based method. Fig 5 shows the performance comparison of the two methods, where the AUCs for diagnosing ultrasound thyroid malignancy are 0.6371 for radiomics based method and 0.7127 for convolutional neural network (VGG16 model) based method. There is significant difference in AUC between the two used methods. Therefore, compared to the radiomics baesd method, the deep learning can provide us the better performance using our ultrasound datasets.

IV. DISCUSSIONS

From the above comparison experiments, the deep learning based method outperformed the radiomics based.

TABLE 1. The classification performance of radiomics and deep learning based methods.

Methods	ACC	SEN	SPC	AUC
Radiomics	66.81%	51.19%	75.77%	63.71%
Deep learning	74.69%	63.10%	80.20%	71.27%

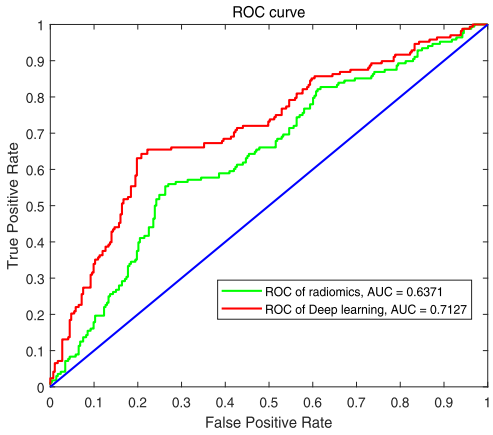


FIGURE 5. Receiver operating characteristic curves for comparison of the classification performance (radiomics vs. deep learning based methods).

The classification of ultrasound images with thyroid nodules by using deep learning could be a better strategy to assess the thyroid nodules, with higher classification accuracy. There are still some limitations that exist in our study. Although the deep learning based method could be better than the radiomics based method, the accuracy of our deep learning method only achieved 74.69% in our study. There could be some reasons for achieving these results. Firstly, for radiomics based method, we did not use other radiomics features such as Fourier transformation, gray-level size zone matrix and neighborhood gray-tone difference matrix, which can express statistical information to some extent. Nevertheless, in future studies, we plan to add more features to increase the throughput of features and further improve the classification performance of radiomics. Secondly, in our deep learning based method, we only used a common CNN model (VGG16) as our pre-trained model and trained it in our relatively small dataset. Previous studies on the impact of dataset size in transfer learning using CNNs demonstrated that a suitable size of dataset could improve the performance of the deep learning models [31]. Moreover, it might be a problem that a pre-trained CNN model was applied rather than building a new deep learning architecture for our specific task [32]. Therefore, we can augment our datasets to have enough number of training samples and modify the deep learning architecture [33] to adapt our ultrasound data directly. The fusion of deep learning based method and radiomics based method is also a focus. Guanghui Han *et al.* used the fusion method to automatic recognize cavity imaging signs in lung computed tomography (CT)images and improved the performance [34]. Last but not least, there have a strategy which uses multimodal images to improve the performance of disease [35]–[38], the application of color doppler flow images

and real-time elastography images were common [36]–[38]. These enlighten us that combining another imaging modality with B-mode may make the deep learning method have a better performance and become more robust.

V. CONCLUSION

In this paper, we compared the classification performance of radiomics and deep learning based methods using ultrasound images with thyroid nodules. The comparison results show that the deep learning based model outperformed radiomics based method. In future work, we plan to supplement the throughput of features and further fine-tune the pre-trained deep learning model to improve the classification performance. Additionally, using multicentre medical data to verify the effectiveness of deep learning models in thyroid nodules detection is achallenging and focus.

ACKNOWLEDGMENT

(Yongfeng Wang and Wenwen Yue contributed equally to this work.)

REFERENCES

[1] A. A. Parsa and H. Gharib, "Thyroid nodule: Current evaluationx and management," in *The Thyroid and Its Diseases: A Comprehensive Guide for the Clinician*, M. Luster, L. Duntas, L. Wartofsky, Eds. Cham, Switzerland: Springer, Jan. 2019, pp. 493–516.

[2] K. D. Miller, R. L. Siegel, C. C. Lin, A. B. Mariotto, J. L. Kramer, J. H. Rowland, K. D. Stein, R. Alteri, and A. Jemal, "Cancer treatment and survivorship statistics, 2016," *CA, Cancer J. Clinicians*, vol. 66, no. 4, pp. 271–289, Jul. 2016, doi: 10.3322/caac.21349.

[3] J. P. Brito, J. C. Morris, and V. M. Montori, "Thyroid cancer: Zealous imaging has increased detection and treatment of low risk tumours," *BMJ*, vol. 347, no. 4, pp. f4706–f4706, Aug. 2013. [Online]. Available: http://www.bookref.com

[4] B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Y. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, J. A. Sosa, D. L. Steward, R. M. Tuttle, and L. Wartofsky, "2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer," *Thyroid.*, vol. 26, no. 1, pp. 1–133, Jan. 2016.

[5] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, and J. S. Suri, "ThyroScreen system: High resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform," *Comput. Methods Programs Biomed.*, vol. 107, no. 2, pp. 233–241, Aug. 2012.

[6] R. Zhu, Z. Wang, Y. Zhang, B. Yu, M. Qi, X. Feng, C. Wu, Y. Cui, L. Huang, F. Li, and F. Zhou, "Integrating five feature types extracted from ultrasonograms to improve the prediction of thyroid papillary carcinoma," *IEEE Access*, vol. 7, pp. 101820–101828, Jul. 2019, doi: 10.1109/ACCESS.2019.2929237.

[7] P. Poudel, A. Illanes, E. J. G. Ataide, N. Esmaeili, S. Balakrishnan, and M. Friebe, "Thyroid ultrasound texture classification using autoregressive features in conjunction with machine learning approaches," *IEEE Access*, vol. 7, pp. 79354–79365, Jun. 2019, doi: 10.1109/ACCESS.2019.2923547.

[8] W. Yang, J. Zhao, Y. Qiang, X. Yang, Y. Dong, Q. Du, G. Shi, and M. B. Zia, "DScGANS: Integrate domain knowledge in training dual-path semi-supervised conditional generative adversarial networks and s3vm for ultrasonography thyroid nodules classification," in *Proc. MICCAI*, Shenzhen, China, Oct. 2019, pp. 558–566.

[9] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016, doi: 10.1148/radiol.2015151169.

[10] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, San Diego, CA, USA, Jun. 2015, pp. 886–893.

[12] M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process.*, vol. 4, no. 2, pp. 172–179, Jun. 1975.

[13] Q. Wang, Q. Li, R. Mi, H. Ye, H. Zhang, B. Chen, Y. Li, G. Huang, and J. Xia, "Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: A cohort study," *J. Magn. Reson. Imag.*, vol. 49, no. 3, pp. 825–833, Mar. 2019.

[14] H. A. Nugroho, M. Rahmawaty, Y. Triyani, and I. Ardiyanto, "Texture analysis for classification of thyroid ultrasound images," in *Proc. Int. Electron. Symp. (IES)*, Denpasar, Indonesia, Sep. 2016, pp. 475–480.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR.*, San Diego, CA, USA, 2015, pp. 1–14. [Online]. Available: https://arxiv.xilesou.top/pdf/1409.1556.pdf

[16] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," 2017, *arXiv:1707.01836*. [Online]. Available: http://arxiv.org/abs/1707.01836

[17] Z. Jin, Y. Zhu, S. Zhang, F. Xie, M. Zhang, Y. Zhang, X. Tian, J. Zhang, Y. Luo, and J. Cao, "Ultrasound computer-aided diagnosis (CAD) based on the thyroid imaging reporting and data system (TI-RADS) to distinguish benign from malignant thyroid nodules and the diagnostic performance of radiologists with different diagnostic experience," *Med. Sci. Monitor*, vol. 26, pp. 1643–3750, Jan. 2020.

[18] M. Aazam, K. A. Harras, and S. Zeadally, "Fog computing for 5G tactile industrial Internet of Things: QoE-aware resource allocation model," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 3085–3092, May 2019.

[19] K. Alex, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, Lake Tahoe, Spain, 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet classification-with-deep-convolutional-neural-networks.pdf

[20] W. Zhang, S. Gupta, X. Lian, and J. Liu, "Staleness-aware Async-SGD for distributed deep learning," in *Proc. IJCAI*, New York, NY, USA, 2016, pp. 1–7. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet classification-with-deep-convolutional-neural-networks.pdf

[21] K. Chen and Q. Huo, "Scalable training of deep learning machines by incremental block training with intra-block parallel optimization and blockwise model-update filtering," in *Proc. ICASSP.*, Shanghai, China, Mar. 2016, pp. 5880–5884.

[22] J. Liu, W. Li, N. Zhao, K. Cao, Y. Yin, Q. Song, H. Chen, and X. Gong, "Integrate domain knowledge in training CNN for ultrasonography breast cancer diagnosis," in *Proc. MICCAI*, Spain, Granada, 2018, pp. 868–875.

[23] A. E. Pablo and T. Michel, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Jan. 2019.

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[25] H. Brunzell and J. Eriksson, "Feature reduction for classification of multidimensional data," *Pattern Recognit.*, vol. 33, no. 10, pp. 1741–1748, Oct. 2000.

[26] A. Iosifidis, A. Tefas, and I. Pitas, "On the optimal class representation in linear discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1491–1497, Sep. 2013.

[27] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *Analyst*, vol. 135, no. 2, pp. 230–267, Oct. 2018.

[28] I. Gatos, S. Tsantis, M. Karamesini, A. Skouroliakou, and G. Kagadis, "Development of a support vector machine–based image analysis system for focal liver lesions classification in magnetic resonance images," *J. Phys., Conf. Ser.*, vol. 633, no. 1, Sep. 2015, Art. no. 012116.

[29] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.

[30] J. Han, X. Ji, X. Hu, D. Zhu, K. Li, X. Jiang, G. Cui, L. Guo, and T. Liu, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.

[31] D. Soekhoe, P. van der Putten, and A. Putten, "On the impact of data set size in transfer learning using deep neural networks," in *Proc. 15th Int. Symp. IDA*, Stockholm, Sweden, 2016, pp. 50–60.

[32] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[33] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.
[34] G. Han, X. Liu, H. Zhang, G. Zheng, N. Q. Soomro, M. Wang, and W. Liu, "Hybrid resampling and multi-feature fusion for automatic recognition of cavity imaging sign in lung CT," *Future Gener. Comput. Syst.*, vol. 99, pp. 558–570, Oct. 2019.
[35] Q. Li, Z. Gao, Q. Wang, J. Xia, H. Zhang, H. Zhang, H. Liu, and S. Li, "Glioma segmentation using a novel unified algorithm in multimodal MRI images," *IEEE Access*, to be published.
[36] C. Shuzhen, "Comparison analysis between conventional ultrasonography and ultrasound elastography of thyroid nodules," *Eur. J. Radiol.*, vol. 81, no. 8, pp. 1806–1811, Aug. 2012.
[37] A. E. Ebeed, M. A. E.-H. Romeih, M. M. Refat, and N. M. Salah, "Role of ultrasound, color Doppler, elastography and micropure imaging in differentiation between benign and malignant thyroid nodules," *Egyptian J. Radiol. Nucl. Med.*, vol. 48, no. 3, pp. 603–610, Sep. 2017.
[38] A. Abbasian Ardakani, A. Bitarafan-Rajabi, A. Mohammadi, S. Hekmat, A. Tahmasebi, M. B. Shiran, and A. Mohammadzadeh, "CAD system based on B-mode and color Doppler sonographic features may predict if a thyroid nodule is hot or cold," *Eur. Radiol.*, vol. 29, no. 8, pp. 4258–4265, Aug. 2019.

YONGFENG WANG is currently pursuing the master's degree with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. Her research interest is health informatics computing.

WENWEN YUE received the M.D. degree in clinical medicine from Binzhou Medical University, China, in 2011. She is currently pursuing the Ph.D. degree with Tongji University. Her research focuses on diagnostic and interventional ultrasound in medicine.

XIAOLONG LI received the M.D. degree in imaging and nuclear medicine from Tongji University, China, in 2017. He is currently an Attending Doctor with Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, School of Medicine, Tongji University. His research focuses on diagnostic and interventional ultrasound in medicine.

SHUYU LIU received the Ph.D. degree from Sun Yat-sen University, in 2010. He is the Director of the Experimental Teaching Center, School of Pharmacy, Sun Yat-sen University, China. His research interests include biological image processing, biomedical engineering, and intelligent medicine.

LEHANG GUO received the M.D. degree in clinical medicine from Tongji University, Shanghai, China, in 2015. He is currently pursuing the Ph.D. degree with Nanjing Medical University, Nanjing, China. He is currently an Attending Doctor with Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, School of Medicine, Tongji University. His research focuses on diagnostic and interventional ultrasound in medicine.

HUIXIONG XU received the M.D. degree in clinical medicine from Tongji Medical University, China, in 1994, and the Ph.D. degree in imaging medicine from the Tongji Medical College, Huazhong University of Science and Technology, China, in 2001. He is currently a Full Professor with Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, School of Medicine, Tongji University, China. His research focuses on diagnostic and interventional ultrasound in medicine.

HEYE ZHANG (Member, IEEE) received the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, China, in 2007. He is currently a Full Professor with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. His research focuses on health informatics computing.

GUANG YANG (Member, IEEE) received the M.Sc. degree in vision imaging and virtual environments from the Department of Computer Science, University College London, in 2006, and the Ph.D. degree in medical image analysis jointly from the CMIC and the Department of Computer Science and Medical Physics, University College London, in 2012. He is currently an Honorary Lecturer with the Neuroscience Research Centre, Cardiovascular and Cell Sciences Institute, St. George's, University of London. He is also an Image Processing Physicist and an Honorary Senior Research Fellow of the Cardiovascular Research Centre, Royal Brompton Hospital, and also with the National Heart and Lung Institute, Imperial College London.