IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 24, 2020, accepted April 8, 2020, date of publication April 14, 2020, date of current version May 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987961

Unsupervised Multi-Discriminator Generative Adversarial Network for Lung Nodule Malignancy Classification

YAN KUANG[1,2], TIAN LAN[1,3], (Member, IEEE), XUEQIAO PENG[1], GATI ELVIS SELASI[1], QIAO LIU[1], (Member, IEEE), AND JUNYI ZHANG[2]

[1] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[2] The 54th Research Institute of CETC, Shijiazhuang 050081, China
[3] Network and Data Security Key Laboratory of Sichuan Province, Chengdu 610054, China

Corresponding author: Tian Lan (lantian1029@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U19B2028 and Grant 61772117, in part by the National Science and Technology Commission Innovation Project under Grant 19-H863-01-ZT-003, in part by the CETC Big Data Research Institute under Grant 10-2018039, in part by the Sichuan Hi-Tech Industrialization Program under Grant 2018GFW0150, and in part by the Fundamental Research Funds for the Central Universities under Grant XGBDFZ09 and Grant ZYGX2019J077.

ABSTRACT Computer-aided diagnosis systems with deep learning frameworks have been used to identify benign and malignant pulmonary nodules in lung cancer diagnosis. It's commonly known that a premise of training complex deep neural nets is the large-scale labeled datasets. However, the abundance of labeled datasets is usually unavailable in many medical image domains. This factor can lead to the poor generalization performance of deep learning models. In this paper, we propose a novel multi-discriminator generative adversarial network model combined with an encoder for the classification of benign and malignant pulmonary nodules. To the best of our knowledge, we are the first to apply unsupervised learning to identify benign and malignant lung nodules. Firstly, we use a multi-discriminator generative adversarial network to build a generative model trained with unlabeled benign lung nodule images. Then an encoder is combined with the trained generative model to establish a mapping of benign pulmonary nodule images to the latent space. The benign and malignant lung nodules are scored by calculating the GAN discriminator feature loss and image reconstruction loss. The model yields high anomaly scores on malignant images and low anomaly scores on benign images. Experimental results show that our method with only a small number of unlabeled datasets could achieve more competitive results compared with other supervised deep learning approaches.

INDEX TERMS Computer-aided diagnosis (CAD), lung nodule, malignancy classification, unsupervised learning, generative adversarial networks.

I. INTRODUCTION

According to the 2015 Global Cancer Statistics, lung cancer is approximately more than 27% of all cancers and causes 19.5% of cancer-related deaths each year [1]. Early lung cancer has no obvious clinical symptoms, and some are only presented in the form of lung nodules. In the majority of cases, it is too late for successful therapy once the patient develops the first symptoms. In recent decades, with the update and development of various clinical examination technologies, especially low-dose spiral computed tomography in lung screening has provided an effective method for the early diagnosis of lung cancer. Because of the large range of nodule shape, texture variation and visual similarities shared by malignant and benign nodules, pulmonary nodules identification has become a research focus [1], [2].

It is a heavy task for doctors to identify lung nodules from a large number of CT images and judge whether they are benign or malignant. Due to the subjective nature of doctors, there is a high chance of misdiagnosis. Thus, computer-aided diagnosis (CAD) systems have been developed to overcome this problem. There are two main classes of CAD systems [3]:

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

detection systems (CADe) and diagnosis systems (CADx). The goal of CADe is to detect the presence of nodules at the initial stage. The aim of CADx is to diagnose the benign and malignant nodules based on the detected nodules, which can assist doctors to make an accurate diagnosis of patients with lung diseases in time [4]–[7].

Recently, deep learning techniques have achieved profound success in many medical image domains, for instance, the classification of benign-malignant lung nodules [8]. Hua *et al.* [9] applied deep convolutional neural networks (DCNN) and deep belief networks (DBN) to lung nodule classification, respectively, and confirmed that deep learning can better distinguish benign and malignant lung nodules. Shen *et al.* [10] proposed to apply multi-crop convolutional neural networks (MC-CNN) for lung nodule malignancy classification and learn the heterogeneous features of lung nodules by learning images of different scales. Xie *et al.* [11] jointly used the nine multi-view knowledge-based collaborative (MV-KBC) deep model to separate malignant from benign nodules by characterizing the nodules' overall appearance, voxel, and shape heterogeneity, respectively. These experiments indicate that deep learning framework has good performance in the classification of benign and malignant nodules.

A prerequisite for training deep learning models is the presence of large-scale labeled datasets. However, labels are especially difficult to obtain. This is due to a number of factors: a) marking medical data usually requires a specially trained physician; b) It is difficult for even experts to mark the lesion boundaries due to the low signal-to-noise ratio in many medical images; c) The annotators have to mark the entire 3D volume of data, which can be expensive and time-consuming. Because of these limitations, CT medical image datasets are usually small, which may lead to over-fitting on the training set, and by extension, poor generalization performance on the test set [12].

In recent years, with the popularization of medical examinations, the demand for processing large amounts of unlabeled medical image data has become more and more urgent. Therefore, it is of great significance to explore the computer-aided diagnosis of pulmonary diseases based on weakly supervised learning, semi-supervised learning, and unsupervised learning in the field of medical imaging. Zhu *et al.* [12] proposed DeepEM, a novel deep 3D ConvNet framework augmented with expectation-maximization (EM), to mine weakly supervised labels in EMRs for pulmonary nodule detection. Feng *et al.* [13] applied a weakly supervised method to pulmonary nodule segmentation. In this method, multiscale learning and global average pooling (GAP) were used to obtain the true nodule location, and then pulmonary nodules were segmented in combination with the iterated conditional mode. Zhang *et al.* [14] presented an unsupervised multi-hidden layer deep belief network to extract the deep features of lung nodule images, and then used extreme learning machine as a classifier to classify the extracted features into benign and malignant ones. However, the practical classification method of benign and malignant pulmonary nodules with unsupervised learning has not been proposed.

Generative adversarial networks (GANs) introduced by Goodfellow *et al.* [15] is one of the most promising unsupervised learning methods. The proposed algorithm is based on GAN. A GAN consists of two adversarial networks, a generator G and a discriminator D. G is a generative network that uses a random to generate an image. D is a discriminating network that discriminates whether an image is real. The formulation of GAN training suffers from training instability and is prone to mode collapse, and thus GAN is hard to train [15]. DCGAN [16] is a better improvement after GAN, which provides a good network topology for GAN training. However, training instability is not fundamentally solved, and the training process of G and D needs to be carefully balanced during training. Unlike DCGAN, WGAN [17] mainly improves GAN from the perspective of the loss function. It uses Wasserstein distance to measure the distance between the generated data distribution and the real data distribution, and theoretically solves the problems of training instability and collapse mode. Gulrajani *et al.* [18] proposed an improved WGAN training procedure replacing weight clipping by gradient penalty to solve the problem that WGAN sometimes generates low quality images.

Here, we proposed an unsupervised method for the classification of benign-malignant lung nodules based on the combination of a multi-discriminator generative adversarial network (MDGAN) and an encoder in the medical images domain. Our model training doesn't require large amounts of labeled data, and only the benign nodules images are required as the training set to classify benign and malignant pulmonary nodules. The idea of our method is motivated by anomaly detection [19]. Anomaly detection is a common application of machine learning algorithms, which is to find objects that are different from normal objects. It is generally required that the data have a "normal" model, and anomaly are considered deviations from this normal model, and the degree of their deviation will be used to calculate anomaly scores. In this paper, we define benign pulmonary nodules as normal data, and malignant pulmonary nodules as anomalous data. Our method is to use a multi-discriminator generative adversarial network and an encoder model to learn the feature distribution of normal data by establishing the mapping of real normal data to the MDGAN latent space and then calculate the MDGAN discriminator feature loss and image reconstruction loss which are used to score benign and malignant lung nodules. In general, our model yields high anomaly scores on malignant images and low anomaly scores on benign images. Our main contributions can be summarized as follows:

1) Our method is the first to successfully introduce unsupervised learning into the classification of lung nodule malignancy. It overcomes the shortcomings of existing deep learning methods that required large labeled datasets for training.
2) The combination of a multi-discriminator generative adversarial network with an encoder is more conducive

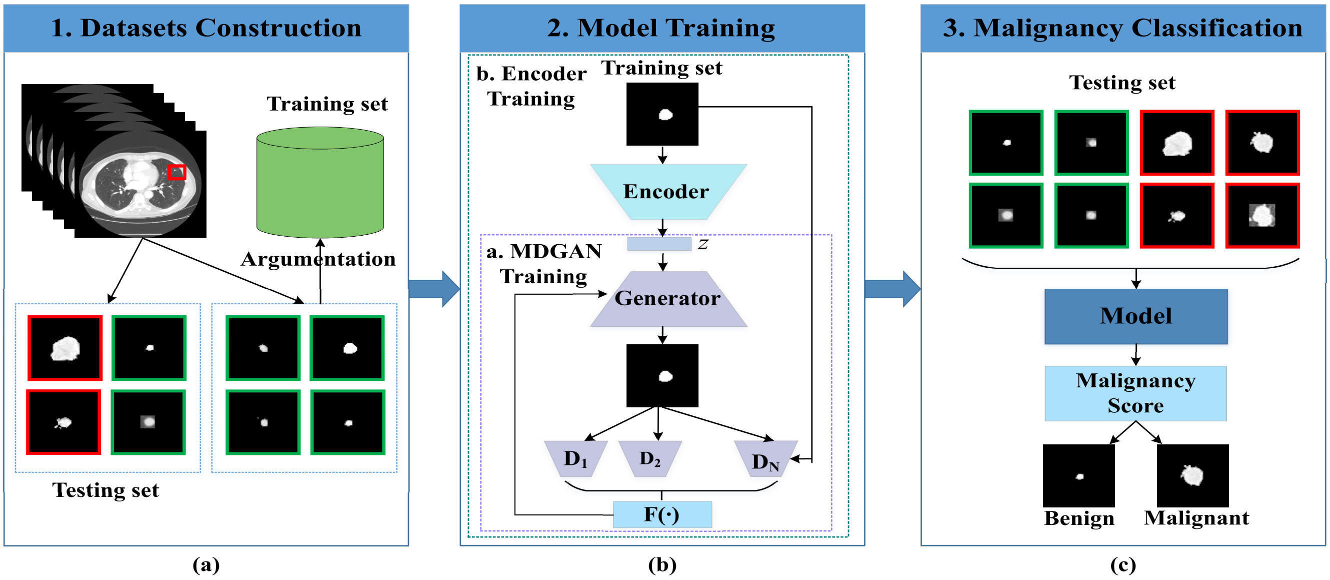

FIGURE 1. Flowchart of proposed approach: (a) Sample datasets construction; (b) Model training; (c) Lung nodule malignancy classification.

to improving the generator performance and better learning the characteristic distribution of pulmonary nodules images.

II. METHODS

An overview of the proposed classification approach of lung nodules is shown in Fig.1, which mainly includes three steps: a) sample datasets construction; b) model training: multi-discriminator generative adversarial network training and MDGAN guided encoder training; c) lung nodule malignancy classification.

Using the publicly available LIDC-IDRI dataset, the information of lung nodules' locations and their malignancy levels are extracted from the XML file and the ROI regions of lung nodules are segmented to form the sample dataset. Benign nodules images without any label are used to train the multi-discriminator GAN network. The trained generator and discriminators are combined with an encoder to map the real unmarked lung nodules images into the latent space. The latent representation is subsequently used as an input to the MDGAN network generator. The corresponding image is generated based on the compressed data representation, making the difference between the real image and the generated image very small, and the degree of their deviation will be used to calculate the abnormal score. In general, our model yields high anomaly scores on malignant images and low anomaly scores on benign images. Benign and malignant classification results can be obtained according to the defined threshold.

Compared with the previous supervised deep learning methods, we propose to apply the idea of anomaly detection based on unsupervised learning to benign and malignant classification of pulmonary nodules. We only need to train benign data to classify benign and malignant nodules with adversarial generation networks and encoders. Furthermore, in order to improve the learning ability of the generator, we propose to use multiple weak discriminators as a whole to provide more positive feedback for the generator. When classifying benign and malignant lung nodules based on abnormal scores, a scoring threshold is needed to divide them into two categories. We propose a formula for calculating the scoring threshold.

A. SAMPLE DATASETS CONSTRUCTION

The LIDC-IDRI dataset contains 1018 cases [20]–[22]. Each case contains a CT image and a corresponding XML annotation file, which provides the nodule contours marked by 4 radiological experts. Nodule diameters in lung nodule images generally range from 3 mm to 30 mm. According to the nodule diameter, lung nodules are divided into non-nodules, nodules smaller than 3 mm, or nodules larger than 3 mm. According to the mainstream screening scheme [10], [12], [22]–[28], we use nodules larger than 3 mm for our experiments. A 64 × 64 pixels ROI region is cropped by reading the XML annotation file, and the improved Threshold Probability Map algorithm [12] is used to segment the lung nodule. According to the malignant degree of a lung nodule in the XML annotation file, the malignant degree of each nodule is evaluated by at least one radiologist and at most four radiologists. For the nodule labeled by at least two radiologists, the mean malignancy level (MML) was taken as the ground truth label by the semantic attribute scores of different radiologists, and the lung nodule was stored as the malignancy by rounding method. If the MML is less than 3, it's labeled as a benign nodule; if it's equal to 3, it's labeled as an uncertain nodule; if it's greater than 3, it's labeled as a malignant nodule. To reduce the impact on the evaluation of the uncertainty of nodal malignancy, we excluded all uncertain

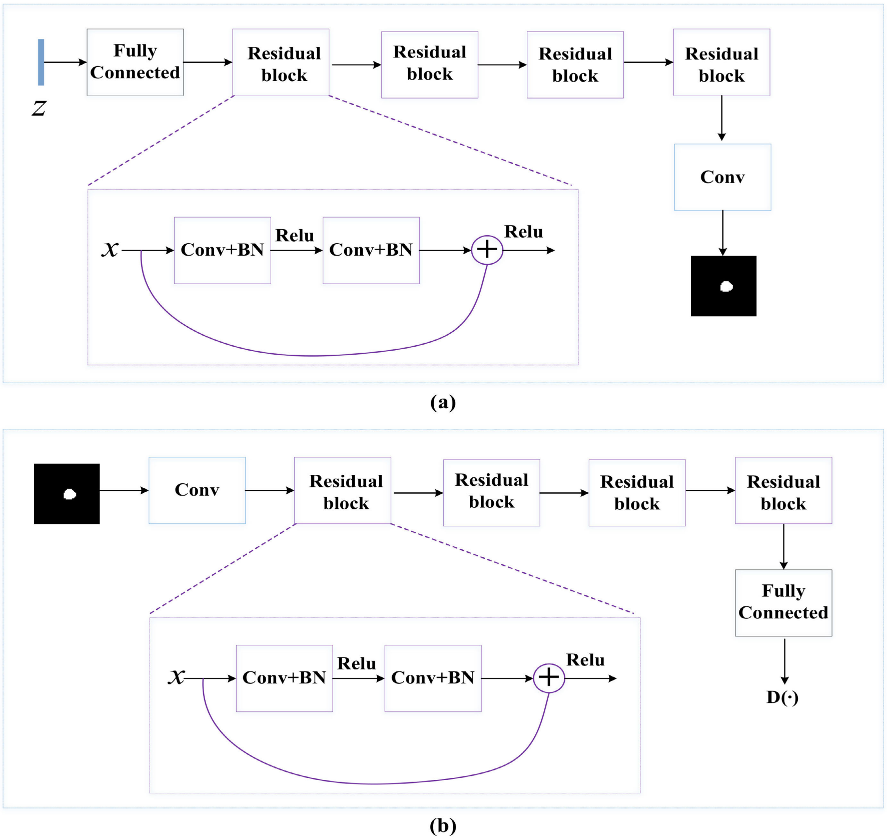

FIGURE 2. The architecture of our MDGAN: (a) Generator architecture; (b) Single discriminator architecture.

pulmonary nodules. Therefore, there were 1375 benign and 930 malignant nodules in Table 1.

TABLE 1. Distribution of malignancy level of each nodule.

Dataset	Benign		Uncertain	Malignant	
Malignancy level	1	2	3	4	5
Number	484	891	1386	673	257

We only utilized unlabeled benign lung nodule images as the training set with unsupervised learning. In the experiments of this method, 5-fold cross-validation is used. The 1375 benign nodules sample set is divided into 5 groups, one group is reserved as the testing set, and the remaining 4 groups are input to the integrated model as the training set. The experiment was repeated 5 times, and the average value is taken as the final result of the experiment. The randomly selected 400 images from 930 malignant nodules images are included in the testing set. Due to the limited amount of data in the training set and in order to advance the generalization of our model, we employed data argumentation, through simple techniques such as cropping, rotating, and flipping input images as in [29]–[31] to create more data and to increase the randomness and diversity in the training set. The test set is maintained without data augmentation and is used directly in the final test step. To expand the training set, the benign nodules are augmented by translating the nodule patches along the x-axis and y-axis with ±2 pixels, and rotating 90°, 180°, and 270° [14]. The training set of 16000 nodules was then obtained.

B. MULTI-DISCRIMINATOR GENERATIVE ADVERSARIAL NETWORK TRAINING

Unsupervised learning extracts sample features by learning data distribution. We utilize Multi-discriminator Generative Adversarial Network (MDGAN) to complete an unsupervised learning task [19]. The model includes two adversarial models: a generator and a group of discriminators as an ensemble. Among them, the generator produces images according to the input noise of latent space, which comes from the compression of the real images. The distribution of generated images is as close as possible to the distribution of real images. On the other hand, the discriminator group determines whether the input image to be identified comes from the generated images or the real images.

We adopted the improved WGAN-GP network as in [18], which has a faster convergence speed than standard WGAN, and provides a stable GAN training method. It can generate higher quality images with almost no tuning. The generator and discriminator use a standard convolutional decoder in Fig.2 and a convolutional encoder combined with the residual network structure, respectively [19].

The generator network is constructed as follows: firstly the input is 1 × 128 noise and it is input to the fully connected layer. The number of neurons in the fully connected layer is 4 × 4 × 512. The next four network blocks are standard residual blocks. The number of convolution kernels is 512-256-128-64, the size of the convolution kernel is 3 × 3, the step size is set to 1, and then it is input to a convolution layer. ReLU activation function is used in the generator of the GAN model and the final output layer of the generator applies a tanh activation function, afterwards, a final 64 × 64 image is produced.

The construction of a single discriminator network is as follows: The input is a 64 × 64 lung nodule image. After processing by a convolution layer, it is input into four network blocks that are standard residual blocks, in which the number of convolutional kernels is 128-256-512-512. Each convolution kernel is 3 × 3 and the step size is set to 1. Finally, it is input to a fully connected network for classification. The training set and validation set are used to iteratively train the adversarial generation network 50,000 times. The loss function is defined as [15]:

$$\min_G \max_D V(D, G) = E_{x \backsim P_{data}(x)} [log (D(x))] + E_{z \backsim P_z(x)} [log (1 - D(G(z)))] \quad (1)$$

Since the goal of the GAN network model is to learn the data distribution, we prefer a more powerful generator than a more accurate discriminator. In [32], training against a far superior discriminator can impede the generator's learning. This is because the generator is unlikely to generate any samples by the discriminator's standards, and so the generator will receive uniformly negative feedback contained in the gradient. The information will weaken the generator's data distribution. Our goal isn't to present a better approximation $max_D V(D, G)$ of to the generator. We need to soften the max operator. Therefore, we need a soft-discriminator to provide generators with positive feedback to promote generator learning. We consider multi-discriminator variants that attempt to better approximate $max_D V(D, G)$ providing a harsher critic to the generator. The generator trains using feedback aggregated over multiple discriminators (See Figure 1). If $F{:=}max$, G trains against the best discriminator. If $F{:=}mean$, G trains against an ensemble.

We use multiple weak discriminators as an ensemble to combat the generator. Therefore, we reformulate the objective of generator G as $\underset{G}{min} \max F(V(D_1,G), \cdots, V(D_N,G))$, and each $D_i(i \in (1, 2, \cdots N))$ is expected to find the maximum value of $V(D_i, G)$. So we sometimes abbreviate $V(D_i, G)$ to V_i, and $F(V(D_1,G), \cdots, V(D_N,G))$ to $F_G(V_i)$. When $F(\cdot)$ is the mean value, the generator G is against the whole composed of multiple discriminators.

At the beginning of training, when $max_D V(D, G)$ may be too harsh for the generator, we explore a variety of functions that allow us to soften it. We can use soft versions of the classical Pythagorean means [32] parameterized by λ:

$$AM_{soft}(V, \lambda) = \sum_i^N w_i V_i \quad (2)$$

where $w_i = e^{\lambda V_i} / \sum_j e^{\lambda V_j}$ with $\lambda \geq 0$, $V_i < 0$.

And we can set λ closer to 0 to use the mean, increasing the chance of providing positive feedback to the generator. Note that we only require continuity to guarantee that computing the softmax is actually equivalent to computing $V(D, G)$ where D is some convex combination of Di. And the minimax objective function of the generator can be written as:

$$\frac{1}{N} \sum_i^N \mathbb{E}_{x \backsim P_G(x)} [log (1 - D_i(x))] = \frac{1}{N} \mathbb{E}_{x \backsim P_G(x)} [log (z)] \quad (3)$$

where $z = \prod_i^N (1 - D_i(x))$. From Eqn.(3), the generator gradient is $\frac{\partial log(z)}{\partial z}$. It's minimized at $z{=}1$ over $z \in (0, 1]$. From this formula, it is clear that $z{=}1$, if and only if $D_i{=}0 \forall i$. So the generator G only needs to fool a single D_i for receiving positive feedback. This result allows the generator to successfully minimize the original generator objective, $log(1-D)$.

C. MDGAN MODEL GUIDED ENCODER TRAINING

The encoder with the trained generative adversarial network model is combined to construct an anomaly scorer. The encoder is used to extract important information on a real pulmonary nodule image. It maps the real image to latent space, while the generator, similar to decoder, maps from latent space to the image space. In the training process of the encoder, the parameters of the generator are fixed whiles the parameters of the encoder will be optimized.

Encoder training is based on the convolutional autoencoder (AE) architecture [19]. Similar to the discriminator, the first input is a 64 × 64 lung nodule image, which is input to the convolution layer. The number of neurons in the fully connected layer is 5 × 5 × 128. The next four network blocks are a standard residual block. The number of convolutional kernels is 128-256-512-512. Each convolution kernel is 3 × 3, and the step size is set to 1. This then serves as an input to the fully connected layer using tanh as an activation function. The output of a trained adversarial generative network connects to the MDGAN guided encoder as input for training. We used the training set to train the pulmonary nodule benign and malignant scorer network to iterate 100,000 times. During the encoder training with fixed parameters of G and D, only the encoder parameters are adapted. We define a loss function for the mapping of new images to the latent space that comprises two components, an image loss and a discrimination loss. The image loss enforces the visual similarity between real input images and generated images. The discrimination loss enforces the residual on discriminator's features. During training, the loss function is:

$$L(x) = L_{image}(x) + L_D(x) \quad (4)$$

where x is the input image, L_{image} is the residual loss measures the visual dissimilarity between the real image and the generated image. This is defined by:

$$L_{\,image}\,(x) = \frac{1}{n}\,||\text{x-G}\,(E\,(x))||^{\,2} \tag{5}$$

where n is the number of pixels in an image. E is the encoder model, and G is the generator model. The loss function for discriminator $L_D(x)$, is defined by:

$$L_D(x) = \frac{\kappa}{n_d}\,\|f\,(x) - f\,(G\,(E\,(x)))\|^{\,2} \tag{6}$$

where the $f\,(\cdot)$ is the standard deviation function based on feature matching, n_d is the dimensionality of the intermediate feature representation, and κ is a weighting factor. In order to simplify experiment as in [19], we use $\kappa = 1.0$.

The anomaly scorer adapts batch training. Each epoch will optimize the model on the data in batches. The Adam optimization method is used when training the network. When a given training epoch is over, the model will stop training, and the model parameters of each epoch will be saved.

D. LUNG NODULE MALIGNANCY CLASSIFICATION

We used the trained MDGAN model guided encoder to classify the benign and malignant lung nodules in the test set. First, the benign and malignant scores of the pulmonary nodules are calculated, and then the classification results based on the threshold are obtained. The implementation steps are as follows: Firstly, a single unknown image x in the test set is taken as an input into the MDGAN model guided encoder. Encoder extracts the important lung nodule feature z which is the input of the generator. The generator produces a corresponding image $G\,(x)$ based on these features. Since only benign pulmonary nodules are used for training and verification, the generated image $G\,(x)$ is the closest generated image of benign pulmonary nodules mapped by encoder and generator. The test image and the generated image are fed into the discriminator and the benign and malignant scores are calculated. According to [19], the final anomaly score can be expressed as:

$$M\,(x) = \frac{1}{n}\|x - G\,(E(x))\|^{\,2} + \frac{\kappa}{n_d}\|f\,(x) - f\,(G\,(E(x)))\|^{\,2} \tag{7}$$

In general, our model yields high anomaly scores on malignant images and low anomaly scores on benign images. According to the distribution of abnormal scoring results from multiple experiments, the abnormal scores of benign pulmonary nodules are generally between 0 and 0.1, while the abnormal scores of malignant pulmonary nodules are generally scattered between 0.1 and 0.7. The difference in abnormal scores between benign and malignant pulmonary nodules is obvious. Therefore, as long as we give a scoring threshold T, these abnormal scores can be divided into two categories. The abnormal score above the threshold T is abnormal, so this image is determined to be a malignant pulmonary nodule. Conversely, it below the threshold is normal, and this image is determined to be a benign pulmonary nodule. After training, our model can calculate the abnormal score value $M\,(X_i)$ of each test data. The $M\,(X_i)$ belongs to between 0 and 1. We propose the scoring threshold formula:

$$\hat{\tau} = \underset{\tau\in(0,1)}{\arg\max} \sum_{i=1}^{n} |\tau - M\,(X_i)| \tag{8}$$

where X_i is the input test image, $M\,(X_i)$ is the abnormal score value obtained by the model, n is the number of test images, τ is the scoring threshold, and $\hat{\tau}$ is an approximation of the scoring threshold. We solve $\hat{\tau}$ by stepping τ from 0 to 1 with a step size of 0.001. Therefore, $\hat{\tau}$ s the scoring threshold T that we need.

III. EXPERIMENTS

A. DATASET AND DATA PREPROCESSING

The Lung Image Database Consortium (LIDC) is a publicly available dataset, which we used to train and test our proposed methods. The database includes 1018 sets of chest CT image data for 1010 cases, each of which includes an image file (.dcm) and the corresponding diagnosis results label (.xml). In order to improve the credibility of the experiment and reduce the complexity of the algorithm, we use 3 to 30 mm lung nodules marked by four radiologists. The XML file of this kind of nodule contains the characteristic information and the complete outline of the nodule [20]–[22].

According to the characteristics of the LIDC dataset, this paper converts the original CT images (.dcm) with a size of 512×512 pixels into portable network graphics (.png) to facilitate fast training. Because the lung nodule area only accounts for 0.04% to 1.37% of the CT image size [14], the entire CT image is directly used as the input data of the classification model, the too-small target will cause the learning process to be unclear, so segmentation is performed to extract the ROI region that contains only lung nodules. We utilize the improved Threshold Probability Map (TPM) algorithm [12] to segment the lung nodules. The complete CT is used to segment the lung parenchyma to remove irrelevant information and the corresponding XML file is read to get the location information and the benign and malignant degree information of pulmonary nodules, the 64×64 pixels rectangular region is intercepted with the pulmonary nodules. Finally, the lung nodules were classified and stored according to the mean malignancy level (MML) as in [11]. The degree of lung nodule malignancy is divided into 5 grades, of which 5 indicates the highest possibility of malignancy, 3 indicates the degree of benign and malignant disease is uncertain, and 1 indicates the lowest possibility of malignancy. If the nodule's malignancy label is less than 3, it is considered as a benign nodule. If its malignancy label is greater than 3, it is considered as a malignant nodule. Otherwise, it is considered as an uncertain nodule.

Statistically, 2305 lung nodule images were extracted by preprocessing, of which 1375 were benign, 930 were malignant and 1380 were uncertain. All experiments in this paper

TABLE 2. Comparison results of different methods for lung nodule classification. B denotes benign nodule and M denotes malignant nodule.

Method	Dataset	Nodules	ACC(%)	SEN(%)	SPE(%)	AUC
Kumar et al.,2015 [4] (Autoencoder + decision tree)	LIDC-IDRI	4323	75.01	83.35	82.78	0.8290
Shen et al.,2017 [10] (Multi-crop CNN)	LIDC-IDRI	880B+495M	87.14	93.00	77.00	0.9300
Xie et al.,2018 [11] (MV+KBC)	LIDC-IDRI	1301B+644M	91.60	86.52	94.00	0.9575
Shen et al.,2019 [34] (Hierarchical semantic CNN)	LIDC-IDRI	3212B+1040M	90.80	93.10	76.30	0.9300
Proposed MDGAN+Encoder	**LIDC-IDRI**	**1375B**	**95.32**	**94.15**	**90.78**	**0.9426**

were run on a system with Python 2.7, Tensorflow 1.10, CUDA 9.0. and Geforce GTX 1060 GPU.

B. EVALUATION

Common performance evaluation indicators of benign and malignant diagnosis models of lung nodules are: a) Accuracy (ACC) in e.q.(9), the proportion of correctly classified samples among all samples; b) Sensitivity (SEN) in e.q.(10), the proportion of correctly classified malignant nodules Proportion of all true malignant nodules; c) Specificity (SPE) in e.q.(11), the proportion of benign nodules correctly classified to all benign nodules. The larger the values of accuracy ACC, sensitivity SEN, and specificity SPE, the lower the missed and misdiagnosed rate the model has, and the better the classification ability [33].

$$Acc = (TP+TN)\,/\,(TP+TN+FP+FN)\,, \quad 0 \leq Acc \leq 1 \tag{9}$$

$$Sen = TP/\,(TP+TN)\,, \quad 0 \leq Sen \leq 1 \tag{10}$$

$$Spe = TN/\,(TN+FP)\,, \quad 0 \leq Spe \leq 1 \tag{11}$$

The receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) is often used to evaluate the pros and cons of a binary classifier [34], [35]. The ROC curve is a graphical method showing the trade-off between the true positive rate and the false positive rate of the classifier [35], where the x-axis is the false positive rate (or 1-true negative rate), the y-axis is the true positive rate, and each point of the curve corresponds to a model summarized by a certain classifier. A good classification model should be as close as possible to the upper left corner of the ROC curve ($TPR = 1, FPR = 0$); the area under the curve (AUC) provides another method to evaluate the average performance of the classification model. The area under the curve (AUC) of the ROC curve provides another method to evaluate the average performance of the classification model. A good model has an AUC value close to 1.

To objectively evaluate the classification performance of benign and malignant classification systems of pulmonary nodules, we also calculated different AUC values to evaluate different classification methods of benign and malignant pulmonary nodules. The larger the AUC value, the better the classification performance.

C. RESULTS

To verify the effectiveness of this method, we compared the experimental results with the current lung nodule classification models in Table 2. As seen from the table, the current convolutional neural network is still the mainstream of deep learning framework for benign and malignant classification of lung nodules [10], [11], [36]. Convolutional neural networks based on improved feature extraction methods can get good results on the LIDC dataset. However, in reality, the clinical image dataset of pulmonary nodules has the problem of a few data and missing labels. The main problem of convolutional neural networks on small data sets is overfitting, which will lead to weak generalization ability of the model. Our model is more applicable to clinical application scenarios because the cases of benign pulmonary nodules are far more than those of malignancy in clinical practice. From Table 2 we can see that our method has much higher ACC and SEN than other methods. Although SPE and AUC have room for improvement, the main goal of our method is to solve the problem with a new unsupervised method. It's a new way of thinking for lung nodule malignancy classification. We test the testing set of 675 images, which only takes approximately 5 seconds, indicating that our method is competitive.

Our model is based on the combination of multi-discriminators GAN network and encoder. It can indeed effectively learn the image features and distribution of benign pulmonary nodules in order to carry out subsequent classification of benign and malignant pulmonary nodules. To verify its availability, this paper sets up 3 variations of architecture as shown in Fig. 3 for the experiment:

1) MDGAN + Encoder architecture: it combines the Encoder with a GAN network based on multiple weak discriminators and sends the generated images and the real images to the discriminator group for discrimination.
2) GAN+Encoder architecture: it uses the Encoder to extract important information of the real image and compresses it into a compressed representation z, and then sends the compressed representation to the generator. The generator produces images based on the compressed representation. The unique discriminator utilizes the generated images and real images for discrimination.
3) GAN architecture: it uses only GAN networks to learn the image characteristics of benign pulmonary nodules, inputs only random noise to the generator, and then sends the image generated by the generator and the real image to the discriminator for discrimination.

The three experiments are based on the same training set and training parameters. After data argumentation,training

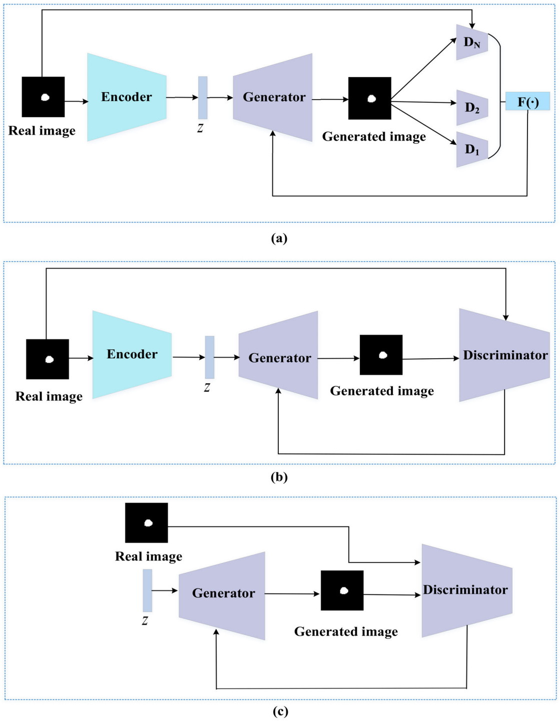

FIGURE 3. The architecture of MDGAN encoder model and self-variants contrast model: (a) MDGAN + encoder architecture, (b) GAN + encoder architecture and (c) GAN architecture.

TABLE 3. Comparison results of different self-variants contrast methods of lung nodule classification.

Method	Dataset	Nodules	ACC(%)	SEN(%)	SPE(%)	AUC(%)
MDGAN + Encoder	**LIDC-IDRI**	**1375B**	**95.32**	**94.15**	**90.79**	**94.26**
GAN+Encoder	LIDC-IDRI	1375B	90.19	86.21	90.32	86.43
GAN	LIDC-IDRI	1375B	87.61	71.83	88.97	78.95

set=16000, testing set=775, batch size=64, epoch=10, learning rate=0.001.

GAN training yields a generator only samples from the normal distribution and maps from images to latent space. However, no coding relationship between real image space and latent space has been established. To establish the coding relationship, we transform the real image received by the encoder into an efficient internal representation. Latent space noise can make the generated image approximately follow the distribution of the real images. It can be seen from the results that after the combination of GAN and Encoder, the classification ability has been significantly improved in Fig.4.

To verify whether the multi-discriminator GAN network model is more conducive to the generator learning data distribution, thereby improving the performance of benign and malignant classification of pulmonary nodules. We set the number of discriminators of the MDGAN model to 1, 3, 5, 7 and 9. Under the same experimental environment, training data, training parameters and test data, the observed experimental results can be seen in Table 4:

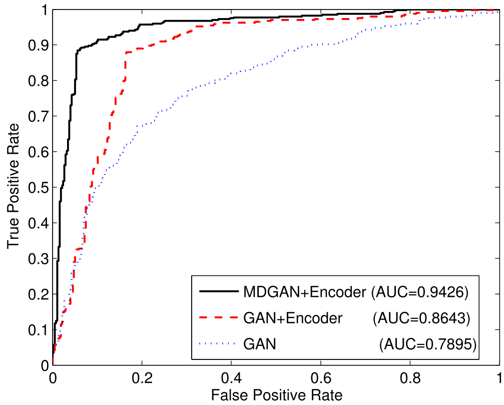

FIGURE 4. ROC curves of the proposed MDGAN + encoder model and self-variants contrast model.

TABLE 4. Comparison results of the different number of discriminators of the MDGAN mode.

Number of Discriminators	ACC(%)	SEN(%)	SPE(%)	Running Time
N=1	91.59	87.91	90.63	13h28m
N=3	92.34	90.97	92.93	10h09m
N=5	**95.32**	**94.15**	**90.79**	**9h31m**
N=7	83.26	89.35	87.63	14h10m
N=9	80.39	71.57	80.71	26h48m

The experimental results show that the increase in the number of discriminators helps to accelerate the convergence of the objective function of the GAN network to a stable state. The convergence speed of the number of discriminators $N = 5$ is about twice that of $N = 1$. But an excessive number of discriminators increase model complexity without improving efficiency.

The problems of unstable training, vanishing gradient, and mode collapse have been widely recognized in previous work with GANs. We use the improved WGAN training procedure [18] for stable GAN training. The improved WGAN solves the problem of unstable training by estimates the Wasserstein distance between the generator and the real data distribution and replacing weight clipping by gradient penalty. Furthermore, we analyzed whether the learned latent representation is smooth [19]. If there were only a few locations in latent space during GAN training, which could allow the generator to generate realistic images, it indicated that the unstable GAN training led to mode collapse. We tested by selecting random z positions in latent space and generating image series from sampling points. According to the experiment, we find that when z changes, the corresponding generated image $G(z)$ also changes continuously. The smooth transitions in the image series indicated that there was no mode collapse.

IV. CONCLUSION

With the issue of inadequate labeled medical image datasets and the cost involved in obtaining labeled sets in clinical, we propose an unsupervised multi-discriminator generative adversarial network combined with an encoder for benign and malignant classification of lung nodules. As far as we know, this is the first time that unsupervised learning has been successfully applied to the classification of benign and malignant lung nodules. The experimental results on the LIDC dataset reveal that compared with other supervised deep learning methods, our proposed approach can achieve better classification results using only unlabeled benign lung nodule images for training. In future work, we will attempt to apply this model to other small datasets or partially labeled datasets. This method also can be extended to pulmonary nodule detection or other disease anomaly detection to reduce the demand for labeled data.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: A Cancer J. for Clinicians*, vol. 65, no. 1, pp. 5–29, Jan. 2015.

[2] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. S. Tavares, "Automatic 3D pulmonary nodule detection in CT images: A survey," *Comput. Methods Programs Biomed.*, vol. 124, pp. 91–107, Feb. 2016.

[3] G. X. Wu and D. J. Raz, "Lung cancer screening," in *Lung Cancer*. Cham, Switzerland: Springer, 2016, pp. 1–23.

[4] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. 12th Conf. Comput. Robot Vis.*, Jun. 2015, pp. 133–138.

[5] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2015, pp. 588–599.

[6] P. P. R. Filho, R. M. Sarmento, G. B. Holanda, and D. de Alencar Lima, "New approach to detect and classify stroke in skull CT images via analysis of brain tissue densities," *Comput. Methods Programs Biomed.*, vol. 148, pp. 27–43, Sep. 2017.

[7] P. P. R. Filho, E. D. S. Reboucas, L. B. Marinho, R. M. Sarmento, J. M. R. S. Tavares, and V. H. C. de Albuquerque, "Analysis of human tissue densities: A new approach to extract features from medical images," *Pattern Recognit. Lett.*, vol. 94, pp. 211–218, Jul. 2017.

[8] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 559–567.

[9] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets Therapy*, vol. 2015, pp. 2015–2022, Aug. 2015.

[10] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017.

[11] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019.

[12] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, "DeepEM: Deep 3D convnets with em for weakly supervised pulmonary nodule detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 812–820.

[13] X. Feng, J. Yang, A. F. Laine, and E. D. Angelini, "Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 568–576.

[14] T. Zhang, J. Zhao, J. Luo, and Y. Qiang, "Deep belief network for lung nodules diagnosed in CT imaging," *Int. J. Performability Eng.*, vol. 13, no. 8, pp. 1358–1370, 2017.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[17] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

[19] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[20] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, and B. Zhao, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Jan. 2011.

[21] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[22] C. Deng, X. Liu, C. Li, and D. Tao, "Active multi-kernel domain adaptation for hyperspectral image classification," *Pattern Recognit.*, vol. 77, pp. 306–315, May 2018.

[23] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *J. Digit. Imag.*, vol. 28, no. 1, pp. 99–115, Feb. 2015.

[24] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung CT images," *J. Digit. Imag.*, vol. 29, no. 4, pp. 466–475, Aug. 2016.

[25] S. Hussein, R. Gillies, K. Cao, Q. Song, and U. Bagci, "TumorNet: Lung nodule characterization using multi-view convolutional neural network with Gaussian process," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 1007–1010.

[26] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sanchez, and B. van Ginneken, "Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.

[27] F. Han, G. Zhang, H. Wang, B. Song, H. Lu, D. Zhao, H. Zhao, and Z. Liang, "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng.*, Oct. 2013, pp. 14–18.

[28] C. Deng, E. Yang, T. Liu, J. Li, W. Liu, and D. Tao, "Unsupervised semantic-preserving adversarial hashing for image search," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4032–4044, Aug. 2019.

[29] H. R. Roth, J. Yao, L. Lu, J. Stieger, J. E. Burns, and R. M. Summers, "A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Computer-Assisted Intervention–(MICCAI)*. Cham, Switzerland: Springer, 2014, pp. 520–527.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[31] H. R. Roth, J. Yao, L. Lu, J. Stieger, J. E. Burns, and R. M. Summers, "Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*. Toulon, France: Springer, 2017, pp. 3–12.

[32] I. Durugkar, I. Gemp, and S. Mahadevan, "Generative multi-adversarial networks," 2016, *arXiv:1611.01673*. [Online]. Available: http://arxiv.org/abs/1611.01673

[33] P.-L. Lin, P.-W. Huang, C.-H. Lee, and M.-T. Wu, "Automatic classification for solitary pulmonary nodule in CT image by fractal analysis based on fractional brownian motion model," *Pattern Recognit.*, vol. 46, no. 12, pp. 3279–3287, Dec. 2013.

[34] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[35] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.

[36] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Syst. Appl.*, vol. 128, pp. 84–95, Aug. 2019.

YAN KUANG received the B.Sc. degree in software engineering from the University of Electronic Science and Technology of China (UESTC), in 2017, where she is currently pursuing the M.S. degree in software engineering. She is involved in research in the area of classification of pulmonary nodules through techniques of digital image processing and machine learning.

TIAN LAN (Member, IEEE) received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009. He is currently an Associate Professor with the School of Information and Software Engineering, UESTC. His current research interests include medical image processing, speech enhancement, and natural language processing.

XUEQIAO PENG is currently pursuing the B.Sc. degree in software engineering with the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. Her current research interests include medical image processing and analysis, machine learning, and deep learning.

GATI ELVIS SELASI received the B.Sc. degree in computer science from the University for Development Studies Tamale at Navrongo Campus, Ghana, in 2017. He is currently pursuing the M.S. degree with the School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include medical image processing and analysis, machine learning, deep learning, and data science.

QIAO LIU (Member, IEEE) received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010. He is currently a Full Professor with the School of Information and Software Engineering, UESTC. His current research interests include natural language processing, machine learning, and data mining.

JUNYI ZHANG received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, in 2012. His current research interests include digital signal processing, machine learning, and data mining.