



Application of Deep Convolutional Neural Networks for Discriminating Benign, Borderline, and Malignant Serous Ovarian Tumors From Ultrasound Images

OPEN ACCESS

Edited by:

Harini Veeraraghavan,
Memorial Sloan Kettering Cancer
Center, United States

Reviewed by:

Abdullah Nazib,
Memorial Sloan Kettering Cancer
Center, United States
Qiang Wang,
University of Edinburgh,
United Kingdom

*Correspondence:

Lisha Qi
qilisha2005@163.com
Wenjuan Ma
mawenjuan@tmu.edu.cn

†These authors have contributed
equally to this work and
share first authorship

Specialty section:

This article was submitted to
Cancer Imaging and
Image-directed Interventions,
a section of the journal
Frontiers in Oncology

Received: 04 September 2021

Accepted: 29 November 2021

Published: 20 December 2021

Citation:

Wang H, Liu C, Zhao Z, Zhang C,
Wang X, Li H, Wu H, Liu X, Li C, Qi L
and Ma W (2021) Application of Deep
Convolutional Neural Networks for
Discriminating Benign, Borderline, and
Malignant Serous Ovarian Tumors
From Ultrasound Images.
Front. Oncol. 11:770683.
doi: 10.3389/fonc.2021.770683

Huiquan Wang^{1†}, Chunli Liu^{1†}, Zhe Zhao¹, Chao Zhang^{2,3}, Xin Wang⁴, Huiyang Li³,
Haixiao Wu^{2,3}, Xiaofeng Liu², Chunxiang Li², Lisha Qi^{2*} and Wenjuan Ma^{2,3*}

¹ School of Electrical and Electronic Engineering, TianGong University, Tianjin, China, ² Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China, ³ The Sino-Russian Joint Research Center for Bone Metastasis in Malignant Tumor, Tianjin, China, ⁴ Department of Epidemiology and Biostatistics, West China School of Public Health, Sichuan University, Chengdu, China

Objective: This study aimed to evaluate the performance of the deep convolutional neural network (DCNN) to discriminate between benign, borderline, and malignant serous ovarian tumors (SOTs) on ultrasound(US) images.

Material and Methods: This retrospective study included 279 pathology-confirmed SOTs US images from 265 patients from March 2013 to December 2016. Two- and three-class classification task based on US images were proposed to classify benign, borderline, and malignant SOTs using a DCNN. The 2-class classification task was divided into two subtasks: benign vs. borderline & malignant (task A), borderline vs. malignant (task B). Five DCNN architectures, namely VGG16, GoogLeNet, ResNet34, MobileNet, and DenseNet, were trained and model performance before and after transfer learning was tested. Model performance was analyzed using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC).

Results: The best overall performance was for the ResNet34 model, which also achieved the better performance after transfer learning. When classifying benign and non-benign tumors, the AUC was 0.96, the sensitivity was 0.91, and the specificity was 0.91. When predicting malignancy and borderline tumors, the AUC was 0.91, the sensitivity was 0.98, and the specificity was 0.74. The model had an overall accuracy of 0.75 for in directly classifying the three categories of benign, malignant and borderline SOTs, and a sensitivity of 0.89 for malignant, which was better than the overall diagnostic accuracy of 0.67 and sensitivity of 0.75 for malignant of the senior ultrasonographer.

Conclusion: DCNN model analysis of US images can provide complementary clinical diagnostic information and is thus a promising technique for effective differentiation of benign, borderline, and malignant SOTs.

Keywords: deep convolutional neural network, deep learning, ultrasound, serous ovarian tumor, transfer learning

INTRODUCTION

Serous ovarian tumors comprise benign, borderline, and malignant lesions which have distinct clinicopathological characteristics, therapeutic schemes, and prognoses (1, 2). Accurate identification of SOTs prior to operation is critical to the development of appropriate treatments to avoid inadequate excision or surgical overtreatment (3). Pathology results revealed by fine-needle aspiration cytology are considered the gold standard for the diagnosis of SOTs before surgery or neoadjuvant chemotherapy. However, this method cannot be used to precisely identify the histological types of SOTs because of inadequate cytologic samples and the heterogeneous nature of the tissue composition (4).

Adnexal ultrasound is currently the first-line imaging modality for diagnosis of SOTs (5, 6). Although US imaging cannot replace biopsies, it can provide additional information that biopsies cannot deliver, such as intra-tumor heterogeneity. Diagnostic analysis of US images depends mainly on the physician's expertise. However, in recent years, the emergence of artificial intelligence has brought hope for more objective and accurate diagnosis (7). Sakshi et al. (8) used a fine-tuned VGG-16 deep learning network in order to detect whether an ovarian cyst is present or not. Wu et al. (9) explored deep learning approaches for ovarian tumor classification based on ultrasound images. Zhang et al. (10) used an image diagnosis system for classifying the ovarian cysts in color ultrasound images. The application of deep convolutional neural network in medical image diagnosis has become a hot research topic (11–14). As of May 2020, more than 50 deep learning-based imaging applications have been approved by the US Food and Drug Administration or European Union, spanning most imaging modalities including X-ray, computerized tomography, magnetic resonance imaging, retinal optical coherence tomography, and ultrasound (15–19).

In the present study, we conducted 2- and 3-class classification task (**Figure 1**) using US images and deep learning methods to classify benign, borderline, and malignant SOTs. The results were compared with a senior sonographers with extensive diagnostic experience.

MATERIAL AND METHODS

Dataset

This study was approved by the Institute Review Board of Tianjin Medical University Cancer Hospital Institutional. Due to its retrospective nature, the informed consent requirement was waived.

A data set of US images in patients with SOTs from Tianjin Medical University Cancer Institute and Hospital(412 imaging

studies, 265 patients) was collected. As shown in **Figure 1**, The patient inclusion criteria were as follows: (1) a histologic diagnosis of benign, borderline, or malignant SOTs between March 2013 and December 2016 (**Figure 2**); (2) availability of diagnostic-quality preoperative US images; and (3) US scanning before neoadjuvant therapy or surgical resection. The exclusion criteria were: (1) no ultrasound results or the ovarian mass was not completely in the images(10 images); and (2) mucinous(37 images), clear cell(32 images), endometrioid(25 images), or metastatic cancer(29 images). The included cases(279 images) were randomly assigned to either the training set (70%) or the validation set (30%).

US imaging was performed using equipment manufactured by Philips (EPIQ5, EPIQ7 and IU22), Samsung (RS80A), and GE Healthcare (LOGIQ E9, LOGIQ S7). The images were collected by transabdominal examination according to standard protocols and analyzed by ultrasound specialists.

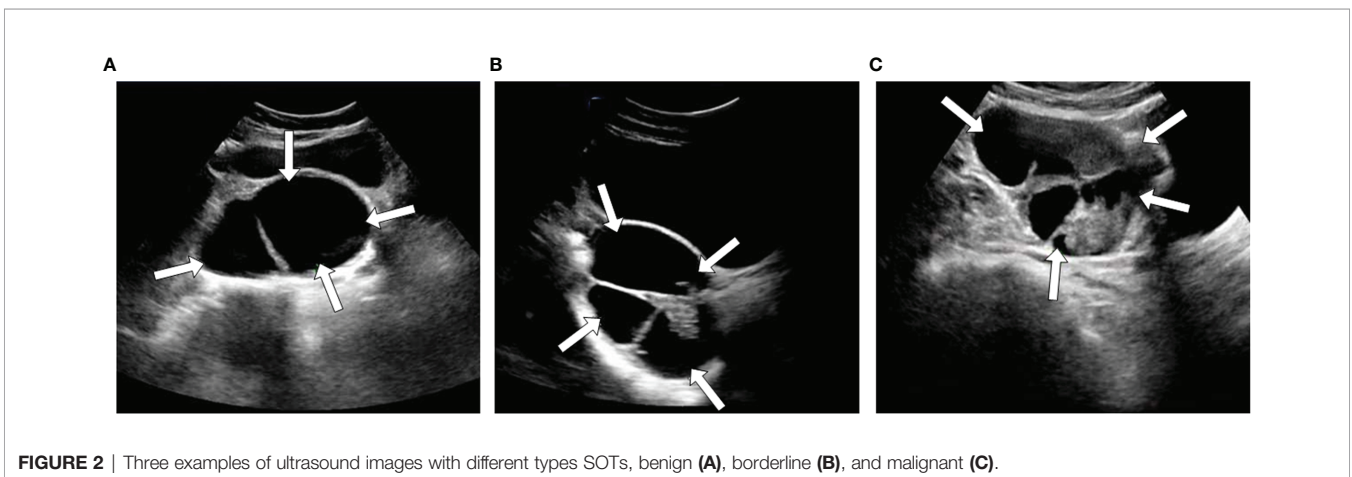
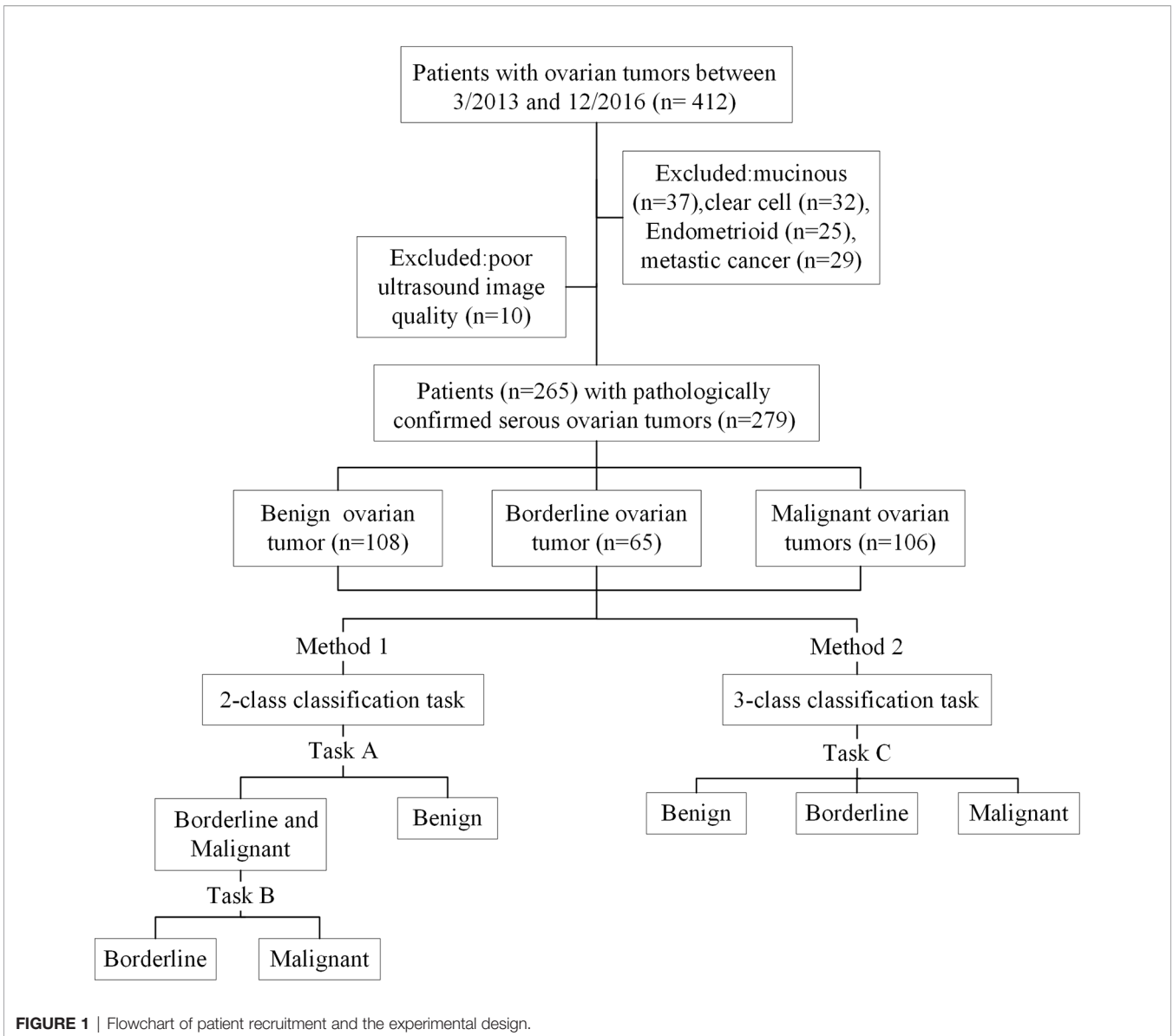
Data Preprocessing

All US images were retrieved from the Picture Archiving and Communication Systems for image segmentation and analysis in the hospital. Lesions were segmented using Image J software (<https://imagej.nih.gov/ij/>) by a sonographer with more than eight years of experience. All images were pre-processed: Due to the limited amount of training data, we used data augmentation techniques for image processing in order to avoid overfitting. The original images were randomly cropped and flipped using data enhancement technology, and all augmented images were resized to 224 * 224 pixels for input to the DCNN model. All pre-processing steps were conducted in Python (version 3.7.3; Python Software Foundation, Wilmington, Del) using the transforms imported from Torchvision (version 0.7.0).

DCNN Model Training and Interpretation

At present, DCNNs are the most well-known type of deep learning architecture in the field of medical image analysis. Given their advantages, five representative DCNN architectures, namely VGG, GoogLeNet, ResNet, MobileNet, and DenseNet (20–24), were used to identify the histological types of SOTs based on US images. In the 2-class classification task, two tasks were trained and validated by the DCNN. The cohort with three classes was split into two-class datasets, and each sub-dataset was then evaluated by two-category classification (**Figure 1**). This yielded the following sub-datasets: benign vs. borderline & malignant (task A) and borderline vs. Malignant (task B). The 3-class classification task (task C) was used to directly identify benign, borderline, and malignant SOTs using the three-category classification DCNN.

During the training phase, the dropout strategy on the fully connected layers with a probability of 0.5 and L2 regularization



strategy on weight and bias were used to prevent the overfitting problem, the initial learning rate was set to 0.0003 with a batch size of 32 and the Adam optimizer was used to update the weights of the neural network. All models were trained for 500 epochs; the learning rate is decayed every 20 epochs with a decay rate of 0.9. An Intel I7-9700K CPU and Nvidia GeForce RTX 2080 GPU were used for models training. In addition, we used the transfer learning methods of these networks for classification in order to compare the diagnostic efficiency of different methods (25–27). The weights of each network were initialized according to the weights from the pretrained model on ImageNet (28). We then fine-tuned the parameters of the fully connected layer of the network on our dataset *via* back propagation. The standard DCNN used comes from the Torchvision (version 0.7.0) package included in the PyTorch software framework. All programs were run in Python version 3.7.3.

To improve the interpretability of our model, we used the method of Class Activation Mapping (CAM) to visualize the important regions leading to the decision of the deep learning model (29). Such a localization map is completely generated by the fully trained network without additional manual annotation. By using a global average pooling layer and visualizing the weighted combination of the resulting feature maps at the penultimate (pre-softmax) layer, we obtained heat maps that explained what parts of an input US images were focused by the DCNN for assigning a diagnostic label. All heat maps were produced using the package OpenCV (version 4.3.0.36).

Statistical Analysis

In the present study, we carried out 3-fold cross validation on different DCNN models, obtains the experimental results and

calculates their indicators. The performance of the 2-class classification task was evaluated by the area under the receiver operating characteristic curve, accuracy, sensitivity, specificity, and F1 score. The performance of the 3-class classification task was evaluated by accuracy, sensitivity, and specificity only. Differences between the AUC values were considered statistically significant when $P < 0.05$. The method described by Hanley and McNeil was used to calculate the 95% confidence interval (CI) of the AUC values (30). These measurements were calculated using the numpy (version 1.16.2) Python library.

RESULTS

Patient Characteristics

A total of 265 patients (median age 51 years, range 15–79 years) with 279 ovarian tumors (108 benign ovarian tumors, 65 borderline ovarian tumors, 106 malignant ovarian tumors), were retrospectively enrolled in this study.

Performance of the Two-Class Classification Task

Table 1 show the performance of the 2-class classification DCNN model using transfer learning and the full training methods on the validation sets. The AUC values of all models for the classification tasks were within the range of 0.827–0.963. In general, the transfer learning methods trained to distinguish benign from non-benign or borderline from malignant SOTs appear to perform better than the full training methods.

When using migration learning methods for SOT classification in the validation sets of Task A (**Figure 3A**),

TABLE 1 | Performance of the two-class classification deep convolutional neural network models in the validation set.

		AUC (± SD)	ACC (± SD)	SEN (± SD)	SPEC (± SD)	F1-Score (± SD)
Transfer learning						
Task A	VGG16	0.897 (± 0.016)	0.871 (± 0.018)	0.931 (± 0.015)	0.771 (± 0.023)	0.843 (± 0.019)
	GoogLeNet	0.924 (± 0.017)	0.883 (± 0.019)	0.828 (± 0.020)	0.972 (± 0.017)	0.894 (± 0.019)
	ResNet34	0.963 (± 0.016)	0.914 (± 0.017)	0.914 (± 0.015)	0.914 (± 0.018)	0.914 (± 0.017)
	MobileNet	0.885 (± 0.018)	0.871 (± 0.018)	0.931 (± 0.015)	0.771 (± 0.021)	0.843 (± 0.019)
	DenseNet	0.877 (± 0.019)	0.871 (± 0.021)	0.983 (± 0.016)	0.686 (± 0.022)	0.808 (± 0.019)
Task B	VGG16	0.865 (± 0.018)	0.882 (± 0.019)	0.944 (± 0.015)	0.714 (± 0.024)	0.813 (± 0.021)
	GoogLeNet	0.896 (± 0.018)	0.860 (± 0.019)	0.917 (± 0.016)	0.762 (± 0.023)	0.832 (± 0.020)
	ResNet34	0.914 (± 0.017)	0.893 (± 0.019)	0.983 (± 0.016)	0.743 (± 0.020)	0.846 (± 0.019)
	MobileNet	0.907 (± 0.017)	0.842 (± 0.021)	0.889 (± 0.019)	0.762 (± 0.021)	0.821 (± 0.020)
	DenseNet	0.898 (± 0.017)	0.825 (± 0.022)	0.806 (± 0.022)	0.857 (± 0.019)	0.831 (± 0.021)
Full training						
Task A	VGG16	0.886 (± 0.018)	0.839 (± 0.020)	0.931 (± 0.016)	0.686 (± 0.023)	0.790 (± 0.020)
	GoogLeNet	0.914 (± 0.017)	0.872 (± 0.019)	0.845 (± 0.022)	0.917 (± 0.017)	0.880 (± 0.019)
	ResNet34	0.909 (± 0.017)	0.893 (± 0.018)	0.966 (± 0.016)	0.771 (± 0.021)	0.858 (± 0.019)
	MobileNet	0.870 (± 0.018)	0.850 (± 0.019)	0.948 (± 0.016)	0.686 (± 0.024)	0.796 (± 0.021)
	DenseNet	0.900 (± 0.018)	0.850 (± 0.020)	0.966 (± 0.017)	0.657 (± 0.025)	0.782 (± 0.021)
Task B	VGG16	0.827 (± 0.021)	0.842 (± 0.021)	0.972 (± 0.016)	0.619 (± 0.025)	0.756 (± 0.022)
	GoogLeNet	0.843 (± 0.020)	0.842 (± 0.021)	0.999 (± 0.016)	0.571 (± 0.024)	0.727 (± 0.021)
	ResNet34	0.905 (± 0.018)	0.882 (± 0.019)	0.983 (± 0.015)	0.714 (± 0.023)	0.827 (± 0.020)
	MobileNet	0.845 (± 0.019)	0.825 (± 0.020)	0.917 (± 0.017)	0.667 (± 0.024)	0.772 (± 0.021)
	DenseNet	0.886 (± 0.018)	0.807 (± 0.020)	0.999 (± 0.017)	0.476 (± 0.025)	0.645 (± 0.023)

Task A: discriminating benign vs. borderline & malignant. Task B: discriminating borderline vs. Malignant.

AUC, area under the receiver operating characteristic curve; ACC, accuracy; SEN, sensitivity; SPEC, specificity; SD, standard deviation.

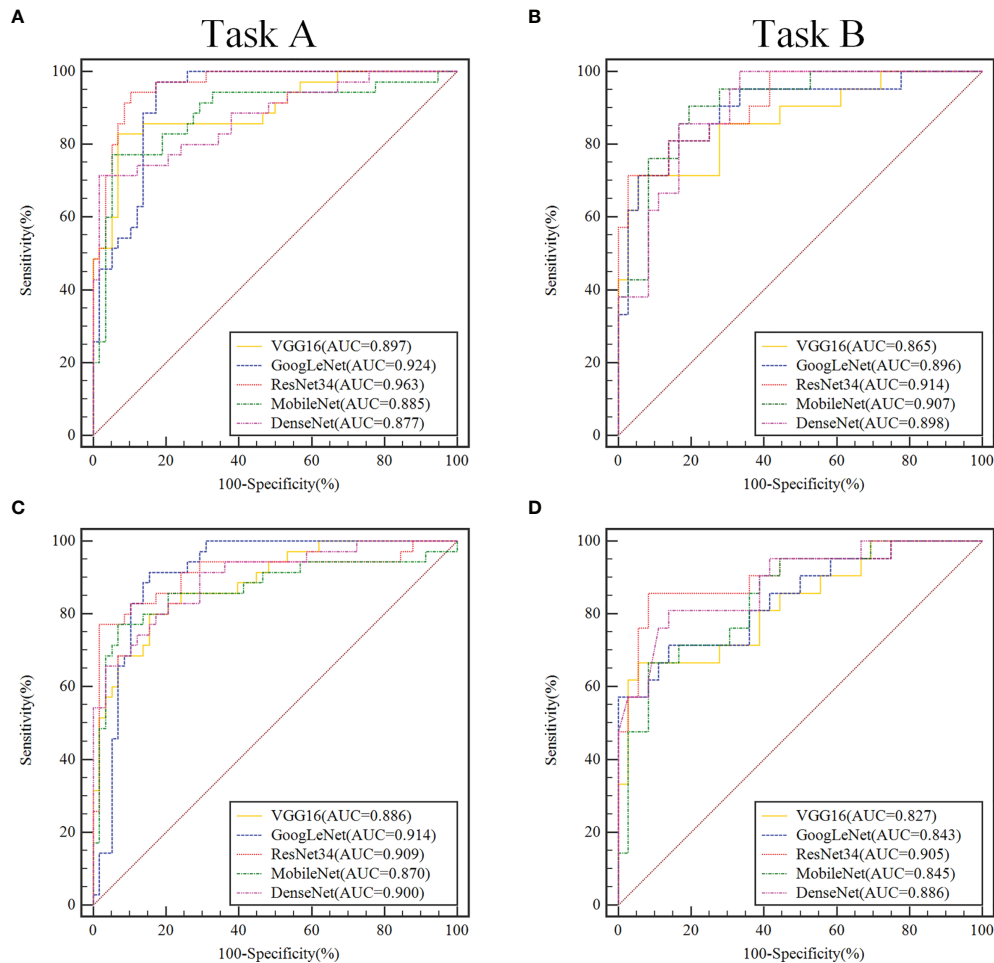


FIGURE 3 | In the validation set, ROC curve analysis of two classification tasks with different convolutional neural network models before and after transfer learning. Task A (**A, C**) discriminating benign vs. borderline & malignant, Task B (**B, D**) discriminating borderline vs. malignant. In the convolutional neural network model, the models that use transfer learning are (**A, B**), and the fully trained models are (**C, D**).

the AUC values of VGG16, GoogLeNet, ResNet34, MobileNet and DenseNet models were more than $0.877 (\pm 0.019)$. When the full training method was used for differential diagnosis (**Figure 3C**), the AUC values of the above models were over than $0.870 (\pm 0.018)$. When using migration learning methods for SOT classification in the validation sets of Task B (**Figure 3B**), the AUC values of five DCNN models were more than $0.865 (\pm 0.018)$. When the full training method was used for differential diagnosis (**Figure 3D**), the AUC values of the above models were over than $0.843 (\pm 0.020)$.

In the validation sets, the ResNet34 models with the transfer learning method performs more comprehensively and better than other models in the 2-class classification task. When classifying benign and non-benign tumors, the AUC was $0.963 (\pm 0.016)$, the sensitivity was $0.914 (\pm 0.015)$, and the specificity was $0.914 (\pm 0.018)$, and the F1-Score was $0.914 (\pm 0.017)$. When predicting malignancy and borderline tumors, the AUC was $0.914 (\pm 0.017)$, the sensitivity was $0.983 (\pm 0.016)$,

and the specificity was $0.743 (\pm 0.020)$, and the F1-Score was $0.846 (\pm 0.019)$.

Performance of the Three-Class Classification Task

Table 2 shows the performance of the 3-class classification DCNN model using transfer learning and the full training methods on the validation sets. The ResNet34 model using transfer learning methods exhibit higher discrimination performance than sonographer with 12 years of experience, with ACC values of $0.753 (\pm 0.019)$ and the sensitivity of this model to malignant tumors reached $0.889 (\pm 0.017)$. The sonographer's overall diagnostic accuracy based on the semantic features of the images for differentiating benign, borderline, and malignant SOTs was $0.667 (\pm 0.021)$ and the sensitivity to malignant tumors reached $0.750 (\pm 0.018)$. Confusion matrix for 3-class classification DCNN models are described in **Supplementary Figure 2**.

TABLE 2 | Performance of the three-class classification deep convolutional neural network models and the senior sonographer in the validation set.

		Class	SEN (± SD)	SPEC (± SD)	ACC (± SD)
Transfer learning					
Task C	VGG16	Class 0	0.750 (± 0.022)	0.825 (± 0.018)	0.699 (± 0.021)
		Class 1	0.409 (± 0.025)	0.873 (± 0.017)	
		Class 2	0.829 (± 0.020)	0.845 (± 0.017)	
	GoogLeNet	Class 0	0.889 (± 0.016)	0.772 (± 0.021)	0.720 (± 0.020)
		Class 1	0.455 (± 0.024)	0.944 (± 0.016)	
		Class 2	0.800 (± 0.020)	0.897 (± 0.017)	
	ResNet34	Class 0	0.889 (± 0.017)	0.754 (± 0.022)	0.753 (± 0.019)
		Class 1	0.455 (± 0.024)	0.958 (± 0.015)	
		Class 2	0.800 (± 0.019)	0.897 (± 0.017)	
	MobileNet	Class 0	0.861 (± 0.018)	0.772 (± 0.021)	0.720 (± 0.020)
		Class 1	0.500 (± 0.023)	0.873 (± 0.017)	
		Class 2	0.714 (± 0.022)	0.931 (± 0.016)	
	DenseNet	Class 0	0.917 (± 0.017)	0.772 (± 0.021)	0.699 (± 0.023)
		Class 1	0.191 (± 0.027)	0.986 (± 0.019)	
		Class 2	0.857 (± 0.019)	0.759 (± 0.021)	
Full training					
Task C	VGG16	Class 0	0.917 (± 0.017)	0.632 (± 0.022)	0.667 (± 0.020)
		Class 1	0.917 (± 0.018)	0.772 (± 0.020)	
		Class 2	0.771 (± 0.019)	0.845 (± 0.021)	
	GoogLeNet	Class 0	0.944 (± 0.016)	0.684 (± 0.020)	0.710 (± 0.021)
		Class 1	0.500 (± 0.024)	0.901 (± 0.018)	
		Class 2	0.600 (± 0.023)	0.966 (± 0.018)	
	ResNet34	Class 0	0.778 (± 0.017)	0.825 (± 0.021)	0.710 (± 0.020)
		Class 1	0.500 (± 0.023)	0.845 (± 0.018)	
		Class 2	0.771 (± 0.017)	0.897 (± 0.022)	
	MobileNet	Class 0	0.861 (± 0.018)	0.719 (± 0.021)	0.688 (± 0.022)
		Class 1	0.409 (± 0.023)	0.916 (± 0.020)	
		Class 2	0.686 (± 0.021)	0.879 (± 0.020)	
	DenseNet	Class 0	0.889 (± 0.018)	0.737 (± 0.022)	0.677 (± 0.023)
		Class 1	0.318 (± 0.025)	0.901 (± 0.019)	
		Class 2	0.686 (± 0.020)	0.862 (± 0.020)	
Doctor diagnoses					
Task C	Doctor	Class 0	0.750 (± 0.018)	0.825 (± 0.022)	0.667 (± 0.021)
		Class 1	0.474 (± 0.023)	0.851 (± 0.019)	
		Class 2	0.684 (± 0.019)	0.818 (± 0.022)	

Task C: discriminating benign vs. borderline vs. malignant tumors.

Class 0: malignant tumors; Class 1: borderline tumors; Class 2: benign tumors.

ACC, accuracy; SEN, sensitivity; SPEC, specificity; SD, standard deviation.

Visualizing and Understanding DCNN

As shown in the heatmaps produced using the means of the CAM results (**Figure 4**), Such a localization map is completely generated by the fully trained Resnet34 model without additional manual annotation. The red and yellow regions represent areas activated by the Resnet34 model and have the greatest predictive significance; the blue and green backgrounds reflect areas with weaker predictive values. The redder the feature color, the greater the possibility of the high DCNN score (**Figure 4A** DCNN score is 0.99, **Figure 4B** DCNN score is 0.71). We compared these findings with clinicians' justifications. For the correctly diagnosed images (**Figures 4A–C**), the DCNN focused on the same areas as the clinicians. However, there were also some images that both the clinicians and DCNNs incorrectly diagnosed (**Figures 4D, E**). We also compared the areas of interest for the Advanced Sonographer and Resnet34, as detailed in the **Supplementary Figure 5**.

DISCUSSION

Predicting the specific histopathological condition from US images of patients with SOTs can help avoid unnecessary surgery on physiological hemorrhagic cysts. In this study using US images and a DCNN, we used 2- and 3-class classification task to distinguish benign, borderline, and malignant SOTs: 2-class classification task involved two task, task A distinguished benign vs. borderline vs. malignant tumors, and task B differentiated borderline vs. malignant tumors; 3-class classification task was a direct three-category classification. The results showed that the performance of the 2-class classification task was better than that of the 3-class classification task, although the process was more complicated. The use of transfer learning further improved the performance of the DCNN models.

In the present study, we also examined the capabilities of fine-tuning in comparison with a training scheme from scratch in the

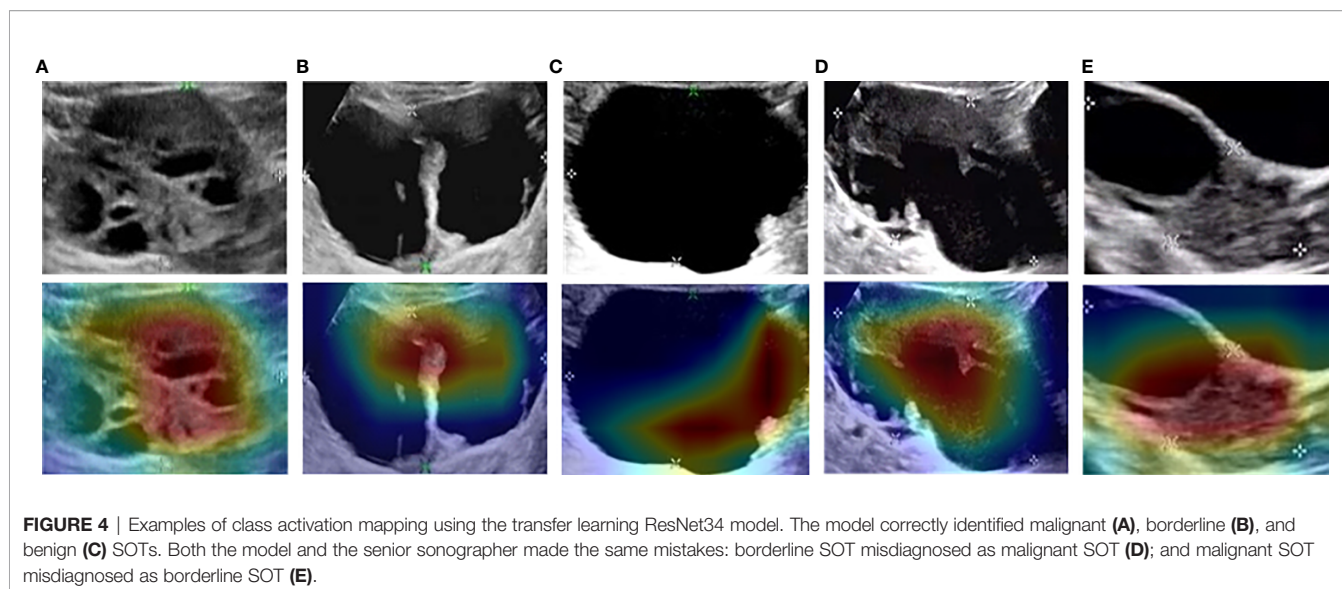


TABLE 3 | The training time of transfer learning and full training method with epoch=500.

		VGG16	GoogLeNet	ResNet34	MobileNet	DenseNet
Transfer learning						
Method1	Step1	890	650	428	385	684
	Step2	584	431	281	259	684
Method2	Step1	912	654	436	387	690
Full training						
Method1	Step1	1791	626	707	567	1213
	Step2	1125	626	465	376	784
Method2	Step1	1776	638	720	581	1216

Unit of time: second.

context of ultrasonic medical image analysis. Our experiments, which were based on five DCNN architectures, demonstrated that transfer learning DCNNs are useful for ultrasonic medical image analysis, performing as well as – and sometimes even outperforming – fully trained DCNNs. Another advantage of fine-tuned DCNNs is the training speed. To demonstrate this advantage, we compared the training times for a fine-tuned DCNN and a DCNN trained from scratch in **Table 3**. However, our results showed that transfer learning using a pre-trained model on ImageNet does not significantly improve the training of all DCNN models. Probably for this reason, ImageNet classes are images of animal, plants, and objects. Transfer learning may be better used for ultrasound image analysis if pre-trained models on medical image dataset are available in the future.

To improve convincingness in our DCNN, we used “interpretable” method(CAM) that explains why they predict what they do. We also provided some heatmaps of the CAMs output, with visual examples of the explanation maps presented in **Figure 4**. Most SOTs are cystic-solid tumors. When identifying malignancy, doctors typically pay attention to the amount of solid area: the more solid components present in the US images, the higher the degree of malignancy. **Figures 3A, B** showed that the DCNN focuses on the solid components of the

tumor, while **Figures 3D, E** showed cases in which both the DCNN and clinician made incorrect decisions in diagnosis. However, the imaging features of borderline and malignant SOTs greatly overlap. Therefore, it can be very difficult to distinguish between these two types of images. In future work, we plan to increase the training of cases to improve the accuracy of diagnosis. Furthermore, the prior knowledge of senior sonographers could also be integrated into the learning process.

Our research is subject to several limitations. Firstly, this was a retrospective study from a single center with a limited cohort size. External multi-center validation in a larger cohort is necessary to perform a better analysis. Secondly, since this was a retrospective study, there were cross marks left by the sonologist who made the diagnosis that could not be removed from the images. However, it can be seen from the CAM images that the DCNN did not pay attention to these cross marks, and they thus did not affect diagnosis (**Figure 3**). Thirdly, the performance reported for the different models may not be the best performance that can be achieved for each task, as performance is related to the hyper-parameters of DCNNs that influence the speed of convergence and final accuracy of the model. Identifying the optimal values for these hyper-parameters is rather difficult because each DCNN is a time-consuming

process, even using high-end GPUs. Finally, US protocols and scanners, acquisition procedures, and parameter selection change over time, which may result in imaging variability and heterogeneity between patient cohorts.

CONCLUSIONS

In conclusion, we demonstrated that DCNNs and transfer learning technology can achieve high accuracy at distinguishing benign, borderline, and malignant SOTs from ovarian ultrasound images. With further verification and model calibration in a larger cohort population, our DCNN-based model could be a promising tool for supporting clinical decision-making.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available. Requests to access these datasets should be directed to Wenjuan Ma, mawenjuan@tmu.edu.cn.

ETHICS STATEMENT

This study was approved by the Institute Review Board of Tianjin Medical University Cancer Hospital Institutional. Due to its

retrospective nature, the informed consent requirement was waived.

AUTHOR CONTRIBUTIONS

WM, HQW, and CLL: study design and methods development. LQ and CXL: lesions identification and marking. HQW, CLL, and ZZ: results inference and implementation of methods. HQW, CLL, and CZ: manuscript writing. HQW, CLL, ZZ, XW, HL, HXW, XL, and WM: manuscript/results proof read and approval. All authors contributed to the article and approved the submitted version.

FUNDING

This study has received funding by National Natural Science Foundation of China (82011530050, 81801781, 81903398, 82072004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.770683/full#supplementary-material>

REFERENCES

- Jayson GC, Kohn EC, Kitchener HC, Ledermann JA. Ovarian Cancer. *Lancet* (2014) 384:1376–88. doi: 10.1016/S0140-6736(13)62146-7
- Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian Cancer Statistics, 2018. *CA Cancer J Clin* (2018) 68:284–96. doi: 10.3322/caac.21456
- Fischerova D, Zikan M, Dundr P, Cibula D. Diagnosis, Treatment, and Follow-Up of Borderline Ovarian Tumors. *Oncologist* (2012) 17:1515–33. doi: 10.1634/theoncologist.2012-0139
- Auersperg N, Wong AS, Choi KC, Kang SK, Leung PC. Ovarian Surface Epithelium: Biology, Endocrinology, and Pathology. *Endocr Rev* (2001) 22:255–88. doi: 10.1210/edrv.22.2.0422
- Willmann JK, Bonomo L, Testa AC, Rinaldi P, Rindi G, Valluru KS, et al. Ultrasound Molecular Imaging With BR55 in Patients With Breast and Ovarian Lesions: First-In-Human Results. *J Clin Oncol* (2017) 35:2133–40. doi: 10.1200/JCO.2016.70.8594
- Spencer JA, Ghattamaneni S. MR Imaging of the Sonographically Indeterminate Adnexal Mass. *Radiology* (2010) 256:677–94. doi: 10.1148/radiol.10090397
- LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539
- Srivastava S, Kumar P, Chaudhry V, Singh A. Detection of Ovarian Cyst in Ultrasound Images Using Fine-Tuned VGG-16 Deep Learning Network. *SN Comput Sci* (2020) 1:81. doi: 10.1007/s42979-020-0109-6
- Wu C, Wang Y, Wang F. Deep Learning for Ovarian Tumor Classification With Ultrasound Images. *Pacific Rim Conference on Multimedia Cham: Springer, Cham* (2018), 395–406. doi: 10.1007/978-3-030-00764-5_36
- Zhang L, Huang J, Liu L. Improved Deep Learning Network Based in Combination With Cost-Sensitive Learning for Early Detection of Ovarian Cancer in Color Ultrasound Detecting System. *J Med Syst* (2019) 43(8):251. doi: 10.1007/s10916-019-1356-8
- Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man Against Machine: Diagnostic Performance of a Deep Learning Convolutional Neural Network for Dermoscopic Melanoma Recognition in Comparison to 58 Dermatologists. *Ann Oncol* (2018) 29(8):1836–42. doi: 10.1093/annonc/mdy166
- Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. *Radiology* (2019) 292(2):331–42. doi: 10.1148/radiol.2019182622
- Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep Learning Predicts Lung Cancer Treatment Response From Serial Medical Imaging. *Clin Cancer Res* (2019) 25(11):3266–75. doi: 10.1158/1078-0432.CCR-18-2495
- Emran MAA, Parvin M, Purang A. A Deep Learning Approach for Real Time Prostate Segmentation in Freehand Ultrasound Guided Biopsy. *Med Image Anal* (2018) 48:107–16.
- Adusumilli S, Hussain HK, Caoili EM, Weadock WJ, Murray JP, Johnson TD, et al. MRI of Sonographically Indeterminate Adnexal Masses. *AJR Am J Roentgenol* (2006) 187(3):732–40. doi: 10.2214/AJR.05.0905
- Jeong YY, Outwater EK, Kang HK. Imaging Evaluation of Ovarian Masses. *Radiographics Rev Publ Radiol Soc North Am Inc* (2000) 20(5):1445–70. doi: 10.1148/radiographics.20.5.g00se101445
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
- Philippe L, Ralph THL, Timo MD, Jurgen P, Evelyn ECDJ, Janita VT, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62.

19. Stefania R, Francesca B, Sara R, Daniela O, Valentina B, Anna C, et al. Radiomics of High-Grade Serous Ovarian Cancer: Association Between Quantitative CT Features, Residual Tumour and Disease Progression Within 12 Months. *Eur Radiol* (2018) 28(11):4849–59.
20. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* (2015) 1409.1556.
21. Szegedy C, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, et al. Going Deeper With Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015) 1–9. doi: 10.1109/CVPR.2015.7298594
22. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) 770–8. doi: 10.1109/CVPR.2016.90
23. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* (2017) 1704.04861.
24. Huang G, Liu Z, Laurens V, Weinberger KQ. Densely Connected Convolutional Networks. *IEEE Comput Soc* (2016) 4700–08. doi: 10.1109/CVPR.2017.243
25. Zhuang F, Qi Z, Duan K, Xi D, He Q. A Comprehensive Survey on Transfer Learning. *P Ieee* (2020) 109(1):43–76. doi: 10.1109/JPROC.2020.3004555
26. Wenhe L, Xiaojun C, Yan Y, Yi Y, Alexander GH. Few-Shot Text and Image Classification via Analogical Transfer Learning. *ACM Trans Intell Syst Technol (TIST)* (2018) 9(6):1–20. doi: 10.1145/3230709
27. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, et al. Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images. *Am J Ophthalmol* (2019) 198:136–45. doi: 10.1016/j.ajo.2018.10.007
28. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009) 248–55. doi: 10.1109/CVPR.2009.5206848
29. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *CVPR IEEE Comput Soc* (2016) 2921–9. doi: 10.1109/CVPR.2016.319
30. Hanley JA, McNeil BJ. A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived From the Same Cases. *Radiology* (1983) 148:839–43. doi: 10.1148/radiology.148.3.6878708

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Liu, Zhao, Zhang, Wang, Li, Wu, Liu, Li, Qi and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.