IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 September 2023, accepted 23 September 2023, date of publication 2 October 2023, date of current version 12 October 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3321408*

RESEARCH ARTICLE

# Accurate Ovarian Cyst Classification With a Lightweight Deep Learning Model for Ultrasound Images

**JUNFANG FAN[1], JUANQIN LIU[1], QILI CHEN[2], WEI WANG[3], (Member, IEEE), AND YANHUI WU[4]**

[1]Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science & Technology University, Beijing 100192, China
[2]Beijing Information Science & Technology University, School of Automation, Beijing 100192, China
[3]School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China
[4]Department of Gynecology, The Third Affiliated Hospital of Beijing University of Chinese Medicine, Beijing 100029, China

Corresponding author: Yanhui Wu (doctor_wyh@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC1511705, in part by the National Natural Science Foundation of China under Grant 61801032 and Grant 62103056, in part by the Key Research and Cultivation Projects of Promote the Classified Development of Universities under Grant 2121YJPY211, and in part by the Project of Construction and Support for High-Level Innovative Teams of Beijing Municipal Institutions under Grant BPHR20220123.

**ABSTRACT** The ovarian cyst is a prevalent disease among women of childbearing age. Early detection of ovaries can effectively prevent the risk of large cysts leading to torsion, infertility, and even progression to ovarian cancer. Ultrasonography is a common method for screening ovarian cysts. However, as the demand for ultrasound has exploded in recent years, doctors' workloads have undoubtedly increased. The ultrasonic image analysis of ovarian cysts using deep learning is aimed at assisting doctors in rapid diagnosis and providing a good diagnostic decision for patients. We proposed a deep learning network for the classification and diagnosis of ovarian cysts, namely Ocys-Net. This method incorporates a reverse bottleneck design strategy and makes full use of global information to improve its feature extraction ability. Meanwhile, the efficient channel attention (ECA) module is used to realize local cross-channel interaction, which pays sufficient attention to pathological information features and effectively makes up for the defects caused by channel dimension reduction. As a lightweight network, the proposed method takes into account the efficient learning performance of the model and is evaluated on our ovarian cyst dataset with high accuracy. The classification accuracy of this network is 95.93%, which has certain practicability in clinical application.

**INDEX TERMS** Deep learning, ovarian cyst, image classification, lightweight, ultrasound.

## I. INTRODUCTION

The ovaries are one of the critical female reproductive organs. They help to maintain female characteristics, regulate metabolism, promote ovum generation, and regulate endocrine. The ovarian cyst is a structure formed inside or on the surface of an ovary. It is a common clinical gynecological condition that is usually caused as a consequence of a disturbance in the hormone levels. This disease is highly prevalent in women of childbearing age and often leads to infertility. The statistics show that around 7% of the women in the world have suffered from symptomatic ovarian cysts [1]. As this disease progresses, ovarian tumors may develop. Therefore, early detection and treatment of cysts are a dire need. Early ovarian cysts are usually detected by ultrasound, computed tomography (CT) scan, and magnetic resonance imaging (MRI) [2]. The examination can determine the status of a cyst by looking at the boundary of the ovarian mass, whether there are clutters of light, spots, and blood flow signals within the mass, and even fluid in the pelvic cavity. Among them, ultrasonography is widely used in the field of gynecology for estimating the size, location, shape, and nature of the mass. It is the best non-invasive diagnostic approach for evaluating the ovarian cysts and other

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

types of ovarian abnormalities. During the ultrasonography process, the doctors rely on their professional knowledge and experience to observe whether the ovarian tumor is a simple fluid area. This helps in determining if the mass is an ovarian cyst or an impure ovarian cyst [3], [4]. As a result, the examination process is extremely subjective and is significantly influenced by the expertise of the doctor. Additionally, the manual examination increases the burden on the doctors. Recently, the contradiction between the increasing demand for ultrasound examination and the shortage of medical resources has led to a heavy workload for ultrasound examination [5], which isn't conducive to the patient's condition and also increases the medical cost.Therefore, the main focus of this study is on the auxiliary diagnosis of the three categories of ''normal pelvic'', ''ovarian cyst'' and ''impure ovarian cyst'', aiming to assist medical staff in making rapid diagnostic decisions.

Medical imaging in the age of big data marked the beginning of a new trend in data management and sharing [6], [7]. The development of medical aid diagnostic systems for analyzing the existing medical image datasets has become a hot trend for clinical and research institutions. However, medical image analysis is still obstructed, mainly due to the lack of theoretical understanding of medical images, which results in limited visual feature extraction and processing. Furthermore, due to the specificity and rigor of medicine, cross-domain solutions are often unable to achieve satisfactory results in the medical field. Recently, deep learning has been advancing at a rapid pace. Some valuable challenges in the medical field have emerged, such as the classification of optical coherence tomography (OCT) scans of retinal diseases [8], the diagnosis and evaluation of corona virus disease 2019(COVID-19) [9], histopathological image classification of liver cancer [10], and classification of breast cancer [11]. Please note that a few alternative methods have also been presented for the automatic detection and the classification of ovarian cysts to lessen the stress on doctors and accelerate the diagnostic procedure.

There are various algorithms presented in the literature for automatically detecting ovarian cysts based on ultrasound images. The authors [12] classified three types of ovarian cysts by using a technique based on the strength of histogram moments and the grayscale co-occurrence matrix, combined with fuzzy mathematics and K-nearest neighbor (KNN) analysis. Rihana et al. [13] achieved a classification accuracy of 90% with the linear support vector machine (SVM), and Nabilah et al. [14] adopted watershed segmentation and contour analysis for distinguishing two different types of ovarian cysts. The researchers [15] used fuzzy logic to distinguish simple and complex types of cysts. Rajendran et al. [16], the authors proposed a new dove-inspired optimization (PIO) for obtaining the optimal threshold of automatic cyst detection and feature extraction and developed an automatic cyst detection system. After that, Akter and Akhter [17] further implemented the screening of ovarian cysts and predicted the ovarian cancer based on machine learning (ML) approaches, such as random forest, KNN, and eXtreme Gradient Boosting (XGBoost). Although these methods improve the accuracy of ovarian cyst identification to varying degrees, they are all based on traditional techniques with low generalization ability and are unable to match the accuracy achieved by the prevailing popular deep learning methods. In addition, the author [2], [18] proposed a deep reinforcement learning framework to classify ovarian cysts. Professional researchers [19] integrated neural network algorithms for migratory learning have been devised to accurately predict polycystic ovarian syndrome form tiny amounts of ovarian ultrasound images. The authors [20] used supervised machine learning algorithms to classify multicellular ovarian syndrome, and the key was to design the GIST-MDR technology to extract features. This type of reinforcement learning approach is more difficult to model the environment and requires a lot of time and effort.

In today's world, a large number of medical images are generated each day. However, it is difficult to use these medical images due to privacy and security concerns [21]. Please note that there are no publicly available datasets comprising ultrasound images of ovarian cysts. In addition, a huge number of clinical ultrasound images are often different in different conditions. The histopathological morphology of individual patients shows a variety of shapes, sizes, textures, and color distributions. Moreover, the confusion among symptoms also increases the difficulty of identifying cysts. For the study of ovarian cysts, the current problem is, on the one hand, the classification accuracy deficit faced by automatic diagnosis. On the other hand, is the difficulty of deployment due to the high complexity of the model. This study mainly attempts to use a lightweight deep learning model to automatically extract key features and obtain a high classification accuracy, which makes the detection of ovarian cysts more accurate and reliable, while using a lightweight model significantly reduces the computational and storage costs [22], and is more convenient to deploy on mobile, which can contribute to the diagnosis and treatment of ovarian cyst-related diseases.

Key contributions to this work are outlined below.

- A lightweight classification and diagnosis network is proposed, and the reverse bottleneck design strategy has a major impact on the feature extraction ability, with the network performance effectively improved.
- Integrated with the efficient channel attention mechanism module, with effectively extract the major pathological characteristics of ovarian cysts. It can provide a good solution for the task of embedded medical image analysis.
- We constructed three types of ovarian cyst datasets based on medical images obtained from professional medical institutions, and used this framework to achieve effective classification of normal pelvic cavity, ovarian cyst and impure ovarian cyst. It provides a good solution for performing medical image analysis.

The remainder of the manuscript is structured in the following way. The associated image classification methods are described in Section II. It is followed by the account of the Ocys-Net in Section III and Section IV presents the experimental results and compares them with other methods. Lastly, Section V brings this work to a conclusion.

## II. RELATED WORKS

Due to the particularity and rigor of medicine, it is very important to improve the accuracy and precision of the automatic diagnosis of ovarian cysts, while considering a lightweight and highly efficient network. There are various networks proposed in the literature for achieving this, such as SqueezeNet [23], MobileNet [24], GhostNet [25], ShuffleNet [26], and MixNet [27]. Please note that group convolution (GC) and depthwise separable convolution effectively solve the mobile deployment problems of deep learning models. These methods provide substantial technological support for deep learning model deployment in the medical industry. Throughout this section, we introduce two kinds of architectures that are particularly relevant to the proposed work, i.e., the ShuffleNet model and attention mechanisms.

### *A. ShuffleNet MODEL*

ShuffleNet, a lightweight mobile terminal convolutional neural network (CNN), adopts pointwise and depthwise (DW) convolutions, thus significantly decreasing the computational costs and ensuring accuracy. One of the core ideas of this network is the use of channel shuffling to design a more powerful architecture comprising multi-set convolution layers, i.e., for the feature map created at the preceding layer, each group's channels are subdivided into multiple subgroups. Subsequently, the following layer provides different subgroups for each group. Its functional realization is presented in Figure 1.

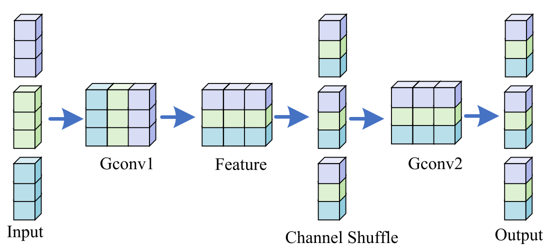


**FIGURE 1. The channel shuffling process, GConv, is a shortened form for group convolution.**

In ShuffleNetV1, designed based on the channel shuffling design concept, $1 \times 1$ pointwise convolutions are adopted to replace the ordinary convolutions. Afterwards, the channel shuffling operation is performed. The pointwise convolutions used for the second time aim to recover the shortcut path of channel dimension matching. In order to improve the memory-intensive operations in ShuffleNetV1, the proposed ShuffleNetV2 [28] effectively balances the relation between the recognition accuracy and computation speed. Please note that ShuffleNetV2 no longer uses pointwise convolutions. Instead, it uses ordinary $1 \times 1$ convolution operations again. In order to keep the number of channels constant, these two branches are concatenated rather than added. In the end, the channels are shuffled to realize the information flow between the two branches. Figure 3(a) and (b) depict the ShuffleNetV2 architecture. Ghosh et al. [29], ShufflenetV2 was used to realize the recognition and classification of organic products among 41 independent categories, and the test accuracy reached 96.24%. Gomes et al. [30] apply the ShuffleNet unit to a semantic segmentation network. This increases the network's lightness and accuracy of semantic segmentation of remote sensing images. The authors [31] use improved ShuffleNetV2 for designing a GCNet household garbage classification system, which effectively promotes academic research and technological implementations in the realm of resources and the environment. It is evident that the ShuffleNet has achieved positive outcomes in various fields of research, and its application in medical research should also have a potential application value. household garbage classification system, which implements mechatronics in realms of environmental and resource management.

### *B. ATTENTION MECHANISM IN MEDICAL IMAGE ANALYSIS*

The attention mechanism (AM) [32] originated from the RNN model commonly used in natural language processing (NLP) and has since been applied to other fields as well. The basic idea is to emphasize the effect of a key input on the output by calculating the weight of the input data [33]. As a general idea, the attention mechanism can be connected to each model for improving performance.

In medical image analysis, people are particularly interested in two forms of attention mechanisms, namely saliency detection and the visual attention model. The well-known detection methods include gradient-weighted class activation mapping (Grad-CAM) [34], class activation mapping (CAM) [35], and saliency mapping (SM) [36]. These methods confirm the basis of judgment by applying them to the learned models. However, they do not directly contribute to improving the performance. The other visual attention model pays more attention to the important features and less attention to the unimportant features. Therefore, the visual attention model significantly improves performance. Currently, the most popular channel attention mechanisms include the compression and excitation (SE) [37] model, and the concurrent spatial and channel 'squeeze & excitation' (scSE) [38] model inspired by this model. In addition, there is a lightweight convolutional block attention module (CBAM) [39], which is compatible with any CNN architecture and conducts end-to-end training by using CNN.

Recently, the research community has presented the effects of different visual attention mechanisms on medical tasks. The authors [40] observe that the addition of an attention mechanism to each convolution block before residual connection in the ResNet improves the accuracy

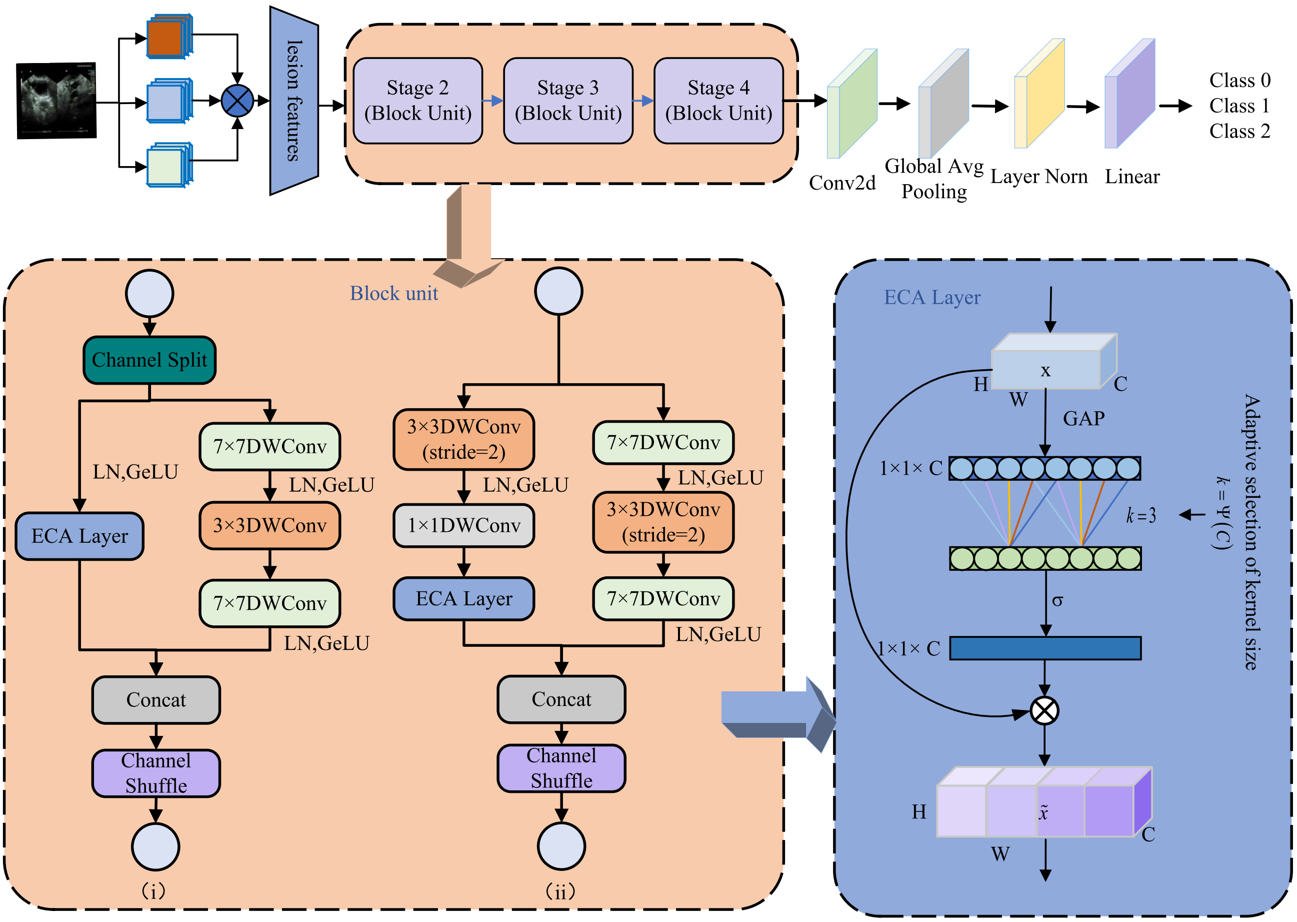


**FIGURE 2.** The architecture of the proposed Ocys-Net: composed of the design of the Block unit and the ECA layer.

of medical vision tasks in skin melanoma images, chest X-ray (CXR) images, brain magnetic resonance imaging (MRI) images, and COVID-19 computer tomography (CT) scans. In the study [41], to effectively improve the performance of abdominal CT image segmentation, the authors integrated the attention mechanism with the classic U-Net.

In the study [42], the researchers improved the accuracy of tumor image segmentation by adding a parallel mixed attention mechanism. The authors [43] added a series of attention gates to jump connections for improving the segmentation performance of brain tumor MRI. The researchers [44] introduced channel attention in a dense upsampling network for detecting tumors in breast X-ray images. In terms of improving the performance of deep convolutional neural networks, the channel attention module has shown enormous promise.

The channel attention module, on the other hand, necessitates channel compression of the input feature graph, and such compression dimension reduction usually adversely affects the dependence between the learning channels. By combining the advantages of ShuffleNetV2 and ECA, the method proposed in this work captures the local cross-channel interaction information on the shortcut branch, adopts a reverse bottleneck design strategy on the main branch, and realizes information flow between two branches based on channel shuffling, thus achieving good classification performance.

## III. METHODS

We created a deep learning network that can automatically extract the shape and texture characteristics of normal pelvic cysts, ovarian cysts, and non-pure ovarian cysts in order to accomplish correct classification. Experts use their professional expertise and experience to diagnose ovarian cysts by observing ultrasonic pictures. To ensure the model's accuracy and dependability, we use clinical specialists' diagnosis results as the gold standard. Meanwhile, the model's output will be professionally assessed by clinical professionals. This effectively secures the model's availability. In this section, we offer a detailed description of the proposed ovarian cyst classification network, including network architecture, basic network components, and design methodologies.

### *A. OVARIAN CYST CLASSIFICATION NETWORK (OCYS-NET)*

Figure 2 depicts the proposed Ocys-Net architecture. The input ultrasound images of the ovaries are 224 × 224 × 3. In this work, the whole network operates in a feedforward

**TABLE 1.** A comparison of Ocys-Net architecture.

| Layer | Output size | K_size | Stride | Repeat | Output channels |
|---|---|---|---|---|---|
| Image | 224×224 | - | - | - | 3 |
| Conv1 | 56×56 | 4×4 | 4 | 1 | 24 |
| Maxpool | 28×28 | 3×3 | 2 | 1 | |
| Stage2 | 14×14 | - | 2 | 1 | 48 |
| | 14×14 | - | 1 | 4 | |
| Stage3 | 7×7 | - | 2 | 1 | 96 |
| | 7×7 | - | 1 | 4 | |
| Stage4 | 4×4 | - | 2 | 1 | 192 |
| | 4×4 | - | 1 | 4 | |
| Conv5 | 4×4 | 1×1 | 1 | 1 | 1024 |
| Globalpool | 1×1 | 7×7 | - | - | - |
| Fc | - | - | - | - | 3 |

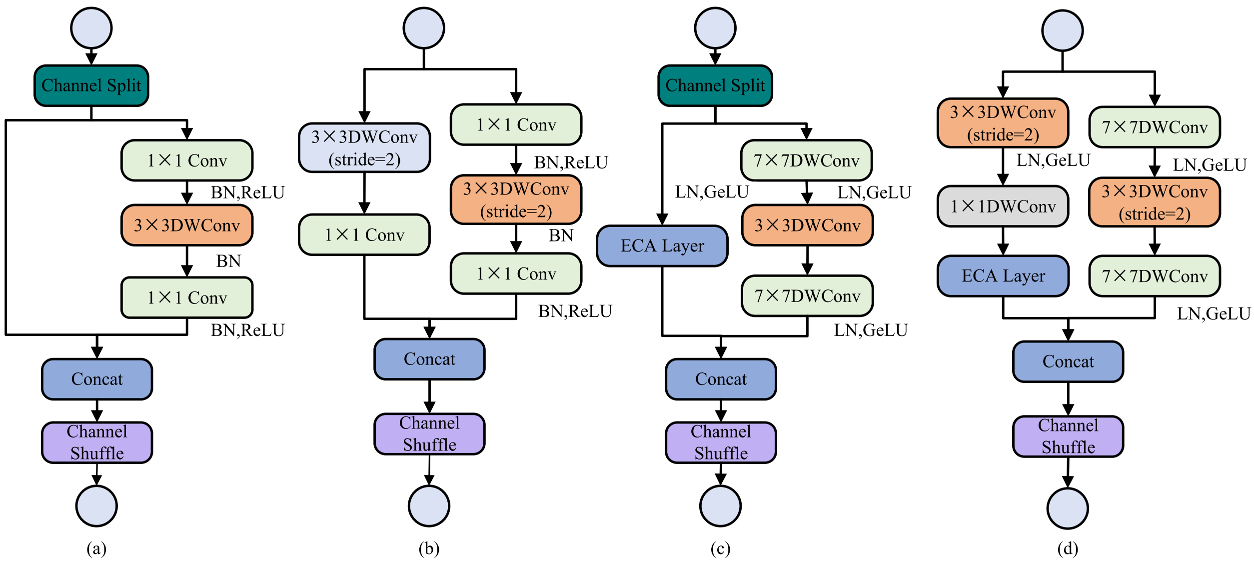


**FIGURE 3.** The building blocks of the ShuffleNetV2 and the network proposed in this work. (a) A basic ShuffleNetV2 unit. (b) The ShuffleNetV2 unit for spatial down sampling. (c) The basic unit of the proposed network. (d) The proposed unit for spatial down sampling. Depthwise convolution is denoted by the symbol DWConv.

mode, and the features extracted from the previous layer are fed to the middle layer by convolutional operation, and then the results are transmitted to the next layer.

For feature extraction, a standard convolution is first employed with a stride of 4 and a convolution kernel of size 4 × 4. This is followed by a downsampling operation based on maximum pooling. In addition, a group normalization layer (GN) is added to prevent overfitting of the network. Subsequently, in the second stage, five block units are used consecutively with 24 feature maps. This block unit is inspired by the improved ShuffleNetV2. In the third stage, five block units are continuously used, and the number of channels is expanded to 48. The fourth stage continues with five block units and expands the number of channels to 96. After these three stages, a convolution layer with a convolution kernel of 1 × 1 and a stride of 1 is connected to ensure the accuracy of classification. Afterwards, a global average pooling layer and layer normalization (LN) are also used to fuse the spatial information, prevent overfitting, and enhance the network's generalization ability Finally, 1024 convolutions with 1 × 1 kernels are used to expand the number of channels. Please note that the large channels and large convolutions used in the aforementioned design were used to obtain the characteristic information of ovarian cysts for performing classification. Most importantly, the design of block units is crucial for this network architecture.

*T*able 1 summarizes the entire network architecture of the proposed Ocys-Net. The main difference between the proposed networked and the ShuffleNetV2 0.5 × is that the first feature extraction convolution layer uses a 4 × 4 large convolution kernel with a step size of 4. The output size caused by the downsampling effect also changes accordingly in stages 2,3, and 4. Each stage is repeated 5 times. This is beneficial to model optimization and achieving the effect of model lightweight.

## *B. BLOCK UNIT*

Figure 3 depicts the block unit structure of the proposed Ocys-Net presented in this work. The inverted bottleneck design is used in comparison to ShuffleNet V2 0.5×, and the ordinary convolution is replaced by the depthwise convolution. In the main branch, we use a large 7 × 7 DW convolution, followed by a 3 × 3 DW convolution in order to reduce dimensions, followed by a 7 × 7 DW convolution. Please note that we abandon the operations performed using a 1 × 1 small convolution kernel to first reduce the dimension and then increase it. This design maintains the same shape of the eigenmatrix output on the main and shortcut branches. This was inspired by the work presented [45] that the Convnext block units would be similar to those in Transformers [46], with a reverse bottleneck design that was thin in the middle and thick at both ends to boost the network's performance. It is noteworthy that we added an ECA attention module in the block unit to assist the process of extracting key features. In addition, the GN layer is more prevalent to avoid overfitting. The GeLU activation function [47], i.e., a smooth variant of ReLU, is replaced by the ReLU activation function to strengthen the effectiveness of the network.

### C. ECA LAYER

The ECA attention mechanism [48] employs 1-dimensional convolution for realizing the local cross-channel interaction strategies and extracting the channel dependencies, thus effectively avoiding the defects caused by dimensionality reduction. The size of the convolution kernel k is adaptively chosen by the ECA module after the convolution features are aggregated using global average pooling (GAP) without dimension reduction. Afterward, it performs 1-dimensional convolution and uses the sigmoid function to learn the channels, thus determining the extent of cross-channel interaction coverage. In this work, the size of the adaptive convolution kernel k is 3, and the ECA's attention mechanism layer is embedded in the shortcut branch of the block unit, which helps to extract important elements of pathological information while also avoiding the defect brought on by dimensionality reduction, as shown in Figure 3 (c) and (d).

## IV. EXPERIMENTS

In this chapter, we describe the dataset and training process, analyze the proposed network and evaluate the results.

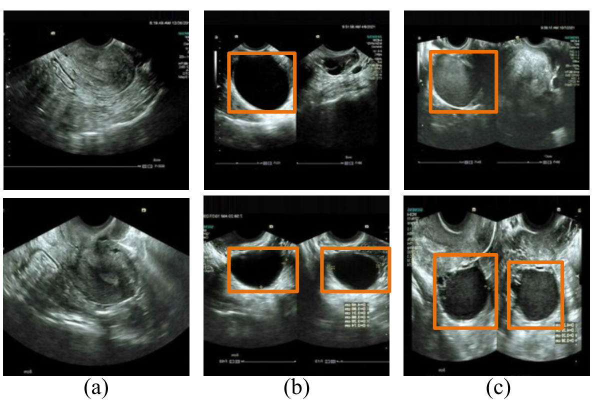


**FIGURE 4. Three categories of samples from the ultrasound image dataset. (a) normal pelvic. (b) ovarian cyst. (c) impure ovarian cyst. The area outlined by the yellow box is the lesion area.**

### A. DATASET

We collaborated with relevant hospitals to obtain ultrasound images of three types of ovaries. There are 750 images in total, including 250 normal pelvic ultrasound images, 250 ovarian cyst ultrasound images, and 250 impure ovarian cyst ultrasound images. These data are collected from the same machine by professional doctors from relevant hospitals and manually annotated one by one. We process these images to obtain three classification datasets, where each input image is 224 × 224 pixels in size. Furthermore, since the data size is small, we perform data augmentation, such as random transformation, to augment the data. And normalize each channel of the image. Scale the pixel value of each channel from the range [0, 255] to the range [−1, 1] to facilitate model training. Three categories of samples from the ultrasound image dataset are shown in Figure 4.

### B. EXPERIMENTAL SETUP AND CONFIGURATION

The hardware used for this experiment includes an Intel (R) Core (TM) i7-10750H CPU and an NVIDIA GeForce RTX2060 GPU. The proposed method is implemented on top of the Pytorch framework. During the training process, the adaptive moment estimation (Adam) optimizer is used to dynamically adjust the parameters. The learning rate is set to 1e-4, the batch size is 4, and the number of training epochs is 300. The training loss is realized by the cross-entropy loss function.

**TABLE 2. Comparison of experimental results for ovarian cyst classification.**

| Model | Parameters | FLOPs | MAdds | Accuracy |
|---|---|---|---|---|
| MobileNetV1 | 12.29M | 587.94M | 1.16G | 91.51% |
| MobileNetV2 | 8.50M | 318.96M | 625.14M | 93.61% |
| MobileNetV3 | 20.92M | 227.71M | 448.69M | 93.42% |
| EfficientNet | 15.30M | 398.02M | 789.29M | 94.64% |
| GhostNet | 14.90M | 149.11M | 292.14M | 92.25% |
| MixNet | 15.77M | 251.82M | 497.23M | 94.61% |
| ShufflenetV2 | 1.32M | 42.63M | 83.41M | 92.12% |
| Ocys-Net (Proposed) | **1.15M** | **13.25M** | **26.23M** | **95.93%** |

### C. COMPARISON WITH OTHER ADVANCED MODELS EXPERIMENT

In order to guarantee the dependability and stability in terms of the performance of the proposed model, the following comparative experiments are designed. During our training process, except for the different network models, the rest of the training parameters remain consistent. Including data set, learning rate, optimizer, batch size, epochs, etc. We compare the proposed Ocys-Net with seven classic lightweight deep learning classification networks, *T*able 2 shows the comparison results.

We compared the proposed model with seven other leading methods, including: MobileNetV1, MobileNetV2, MobileNetV3, EfficientNet, GhostNet, MixNet and ShufflenetV2 methods, as shown in *T*able 2. First, the proposed network shows the effect of being lightweight. Please note that the Params, Flops, and Madd parameters of the proposed network are comparable with those of MobileNet, EfficientNet [49], MixNet, and other networks, with varying degrees of advantage. This proves the high efficiency of network

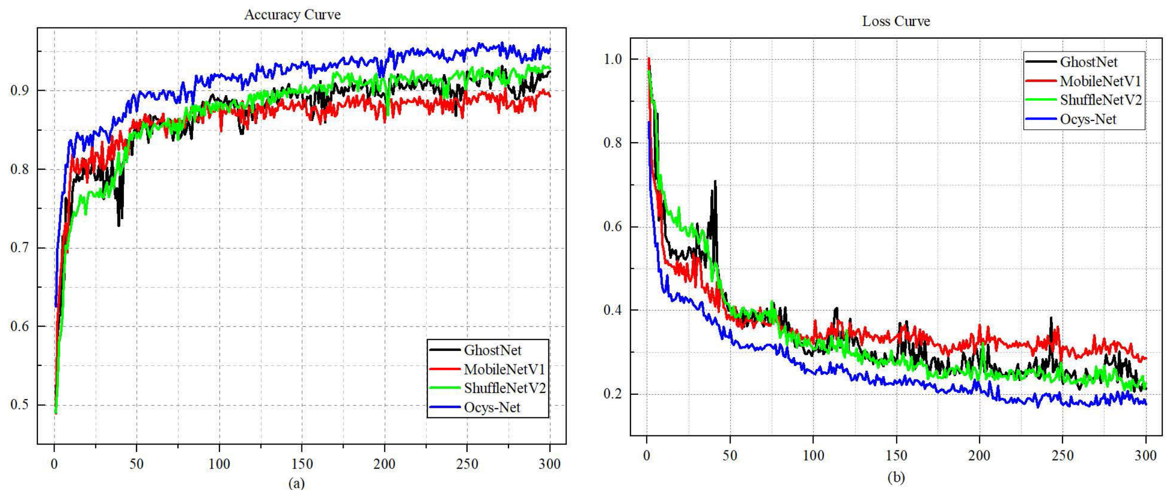


**FIGURE 5. The training curve of different models. (a) Curves of training accuracy. (b) Curves of training loss.**

operation. Furthermore, the results show that MobileNetV1 has an accuracy of 91.51%, while the improved MobileNetV2 and MobileNetV3 have an accuracy of about 93%. The Top1 accuracy of GhostNet is approximately 0.74% higher as compared to MobileNetV1, which benefits from GhostNet's taking advantage of cheap operations to get more feature maps. Due to channel mixing, the ShuffleNetV2 model enables multiple convolutional layers to build a more powerful structure, which further improves the accuracy and ensures the lightness of the network and achieves an accuracy of 92.12%.

In addition, the EfficientNet and MixNet models achieve accuracy rates of 94.64% and 94.61%, respectively. The proposed method shows a big improvement over ShufflenetV2, i.e., an improvement of 3.81%. It's noteworthy that the proposed network for this challenge achieves superior performance, and its classification accuracy is the highest among the aforementioned methods, i.e., 95.93%. In this work, the large convolution kernel, reverse bottleneck design, and attention mechanism are boldly adopted in the block unit. According to the *T*able 2, this operation significantly enhances the network's classification performance.

Figure 5 (a) and (b) depict a visual comparison of training accuracy and training loss among several related classification models and the proposed Ocys-Net. The network proposed in this work has lower loss and faster convergence after 300 training epochs as compared to the other models, and better results are obtained under the same conditions.

## D. ABLATION EXPERIMENT OF DIFFERENT MODULES

We designed the ablation experiment with two important points. The first is the effect of the Block module developed in this study on performance, the second is the influence of the added effectiveness of the attention module (ECA).We used the pre-trained ShuffleNetV2 0.5 $\times$ network as the baseline and then test the training effect of adding different attention mechanism modules, such as SE, CBAM, polarized self-attention (PSA) [50], and ECA. Please note that the Baseline_block refers to the network that uses the block proposed in this work, but removes the ECA layer. *T*able 3 presents the ablation results.

**TABLE 3. The results of ablation experiment.**

| Model | Params | Flops | Madd | Accuracy |
|---|---|---|---|---|
| Baseline | 1.32M | **42.63M** | **83.41M** | 92.12% |
| Baseline+SE | 1.36M | 42.64M | 83.43M | 91.94% |
| Baseline+CBAM | 1.33M | 42.73M | 83.61M | 92.68% |
| Baseline+ECA | 1.32M | 42.63M | 83.41M | 93.35% |
| Baseline+PSA | 1.39M | 44.69M | 87.54M | 92.79% |
| Baseline_block | 1.15M | 42.90M | 84.72M | 93.16% |
| Baseline_block+SE | 1.19M | 42.93M | 84.81M | 93.06% |
| Baseline_block+CBAM | 1.16M | 43.03M | 84.99M | 94.32% |
| Baseline_block+PSA | 1.22M | 44.99M | 88.92M | 94.51% |
| Baseline_block+ECA(Proposed) | **1.15M** | 42.92M | 84.78M | **95.93%** |

As seen in *T*able 3, the baseline represents the ShufflenetV2 backbone network with an accuracy of 92.12%, whereas the Baseline_block represents the block layer without our attention mechanism. It is evident from the results that classification accuracy has improved. This demonstrates that the large convolution kernel and the reverse bottleneck design with a thin middle and two thick ends improve the network performance, and the Top1 accuracy increases from 92.12% to 93.16%, i.e., an increase of 1.04%. Second, the performance gains achieved due to the addition of various attention modules are compared for both the baseline backbone network and the baseline block network. The results show that the use of SE attention module in this challenge reduces the performance of the network, and the precision reduces by approximately 0.1%. The network's performance is enhanced by 1.16% and 1.35%, respectively, by the addition of CBAM and PSA. The results show that the ECA attention mechanism performs better, and the accuracy increases by 2.77%. Generally, the accuracy of the network designed in this experiment has increased by 3.81%. The results of this ablation experiment provide substantial and

**TABLE 4.** Classification performance comparison results of various categories.

| Methods | Normal pelvic | | | Ovarian cyst | | | Impure ovarian cyst | | |
|---|---|---|---|---|---|---|---|---|---|
| | precision | recall | f1-score | precision | recall | f1-score | precision | recall | f1-score |
| MobileNetV1 | 0.9357 | 0.8572 | 0.8948 | 0.8751 | 0.9126 | 0.894 | 0.8528 | 0.8858 | 0.869 |
| MobileNetV2 | 0.9586 | 0.9222 | 0.9448 | 0.8897 | 0.9637 | 0.9344 | 0.9345 | 0.9065 | 0.9299 |
| MobileNetV3 | 0.9617 | 0.9261 | 0.9436 | 0.9087 | 0.9604 | 0.9338 | 0.9341 | 0.9182 | 0.9261 |
| EfficientNet | 0.9542 | 0.9142 | 0.9337 | 0.947 | 0.9797 | 0.953 | 0.9393 | 0.946 | 0.9427 |
| GhostNet | 0.9483 | 0.9162 | 0.936 | 0.8698 | 0.9776 | 0.9203 | 0.9551 | 0.8795 | 0.9157 |
| MixNet | **0.9652** | 0.9301 | 0.9569 | 0.9312 | **0.9898** | 0.9596 | 0.9548 | 0.9496 | 0.952 |
| ShuffleNetV2 | 0.9571 | 0.8902 | 0.9224 | 0.8926 | 0.9797 | 0.9341 | 0.919 | 0.8975 | 0.9081 |
| Ocys-Net(Proposed) | 0.964 | **0.9621** | 0.9630 | **0.9505** | 0.9856 | **0.9629** | **0.9639** | **0.9548** | **0.9527** |

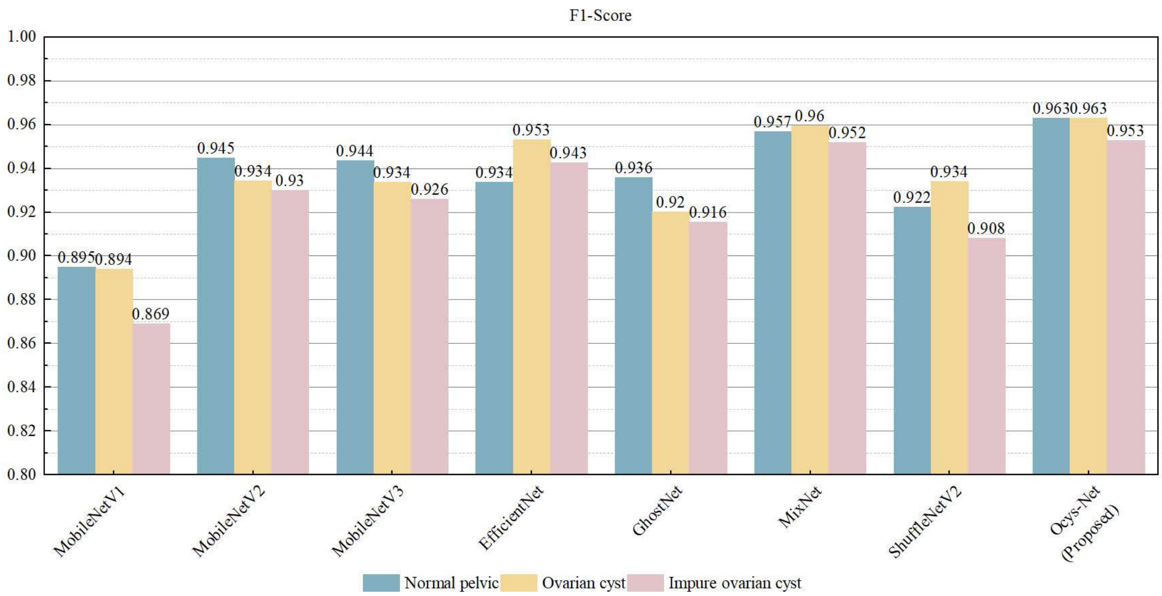


**FIGURE 6.** F1-Score evaluation index results of three categories.

compelling evidence that the inclusion of an ECA layer in the Block module, as designed in this experiment, is indeed the correct and optimal choice.

### *E. DIFFERENT CATEGORIES OF CLASSIFICATION EXPERIMENTS*

This experiment is mainly used to evaluate the performance of the classifier for each category, including classification accuracy, recall rate, F1-Score, and confusion matrix. The evaluation indicators used in this work are shown as follows:

$$Accuracy = TP + TN/(TP + TN + FP + FN), \quad (1)$$

$$Precision = TP/(TP + FP), \quad (2)$$

$$Recall = TP/(TP + FN), \quad (3)$$

$$F1{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (4)$$

where $TP$ and $FP$ represent the quantity of true and false positives, respectively, and and represent the quantity of true and false negatives, respectively.

As presented in Table 4, the MixNet approach demonstrates the highest precision of 0.9652 for the normal pelvic category. Comparatively, the precision of the proposed method is measured at 0.9416, indicating a marginal decrease of 0.0236 in comparison to the MixNet approach. Similarly, the recall rate of the proposed method reaches 0.9621, surpassing MixNet by 0.032. Turning to the ovarian cyst category, it is noteworthy that MixNet achieves the highest recall rate, while the proposed method exhibits superior precision and F1-Score. In the context of impure ovarian cyst instances, the proposed methodology attains the highest levels of precision, recall, and F1-Score outcomes. In terms of performance metrics holistically, the method introduced in this study demonstrates a substantial advantage over alternative approaches, manifesting balanced accuracy across all three recognition categories. Moreover, the F1-Score, as an evaluation metric, stands as a more representative measure than the other two, given its nature as a weighted amalgamation of recall rate and accuracy. Significantly, as depicted in Figure 6, this study achieves the highest F1-Score across all three categories, underscoring the robustness of the proposed methodology.

The confusion matrices obtained by various methods are depicted in Figure. 7. The normal pelvic cavity, ovarian cyst, and impure ovarian cyst are represented by "1", "2" and "3", respectively. The diagonal line shows the classification accuracy of each category. According to the confusion matrix, the classification accuracy of ovarian cysts was relatively high, while the classification accuracy of normal pelvic and impure ovarian cyst was low. This is

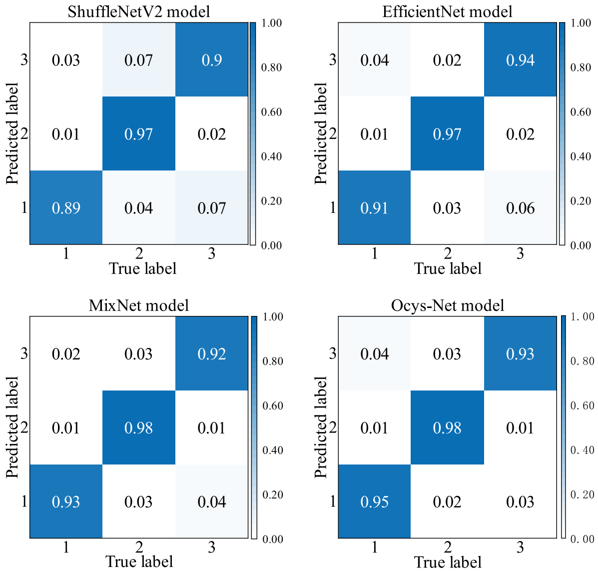


**FIGURE 7.** **The confusion matrices of different models, including ShuffleNetV2, EfficientNet, MixNet and Ocys-Net models.**

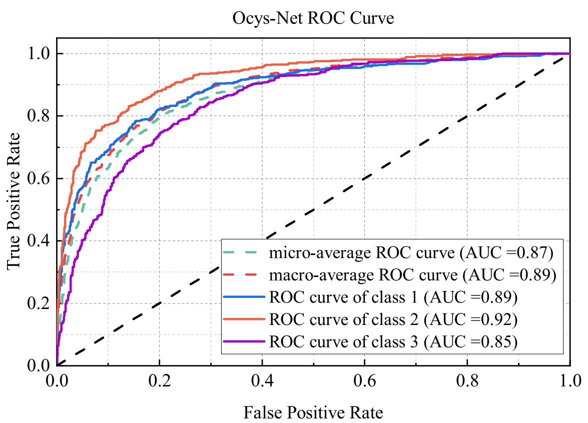


**FIGURE 8.** **The ROC curves for three different categories in Ocys-Net model.**

because there is less difference in the presentation of symptoms between the normal pelvic category and the non-pure ovarian cyst category in the image, so there is more potential for confusion. In this work, 3% of normal pelvic cavities are mistaken for impure ovarian cysts, and 4% of impure ovarian cysts are mistaken for normal pelvic cavities. The MixNet method misclassified 4% of normal pelvic cavities as impure ovarian cysts, and the EfficientNet and ShuffleNetV2 error rates are as high as 6% and 7%, respectively. In general, the classification accuracy of the proposed method for each category is higher and more balanced.

This study further evaluates the performance of the proposed algorithm and draws the receiver operating characteristic (Receiver Operating Characteristic, ROC) curve. This curve is widely used to evaluate the performance of classifiers and reliably distinguish between positive and

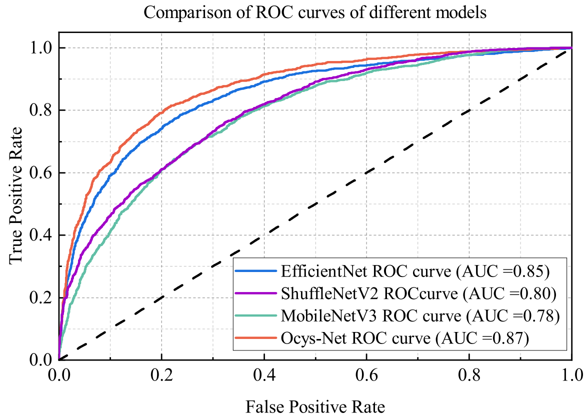


**FIGURE 9.** **The ROC curve comparison of four major lightweight models.**

negative classes. The ROC curve of the Ocys-Net model proposed in this chapter is shown in Figure 8, and the performance is evaluated by the area index (AUC) of the curve. In the ovarian cyst detection task, the algorithm has the best recognition effect on the category of ovarian cysts, and its AUC value reaches 0.92, indicating that the algorithm has high accuracy and sensitivity. This was followed by the normal pelvis category with an AUC of 0.89, and the worst classifier was the non-pure ovarian cyst category with an AUC of 0.85. In addition, this study also calculated the AUC values of the micro-average (micro-average) ROC curve and the macro-average (macro-average) ROC curve, which were 0.87 and 0.89, respectively. This shows that the algorithm performed well in terms of overall classification ability and can effectively distinguish different ovarian cyst types. Therefore, the Ocys-Net model has great practical application value and provides a useful exploration for the development of the field of medical image recognition.

Figure 9 displays the microscopic average ROC curve comparison results of four different classification models used in ovarian cyst recognition tasks. The performance of each classifier was evaluated based on the area under the curve (AUC) metric. The results showed that MobileNetV3 had the poorest performance, with an AUC value of 0.78. ShuffleNetV2 and EfficientNet had slightly better AUC values of 0.80 and 0.85, respectively. However, the Ocys-Net proposed in this paper stood out with an AUC value of 0.87, indicating its stronger classification ability and robustness. Therefore, the Ocys-Net model has greater potential for practical applications and provides strong support for research and practice in the field of medical image recognition.

## V. DISCUSSION AND CONCLUSION

In this paper, an efficient and lightweight classification and diagnosis network for ovarian cysts, Ocys-Net, is proposed to realize the automatic classification of normal pelvic, ovarian, and non-pure ovarian cysts. This method can help doctors make quick decisions and greatly reduce the risk of missed diagnoses and misdiagnoses. Firstly, the model improves the performance of the block cell network by adopting the inverse bottleneck design with a large convolution kernel, a thin middle, and two thick sides. Second, ECA focuses on modules that extract key characteristics more efficiently, thereby further improving performance. In order to ensure the accuracy and reliability of the model, clinical experts make use of their professional knowledge and experience to diagnose ovarian cysts by observing ultrasound images, with the diagnosis results of clinical experts as the gold standard. At the same time, the model output is also professionally evaluated by clinical professionals, thus effectively ensuring the usability of the model. This method can effectively reduce the workload of doctors and is of great significance for the diagnosis and treatment of ovarian cyst-related diseases. A large number of experiments have verified the effectiveness of this model, which not only improves the engineering practicability of ovarian cyst classification but also introduces a new way of thinking about medical image analysis.

Despite initial success in identifying ovarian ultrasound images, the method's use is confined to three specific types of ovarian cysts, and the dataset size is limited. This suggests that we still need to broaden the research to include more forms of cysts. Furthermore, given the need for both doctors and patients to fully understand the rationale behind model diagnosis or classification decisions, the issue of explainability is of critical importance in the medical field. Although the accuracy of the model in identifying ovarian cysts is excellent, its interpretability remains a great challenge. Therefore, we will work on solving this problem to ensure that the decision-making process of the model can be clearly explained and understood. In the future, our research direction will focus on developing more robust systems designed to identify more types of ovarian cysts while improving the interpretability of the model. We believe that this will help drive the practical application of smart medical technology and improve the accuracy and safety of medical diagnostics to better meet the needs of patients and doctors.

## REFERENCES


[1] M. Henes, T. Engler, F.-A. Taran, S. Brucker, K. Rall, B. Janz, and B. Lawrenz, "Ovarian cyst removal influences ovarian reserve dependent on histology, size and type of operation," *Women's Health*, vol. 14, May 2018, Art. no. 1745506518778992.

[2] C. Narmatha, P. Manimegalai, J. Krishnadass, P. Valsalan, S. Manimurugan, and M. Mustafa, "Ovarian cysts classification using novel deep reinforcement learning with Harris hawks optimization method," *J. Supercomput.*, vol. 79, no. 2, pp. 1374–1397, Feb. 2023.

[3] P. Glanc, N. Brofman, S. Salem, A. Kornecki, J. Abrams, and D. Farine, "The prevalence of incidental simple ovarian cysts $\geq 3$ cm detected by transvaginal sonography in early pregnancy," *J. Obstetrics Gynaecol. Canada*, vol. 29, no. 6, pp. 502–506, Jun. 2007.

[4] S. Srivastava, P. Kumar, V. Chaudhry, and A. Singh, "Detection of ovarian cyst in ultrasound images using fine-tuned VGG-16 deep learning network," *Social Netw. Comput. Sci.*, vol. 1, no. 2, pp. 1–8, Mar. 2020.

[5] Y. Wang, X. Ge, H. Ma, S. Qi, G. Zhang, and Y. Yao, "Deep learning in medical ultrasound image analysis: A review," *IEEE Access*, vol. 9, pp. 54310–54324, 2021.

[6] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsaftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, K. Huang, K. S. Nikita, B. P. Veasey, M. Zervakis, J. H. Saltz, and C. S. Pattichis, "AI in medical imaging informatics: Current challenges and future directions," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020.

[7] S. D. Arul and M. Iyapparaja, "Social Internet of Things using big data analytics and security aspects—A review," *Electron. Government, Int. J.*, vol. 16, nos. 1–2, pp. 137–154, 2020.

[8] S. Najeeb, N. Sharmile, M. S. Khan, I. Sahin, M. T. Islam, and M. I. H. Bhuiyan, "Classification of retinal diseases from OCT scans using convolutional neural networks," in *Proc. 10th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2018, pp. 465–468.

[9] I. Rodrigues, G. L. Santos, D. F. H. Sadok, and P. T. Endo, "Classifying COVID-19 positive X-ray using deep learning models," *IEEE Latin Amer. Trans.*, vol. 19, no. 6, pp. 884–892, Jun. 2021.

[10] C. Sun, A. Xu, D. Liu, Z. Xiong, F. Zhao, and W. Ding, "Deep learning-based classification of liver cancer histopathology images using only global labels," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 6, pp. 1643–1651, Jun. 2020.

[11] G. Murtaza, L. Shuib, A. W. A. Wahab, G. Mujtaba, G. Mujtaba, H. F. Nweke, M. A. Al-garadi, F. Zulfiqar, G. Raza, and N. A. Azmi, "Deep learning-based breast cancer classification through medical imaging modalities: State of the art and research challenges," *Artif. Intell. Rev.*, vol. 53, no. 3, pp. 1655–1720, Mar. 2020.

[12] A. S. M. Sohail, P. Bhattacharya, S. P. Mudur, S. Krishnamurthy, and L. Gilbert, "Content-based retrieval and classification of ultrasound medical images of ovarian cysts," in *Proc. IAPR Workshop Artif. Neural Netw. Pattern Recognit.*, 2010, pp. 173–184.

[13] S. Rihana, H. Moussallem, C. Skaf, and C. Yaacoub, "Automated algorithm for ovarian cysts detection in ultrasonogram," in *Proc. 2nd Int. Conf. Adv. Biomed. Eng.*, Sep. 2013, pp. 219–222.

[14] A. Nabilah, R. Sigit, T. Harsono, and A. Anwar, "Classification of ovarian cysts on ultrasound images using watershed segmentation and contour analysis," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2020, pp. 513–519.

[15] A. M. Parekh and N. B. Shah, "Classification of ovarian cyst using soft computing technique," in *Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2017, pp. 1–5.

[16] S. Rajendran and U. M. Sankareswaran, "A novel pigeon inspired optimization in ovarian cyst detection," *Current Med. Imag. Rev.*, vol. 12, no. 1, pp. 43–49, Jan. 2016.

[17] L. Akter and N. Akhter, "Ovarian cancer prediction from ovarian cysts based on TVUS using machine learning algorithms," in *Proc. Int. Conf. Big Data, IoT, Mach. Learn.* Cham, Switzerland: Springer, 2022, pp. 51–61.

[18] M. A. Rahman, R. C. Muniyandi, K. T. Islam, and M. M. Rahman, "Ovarian cancer classification accuracy analysis using 15-neuron artificial neural networks model," in *Proc. IEEE Student Conf. Res. Develop. (SCOReD)*, Oct. 2019, pp. 33–38.

[19] C. Gopalakrishnan and M. Iyapparaja, "ITL-CNN: Integrated transfer learning-based convolution neural network for ultrasound PCOS image classification," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 36, no. 16, Dec. 2022, Art. no. 2240002.

[20] C. Gopalakrishnan and M. Iyapparaja, "Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning," *Int. J. Syst. Assurance Eng. Manag.*, vol. 12, pp. 1–8, Aug. 2021.

[21] M. Iyapparaja, "Effective feature selection using hybrid GA-EHO for classifying big data SIoT," *Int. J. Web Portals*, vol. 12, no. 1, pp. 12–25, Jan. 2020.

[22] S. K. Mani and I. Meenakshisundaram, "Improving quality-of-service in fog computing through efficient resource allocation," *Comput. Intell.*, vol. 36, no. 4, pp. 1527–1547, 2020.

[23] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.

[24] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.

[25] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.

[26] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[27] M. Tan and Q. V. Le, "MixConv: Mixed depthwise convolutional kernels," 2019, *arXiv:1907.09595*.

[28] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 116–131.
[29] S. Ghosh, Md. J. Mondal, S. Sen, S. Chatterjee, N. K. Roy, and S. Patnaik, "A novel approach to detect and classify fruits using ShuffleNet V2," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 163–167.
[30] R. Gomes, P. Rozario, and N. Adhikari, "Deep learning optimization in remote sensing image segmentation using dilated convolutions and ShuffleNet," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2021, pp. 244–249.
[31] Z. Chen, J. Yang, L. Chen, and H. Jiao, "Garbage classification system based on improved ShuffleNet V2," *Resour., Conservation Recycling*, vol. 178, Mar. 2022, Art. no. 106090.
[32] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[35] W. Li, O. H. Jafari, and C. Rother, "Deep object co-segmentation," in *Computer Vision–ACCV*. 2019, Berlin, Germany: Springer, pp. 638–653.
[36] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.
[37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[38] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Granada, Spain: Springer, 2018, pp. 421–429.
[39] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[40] A. Rao, J. Park, S. Woo, J.-Y. Lee, and O. Aalami, "Studying the effects of self-attention for medical image analysis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 3409–3418.
[41] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[42] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, "Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2880–2889, Apr. 2021.
[43] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate ResU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.
[44] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. Dagan Feng, and S. Wang, "AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Phys. Med. Biol.*, vol. 65, no. 5, Feb. 2020, Art. no. 055005.
[45] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.
[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[47] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
[48] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[49] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[50] H. Liu, F. Liu, X. Fan, and D. Huang, "Polarized self-attention: Towards high-quality pixel-wise regression," 2021, *arXiv:2107.00782*.

**JUNFANG FAN** received the Ph.D. degree in flight vehicle design from the Beijing Institute of Technology. He is a professor. He has been working at the School of Automation, Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science and Technology University, since 2009, focusing on intelligent guidance and control.

**JUANQIN LIU** received the master's degree in artificial intelligence from Beijing Information Science & Technology University, in June 2023. She is mainly engaged in deep learning and medical image processing related research.

**QILI CHEN** received the Ph.D. degree from the Beijing University of Technology, in 2014, focusing on pattern recognition and intelligent systems. She has been with the Department of Intelligent Science and Technology, Beijing Information Science & Technology University, since September 2014. She is an Associate Professor.

**WEI WANG** (Member, IEEE) received the Ph.D. degree in aircraft design from the Beijing Institute of Technology. He has been with the School of Aerospace Engineering, Beijing Institute of Technology, since 2009, focusing on the general design of aircraft and aircraft guidance and control. He is an Associate Professor.

**YANHUI WU** received the bachelor's degree from Hebei Medical University and the master's degree from the Beijing University of Chinese Medicine. He is an Associate Chief Physician. He focuses on gynecology clinical and scientific research by integrating traditional Chinese and Western medicine treatment of gynecological diseases and uterine and laparoscopic surgery.