IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 January 2023, accepted 20 January 2023, date of publication 27 January 2023, date of current version 1 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3240216

RESEARCH ARTICLE

Classification of Liver Fibrosis From Heterogeneous Ultrasound Image

YUNSANG JOO[1], HYUN-CHEOL PARK[2], O-JOUN LEE[3], CHANGHAN YOON[4,5], MOON HYUNG CHOI[6], AND CHANG CHOI[1], (Senior Member, IEEE)

[1]Department of Computer Engineering, Gachon University, Seongnam, Sujeong 13120, Republic of Korea
[2]Department of Artificial Intelligence, Gachon University, Seongnam, Sujeong 13120, Republic of Korea
[3]Department of Artificial Intelligence, The Catholic University, Seoul, Jongno 03083, Republic of Korea
[4]Department of Biomedical Engineering, Inje University, Gimhae, Inje 50834, Republic of Korea
[5]Department of Nanoscience Engineering, Inje University, Gimhae, Inje 50834, Republic of Korea
[6]Department of Radiology, The Catholic University, Seoul, Banpo 06591, Republic of Korea

Corresponding author: Chang Choi (changchoi@gachon.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant by the Korea Government [Ministry of Science and ICT (MSIT)] under Grant 2021R1A2B5B02087169, and in part by the Gachon University Research Fund of 2021 under Grant GCU-202103440001.

ABSTRACT With the advances in deep learning, including Convolutional Neural Networks (CNN), automated diagnosis technology using medical images has received considerable attention in medical science. In particular, in the field of ultrasound imaging, CNN trains the features of organs through an amount of image data, so that an expert-level automatic diagnosis is possible only with images of actual patients. However, CNN models are also trained on the features that reflect the inherent bias of the imaging machine used for image acquisition. In other words, when the domain of data used for training is different from that of data applied for an actual diagnosis, it is unclear whether consistent performance can be provided by the domain bias. Therefore, we investigate the effect of domain bias on the model with liver ultrasound imaging data obtained from multiple domains. We have constructed a dataset considering the manufacturer and the year of manufacturing of 8 ultrasound imaging machines. First, training and testing were performed by dividing the entire data, in a commonly used method. Second, we have utilized the training data constructed according to the number of domains for the machine learning process. Then we have measured and compared the performance on internal and external domain data. Through the above experiment, we have analyzed the effect of domains of data on model performance. We show that the performance scores evaluated with the internal domain data and the external domain data do not match. We especially show that the performance measured in the evaluation data including the internal domain was much higher than the performance measured in the evaluation data consisting of the external domain. We also show that 3-level classification performance is slightly improved over 5-level classification by mitigating class imbalance by integrating similar classes. The results highlight the need to develop a new methodology for mitigating the machine bias problem so that the model can work correctly even on external domain data, as opposed to the usual approach of constructing evaluation data in the same domain as the training data.

INDEX TERMS Domain bias, multi-domain learning, ultrasonography, liver fibrosis.

I. INTRODUCTION

Ultrasound (US) images, which can be obtained without harmful radiation being applied to the human body, are mainly used in the medical field. In abdominal radiology, US images are most used in the continuous observation of patients with liver cirrhosis or chronic hepatitis to detect hepatocellular carcinoma and evaluate the degree of liver fibrosis [1]. US images are taken using the reflected wave of a sound wave pulse [2]. Unlike superficial organs, such as the breasts and thyroid gland, the liver is located deep inside the human body. Therefore, during the process of

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

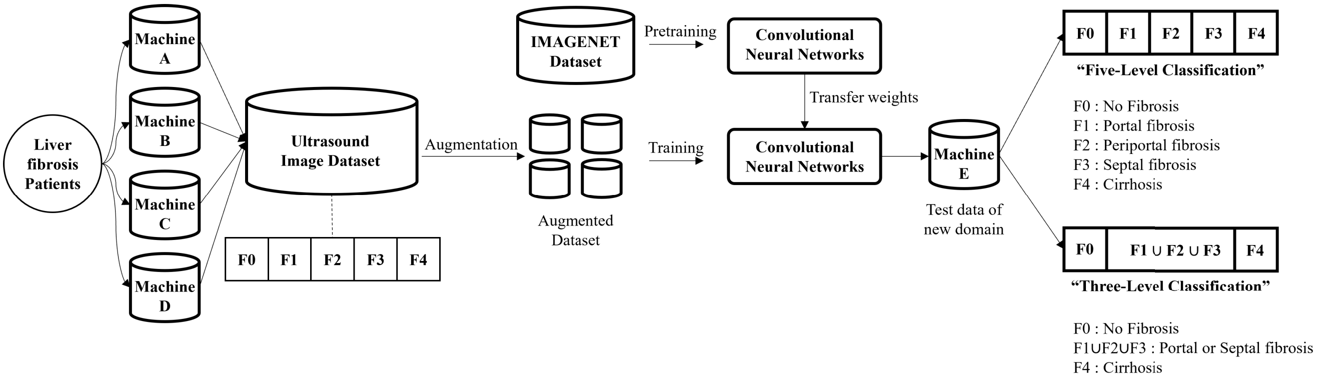

FIGURE 1. Schematic diagram of convolutional neural network using multi-domain datasets. We trained VGGNet, ResNet, DenseNet, EfficientNet, and ViT. The data from machine E is not used for training the models. Models were evaluated using five and three-level classifications.

transmitting and receiving signals inside the human body, which has numerous obstacles, the signals are weakened and the probability of exposure to various noises increases. Compared to other medical imaging technologies, diagnosis using US imaging tends to be highly dependent on the competence of an expert.

For objective diagnoses, research on US imaging diagnosis based on the use of deep convolutional neural networks (DCNNs) has been actively conducted. A DCNN is an algorithm mainly applied in imaging and has shown excellent performance in various applications, such as image segmentation and classification. US imaging diagnosis using DCNNs is objective as it eliminates individual differences in disease diagnosis and shows performance comparable to that of radiologists. Existing automated classification models were trained and evaluated using images acquired from machines limited to a specific domain. There are many types of US imaging machines, and each machine has its own noise. A model trained on images acquired from a single device is biased toward the characteristics of the corresponding imaging device. In other words, only images acquired from the same domain as the data used for learning are correctly diagnosed. However, US images used in most studies are either acquired by a single machine or used without consideration of the imaging machines [3]. Some studies have reported that this may not work effectively for images acquired from infrequently used machines [4], [5], [6]. It is difficult to guarantee the level of performance when US images obtained using a new device are used for diagnosis. Therefore, application and analysis of DCNN learning using multi-domain data is required for generalized automatic diagnosis.

In this study, liver US images obtained from 8 different US instruments were used to analyze the machine bias problem. Considering that there are several different types of US equipment (and of different ages), multi-domain data is expected to reflect real clinical situations. The liver US image data set consists of five stages of cirrhosis according to the METAVIR scoring system: no fibrosis (F0), portal fibrosis (F1), periportal fibrosis (F2), septal fibrosis (F3) and cirrhosis (F4). We use VGGNet, ResNet, DenseNet, EfficientNet and ViT, which are deep learning models mainly used for image classification. US images obtained from eight different US machines were used for the learning of each model. Finally, diagnostic validity was tested using images obtained from a new domain machine. The classification performance for each class and the effectiveness of class unification for similar symptoms were evaluated.

TABLE 1. CNN-based diagnosis study using data collected with a single US machine.

Reference	Machine	Task	Performence
[9]	ACUSON S1000	Fatty Liver Disease	0.906
[10]	ACUSON S2000	Breast Tumor	0.741
[11]	Sonosite X-Porte	Pleural Effusion	0.911
[12]	Aixplorer	Liver Fibrosis	0.937

A. RELATED WORKS

In this section, we describe related works on AI-based US image classification. Some related research works were searched from PubMed or Google Scholar engine. A variety of disease classification studies have been conducted through CNN-based deep-learning models using organ ultrasound images [7]. In many studies, deep learning models successfully diagnose diseases by training features from ultrasound images [8]. The dataset used in the study consists of images collected from single or multiple ultrasound machines. The research mentioned in Table 1 proposes an automated disease diagnosis system using data collected from a single ultrasound machine. Reddy et al. [9] proposed a framework using convolutional neural networks and transfer learning to improve the accuracy of fatty liver disease classification using ultrasound images. They validated fatty liver disease classification performance with 90.6% accuracy using the VGG-16 pre-trained with the ImageNet dataset. Ultrasound images used in the experiment were collected with Siemens's ACUSON S1000. Zhou et al. [10] proposed a new multi-task learning framework for tumor segmentation and classification in breast ultrasound images. The proposed framework verified the breast tumor segmentation and classification performance with a mean dice similarity coefficient of 0.778 and an accuracy of 74.1%. Ultrasound images used in the experiment were collected with Siemens's ACUSON S2000. Tsai et al. [11] developed a deep learning-based automated system for the automatic detection of pleural effusion in lung ultrasound images. For efficient and stable classification, a regularized spatial transformer network (Reg-STN) structure was proposed. The proposed system verified the classification performance of pleural effusion with an accuracy of 91.12%. Ultrasound images used in the experiment were collected with FujiFilm's Sonosite X-Porte. Xue et al. [12] proposed a multimodal ultrasound imaging-based radiomics transfer learning method that combines image information of gray scale modality and elastogram modality to classify liver fibrosis in liver ultrasound images. Using a model pre-trained with the ImageNet dataset, they compared models with and without transfer learning. Liver fibrosis grading performance was validated with an area under the roc curve(AUC) of 93.7%. Ultrasound images used in the experiment were collected with SuperSonic Imagine's Aixplorer. Automated diagnostic systems using data collected from a single instrument have been validated with meaningful performance. However, it has not been validated with data collected with external domains, and validation on external domain data can not guarantee consistent performance with validation on internal domain data.

TABLE 2. CNN-based diagnosis study using data collected with multiple US machines.

Reference	Machine	Task	Performance
[13]	EPIQ 7 Aplio XG	Liver Kidney Spleen Pancreas Gallbladder	0.779
[14]	LOGIQ E9 LOGIQ P3	Chronic Kidney Disease	0.856
[15]	DC-70 Exp MyLab Alpha Aplio XV Ultrasound Probes	COVID-19	0.75
[16]	ACUSON Oxana ACUSON S3000 Apolio 500 LOGIQ E9 EPIQ 7	Thyroid Nodules Breast Lesions	0.863 0.865
[17]	IU22 ATL UM-9 HDI HDI-3000 HDI-5000 LOGIQ E9 ACUSON Sequoia 128XP	Liver Fibrosis	0.764
[18]	IU22 EPIQ 7 MyLab 50 HI VISION Ascendus ALOKA Prosound F75 Voluson E8 LOGIQ E9 Vivid E9 ACUSON S2000	Thyroid Nodules	0.8732
[19]	Voluson 730 Voluson 730 expert Volusion E6 Volusion E8 Volusion E10 ALOKA SSD-a10 ACUSON S2000 TUS-X200 UGEO WS80A EPIQ 7	Fetal Brain	0.963

The study mentioned in Table 2 proposed an automatic diagnosis system using data collected from multiple ultrasound machines. Cheng and Malhi [13] evaluated the performance of transfer learning using VGGNet and CaffeNet pre-trained with the ImageNet dataset for the classification of abdominal ultrasound images. Abdominal ultrasound images were classified into 11 categories, and classification performance was verified with an accuracy of up to 77.9%. Ultrasound images used in the experiment were collected with Philips's EPIQ 7 and Toshiba's Aplio XG. Kuo et al. [14] evaluated the performance using a pre-trained ResNet on ImageNet to automatically diagnose chronic liver disease from renal ultrasound images. The classification performance was verified with an accuracy of 85.6%. Ultrasound images used in the experiment were collected with GE's LOGIQ E9 and LOGIQ P3. Roy et al. [15] presented a new deep network derived from spatial transformer networks for lung ultrasound image segmentation. They verified the segmentation performance of imaging biomarkers of COVID-19 in lung ultrasound images with a Dice score of 0.75. Ultrasound images used in the experiment were collected with Mindray's DC-70 Exp, Esaote's MyLab Alpha, Toshiba's Aplio XV, and ATL's Ultrasound Probes. Zhu et al. [16] developed and evaluated TNet and BNet using VGG-19 pre-trained on the ImageNet dataset to classify thyroid nodules and breast lesions in ultrasound images. The classification performance of thyroid nodules and breast lesions was verified with an accuracy of 86.3% and 86.5%, respectively. Ultrasound images used in the experiment were collected with Siemens' ACUSON Oxana and ACUSON S3000, Toshiba's Aplio 500, GE's LOGIQ E9, and Philips' EPIQ 7. Lee et al. [17] evaluated METAVIR score prediction performance with VGGNet pre-trained on ImageNet in liver ultrasound images. The classification performance of liver fibrosis was verified with an accuracy of 76.4%. Ultrasound images used in the experiment were collected with Philips' IU22, ATL UM-9 HDI, HDI-3000, HDI-5000 and GE's LOGIQ E9, and Siemens' ACUSON Sequoia, 128XP. Wang et al. [18] proposed a deep learning method for diagnosing thyroid nodules using multiple ultrasound images as inputs in one examination. The proposed system verified the classification performance of thyroid nodules with an accuracy of 87.32%. Ultrasound images used in the experiment were collected with Philips' IU22, EPIQ 7 and Esaote's MyLab 50, Hitachi's HI VISION Ascensus, ALOKA Prosound F75, and GE's

TABLE 3. Details of the data set used in the experiment. It was collected with 8 US machines of various companies and years of manufacture and represents the number of patients at each stage of liver fibrosis.

Machine	Manufacturer	Number of patients	Manufactured year	Fibrosis				
				F0	F1	F2	F3	F4
EUB-7500	Hitachi	182	2009	43	27	28	29	55
IU22	Philips	176	2009	88	13	7	16	52
ACUSON S2000	Siemens	106	2009	19	8	6	13	60
ACUSON Sequoia	Siemens	65	2009	13	2	4	7	39
LOGIQ E9	GE	108	2013	46	4	6	17	35
ALOKA Prosound-F75	Hitachi	129	2016	31	29	28	11	30
LOGIQ E10	GE	127	2019	38	29	11	27	22
LOGIQ S8	GE	51	2019	13	10	11	8	9

Voluson E8, LOGIQ E9, Vivid E9, and Siemens' ACUSON S2000. Xie et al. [19] evaluated the performance of a deep learning algorithm to segment and classify as normal or abnormal in fetal brain ultrasound images. Segmentation and classification performance was verified with a dice score of 0.941 and an accuracy of 96.3%. Ultrasound images used in the experiment were collected with GE's Volusion 730, Volusion 730 expert, Volusion E6, Volusion E8, Volusion E10, and Hitachi's ALOKA SSD-a10 and Siemens' ACUSON S2000, Toshiba's TUS-X200, Samsung's UGEO WS80A, and Philips's EPIQ 7. Experiments using data collected with multiple ultrasound machines constituted training and validation data with images collected in the same domain. Such a model may have generalized diagnostic performance, but no study has yet directly analyzed the domain bias. The domain bias problem of ultrasound images is a common problem, but there are not many studies that have compared the performance of internal and external verification data. In this paper, data collected with various ultrasound machines were reconstructed according to the number of machines and verified with images collected from internal or external domains.

II. METHODS

A. DATA SOURCE

US images from a tertiary university hospital (Seoul St.Mary's Hospital) were used for training and validation. Data from another university hospital (Eunpyeong St. Mary's Hospital) were used for testing. This study was licensed by the institutions of both hospitals (KC20RISI0869 and PC20RISI0229). The training/validation dataset consisted of US images acquired from eight different machines (mainly by six manufacturers), four to fifteen years old (Table 1), whereas the test dataset consisted of US images from two machines, which were three years old. Data from patients who underwent liver biopsy or liver resection at Seoul St. Mary's Hospital between 2011 and 2020 and Eunpyeong St. Mary's Hospital between 2019 and 2020 are included. In the case of a contracted liver or ascites, non-invasive methods, such as transient elastic angiography, to evaluate liver fibrosis are error-prone. Among them, data from patients who had a US liver examination within 3 months prior to biopsy or surgery were included in this study, and data from 766 patients in the training/validation set and 189 patients in the test set were included. A radiologist with 11 years of experience in US abdominal imaging reviewed all images and selected liver images using a convex probe commonly applied to the abdominal organs. Doppler US images and images showing biopsy needles were excluded. In this study, for automatic diagnosis of liver fibrosis, METAVIR scores were used to classify the status of US images. The METAVIR score used to evaluate fibrosis consists of five grades: F0, F1, F2, F3, and F4, where F0 is a clear image without fibrosis, F1 is portal fibrosis without septum and minor abnormal areas, and F2 is portal fibrosis with fewer septa and abnormalities in a wider area than F1, F3 indicates many septa and no cirrhosis and significant abnormalities, and F4 indicates liver cirrhosis in sharp contrast to the normal region. F1, F2, and F3 are the initial stages of cirrhosis, and it is difficult to discriminate between abnormal regions in these stages [20]. A visual identification of each METAVIR score using US images depends on the empirical factors of the radiologist [21]. Using US images labeled by METAVIR score, we experimented with five-level classification: F0, F1, F2, F3, and F4. In addition, we grouped classes of similar stages and performed experiments with three-stage classification: normal conditions (F0), portal fibrosis (F1, F2, and F3), and cirrhosis (F4). Classes F1, F2, and F3 can partake in single group because the boundary between these levels is ambiguous. Finally, during the experiment, we considered the effectiveness of an automated diagnosis through both five and three-level classification experiments.

B. DATA BALANCE AND PRE-PROCESSING

Liver US images were collected with eight machines, and the distribution of fibrosis stages is shown in Table 4. Data were annotated by a radiologist with specialized knowledge in this field. When the model is being trained, the class distribution ratio in the data set must be considered [22]. Unbalanced data causes overfitting or underfitting of certain

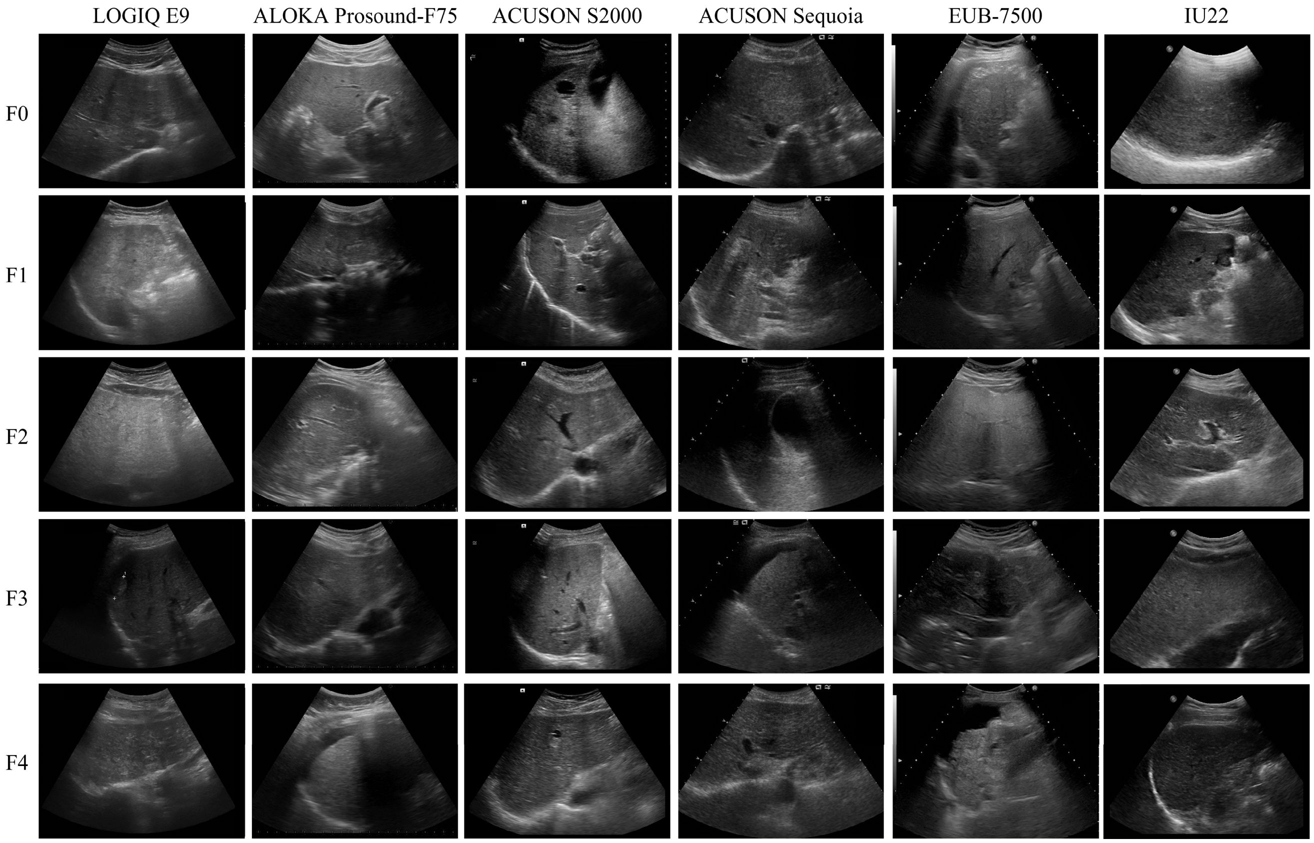

FIGURE 2. Examples of US images by US machine and stages of liver fibrosis. GE's LOGIQ E9, Hitachi's ALOKA Prosound-F75, EUB-7500, Siemens' ACUSON S2000, ACUSON Sequoia, and Philips' IU22 were used.

TABLE 4. The number of data samples per fibrosis stage. Among the advanced stages of liver fibrosis, relatively few intermediate stages, F1, F2, and F3 are present. We perform three-level classification by grouping intermediate stages.

METAVIR SCORE	The number of images
F0	2114
F1	861
F2	793
F3	857
F4	1698
Total	6323

classes when training a model [23], [24]. In particular, data on diseases that are difficult to detect at an early stage, such as cirrhosis, the number of samples that are progressing to malignancy is relatively scarcer than the number of benign or malignant samples. In general, F0 and F4 are easily obtained, and such data occupy more than half of the dataset. Class F4, in which the cirrhosis of the liver has progressed significantly, accounted for 27% of the dataset. However, the distributions of F1, F2, and F3 were relatively low because only a few patients were tested during the early stages of cirrhosis. The proportions of F1, F2, and F3 in our dataset are 13% each. Although the distribution of classes is unbalanced in the five-level classification, the distribution of classes in the three-level classification is relatively uniform at 33%, 39%, and 27%, respectively.

Data augmentation was performed to mitigate the class imbalance problem of the five-level classification [25]. Data augmentation of images using computer vision methods resulted in efficient training from limited datasets [26]. Translation, rotation, flipping, cropping, noise generation, and color jitter are used for such augmentation [27], [28]. However, if data augmentation is applied incorrectly, the inherent meaning of the original data may be damaged by the applied augmentation method. For example, if the data of a liver cirrhosis image is augmented through cropping, the cropped image can be considered normal if the cropped area is local. In this case, the augmented data may correspond to erroneous data that does not include cirrhosis. Thus, the augmentation method should be considered to preserve the inherent meaning. The liver US image has a pixel resolution of 800 × 600 and is a circular sector in shape. Random cropping was not applied to avoid damaging the liver fibrosis area. Random horizontal flips were applied to create geometric diversity. Considering the fan shape, flipping and rotation were not applied. Images resized to 224 × 224 pixel resolution and pixel values normalized were used for model training.

C. MODELS

The models were trained using VGGNet-16 [29], ResNet-50 [30], DenseNet-121 [31], EfficientNet-B0 [32], and ViT [33]. Each model commonly consists of an encoder f(θ) and a classifier g(θ). The encoder f(θ) extracts mid-level features through convolution, and the classifier g(θ) classifies the final features as class. The output values of g were normalized to probabilities using the softmax function. The objective cross-entropy function was configured such that the probability of the target class was maximized. Finally, the parameter θ was trained to optimize the objective function.

1) VGGNET

VGGNet acquired second place in the 2014 ILSVRC. With the advent of VGGNet, the depth of the network can be increased. VGGNet succeeded in network training with a depth more than twice that of AlexNet's 8-layer model and reduced the error rate of AlexNet by half in the ImageNet challenge. Models before VGGNet showed good performance by including 11 × 11 filters or 7 × 7 filters with relatively large receptive fields. However, VGGNet used a 3 × 3 kernel size filter to reduce the number of training parameters and increase the nonlinearity due to many rectified linear units. VGGNet was increasingly used for transfer learning because it was structurally simple and easy to understand.

2) RESNET

ResNet won the 2015 ILSVRC. Microsoft developed it with a layer depth of about 7, making ResNet as deep as that of Google's similar solution GoogleNet. ResNet uses a residual block to solve gradient loss and explosion. The residual block uses shortcuts to add input values to output values. Existing neural networks are trained so that $H(x) = x$, but ResNet is trained so that $F(x)$ becomes 0 by defining $H(x) = F(x) + x$. At this time, if this equation is differentiated, the added x becomes 1, solving the problem of gradient loss. The number of parameters and the complexity of the network were reduced by using a bottleneck design with a 1 × 1 convolution layer added.

3) DENSENET

DenseNet was introduced at CVPR 2017. DenseNet solves the vanishing gradient problem in a slightly different way from ResNet, and can achieve high performance even in low-depth networks. ResNet combines input values of previous layers through add operations, whereas DenseNet improves information flow by connecting all layers through concatenating operations. The vanishing gradient problem is alleviated by directly passing the values of the initial feature map to the values of the last feature map. Since the concatenation operation requires the size of the feature map to be the same, a dense block is introduced to make the size of the connected feature map constant. Similar to ResNet, the amount of computation is reduced by using a bottleneck layer that controls the input value channel.

4) EFFICIENTNET

EfficientNet was introduced at ICML in 2019. To improve the performance of the model, the depth, width, and resolution of the model were adjusted. As the depth increases, more complex features can be captured, but it becomes difficult to learn due to the problem of vanishing gradient. Increasing the width of each layer increases the accuracy, but the amount of computation increases in proportion to the square. If the resolution of the input image is increased, detailed features can be learned, but the amount of computation increases in proportion to the square. When all three (depth, width, and resolution) are increased to a certain extent, the size of the model increases, but the accuracy decreases. Unlike the existing method of manually adjusting these three parameters, EfficientNet achieved state-of-the-art performance with a smaller model by applying a complex scaling method that can be automatically adjusted.

5) VIT

ViT was introduced at ICLR in 2020. Transformers have been limitedly applied to the field of natural language processing, where input data have one-dimensional sequences, and the field of computer vision, where input data have three-dimensional sequences. ViT was the first to introduce transformers to computer vision, and it showed performance similar to or higher than that of state-of-the-art models. ViT uses a three dimension sequence converter by dividing the image into patches and using the same concept as the token of NLP. In this way, the computer vision task does not depend on the CNN structure and can achieve better performance than state-of-the-art models at about one-fifteenth of the computational cost. In this experiment, we used ViT-B/16 trained with ImageNet 1k data and applied it with a patch size of 16 × 16.

D. TRANSFER LEARNING

We applied transfer learning to model training [34]. Learning models from the scratch is valid only when the training data samples are more than 5000 per class [35]. However, the sharing of medical data from hospitals has been stopped according to the personal information protection act [36]. In addition, the number of patients and statistics on disease are limited locally. The restriction of training data causes bias and model overfitting or underfitting [37]. Transfer learning complements parameter optimization with small amounts of training data using models trained on a wide range of data sets from different domains [38]. In this study, a model pre-trained with ImageNet was used to fine-tune the model with the liver US images. During pretraining with ImageNet of 1000 classes, the model learns to extract high-level features from images. Therefore, the pre-trained model's convolutional filters are more optimized than a model trained from scratch when learning new data. It would be ideal to use a model pre-trained by ultrasound imaging, but it is difficult to acquire a quantity comparable to ImageNet for reasons such as patient privacy issues. In particular, in most medical image classification studies, such as US, MRI, CT, and endoscopic images, the performance of transfer learning using a pre-trained model with ImageNet has been verified to be effective [7], [12], [13], [39]. To perform fine-tuning, we adapted the output layer configuration to the number of classes in a given dataset. We classified the liver US images were classified into 5 detailed stages and 3 grouped stages. When the post-trained data set and the pre-trained data set are similar, the model can obtain effective results even if the convolutional layer is frozen. However, in post-training with US images, all parameters are retrained because the intrinsic properties are different from ImageNet and medical US images.

TABLE 5. Dataset configuration to evaluate model performance without domain distinction. Data obtained from the six types of machines were split in an 8:2 ratio for training and testing.

Machines	Class		The number of images
LOGIQ E9 ALOKA Prosound-F75 ACUSON S2000 ACUSON Sequoia EUB-7500 IU22	Train	F0	1628
		F1	626
		F2	581
		F3	608
		F4	1281
	Test	F0	411
		F1	160
		F2	148
		F3	156
		F4	325

E. MODEL TRAINING

In this section, we describe the details of the model training. To evaluate the performance of multi-domain learning, training and test data were obtained from eight different US machines. The first training dataset contains data from LOGIQ E9 and ALOKA Prosound-F75 of similar age. The second training dataset included data collected from the LOGIQ E9, ALOKA Prosound-F75, ACUSON S2000, and ACUSON Sequoia. The third training dataset included data collected with the LOGIQ E9, ALOKA Prosound-F75, ACUSON S2000, ACUSON Sequoia, EUB-7500, and IU22. The test data consisted of: 1) Validation dataset without distinction of the domain, 2) validation datasets collected with LOGIQ E9 and ALOKA Prosound-F75 for evaluation in the internal domain; 3) Validation datasets collected with LOGIQ E10 and LOGIQ S8 for evaluation in external domains. The ratio of the training and validation sets was 8:2. The experimental data composition is shown in Table 5, Table 6. All models used in the experiment were pre-trained using the ImageNet dataset [40]. Cross-entropy loss with negative log-likelihood was used as the loss function for the training phase. The optimization algorithm and learning-rate scheduler used the Adam optimizer and cosine annealing LR, respectively. The batch size was set to 64 and the initial learning rate was started at 0.001 and adjusted to a value close to zero with 50 epoch cycles by the scheduler.

TABLE 6. Dataset configuration to evaluate model performance on internal and external domain data. The model is trained on data acquired from two, four, and six devices, respectively, and evaluated in two domains same as an internal domain and against data acquired from two external domains.

Machines		Class		The number of images
2 Machines	LOGIQ E9 ALOKA Prosound-F75	Train	F0	443
			F1	137
			F2	176
			F3	188
			F4	296
4 Machines	LOGIQ E9 ALOKA Prosound-F75 ACUSON S2000 ACUSON Sequoia		F0	623
			F1	225
			F2	234
			F3	275
			F4	652
6 Machines	LOGIQ E9 ALOKA Prosound-F75 ACUSON S2000 ACUSON Sequoia EUB-7500, IU22		F0	1628
			F1	626
			F2	581
			F3	608
			F4	1281
Internal domain machines	LOGIQ E9 ALOKA Prosound-F75	Test	F0	112
			F1	35
			F2	45
			F3	48
			F4	76
External domain machines	LOGIQ E10 LOGIQ S8		F0	75
			F1	75
			F2	64
			F3	93
			F4	92

F. EVALUATION METRICS

The diagnostic model was evaluated on the cirrhosis images of the test set using the metrics of accuracy (1), precision (2), recall (3), and F1-score (4) [41]. Here, TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Accuracy is defined as the number of correctly predicted data points divided by the total number of data points. Precision is defined as the proportion of data that are actually positive among the data predicted as positive. Recall is defined as the ratio of the data predicted to be positive to the actual positive data. F1-score is the harmonic mean of precision and recall.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1-score = 2*\frac{Precision*Recall}{Precision+Recall} \tag{4}$$

III. RESULTS

Table 7 shows the five-level classification performance of the models consisting of F0/F1/F2/F3/F4 as listed in Table 3. ResNet had the highest accuracy at 85.92%, and the average accuracy of the five models was 84.37%. Among all the models, the classification performances of F0 and F4 were higher than those of F1, F2 and F3. This result indicated that the classification of F1, F2, and F3 was weak. Thus, the classes with relatively small training data were underfitted due to data imbalance.

TABLE 7. Five-level classification performance of the model on datasets constructed without distinction of domains. Precision, Recall, and F1-Score of F0/F1/F2/F3/F4 were measured.

Model	Accuracy	Class	Precision	Recall	F1-score
VGGNet	0.8317	F0	0.8637	0.8637	0.8637
		F1	0.7848	0.7750	0.7799
		F2	0.8605	0.7500	0.8014
		F3	0.7383	0.7051	0.7213
		F4	0.8442	0.9169	0.8791
ResNet	**0.8592**	F0	0.8744	0.8978	0.8860
		F1	0.8313	0.8313	0.8313
		F2	0.9213	0.7905	0.8509
		F3	0.7548	0.7500	0.7524
		F4	0.8780	0.9077	0.8926
DenseNet	0.8417	F0	0.8697	0.8929	0.8812
		F1	0.7939	0.8187	0.8062
		F2	0.8740	0.7500	0.8073
		F3	0.7939	0.6667	0.7247
		F4	0.8366	0.9138	0.8735
EfficientNet	0.8517	F0	0.8802	0.8759	0.8780
		F1	0.8506	0.8187	0.8344
		F2	0.8483	0.8311	0.8396
		F3	0.7852	0.7500	0.6721
		F4	0.8484	0.8954	0.8713
ViT	0.8342	F0	0.8204	0.9002	0.8585
		F1	0.8456	0.7875	0.8155
		F2	0.8759	0.8108	0.8421
		F3	0.7812	0.6410	0.7042
		F4	0.8507	0.8769	0.8636

Table 8 shows the three-level classification performance of the models. Compared to the five-level classification, the classification performances of F0 and F4 were slightly decreased and the classification performance of F123 was slightly improved, which increased the overall model performance. The five-level classification model showed weak performance in F1/F2/F3, but the F1 score was relatively uniform for the three-level classification model. ViT had the highest accuracy at 87.92% and the average accuracy of the five models was 86.45%.

A. PERFORMANCE OF THE MODELS ON THE INTERNAL AND EXTERNAL DOMAIN DATASET

The models were individually trained on three different domain data and evaluated using internal and external data.

TABLE 8. Three-level classification performance of the model on datasets constructed without distinction of domains. Combine F1, F2, and F3 into one class and measure the precision, recall, and F1-score of F0/F123/F4.

Model	Accuracy	Class	Precision	Recall	F1-score
VGGNet	0.8542	F0	0.8721	0.8127	0.8413
		F123	0.8320	0.8750	0.8529
		F4	0.8631	0.8769	0.8716
ResNet	**0.8792**	F0	0.8859	0.8686	0.8771
		F123	0.8691	0.8728	0.8710
		F4	0.8852	0.9015	0.8933
DenseNet	0.8583	F0	0.8813	0.8491	0.8649
		F123	0.8552	0.8276	0.8412
		F4	0.8366	0.9138	0.8735
EfficientNet	0.8775	F0	0.8916	0.8808	0.8862
		F123	0.8599	0.8599	0.8599
		F4	0.8848	0.8985	0.8916
ViT	0.8533	F0	0.8585	0.8564	0.8575
		F123	0.8525	0.8470	0.8497
		F4	0.8480	0.8585	0.8532

TABLE 9. Classification performance of the models on the internal domain dataset.

The number of machines	Model	Accuracy	
		5-class	3-class
2 machines	VGGNet	0.8070	0.8544
	ResNet	0.8259	0.8703
	DenseNet	0.8133	0.8481
	EfficientNet	0.8101	**0.8734**
	ViT	**0.8354**	0.8418
4 machines	VGGNet	0.8070	0.8418
	ResNet	**0.8513**	0.8639
	DenseNet	0.8259	0.8576
	EfficientNet	0.8196	**0.8671**
	ViT	0.8449	0.8418
6 machines	VGGNet	0.7943	0.8513
	ResNet	0.8196	**0.8703**
	DenseNet	0.8165	0.8576
	EfficientNet	0.8101	0.8576
	ViT	**0.8323**	0.8291

Table 9 shows the performance of the multi-domain training model on the data collected with LOGIQ E9 and ALOKA Prosound-F75 of the internal domains. In the five-level classification, the models trained using data from two, four, and six machines had an accuracy of 83.54%, 85.13%, and 83.23%, respectively. In the three-level classification, the models trained using data from two, four, and six machines had an accuracy of 87.34%, 86.71%, and 87.03%, respectively. Table 10 shows the performance of the multi-domain training model on the data collected from the LOGIQ E10 and LOGIQ S8 of the external domains. In the five-level classification, the accuracies were 27.32%, 27.32%, and 26.32% for models trained using data from two, four, and six machines, respectively. In the three-level classification, the accuracies

TABLE 10. Classification performance of the models on the external domain dataset.

The number of machines	Model	Accuracy	
		5-class	3-class
2 machines	VGGNet	**0.2732**	**0.4737**
	ResNet	0.1980	0.4662
	DenseNet	0.2155	0.4561
	EfficientNet	0.2381	0.4110
	ViT	0.2581	0.4110
4 machines	VGGNet	0.2431	0.4586
	ResNet	0.2431	0.4135
	DenseNet	0.2531	**0.4687**
	EfficientNet	0.2707	0.4010
	ViT	**0.2732**	0.4586
6 machines	VGGNet	0.2281	**0.4586**
	ResNet	**0.2632**	0.4536
	DenseNet	0.2331	0.4286
	EfficientNet	0.2431	0.3810
	ViT	0.2281	0.4386

were 47.37%, 46.87%, and 45.86% for models trained using data from two, four, and six machines, respectively.

IV. DISCUSSION

A deep learning-based automatic diagnosis model using ultrasound images classifies ultrasound images to be used for diagnosis by utilizing the feature extraction function acquired through a large number of ultrasound images. Thus, if the training data was acquired only for a specific device, there is a possibility that the model was trained by reflecting the bias of the device. That is, it may show different results from the expected performance depending on which device the image is acquired from. In general, to evaluate a classification model is trained and evaluated by dividing a portion of the data. We focus on whether there is a performance difference depending on which equipment the images used for learning and evaluation are acquired. To evaluate from internal domain data, training data consisting of multi-domains is constructed, and images of the same domain are used for evaluation. When measuring the performance of a model for evaluation on external domain data, the image acquired from specific equipment is excluded and trained, and the excluded image is used for evaluation. Generally, when images acquired from all domains were trained and evaluated at a certain ratio, the classification performance was higher than when evaluated in the internal domain. When evaluated in the internal domain, the classification performance was higher than when evaluated in the external domain. In the case of learning with data composed of two or more domains, when evaluated from data separated from the training data, it was measured higher than when evaluated with a single domain. As such, it was found that when multiple domains were merged into one training and evaluation set, a new domain was formed and reflected in the model. Unless the model is acquired and evaluated on a specific machine, more generalized training data should be used so that the model does not learn the bias of a specific machine.

V. CONCLUSION

When the data used for training is not generalized, deep learning models train with biases in the data. Since the model depends mostly on the training data, the inclusion of bias in the training data in the results is unavoidable. Especially medical US data is more prone to potential bias depending on the skill level of the experts collecting it or the type of machine collecting the data. In this study, We demonstrated the application of deep learning for the automatic diagnosis of cirrhosis using liver US images to analyze it. To alleviate the problem of partial or insufficient data for a specific class, we used transfer learning and data augmentation methods. Additionally, classification was performed by grouping F1/F2/F3 patients with similar symptoms. It is possible to obtain a more effective performance than the 5-level classification in the 3-level classification (normal/progressive/severe). We also utilized internal and external domain data to analyze machine bias. The trained model classified it correctly from internal domain data acquired on the same machine as the data used for training. On the other hand, external domain data acquired from new machines showed biased results that failed to classify them properly. This means that the deep learning model has not yet been trained to generalize enough to classify images acquired by the new machine. Thus, it is interpreted that the model has formed a bias that leads to a particular outcome of the restricted domain data. The results of these experiments could potentially be useful in alleviating the bias problem that is unavoidably caused by limited machinery.

REFERENCES

[1] A. Tang, G. Cloutier, N. M. Szeverenyi, and C. B. Sirlin, "Ultrasound elastography and MR elastography for assessing liver fibrosis: Part 1, principles and techniques," *Amer. J. Roentgenol.*, vol. 205, no. 1, pp. 22–32, 2015, doi: 10.2214/AJR.15.14552.

[2] U. Jung and H. Choi, "Active echo signals and image optimization techniques via software filter correction of ultrasound system," *Appl. Acoust.*, vol. 188, Jan. 2022, Art. no. 108519, doi: 10.1016/j.apacoust.2021.108519.

[3] C. DeBrusk, "The risk of machine-learning bias (and how to prevent it)," MIT Sloan Manage. Rev., 2018. [Online]. Available: https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it

[4] M. Blaivas, L. N. Blaivas, and J. W. Tsung, "Deep learning pitfall: Impact of novel ultrasound equipment introduction on algorithm performance and the realities of domain adaptation," *J. Ultrasound Med.*, vol. 41, no. 4, pp. 855–863, Apr. 2022, doi: 10.1002/jum.15765.

[5] W. K. Moon, Y.-W. Lee, H.-H. Ke, S. H. Lee, C.-S. Huang, and R.-F. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 190, Jul. 2020, Art. no. 105361, doi: 10.1016/j.cmpb.2020.105361.

[6] Z. Cao, L. Duan, G. Yang, T. Yue, and Q. Chen, "An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures," *BMC Med. Imag.*, vol. 19, no. 1, p. 51, Jul. 2019, doi: 10.1186/s12880-019-0349-x.

[7] M. Dan, L. Zhang, G. Cao, W. Cao, G. Zhang, and H. Bing, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," *IEEE Access*, vol. 5, pp. 5804–5810, 2017, doi: 10.1109/ACCESS.2017.2689058.

[8] S. Liu, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019, doi: 10.1016/j.eng.2018.11.020.

[9] D. S. Reddy, R. Bharath, and P. Rajalakshmi, "A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging," in *Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–5, doi: 10.1109/HealthCom.2018.8531118.

[10] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, and D. Shen, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918, doi: 10.1016/j.media.2020.101918.

[11] C.-H. Tsai, "Automatic deep learning-based pleural effusion classification in lung ultrasound images for respiratory pathology diagnosis," *Phys. Medica*, vol. 83, pp. 38–45, Mar. 2021, doi: 10.1016/j.ejmp.2021.02.023.

[12] L.-Y. Xue, Z.-Y. Jiang, T.-T. Fu, Q.-M. Wang, Y.-L. Zhu, M. Dai, W.-P. Wang, J.-H. Yu, and H. Ding, "Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis," *Eur. Radiol.*, vol. 30, no. 5, pp. 2973–2983, May 2020, doi: 10.1007/s00330-019-06595-w.

[13] P. M. Cheng and H. S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," *J. Digit. Imag.*, vol. 30, no. 2, pp. 234–243, 2017, doi: 10.1007/s10278-016-9929-2.

[14] C.-C. Kuo, C.-M. Chang, K.-T. Liu, W.-K. Lin, H.-Y. Chiang, C.-W. Chung, M.-R. Ho, P.-R. Sun, R.-L. Yang, and K.-T. Chen, "Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning," *NPJ Digit. Med.*, vol. 2, no. 1, p. 29, Apr. 2019, doi: 10.1038/s41746-019-0104-2.

[15] S. Roy, "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2676–2687, Aug. 2020, doi: 10.1109/TMI.2020.2994459.

[16] Y.-C. Zhu, A. AlZoubi, S. Jassim, Q. Jiang, Y. Zhang, Y.-B. Wang, X.-D. Ye, and H. Du, "A generic deep learning framework to classify thyroid and breast lesions in ultrasound images," *Ultrasonics*, vol. 110, Feb. 2021, Art. no. 106300, doi: 10.1016/j.ultras.2020.106300.

[17] J. H. Lee, I. Joo, T. W. Kang, Y. H. Paik, D. H. Sinn, S. Y. Ha, K. Kim, C. Choi, G. Lee, J. Yi, and W.-C. Bang, "Deep learning with ultrasonography: Automated classification of liver fibrosis using a deep convolutional neural network," *Eur. Radiol.*, vol. 30, no. 2, pp. 1264–1273, Feb. 2020, doi: 10.1007/s00330-019-06407-1.

[18] L. Wang, L. Zhang, M. Zhu, X. Qi, and Z. Yi, "Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101665, doi: 10.1016/j.media.2020.101665.

[19] H. N. Xie, N. Wang, M. He, L. H. Zhang, H. M. Cai, J. B. Xian, M. F. Lin, J. Zheng, and Y. Z. Yang, "Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal," *Ultrasound Obstetrics Gynecol.*, vol. 56, no. 4, pp. 579–587, Oct. 2020, doi: 10.1002/uog.21967.

[20] A. Tang, G. Cloutier, N. M. Szeverenyi, and C. B. Sirlin, "Ultrasound elastography and MR elastography for assessing liver fibrosis: Part 2, diagnostic performance, confounders, and future directions," *Amer. J. Roentgenol.*, vol. 205, no. 1, pp. 33–40, 2015, doi: 10.2214/AJR.15.14553.

[21] K. Patel and G. Sebastiani, "Limitations of non-invasive tests for assessment of liver fibrosis," *JHEP Rep.*, vol. 2, no. 2, Apr. 2020, Art. no. 100067, doi: 10.1016/j.jhepr.2020.100067.

[22] A. Smith, K. Baumgartner, and C. Bositis, "Cirrhosis: Diagnosis and management," *Amer. Family Physician*, vol. 100, no. 12, pp. 759–770, Dec. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/31845776

[23] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci.*, vol. 513, pp. 429–441, Mar. 2020, doi: 10.1016/j.ins.2019.11.004.

[24] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

[25] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.

[26] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[27] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[28] N. Parmar, "Image transformer," in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 80, Jul. 2018, pp. 4055–4064. [Online]. Available: https://proceedings.mlr.press/v80/parmar18a.html

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE CVPR*, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[32] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[34] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2018, pp. 270–279, doi: 10.1007/978-3-030-01424-7_27.

[35] M. Shaha and M. Pawar, "Transfer learning for image classification," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 656–660, doi: 10.1109/ICECA.2018.8474802.

[36] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[37] I. Bilbao and J. Bilbao, "Overfitting problem and the over-training in the era of data: Particularly for artificial neural networks," in *Proc. 8th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2017, pp. 173–177, doi: 10.1109/INTELCIS.2017.8260032.

[38] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019, doi: 10.1186/s40537-019-0276-2.

[39] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020, doi: 10.3390/app10093233.

[40] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," 2020, *arXiv:2012.00364*.

[41] L. Alzubaidi, J. Zhang, A. J. Humaidi, and A. Al-Dujaili, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

YUNSANG JOO received the B.S. degree in computer engineering from Gachon University, Seongnam, South Korea, in 2022. His research interests include machine learning, artificial intelligence, computer vision, and medical imaging analysis.

HYUN-CHEOL PARK received the B.S. and M.S. degrees in computer engineering from Chosun University, Gwangju, South Korea, in 2015 and 2017, respectively, and the Ph.D. degree in IT convergence engineering from Gachon University, Seongnam, South Korea, in 2022. He is currently working as a Research Professor with the Department of AI Software, Gachon University. His research interests include machine learning, artificial intelligence, computer vision, and medical imaging analysis.

O-JOUN LEE received the B.Eng. degree in software science from Dankook University, in 2015, and the Ph.D. degree in computer science and engineering from Chung-Ang University, in 2019. He has been an Assistant Professor with The Catholic University of Korea, Republic of Korea, since September 2021. Also, he was a full-time Researcher with the Pohang University of Science and Technology, Republic of Korea, from September 2019 to August 2021. He has applied the networked data analysis models and methods to various unstructured data, such as social media, bibliographic data, medical knowledge base, and traffic flow data. His research interests include networked data analysis based on unsupervised/self-supervised representation learning and graph convolutional networks.

CHANGHAN YOON received the M.S. and Ph.D. degrees in electronic engineering from Sogang University, Seoul, South Korea, in 2009 and 2013, respectively. He was a Postdoctoral Research Associate with NIH Resource Center for Medical Ultrasonic Transducer Technology, University of Southern California, Los Angeles, CA, USA and the Georgia Institute of Technology, Atlanta, GA, USA. He is currently an Assistant Professor of biomedical engineering with Inje University, Gyengnam, South Korea. His current research interests include medical ultrasound and photoacoustic imaging systems and their clinical applications and ultrasound microbeams.

MOON HYUNG CHOI received the B.S., M.S., and Ph.D. degrees in radiology from The Catholic University of Korea, Seoul, South Korea, in 2009, 2016, and 2017, respectively. She is currently working as an Assistant Professor at the Department of Radiology, Eunpyeong St. Mary's Hospital, The Catholic University of Korea. Her research interests include prostate imaging, hepatobiliary pancreas imaging, and abdominal/urinary imaging.

CHANG CHOI (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from Chosun University, in 2005, 2007, and 2012, respectively. He has been an Assistant Professor with Gachon University, since 2020. He has authored more than 50 publications, including papers in prestigious journals/conferences, such as *IEEE Communications Magazine*, IEEE Transactions on Industrial Informatics, IEEE Transactions on Information Forensics and Security, IEEE Transactions on Sustainable Computing, IEEE Internet of Things Journal, *Information Sciences*, and *Future Generation Computer Systems*. His research interests include intelligent information processing, semantic web, smart IoT systems, and intelligent system security. He received academic awards from the Graduate School of Chosun University, in 2012. He also received the Korean Government Scholarship for graduate students (Ph.D. course) in 2008. He has served or is currently serving on the organizing or program committees of international conferences and workshops, such as ACM RACS, EAIBDTA, IE, ACM SAC, and IEEE CCNC/SeCHID. He has also served as a Guest Editor for high profile journals, such as IEEE Transactions on Industrial Informatics, *Future Generation Computer Systems*, *Applied Soft Computing*, *Multimedia Tools and Applications*, *Journal of Ambient Intelligence and Humanized Computing*, *Concurrency and Computation: Practice and Experience*, *Sensors*, and *Autosoft*.