IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 October 2023, accepted 12 December 2023, date of publication 18 December 2023,
date of current version 26 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3344579

RESEARCH ARTICLE

Classification of Companion Animals' Ocular Diseases: Domain Adversarial Learning for Imbalanced Data

MARY G. NAM[1] **AND SUH-YEON DONG**[2], **(Member, IEEE)**
[1]Department of Mathematics, Sookmyung Women's University, Seoul 04310, South Korea
[2]Division of Artificial Intelligence Engineering, Sookmyung Women's University, Seoul 04310, South Korea

Corresponding author: Suh-Yeon Dong (sydong@sookmyung.ac.kr)

This work was supported in part by the Technology development Program funded by the Ministry of SMEs and Startups (MSS), South Korea, under Grant S3246381; in part by the National Research Foundation of Korea (NRF); and in part by the Ministry of Science and ICT (MSIT), South Korea, under the ICT Challenge and Advanced Network of HRD (ICAN) Program supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP) under Grant IITP-2022-RS-2022-00156299.

ABSTRACT In contrast to the widespread implementation of computer-aided diagnosis of human diseases, the limited availability of veterinary image datasets has hindered its application in animals. Additionally, while most medical imaging data are captured in clinical settings, such as optical coherence tomography and fundus photography, diagnosis based on digital camera or smartphone images can be more beneficial for pet owners. This study specifically focuses on achieving generalization between screening environments, aiming to accurately diagnose diseases using casual images obtained by pet owners, despite the majority of training images being captured with specialized equipment in hospitals. Given these challenges and the significant role of computer-aided diagnosis in veterinary science, this study aims to develop a practical deep-learning framework for classifying ocular surface disease images in companion animals. The dataset used in this study consists of diverse ocular disease images of canines and felines obtained through slit lamps and digital cameras. The proposed approach includes two layers of labels for multitask learning and a gradient reversal layer based on normalized feature maps. We achieved 84.7% and 65.4% accuracy for the total dataset of canine and feline, respectively. For the camera domain in particular, canines and felines reached 86.2% and 73.2% accuracy, respectively.

INDEX TERMS Computer-aided diagnosis, animal ocular disease, multitask learning, domain adversarial learning.

I. INTRODUCTION

In recent years, the rise of computer-aided diagnosis (CAD) has improved the field of ophthalmology medicine to assist healthcare professionals. By analyzing various ocular imaging modalities such as optical coherence tomography (OCT), fundus photography, and visual field tests, CAD systems assist ophthalmologists in detecting subtle abnormalities that might be difficult to identify with the naked eye.

However, in contrast to the wide implementation of CAD in diagnosing human diseases, the limited accessibility of veterinary image datasets has hindered its application for animals [1]. Obtaining consent from patients is challenging, and veterinarians avoid taking images during visits because of the busy environment in hospitals. Despite the shortage of available data, deep-learning (DL) applications can be in high demand for both veterinarians and pet guardians in various aspects [2]. According to a survey on low-income pet guardians from the Vancouver Humane Society, all 12 participants stated that their financial situation was negatively affected by COVID-19 and that the limited vet services, with high examination cost, exacerbated their condition [3]. Many veterinary clinics also suffer from a shortage of working veterinarians. Failing to monitor health

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas.

conditions consistently can be detrimental to animals, and some diseases can be transmissible to humans [4], [5], [6]. Especially, diseases affecting the cornea and lens are common in small animals such as dogs and cats, often leading to irreversible vision loss [7], [8].

In light of these challenges and the critical role of CAD in veterinary science, DL applications have the potential to enhance the accuracy and efficiency of diagnosis, ultimately leading to timely interventions and improved outcomes for animal patients. This study specifically focuses on achieving generalization between screening environments, aiming to accurately diagnose diseases using casual images obtained by pet owners, despite the majority of training images being captured with specialized equipment in hospitals. To address this, our proposed method involves learning disease-focused, domain-invariant features.

The remainder of this article is structured as follows: Section II discusses related works on computer-aided diagnosis in ophthalmology and veterinary science. Next, in Section III, we present our proposed method in detail. Section IV presents the experimental settings and results of our study. Finally, in Section V, we discuss the remaining limitations.

II. RELATED WORK

A. TRANSFORMER-BASED IMAGE ANALYSIS

For an extended duration, convolutional neural networks (CNNs) have been robust in CAD using medical images not only because of their remarkable ability to analyze data but also because of their computational efficiency [9], [10]. However, a CNN is prone to overfitting due to its image-specific inductive bias and has a low capability for capturing spatial information over a long distance. Afterward, with the robustness of Transformer for natural language processing(NLP), transformer-based models have been applied in various tasks for computer vision and have produced state-of-the-art performance [11], [12], [13], [14]. Vision Transformer (ViT), proposed in 2020, can capture the global context in an image; it may require a large dataset that is generally impractical to obtain in the medical field [15]. Moreover, executing global self-attention increases the computational complexity to be quadratic to the input image size.

Recently, the Swin-Transformer was proposed to address the issues of ViT by conducting self-attention in local windows containing small patches [16]. The patch size varies with the depth of the transformer layers, as they start with the smallest size and gradually merge with neighboring patches in deeper layers. The Swin-Transformer combines the advantages of the ViT and convolutional networks; its self-attention mechanism based on shifting windows reduces the computational complexity to be linear to the input image size, and it can capture both local and global information in an image [17].

In medical imaging tasks, their applications have been acknowledged, particularly for segmentation and classification [18], [19]. Lei, Z et al. performed lung segmentation and classification [18], [19], [20]. Lei executed lung segmentation using UNet, followed by identifying COVID-19 from the segmented image using a Swin-Transformer as the backbone architecture, which outperformed CNN models. Ali et al. proposed Swin UNETR, which uses a Swin-Transformer as the encoder and CNN as the decoder for the semantic segmentation of brain tumors, outperforming nnU-Net and TransBTS [18], [21], [22].

We established the first Swin-Transformer-based framework to identify eye diseases in companion animals, which outperforms multiple powerful convolutional networks.

B. DEEP-LEARNING FOR OPTHALMOLOGY

There has been increasing attention given to studies on the application of DL algorithms for the classification of eye-related diseases based on image analysis. Junayed et al. introduced CataractNet, a deep neural network for automatic cataract detection in fundus images [23]. The proposed network achieves superior performance compared to existing methods, with an average accuracy of 99.13% while reducing computational cost and running time. Li et al. proposed a DL system for classifying keratitis, other corneal abnormalities, and normal corneas based on slit-lamp images [22]. Christopher identified glaucomatous damage in optic nerve head (ONH) fundus images [24]. The study of Aranha covered cataracts, diabetic retinopathy, excavation and blood vessels, but individual binary classification networks were used to distinguish between normal and abnormal images [25]. Most of the research in the field of ophthalmological image analysis was confined to some selected set of diseases, e.g. cataracts, corneal diseases, and glaucoma [23]. In contrast, we use a single model to identify multiple diseases that arise in distinct locations.

Hao et al. presented a novel hierarchical framework for classifying fine-grained corneal diseases from ocular surface slit-lamp images [26]. This study approaches the fine-grained classification by hierarchical labels, which was employed in our work. Chea et al. used ResNet-50 to simultaneously classify diabetic retinopathy, glaucoma, and age-related macular degeneration in fundus photographs with peak and average accuracies of 91.16% and 85.79%, respectively [27]. There are also existing work regarding artificial intelligence technology using digital camera or smartphone images [5], [28]. The latter approaches simplify the complicated operation of screening processes to potentially suggest a practical solution.

C. DEEP-LEARNING FOR ANIMALS

For the application of DL in veterinary sciences, notably, most medical imaging data are captured in clinical settings such as X-rays, magnetic resonance imaging (MRI) scans, or computed tomography (CT) scans. Ergün et al. utilized deep neural networks and support vector machines to determine dog maturity, date fractures, and detect fractures in X-Ray images of long bones [29]. By integrating data augmentation techniques, the ResNet-50 model achieved 0.80, 0.83, and 0.89 accuracy for each task. Banzato et al. proposed a framework of DenseNet121 and ResNet50 pre-trained on a large-scale dataset of everyday images, ImageNet, and fine-tuned on canine images to classify canine thoracic radiographs [30], [31]. Dumortier et al. used ResNet50V2 pre-trained on ImageNet, fine-tuned on human chest X-rays, and fine-tuned again on feline thoracic radiograph images [32].

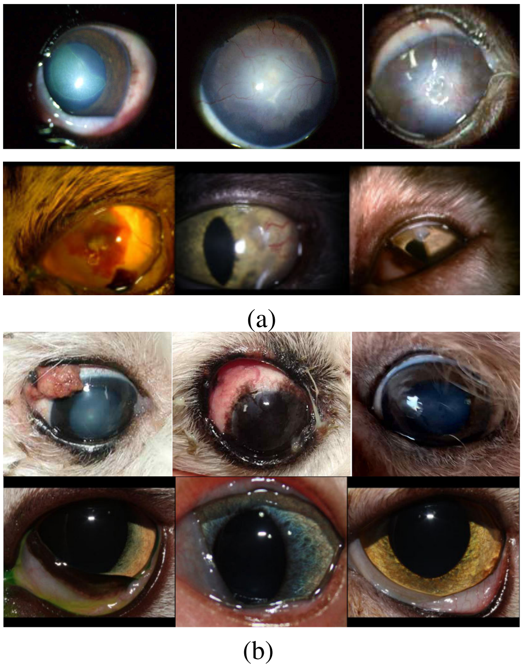
Some existing work utilized external surface images. Kim et al. assessed the condition of dogs' ocular surfaces to detect dry eye disease using an object detection model, YOLOv5 [33]. By analyzing ocular surface video images, the method achieved 0.995 mean average precision, showing promise for object detection in veterinary medicine. Kim et al. detected the severity of corneal ulcers in canine eye images with an accuracy in the range of 90%–100% for all experimented CNN models, Inception, ResNet, and VGG, pre-trained on ImageNet and fine-tuned weights of the fully connected layer [34]. However, the aforementioned studies primarily rely on slit-lamp images, which are not readily accessible for pet guardians. Additionally, the classification models in these studies are trained to assess the severity of a specific disease, rather than to differentiate between various diseases. In contrast, to the best of our knowledge, this is the first study that employs external ocular surface images using digital camera images of both canines and felines.

D. DOMAIN ADVERSARIAL LEARNING

Transfer learning is a well-known technique to improve performance on a small dataset by using a pre-trained model on a large-scale dataset [35], [36]. However, simply fine-tuning a pre-trained network has limited effect when the large and small dataset hold dissimilar characteristics and distributions [37], [38], [39]. On the other hand, domain adversarial learning is more useful when the source and target domains exhibit significant differences [40].

The main idea behind domain adversarial learning involves training a feature extractor, domain classifier, and task classifier simultaneously [41]. The feature extractor learns domain-invariant features from input images, while the domain classifier tries to differentiate between source and target domain images based on those features. The gradient reversal layer or domain confusion loss helps the feature extractor generate features that confuse the domain classifier, enhancing domain adaptation [42].

In recent studies, transfer learning and domain adaptation techniques have been shown to assist in heterogeneous data analysis, including applications in medical image analysis [43], [44], [45], [46]. This approach addresses domain shifts caused by clinics, filming instruments, and physical characteristics. Aranha et al. propose an ensemble of convolutional neural networks trained on high-quality fundus images to diagnose eye conditions [25] which was validated using low-quality images acquired by low-cost equipment.


(a)

(b)

FIGURE 1. Difference of slit-lamp (a) and camera (b) images.

TABLE 1. Companion animal eye disease dataset distribution.

Coarse	Cornea		Lens	
Fine	Cornea ulcer	Keratitis	Cataract	Nuclear sclerosis
Total	1,727	2,038	2,575	10,798
camera	0	117	271	712
slit-lamp	1,727	1,921	2,304	10,086

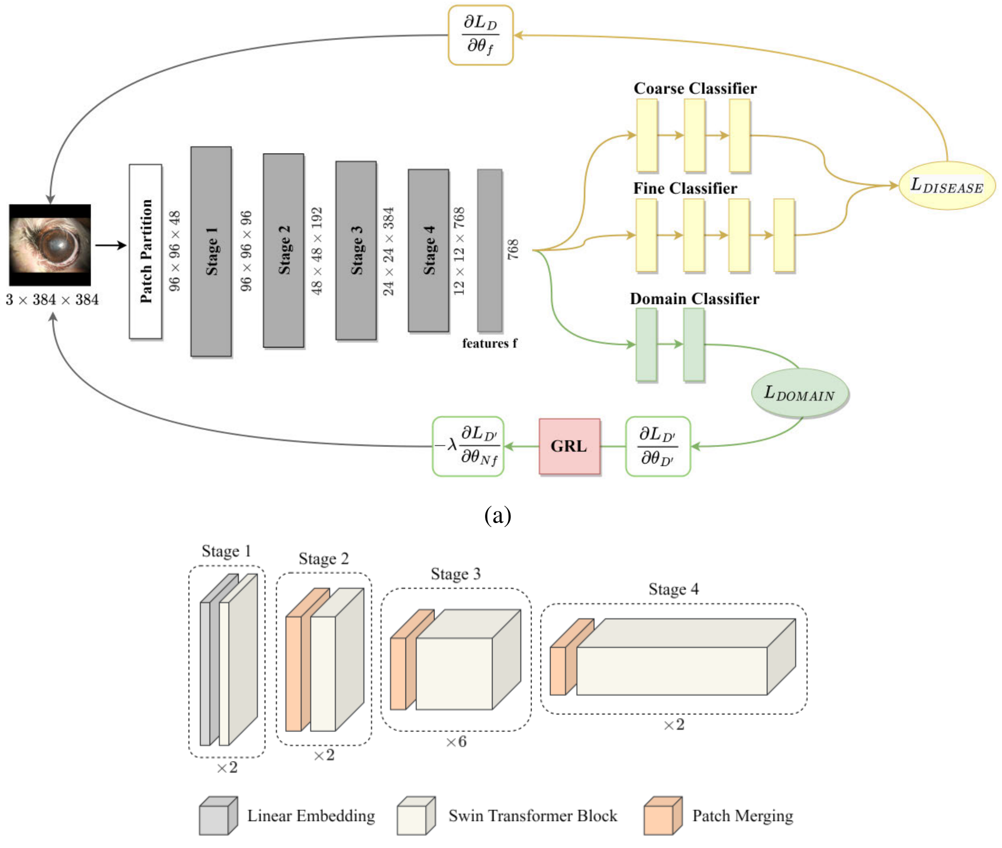
(a) Canine dataset

Coarse	Cornea		Eyelids
Fine	Corneal sequestrum	Cornea ulcer	Blephalitis
Total	3,511	3,531	1,076
camera	128	143	169
slit-lamp	3,383	3,388	907

(b) Feline dataset

This study obtained comparable results to the state-of-the-art to reach accuracies of 87.4%, 90.8%, 87.5%, 79.1% to classify cataract, diabetic retinopathy, excavation and blood vessels, respectively. Bevan et al. used two techniques, namely "Learning Not to Learn" and "Turning a Blind Eye", to remove the bias of artifacts on skin lesion images, for example, surgical markings or rulers [47], [48], [49]. Another issue was the inconsistency between clinical and dermoscopic images for the same lesion; therefore, they aimed to generalize various imaging methods.

Our work uses a gradient reversal layer to alleviate the domain shift between distinct filming instruments, inspired by unsupervised domain adversarial learning [42]. A gradient reversal layer is a component of a DL model to learn domain-invariant representations by reversing the gradients during the backpropagation process. We additionally integrated a method to address the issue of unbalanced domains



(a)

(b)

FIGURE 2. (a) The overview summarizes our work. Once a feature tensor is extracted from the input image, it passes three classifiers: a coarse classifier, a fine classifier, and a domain classifier. The coarse and fine classifier induces the loss $L_{DISEASE}$ which is optimized according to the original parameters, θ_f. Domain classifier induces the loss L_{DOMAIN}, and the gradient passes through a gradient reversal layer (GRL). L_{DOMAIN} is optimized according to the balanced feature weights, θ_{Nf}. (b) The detailed blocks of feature extraction.

by altering classification weights. This was implemented using a τ-normalized classifier, which is a strategy of re-balancing decision boundaries, used for long-tailed recognition [50]. The norms of the weights are associated with the class cardinalities, while following class-balanced sampling, the classifier weight norms tend to be more uniform.

III. METHOD

A. DATASET

This study used a public dataset named Pet Eye Disease Data (AI-Hub, South Korea). All data can be accessed through "AI-Hub [51]." The dataset comprises images depicting diseased eyes from over 5,000 pets, including both canine and feline subjects. Each image is standardized to a size of 400×400 pixels, with corresponding metadata presented in the JSON format. The samples contained twelve commonly raised canine breeds and six feline breeds of companion animals. From the dataset of 12 diseases, we selected a subset of four based on their significance in small animal health. Among the metadata, we used device labels, which indicate the type of filming instrument used for each image. Table 1 shows the distribution of the domains (i.e., devices), where the majority are slit-lamp (SL) images. Canine corneal ulcer images, in particular, are entirely SL images. Distinct devices induce differences in the final image, such as lighting conditions or hues, as shown in Fig. 1. The most explicit distinction is that SL images have light exerted on the central area, whereas camera images are captured without special lighting. The canine and feline datasets were seperated for all experiments, and the train, validation, and test sets were randomly divided with a ratio of 7.5:1.5:1.5.

B. MULTITASK LEARNING

In our work, the main objective is to achieve a fine-grained classification of the disease classes. To enhance the generalization of the main task, we incorporated additional information with a hierarchical layer of labels. Disease classes were grouped into coarse categories under the guidance of veterinarians, as shown in Table 1. Coarse classes are labeled according to the location where the corresponding diseases occur in the eyes, e.g., cornea, lens, and eyelids. Utilizing two layers of labels, multitask learning (MTL) [52] was implemented in which the input image tensor and backbone model architecture for feature extraction are shared. In this work, the Swin-Transformer [16] was implemented as the architecture for feature extraction, while the predictions of both the class levels were made independently from the coarse classifier and fine classifier. Each classifier induced L_C and L_F, respectively, and the model parameters were optimized according to minimize $L_{DISEASE} \overset{\text{def}}{=} (1-\alpha) \times L_C + \alpha \times L_F$.

This work weighted fine classification heavier; the value of α was set to 0.7. Given that several ocular diseases can simultaneously occur in an eye, using hard targets may cause errors that are inconsistent with actual probabilities. Hence, this issue was mitigated via label smoothing [53] of the coarse and fine predictions. Suppose for true labels y_k and predicted labels p_k for the kth class, the hard target values are 1 for the true class and 0 for the rest. Subsequently, we modify y_k into a smoothed y_k^{LS} using the formula below.

$$y_K^{LS} = y_k(1-\delta) + \frac{\delta}{K} \tag{1}$$

where K is the number of classes, δ is the weight of the smoothing labels, which was fixed at 0.1 throughout the experiments.

C. BALANCED DOMAIN ADVERSARIAL TRAINING

We propose balanced domain adversarial training (BDAT) for computational solutions that encourage the learning in the minor domain, as demonstrated in Figure 2(a). The input image utilized for disease classification is simultaneously inputted into the domain classifier, engaging in a binary classification task that distinguishes between camera and slit-lamp images. During the training phase, the gradient of the domain classifier is multiplied by a negative scalar $-\lambda_p$ through the gradient reversed layer (GRL). Accordingly, the parameters of the feature mapping ultimately maximize the L_{DOMAIN}, thus avoiding latent features that determine the domains.

Considering the noise of the domain classifier during earlier training stages, the scalar value λ_p is adjusted every epoch as below:

$$\lambda_p = \frac{2}{1+e^{-(10\times p)}} - 1 \tag{2}$$

$$p = \frac{batch\_idx + epoch \times max\_batches}{max\_epoch \times max\_batches} \tag{3}$$

Using the formula, λ initiates from 0 and gradually increases to converge to 1. This indicates that the influence of GRL is minimized during the initial training stage, and as the model starts to learn domain features, GRL reverses it accordingly.

Since approximately 95% of the dataset consists of SL images, relying solely on GRL would lead to a strong bias of the domain classifier towards SL images. Therefore, the decision boundary of the domain classifier was altered by normalizing the last feature map before passing GRL. We additionally integrated a method to address the issue of unbalanced domains by altering classification weights. Methodically, consider the feature map of the last layer of the domain classifier $W = \{w_j\}$, where w_j are classifier weights associated with domain j. The normalized weights $\tilde{W} = \{\tilde{w_j}\}$ are generated by:

$$\tilde{w_i} = \frac{w_i}{\|w_i\|^\tau} \tag{4}$$

$\tau \in [0.8, 1.2]$ is a hyperparameter empirically optimized, where ||·|| denotes the L2 norm. This process diminishes the latent features of domain spaces, accordingly the disease features are strengthened.

IV. RESULTS

A. EXPERIMENTAL SETUP AND EVALUATION METRICS

Each image was resized to 384×384 pixels, and the pixel values were normalized to the 0–1 range. Empirical results showed that traditional augmentation techniques (e.g. flipping, rotating, color manipulation, etc.) can compromise the preservation of semantic information in the original image. Consequently, we excluded augmentation. The models were trained with the AdamW optimizer and cosine annealing learning rate scheduler, where the learning rate values were cycled from 0 to the specified learning rate every 50 epochs. The batch size was set to 4, and the learning rate to 1×10-5. Canine and feline images were trained for 100 epochs and 50 epochs, respectively, according to the magnitude of image datasets. All the experiments were trained via PyTorch and implemented with an Intel(R) Core(TM) i7-10700KF CPU 3.80 GHz and NVIDIA GeForce RTX 3090 Ti graphics processing unit (GPU).

To evaluate the performance of each proposed method, various metrics were utilized. Firstly, the accuracy of each domain was measured to assess the impact of BDAT. To account for the performance of the network for the entire dataset, accuracy, recall, precision, and F1-score were used. These metrics were formulated as accuracy $= \frac{(TP+TN)}{(TP+TN+FP+FN)}$, recall $= \frac{(TP)}{(FN+TP)}$, precision $= \frac{(TP)}{(TP+FP)}$, specificity $= \frac{(TN)}{(TN+FP)}$, F1-Score $= \frac{2*(Precision*Recall)}{(Precision*Recall)}$. TP, TN, FP, and FN stand for true-positive, true-negative, false-positive, and false-negative instances, respectively. Accuracy and F1-Score serve as effective metrics derived from the outcome of performance evaluation. Accuracy assesses the degree of expected correctness in the results. Precision measures the how the measurement is correctly predicted. Recall evaluates the correctness of the outcomes. F1-Score utilizes precision to compute the overall average of all these values.

B. PERFORMANCE RESULTS

We compared the performance of Swin-Tiny with three different classifying model architectures to validate the ability for feature extraction: Densenet-121 [54], Resnet50 [55], and ConvNeXT-Base [56]. Table 2 shows that Swin-Transformer holds the highest metric values. ConvNeXT was introduced in 2022 as a modernized version of ConvNet, that competes with Swin-Transformer while reserving the simple and efficient traits of ConvNets. In this research, Swin-Transformer not only resulted in higher accuracy than ConvNeXT for both

TABLE 2. GRL results for convolutional networks.

	Architecture	Method	Domain accuracy		Total dataset				# FLOPS
			camera	slit-lamp	accuracy	recall	precision	F1-score	
Canine	Swin-T	MTL +GRL	**0.822**	**0.843**	**0.841**	**0.788**	**0.767**	**0.775**	4.5B
	ConvNeXT-B		0.713	0.804	0.798	0.699	0.707	0.703	15.4B
	Resnet50		0.776	0.813	0.810	0.748	0.722	0.734	4.1B
	Densenet121		0.753	0.833	0.828	0.768	0.745	0.756	2.9B
Feline	Swin-T	MTL +GRL	**0.718**	**0.650**	**0.654**	**0.584**	**0.689**	**0.606**	4.5B
	ConvNeXT-B		0.704	0.573	0.581	0.528	0.576	0.542	15.4B
	Resnet50		0.662	0.602	0.605	0.536	0.604	0.551	4.1B
	Densenet121		0.704	0.603	0.609	0.590	0.602	0.584	2.9B

TABLE 3. Canine results for balanced domain adversarial training.

	Domain accuracy		Total dataset			
	camera	slit-lamp	accuracy	recall	precision	F1-score
Base	0.821	0.821	0.819	0.743	0.749	0.737
MTL	0.799	0.837	0.837	**0.791**	0.751	0.767
GRL	0.822	0.843	0.841	0.788	0.767	**0.775**
τ 1.2	0.839	0.836	0.837	0.755	**0.780**	0.765
τ 1.1	0.851	0.818	0.820	0.770	0.757	0.763
τ 1.0	**0.862**	0.829	0.841	0.763	0.779	0.770
τ 0.9	0.845	**0.847**	**0.847**	0.767	0.779	0.772
τ 0.8	0.851	**0.847**	**0.847**	0.773	0.772	0.772

TABLE 4. Feline results for balanced domain adversarial training.

	Domain accuracy		Total dataset			
	camera	slit-lamp	accuracy	recall	precision	F1-score
Base	0.634	0.624	0.624	0.575	0.599	0.583
MTL	0.662	0.640	0.641	0.575	0.632	0.588
GRL	0.718	**0.650**	**0.654**	**0.584**	0.689	**0.606**
τ 1.2	0.676	0.624	0.627	0.568	0.661	0.591
τ 1.1	0.690	0.631	0.635	0.552	**0.720**	0.578
τ 1.0	0.704	0.609	0.614	0.531	0.677	0.548
τ 0.9	**0.732**	0.612	0.619	0.539	0.698	0.559
τ 0.8	0.662	0.620	0.622	0.554	0.640	0.573

canine and feline, but was also more efficient by owning 71% fewer floating point operations (FLOPs) than ConvNeXT.

Table 3 and Table 4 present the results of BDAT compared to the use of MTL or BDAT. The baseline refers to the scenario where only feature extraction and fine classification are performed. When using MTL exclusively, there was an improvement in performance for the full dataset, but the accuracy in the camera domain decreased for canines. In the case of felines, all metrics improved except for recall, which remained equivalent to the baseline.

In addition to MTL, incorporating GRL without feature normalization further improved the performance for both the canine and feline datasets. We observed improvements in the camera domain, with a larger scale increase compared to SL for both canines and felines. This indicates that reversing the features of the domain classifier was beneficial for disease classification. However, the increment in camera accuracy for canines was minor and not yet significant.

Accordingly, we investigated the effect of BDAT with various τ values in Equation (4). For canines, τ 1.0 achieved the highest camera accuracy, while τ 0.9 and τ 0.8 had the highest accuracy for the total dataset. All experiments with varying τ values not only exceeded the camera accuracy of the baseline, MTL, and GRL, but also outperformed SL accuracy for all values of τ excluding 0.9. Most of the other metrics were also highest when using BDAT.

The overall results for the feline exhibited differently, but the objective of improving the camera domain was confirmed. BDAT with τ 0.9, achieved the highest camera accuracy, surpassing the baseline, MTL, and GRL. However, the SL domain experienced a decline in competence, specifically a 3.8% decrease compared to GRL and a 2.8% decrease compared to MTL. This led to a decrease in other metrics due to the dominant influence of the large volume of SL images in determining the overall performance.

V. CONCLUSION AND LIMITATION

Through this study, we verified the feasibility of a DL framework for classifying companion animals' ocular surface disease images. Our proposed network labels the initial input image into two groups: coarse and fine classes. By adopting MTL and BDAT, the classification network obtained the ability to capture fine disease-related features and disregard the discrepancy between equipment-related ones simultaneously. Using two layers of hierarchical layers of labels enhanced the accuracy of disease classification, whereas it showed limited development in the minor domain. We overcame the discrepancy between clinical and practical screening environments through GRL and further addressed data imbalance through feature normalization. As mentioned in Section III, the proportion of the camera domain is between 3% and 5%. Despite the scarcity, the developments made by BDAT were 6.3% for canines, 7% for felines. Through this practice, we believe a wide range of collected data derived from various environments can be utilized for a constrained task, which suggests an efficient and realistic solution for medical imaging application studies.

Furthermore, most of the research in the field of ophthalmological image analysis were confined to some selected set of diseases, e.g. cataracts, corneal diseases, and glaucoma [23]. Christopher identified glaucomatous damage in optic nerve head (ONH) fundus images [24]. Kim detected severity in corneal ulcers, and Hao classified corneal diseases [26], [34]. The study of Aranha covered cataracts, diabetic retinopathy, excavation and blood vessels, but individual binary classification networks were used to distinguish between normal and abnormal images [25]. In contrast, we use a single model to identify multiple diseases, being the first study to detect a wide range of ophthalmological diseases in both canines and felines.

Canine disease predictions resulted in a high-level accuracy above 80%. Note that it is difficult to objectively compare the metric values with related works because the dataset was not previously used. Moreover, although the outcome of SL accuracy for BDAT felines did not align with that of canines, the benefit of BDAT was significantly shown in camera accuracy. We are willing to extend this case with more collected data.

However, this study carries some limitations. First, domain adaptation was fully supervised for our study. While we succeeded in adapting within our dataset's domains, our network is not generalized in unseen domains other than SL or cameras. Unsupervised domain adaptation or domain generalization methods could be adopted to target unexpected domains. Also, our work does not include tests on human veterinarians. Comparing the performance of AI and human diagnosis can enlighten the appropriate area for the application of DL in veterinary studies. Lastly, in this study, a selected subset of the AIHub dataset was utilized, focusing on disease classes that are frequent ailments observed in small animals. This subset provided a representative sample for the research objectives while optimizing computational resources and ensuring the timely completion of the training process. In future studies, it would be beneficial to expand the dataset to include a wider range of disease classes and incorporate additional data from different sources.

REFERENCES

[1] Y. Liu and S. Sun, "SagaNet: A small sample gated network for pediatric cancer diagnosis," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, M. Meila and T. Zhang, Eds. Jul. 2021, pp. 6947–6956.

[2] P. Arsomngern, N. Numcharoenpinij, J. Piriyataravet, W. Teerapan, W. Hinthong, and P. Phunchongharn, "Computer-aided diagnosis for lung lesion in companion animals from X-ray images using deep learning techniques," in *Proc. IEEE 10th Int. Conf. Awareness Sci. Technol. (iCAST)*, Oct. 2019, pp. 1–6.

[3] A. Morris, H. Wu, and C. Morales, "Barriers to care in veterinary services: Lessons learned from low-income pet guardians' experiences at private clinics and hospitals during COVID-19," *Frontiers Veterinary Sci.*, vol. 8, Oct. 2021, Art. no. 764753.

[4] L. Guardabassi, "Pet animals as reservoirs of antimicrobial-resistant bacteria: Review," *J. Antimicrobial Chemotherapy*, vol. 54, no. 2, pp. 321–332, Jul. 2004.

[5] W. M. D. Wan Zaki, H. A. Mutalib, L. A. Ramlan, A. Hussain, and A. Mustapha, "Towards a connected mobile cataract screening system: A future approach," *J. Imag.*, vol. 8, no. 2, p. 41, Feb. 2022.

[6] C. Weingart, B. Kohn, M. Siekierski, R. Merle, and M. Linek, "Blepharitis in dogs: A clinical evaluation in 102 dogs," *Veterinary Dermatol.*, vol. 30, no. 3, p. 222, Jun. 2019.

[7] K. N. Gelatt, "Diseases and surgery of the canine cornea and sclera," *Animal Eye Res.*, vol. 21, nos. 3–4, pp. 105–113, 2002.

[8] J.-Y. Kim, K.-H. Kim, D. L. Williams, W.-C. Lee, and S.-W. Jeong, "Epidemiological and clinical features of canine ophthalmic diseases in Seoul from 2009 to 2013," *J. Veterinary Clinics*, vol. 32, no. 3, pp. 235–238, Jun. 2015.

[9] 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, pp. 1–18, Dec. 2019.

[10] A. Rehman, M. Ahmed Butt, and M. Zaman, "A survey of medical image analysis using deep learning approaches," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021, pp. 1334–1342.

[11] 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. 5998–6008.

[12] T. Wolf et al., "HuggingFace's transformers: State-of-the-art natural language processing," 2019, *arXiv:1910.03771*.

[13] K. Wu, H. Peng, M. Chen, J. Fu, and H. Chao, "Rethinking and improving relative position encoding for vision transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10013–10021.

[14] L. Yuan, Y. Chen, T. Wang, W. Yu, Y. Shi, Z. Jiang, F. E. H. Tay, J. Feng, and S. Yan, "Tokens-to-Token ViT: Training vision transformers from scratch on ImageNet," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 538–547.

[15] 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*.

[16] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[17] D.-Z. Zhao, X.-K. Wang, T. Zhao, H. Li, D. Xing, H.-T. Gao, F. Song, G.-H. Chen, and C.-X. Li, "A Swin Transformer-based model for mosquito species identification," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, Nov. 2022.

[18] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1748–1758.

[19] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," 2022, *arXiv:2201.09873*.

[20] L. Zhang and Y. Wen, "A transformer-based framework for automatic COVID19 diagnosis in chest CTs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 513–518.

[21] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, ''NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,'' *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[22] Z. Li, J. Jiang, K. Chen, Q. Chen, Q. Zheng, X. Liu, H. Weng, S. Wu, and W. Chen, ''Preventing corneal blindness caused by keratitis using artificial intelligence,'' *Nature Commun.*, vol. 12, no. 1, pp. 1–12, Jun. 2021.

[23] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, ''Cataract-Net: An automated cataract detection system using deep learning for fundus images,'' *IEEE Access*, vol. 9, pp. 128799–128808, 2021.

[24] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, ''Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs,'' *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, Nov. 2018.

[25] G. D. A. Aranha, R. A. S. Fernandes, and P. H. A. Morales, ''Deep transfer learning strategy to diagnose eye-related conditions and diseases: An approach based on low-quality fundus images,'' *IEEE Access*, vol. 11, pp. 37403–37411, 2023.

[26] H. Gu, Y. Guo, L. Gu, A. Wei, S. Xie, Z. Ye, J. Xu, X. Zhou, Y. Lu, X. Liu, and J. Hong, ''Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs,'' *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Oct. 2020.

[27] N. Chea and Y. Nam, ''Classification of fundus images based on deep learning for detecting eye diseases,'' *Comput., Mater. Continua*, vol. 67, no. 1, pp. 411–426, 2021.

[28] C.-J. Lai, P.-F. Pai, M. Marvin, H.-H. Hung, S.-H. Wang, and D.-N. Chen, ''The use of convolutional neural networks and digital camera images in cataract detection,'' *Electronics*, vol. 11, no. 6, p. 887, Mar. 2022.

[29] G. B. Ergün and S. Güney, ''Classification of canine maturity and bone fracture time based on X-ray images of long bones,'' *IEEE Access*, vol. 9, pp. 109004–109011, 2021.

[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ''ImageNet: A large-scale hierarchical image database,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[31] T. Banzato, M. Wodzinski, S. Burti, V. L. Osti, V. Rossoni, M. Atzori, and A. Zotti, ''Automatic classification of canine thoracic radiographs using deep learning,'' *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, Feb. 2021.

[32] L. Dumortier, F. Guépin, M.-L. Delignette-Ḿuller, C. Boulocher, and T. Grenier, ''Deep learning in veterinary medicine, an approach based on CNN to detect pulmonary abnormalities from lateral thoracic radiographs in cats,'' *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, Jul. 2022.

[33] J. Y. Kim, M. G. Han, J. H. Chun, E. A. Huh, and S. J. Lee, ''Developing a diagnosis model for dry eye disease in dogs using object detection,'' *Sci. Rep.*, vol. 12, no. 1, p. 21351, Dec. 2022.

[34] J. Y. Kim, H. E. Lee, Y. H. Choi, S. J. Lee, and J. S. Jeon, ''CNN-based diagnosis models for canine ulcerative keratitis,'' *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, Oct. 2019.

[35] L. Torrey and J. Shavlik, ''Transfer learning,'' in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.

[36] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, ''A comprehensive survey on transfer learning,'' *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[37] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, ''Balanced distribution adaptation for transfer learning,'' in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1129–1134.

[38] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, ''Deep subdomain adaptation network for image classification,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[39] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, ''Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery,'' *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702–8712, Nov. 2020.

[40] C. M. Scannell, A. Chiribiri, and M. Veta, ''Domain-adversarial learning for multi-centre, multi-vendor, and multi-disease cardiac MR image segmentation,'' in *Proc. 11th Int. Workshop Stat. Atlases Comput. Models Heart (STACOM)*, Lima, Peru. Cham, Switzerland: Springer, 2021, pp. 228–237.

[41] T. K. Yoo and J. Y. Choi, ''Outcomes of adversarial attacks on deep learning models for ophthalmology imaging domains,'' *JAMA Ophthalmol.*, vol. 138, no. 11, pp. 1213–1215, Nov. 2020.

[42] Y. Ganin and V. Lempitsky, ''Unsupervised domain adaptation by backpropagation,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[43] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, ''Contrastive adaptation network for unsupervised domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4888–4897.

[44] Y. Yang and S. Soatto, ''FDA: Fourier domain adaptation for semantic segmentation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4084–4094.

[45] C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, ''Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation,'' in *Proc. 33rd AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 865–872.

[46] J. Yang, N. C. Dvornek, F. Zhang, J. Chapiro, M. Lin, and J. S. Duncan, ''Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 255–263.

[47] P. J. Bevan and A. Atapour-Abarghouei, ''Skin deep unlearning: Artefact and instrument debiasing in the context of melanoma classification,'' 2021, *arXiv:2109.09818*.

[48] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, ''Learning not to learn: Training deep neural networks with biased data,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9004–9012.

[49] M. Alvi, A. Zisserman, and C. Nellåker, ''Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 1–16.

[50] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, ''Decoupling representation and classifier for long-tailed recognition,'' 2019, *arXiv:1910.09217*.

[51] AIHub. (2022). Pet Eye Disease Data. [Online]. Available: https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=562

[52] R. Caruana, ''Multitask learning,'' *Mach. Learn.*, vol. 28, pp. 41–75, Dec. 1997.

[53] R. Müller, S. Kornblith, and G. E. Hinton, ''When does label smoothing help?'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 1–10.

[54] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[55] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[56] 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.

MARY G. NAM received the double B.S. degree in mathematics from the Department of IT Engineering and Sookmyung Women's University, Seoul, in 2023.

Since 2022, she has been a Research Intern with the HCI Laboratory, IT Engineering Department, Soookmyung Women's University. She is currently with the Computer Science Department, Rice University. Her research interests include deep-learning and AI-aided diagnosis.

SUH-YEON DONG (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2010, 2011, and 2016, respectively.

She is currently an Associate Professor with the Department of Information Technology Engineering, Sookmyung Women's University, Seoul, South Korea. Her research interests include machine-learning-based biosignal processing and cognitive neuroscience.