



# A difficulty-aware and task-augmentation method based on meta-learning model for few-shot diabetic retinopathy classification

Xueyao Liu<sup>1#^</sup>, Xueyuan Dong<sup>1#</sup>, Tuo Li<sup>2</sup>, Xiaofeng Zou<sup>2</sup>, Chen Cheng<sup>1</sup>, Zekun Jiang<sup>1</sup>, Zhumin Gao<sup>1</sup>, Sixu Duan<sup>1</sup>, Meirong Chen<sup>3</sup>, Tingting Liu<sup>4</sup>, Pu Huang<sup>1</sup>, Dengwang Li<sup>1</sup>, Hua Lu<sup>1</sup>

<sup>1</sup>Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine, School of Physics and Electronics, Shandong Normal University, Jinan, China; <sup>2</sup>Shandong Yunhai Guochuang Cloud Computing Equipment Industry Innovation Co., Ltd., Jinan, China; <sup>3</sup>Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China; <sup>4</sup>Eye Hospital of Shandong First Medical University (Shandong Eye Hospital), Jinan, China

*Contributions:* (I) Conception and design: X Liu, X Dong, P Huang, D Li, H Lu; (II) Administrative support: H Lu, D Li, P Huang; (III) Provision of study materials or patients: X Liu, X Dong, P Huang, T Li, X Zou; (IV) Collection and assembly of data: C Cheng, Z Jiang, Z Gao, S Duan, T Li, X Zou; (V) Data analysis and interpretation: X Liu, X Dong, M Chen, T Liu, H Lu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work as co-first authors.

*Correspondence to:* Hua Lu, PhD; Dengwang Li, PhD; Pu Huang, PhD. Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine, School of Physics and Electronics, Shandong Normal University, No. 1, University Road, Science Park, Changqing District, Jinan 250358, China. Email: hualu@sdsnu.edu.cn; dengwang@sdsnu.edu.cn; pu.wong@139.com.

**Background:** Accurate classification techniques are essential for the early diagnosis and treatment of patients with diabetic retinopathy (DR). However, the limited amount of annotated DR data poses a challenge for existing deep-learning models. This article proposes a difficulty-aware and task-augmentation method based on meta-learning (DaTa-ML) model for few-shot DR classification with fundus images.

**Methods:** The difficulty-aware (Da) method operates by dynamically modifying the cross-entropy loss function applied to learning tasks. This methodology has the ability to intelligently down-weight simpler tasks, while simultaneously prioritizing more challenging tasks. These adjustments occur automatically and aim to optimize the learning process. Additionally, the task-augmentation (Ta) method is used to enhance the meta-training process by augmenting the number of tasks through image rotation and improving the feature-extraction capability. To implement the expansion of the meta-training tasks, various task instances can be sampled during the meta-training stage. Ultimately, the proposed Ta method was introduced to optimize the initialization parameters and enhance the meta-generalization performance of the model. The DaTa-ML model showed promising results by effectively addressing the challenges associated with few-shot DR classification.

**Results:** The Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 blindness detection data set was used to evaluate the DaTa-ML model. The results showed that with only 1% of the training data (5-way, 20-shot) and a single update step (training time reduced by 90%), the DaTa-ML model had an accuracy rate of 89.6% on the test data, which is a 1.7% improvement over the transfer-learning method [i.e., residual neural network (ResNet)50 pre-trained on ImageNet], and a 16.8% improvement over scratch-built models (i.e., ResNet50 without pre-trained weights), despite having fewer trainable parameters (the parameters used by the DaTa-ML model are only 0.47% of the ResNet50 parameters).

<sup>^</sup> ORCID: 0000-0001-9473-8097.

**Conclusions:** The DaTa-ML model provides a more efficient DR classification solution with little annotated data and has significant advantages over state-of-the-art methods. Thus, it could be used to guide and assist ophthalmologists to determine the severity of DR.

**Keywords:** Diabetic retinopathy classification (DR classification); few-shot; meta-learning; difficulty-aware (Da); task-augmentation (Ta)

Submitted Apr 24, 2023. Accepted for publication Nov 09, 2023. Published online Jan 02, 2024.

doi: 10.21037/qims-23-567

View this article at: <https://dx.doi.org/10.21037/qims-23-567>

## Introduction

Diabetic retinopathy (DR) is a microvascular complication of diabetes, and one of the leading causes of blindness and vision loss worldwide (1). The early and accurate classification of DR is crucial for appropriate and timely treatment (2). Fundus images are widely used in clinical diagnosis. Under the international protocol (3,4), there are five categories of DR: (I) non-DR; (II) mild non-proliferative DR (NPDR); (III) moderate NPDR; (IV) severe NPDR; and (V) proliferative DR (PDR) (3,5). *Figure 1* shows the five categories of fundus images.

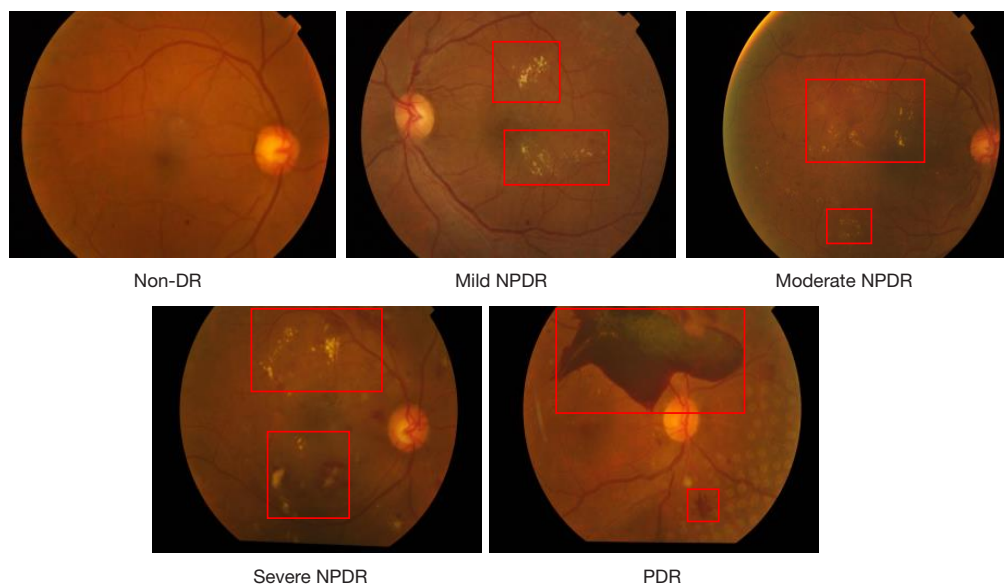
In general, the severity of DR is diagnosed by ophthalmologists based on their clinical experience. However, computer-aided classification technology can significantly save time and improve the efficiency and accuracy of DR classification (6). Currently, research in this area can be broadly categorized into two types: (I) hands-on engineering; and (II) deep-learning methods. The first type, hands-on engineering, uses traditional methods, such as Gabor filters, scale-invariant feature transform, the histogram of oriented gradient, and the local binary pattern, to extract features (7-10). For example, Shahin *et al.* employed morphological processing to extract pathological features, calculated entropy and homogeneity, input these into a neural network, and achieved 88% sensitivity and 100% specificity (11). Casanova *et al.* proposed a random-forest algorithm that achieved over 90% accuracy in DR classification and assessed DR risk based on fundus and systematic data (12). However, these methods produce weak feature representations and can be adversely affected by characteristic factors (13).

The second type, deep-learning methods, has been more widely used in image processing, including DR classification, with convolutional neural networks (CNNs) being particularly popular (14). For example, García *et al.* used the Visual Geometry Group 16 (VGG16) structure

to classify fundus images of left and right eyes, achieving 93.65% specificity, 54.47% sensitivity, and 83.68% accuracy (15). Shanthi and Sabeenian proposed an improved AlexNet architecture that boosted the model accuracy to 96.25% based on the Messidor data set (16). Deep-learning methods enable end-to-end classification without manual feature extraction, but they require large amounts of labeled training data and expensive computational resources to achieve human-level performance (17,18). Similarly, the creation of large data sets with annotations involves skilled labor, which is costly and time-consuming. In addition, CNNs are prone to overfitting when training data are insufficient; however, such data can be challenging to obtain due to privacy and healthcare laws (19,20).

Given the negative effects of limited data on deep-learning models, Gao *et al.* used randomly rotated and flipped fundus images to expand the data (21). Zhou *et al.* introduced the DR generative adversarial network, which generates fundus images for data augmentation, but the generated data were often poor quality and lacked diversity (22). Transfer learning has been effective in addressing limited data by learning from a source domain to improve performance on a target domain (23,24). However, this method may not work well in complex target domains with insufficient source data.

Recently, meta-learning has attracted attention as a way to rapidly adapt to new tasks using small samples by learning internal representations from multiple classification tasks (25-28). Li *et al.* achieved an area under the curve (AUC) of 83.3% with only five samples per category using meta-learning on the International Skin Imaging Collaboration (ISIC) 2018 skin lesion classification data set (29). Yuan *et al.* proposed an active meta-learning method that focuses on difficult tasks using Bayesian dropout uncertainty estimation, which achieved an accuracy of 90% on a new brain cell type classification task with only 1% of training



**Figure 1** The five categories of fundus images. The red rectangle marks the lesion area. DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.

data and one update step (30).

In this study, we proposed the difficulty-aware and task-augmentation method based on meta-learning (DaTa-ML) model for few-shot DR classification with fundus images. The DaTa-ML model can find an initialization point, and the subsequent fine tuning of the target data set can achieve more accurate results than other state-of-the-art methods. The proposed DaTa-ML model:

- (I) Has a strong learning ability and can be applied to few-shot DR classification;
- (II) Can improve the effectiveness of the meta-training stage and optimize the initialization parameters;
- (III) Can adapt to the classification task of DR more quickly by using prior knowledge learned from multiple classification tasks.

The rest of this article is arranged as follows: the “Methods” section describes the proposed method in detail; the “Results” section sets out the experimental results and compares the DaTa-ML model with other models; the “Discussion” section discusses the proposed method, and finally; and the “Conclusions” section concludes the paper. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-567/rc>).

## Methods

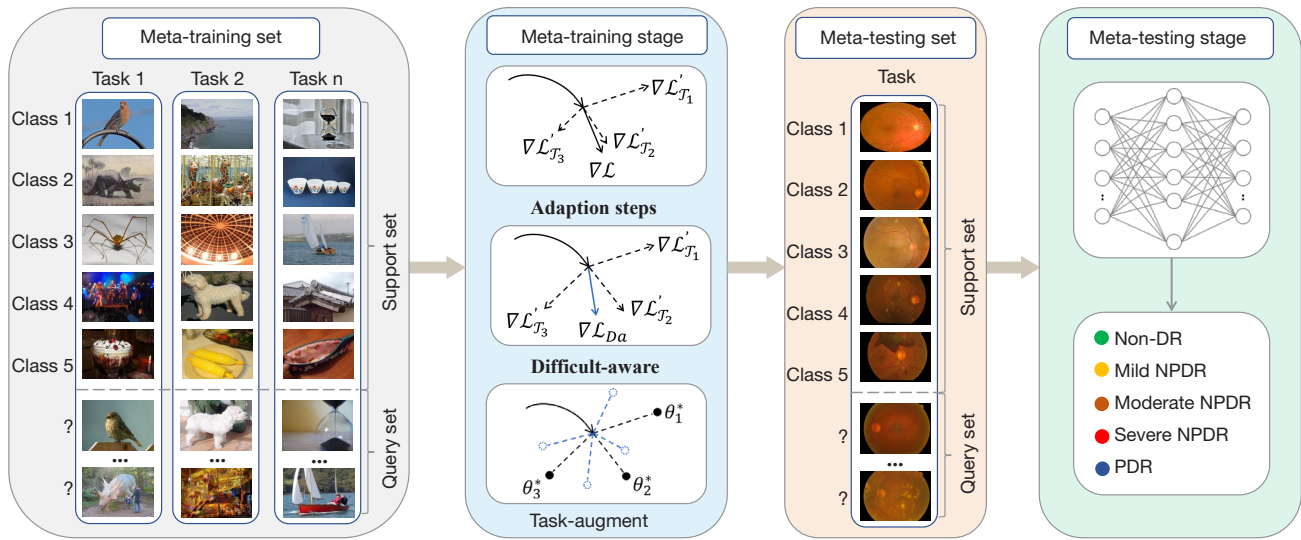
The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Review Board of Shandong Normal University, and the requirement of individual consent for this retrospective analysis was waived.

Due to the limited annotated DR training data, DR classification is a few-shot classification problem. To address this problem, the DaTa-ML model was proposed for few-shot DR classification using fundus images (see *Figure 2*).

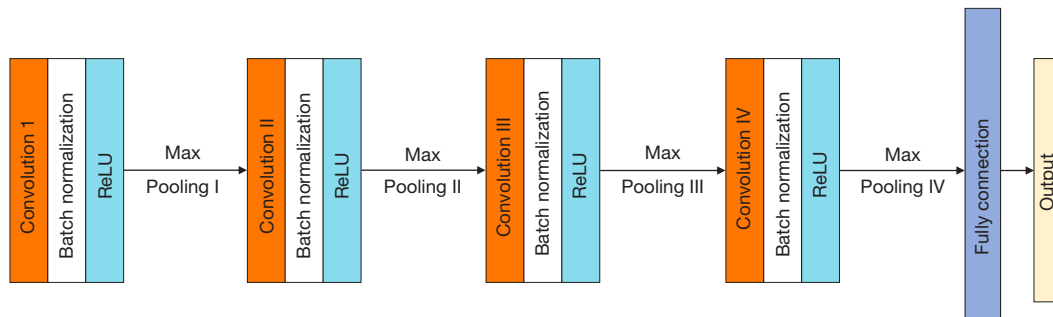
### *The architecture of meta-learner*

Transfer learning is a widely used strategy to initialize model parameters; however, well-known pre-trained models, such as VGG network (VGGNet) and residual neural network (ResNet), are usually require large amounts of data and not suitable for solving few-shot problems. Meta-learning emphasizes data divergence by using various support sets during the meta-training stage, which can improve a model’s generalization performance (31).

**Table S1** shows the parameter settings of the small and lightweight meta-learner model, which uses only



**Figure 2** The DaTa-ML model. The images in the meta-training set are from the natural image data set ImageNet (<https://www.image-net.org/>). DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.



**Figure 3** The network architecture of the meta-learner model. ReLU, rectified linear unit.

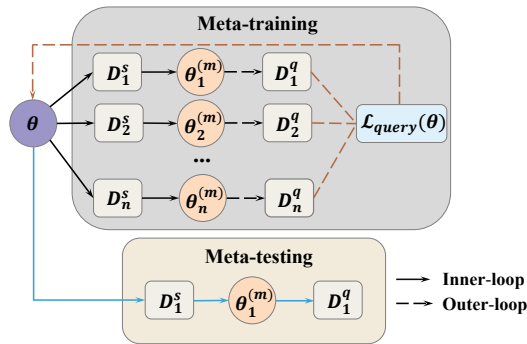
four convolutional layers. As *Figure 3* shows, the feature-extraction network consists of four parts, each of which includes a convolution layer, a batch-normalization layer, a rectified linear unit (ReLU) layer, and a maximum pooling layer. Fundus images are used as input, and four successive convolutions and maximum pooling are performed. The resulting feature map is then sent to the full connection layer for DR classification.

Model-agnostic meta-learning (MAML) can achieve excellent performance on unseen tasks by optimizing the model initialization parameters. Our training strategy is inspired by MAML to obtain robust meta-learner (32-36). For simplicity, [Table S2](#) lists a series of new terms for MAML.

The meta-training includes the optimization of the inner-loop and the outer-loop, where  $D_r$  is divided into “n” meta-training tasks  $\{T_i\}_n$ . In the inner-loop, each meta-training task trains its own model and yields an independent loss value. In the outer-loop, the meta-learner model updates once by aggregating the loss results across all the meta-training tasks. In this way, the meta-learning model can adapt more easily to new tasks through the learning of different tasks.

The meta-learner model with parameter  $\theta$  is denoted by  $f_{\theta}$ , the distribution  $P_r(\mathcal{T})$  over the meta-training tasks, and  $T_i \sim P_r(\mathcal{T})$ .

For the inner-loop optimization, the support set of each meta-training task is randomly sampled from  $D_r$  and fed



**Figure 4** The inner-loop and outer-loop optimization procedure for the DaTa-ML model at the  $m$ -th step. DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

into the meta-learner model at each iteration. Thus, the meta-learner is neither task nor category-specific, and the random sampling of model learning treats each category equally. At the beginning, the initial meta-learner model parameter  $\theta$  is updated to  $\theta_i^{(1)}$  with  $\{\mathcal{T}_i\}_n$ , which is represented as:

$$\theta_i^{(1)} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta}) \quad [1]$$

where  $\alpha$  denotes the learning rate of the inner-loop. At the  $m$ -th step, the model parameters are the sum of all  $m$  steps of gradient descent. The formula is as follows:

$$\theta_i^{(m)} = \theta_i^{(m-1)} - \alpha \nabla_{\theta_i^{(m-1)}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(m-1)}}) = \theta - \sum_{\varepsilon=0}^{m-1} \alpha \nabla_{\theta_i^{(\varepsilon)}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(\varepsilon)}}) \quad [2]$$

where  $\varepsilon \in [0, m]$ , and it denotes the index of the iteration, and  $f_{\theta_i^{(m)}}$  represents that the learner model is trained in  $m$  steps.

The outer-loop optimization is performed on the query set, where the overall loss of the query set is the sum of the losses of each meta-training task after  $m$  iterations. The formula is as follows:

$$\mathcal{L}_{query}(\theta) = \sum_{\mathcal{T}_i \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(m)}}) = \sum_{\mathcal{T}_i \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i} \left( f_{\left( \theta, \sum_{\varepsilon=0}^{m-1} \alpha \nabla_{\theta_i^{(\varepsilon)}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(\varepsilon)}}) \right)} \right) \quad [3]$$

The meta goal is to minimize the total loss function and obtain the ground-truth parameter with stronger generalization ability. At this point, the meta-learning parameter is updated as follows:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \mathcal{L}_{query}(\theta) \quad [4]$$

where the learning rate of the outer-loop is denoted by  $\beta$ . The parameters  $\theta$  of the meta-learner model are fine-tuned and the performance is evaluated on the support set and

query set of the meta-test task, respectively. Figure 4 shows the inner-loop and outer-loop optimization process for the DaTa-ML model at the  $m$  step.

For the few-shot classification problem, as it is a discrete classification task with cross loss, the loss function of the model is:

$$\mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(c)}}) = - \sum_{x^{(j)}, y^{(j)} \sim \mathcal{T}_i} y^{(j)} \log f_{\theta_i^{(c)}}(x^{(j)}) + (1 - y^{(j)}) \log(1 - f_{\theta_i^{(c)}}(x^{(j)})) \quad [5]$$

Here, the input and output of the  $j$ -th sample randomly chosen from  $\mathcal{T}_i$  are denoted by  $x^{(j)}$  and  $y^{(j)}$ . The training details for algorithm 1 are outlined in Table 1.

### Task augmentation

The initialization parameters of a model determine its generalization performance on unseen tasks. However, meta-learning models cannot extract more valuable features with fewer meta-training tasks. To optimize the model's initialization parameters, we proposed a rotation-based Ta method that increases the quantity of meta-training tasks.

In this study, the total quantity of the meta-training set was increased by rotating all the images by 90, 180, and 270 degrees. Different task instances can be sampled during the meta-training stage to implement meta-training task expansion. The proposed Ta method can optimize the initialization parameters of the model and improve the generalization of meta-learner.

### Difficulty-aware (Da)

As the meta-learner model continuously learns based on the average estimate of  $\mathcal{T}_i$  sampled by  $P_r(\mathcal{T})$ , it may struggle with difficult tasks. Thus, we introduced a Da method to focus on these tasks. We constructed a Da function that automatically reduces the weight of simple tasks by dynamically adjusting the cross-entropy loss function. The Da function is as follows:

$$\mathcal{L}_{Da} = \sum_{x^{(j)}, y^{(j)} \sim \mathcal{T}_i} -\mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(c)}})^{\omega} \log \left( \max \left( \lambda, 1 - \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(c)}}) \right) \right) \quad [6]$$

where  $\omega$  is the dynamic factor, and  $\lambda$  is an extremely small positive value satisfying  $\max(\lambda, 1 - \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i^{(c)}})) > 0$ . Eq. [6] introduces a dynamically scaled cross-entropy loss over the learning tasks that intelligently down-weights simple tasks and shifts focus towards difficult tasks. To rephrase, the Da meta-loss method dynamically reduces the contribution of easy tasks, concentrating more on difficult tasks and thus

**Table 1** The training details

---

Algorithm 1 for model training process

Require:  $P_r(\mathcal{T})$ : distribution over tasks

Require:  $\alpha, \beta$ : the learning rates of the inner-loop and outer-loop

Require:  $m$ : number of steps for gradient update

(I) Randomly initialize  $\theta$

(II) While not done do

(III) Sample batch of tasks  $\mathcal{T}_i \sim P_r(\mathcal{T})$

(IV) For all  $\mathcal{T}_i$  do

(V) Sample  $\kappa$  datapoints  $D_S = \{x^{(j)}, y^{(j)}\}$  from  $\mathcal{T}_i$

(VI) For  $\varepsilon = 0 \rightarrow m - 1$  do

(VII) Evaluate  $\nabla_{\theta^{(m-1)}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta^{(m-1)}})$  using  $D_S$  and  $\mathcal{L}_{\mathcal{T}_i}$  in Eq. [5]

(VIII) Update adapted parameters with gradient descent:  
 $\theta_i^{(m)} = \theta_i^{(m-1)} - \alpha \nabla_{\theta^{(m-1)}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta^{(m-1)}})$

(IX) End for

(X) Sample datapoints  $D_q = \{x^{(j)}, y^{(j)}\}$  from  $\mathcal{T}_i$

(XI) End for

(XII) Evaluate  $\nabla_{\theta} \sum_{\mathcal{T}_i \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta^{(m)}})$  on  $D_q$

(XIII) Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta^{(m)}})$

(XIV) End while

---

heightens the importance of optimizing misclassified tasks. The updated parameters are expressed as follows:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \mathcal{L}_{D_q} \quad [7]$$

### Ablation study

The proposed DaTa-ML model leverages the Da method to focus on complex tasks. To verify the benefits of this method, we compare DaTa-ML with Da based meta-learning (Da-ML) (i.e., with no Ta method) and Ta based meta-learning (Ta-ML) (i.e., with no Da method). We also included MAML as a baseline to demonstrate the advantages of both the Da and Ta methods.

## Results

### Data set

The few-shot DR classification data set included two

subsets. Subset 1 was the Mini-ImageNet data set, which comprised 100 categories with 60,000 color images (37). Subset 2 was the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 Blindness Detection data set (38), which comprised 3,662 fundus images from multiple clinics with different cameras, including non-DR [1,805], mild NPDR [370], moderate NPDR [999], severe NPDR [193], and PDR [295] fundus images. Subsets 1 and 2 were used for the meta-training and meta-testing, respectively.

### Experimental setting

Considering the distinctions in the training methods between the other deep learning-based models and our proposed model, the DaTa-ML model was trained in the manner of the N-way K-shot (N=5 and K=1, 5, 10, and 20) (39). The inner-loop learning rate and outer-loop learning rate were set to 0.01 and 0.001, respectively, and the model was trained for 10,000 iterations on a meta-training set consisting of 500 tasks. The APTOS 2019 Blindness Detection data set was used for both the meta-training and meta-testing, with data augmentation generating 2,000 images per class. The images used for both the training and testing comprised five categories and were completely different, and their intersection was the empty set. Each experiment was performed five times, with early stopping used to avoid overfitting.

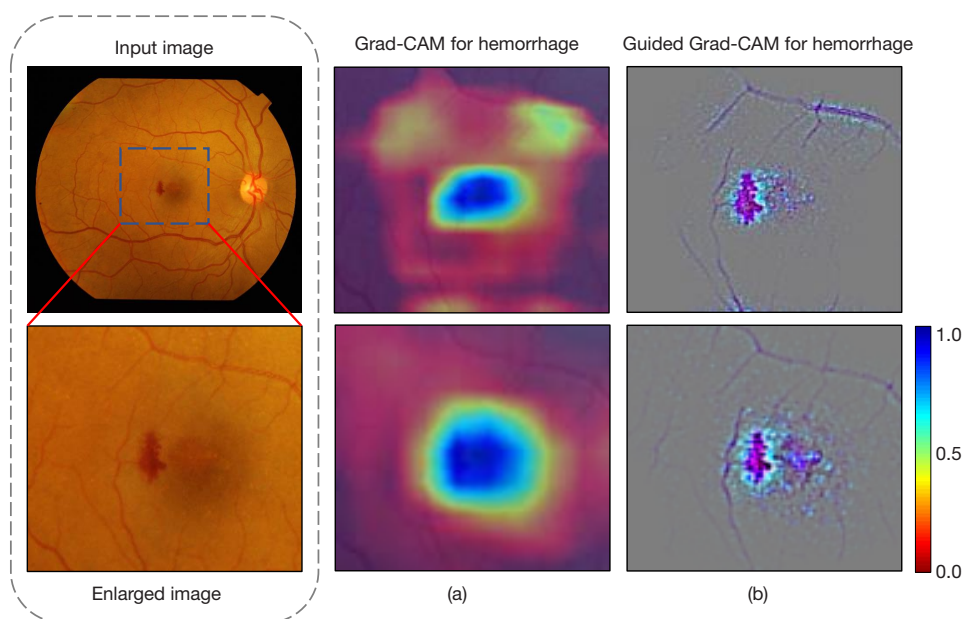
The classification method was implemented on an Intel (R) Core (TM) i7-8700k CPU@3.70 GHz desktop and then transferred to the NVIDIA Tesla P100 GPU. The algorithm was programmed using the TensorFlow library.

### Evaluation measures

For the dichotomous problem, we applied commonly used evaluation indexes (i.e., accuracy, precision, recall, and F1-score) to evaluate the performance of the DaTa-ML model. The higher the value of these metrics, the better the performance. However, for the DaTa-ML model, the fundus images were classified into five categories. Macro-averaging was used as an evaluation indicator; that is, we first statistically assessed each category and then arithmetically averaged each category. The macro-averaging values are calculated as shown in Eqs. [8,9]:

$$\text{Macro\_Precision} = \frac{1}{n} \sum_{i=1}^n p_i \quad [8]$$

$$\text{Macro\_Recall} = \frac{1}{n} \sum_{i=1}^n R_i \quad [9]$$



**Figure 5** Visualization of Grad-CAM on a random selected example. The input image is shown in the left. The first and second rows are the heat maps for the MAML and DaTa-ML models, respectively. For the heatmap, the cool colors represent larger activation. (a) The category-related discriminating features located by Grad-CAM; (b) the highlighted images obtained by guided backpropagation combined with Grad-CAM. Grad-CAM, Gradient-weighted Class Activation Mapping; MAML, model-agnostic meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

The Macro-F1 score is defined as shown in Eq. [10] (40,41). The Macro-F1 score also lies between 0.0 and 1.0, with the smallest value [0] indicating the worst performance of the classifier, and the highest value [1] indicating the best performance of the classifier. The equation is expressed as follows:

$$\text{Macro\_F1} = 2 \frac{\text{Macro\_Precision} \times \text{Macro\_Recall}}{\text{Macro\_Precision} + \text{Macro\_Recall}} \quad [10]$$

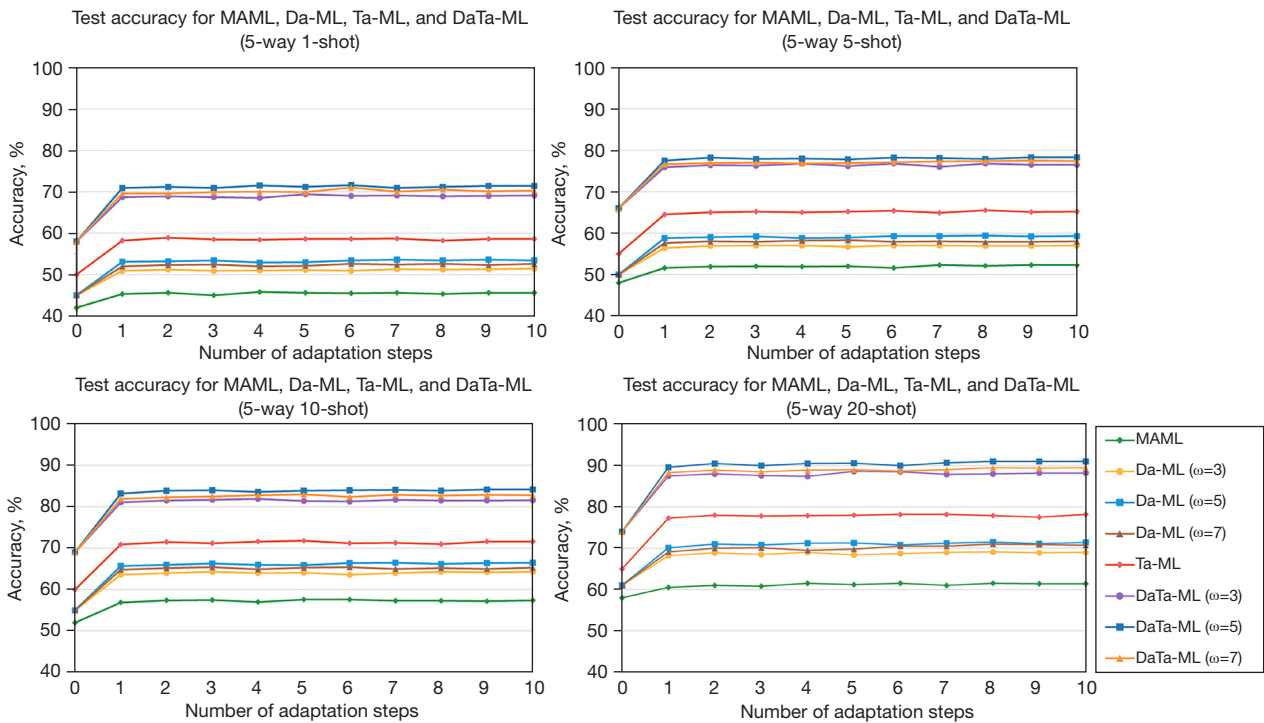
### Comparisons of the MAML, Da-ML, Ta-ML, and DaTa-ML models

To better understand the performance of the DaTa-ML model, Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations were used to show that the model effectively focuses on hemorrhage areas and reduces background noise in raw features (42). *Figure 5* displays the highlighted results. Additionally, guided Grad-CAM showed that the DaTa-ML model weakens unrelated feature areas, such as blood vessels.

*Figure 6* presents the adaptation processes of the MAML, Da-ML, Ta-ML, and DaTa-ML models under 1-, 5-,

10-, and 20-shot. MAML was not effective at adapting to DR classification. Conversely, after a gradient update, the accuracy of the DaTa-ML model significantly improved, indicating the rapid adaptation for meta-test tasks. Further, the accuracy continued to increase as the gradient update steps increased.

The experimental results are presented in *Tables 2-5*. Notably, the DaTa-ML model ( $\omega=5$ ) under 1-, 5-, 10-, and 20-shot outperformed baseline MAML with an accuracy of 0.709 [95% confidence interval (CI): 0.708–0.710], 0.775 (95% CI: 0.772–0.777), 0.832 (95% CI: 0.830–0.834), and 0.896 (95% CI: 0.895–0.897), respectively. The accuracy improvement was 25.6%, 25.9%, 26.3%, and 29.1%, respectively. The accuracy of the Da-ML ( $\omega=5$ ) under 1-, 5-, 10-, and 20-shot was 0.531, 0.588, 0.657, and 0.701, respectively, which was 7.8%, 7.2%, 8.8%, and 9.6% higher than the MAML model, respectively. Similarly, the Ta-ML model under 1-, 5-, 10-, and 20-shot achieved an accuracy of 0.582, 0.645, 0.709, and 0.773, respectively, which was 12.9%, 12.9%, 14.0%, and 16.8% higher than the MAML, respectively. These results also indicate that the accuracy of the DaTa-ML model improved as the amount of training



**Figure 6** The adaptation processes of the MAML, Da-ML, Ta-ML, and DaTa-ML models under 1-shot (top left), 5-shot (top right), 10-shot (bottom left), and 20-shot (bottom right). MAML, model-agnostic meta-learning; Da-ML, difficulty-aware based meta-learning; Ta-ML, task-augmentation based on meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

**Table 2** Experiment results of 5-way, 1-shot

Settings	Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1
1-shot (0.05%)	MAML (baseline)	0.453	0.456	0.453	0.454
	Da-ML ( $\omega=3$ )	0.509	0.509	0.509	0.509
	Da-ML ( $\omega=5$ )	0.531	0.532	0.531	0.531
	Da-ML ( $\omega=7$ )	0.520	0.521	0.520	0.520
	Ta-ML	0.582	0.582	0.582	0.582
	DaTa-ML ( $\omega=3$ )	0.687	0.685	0.687	0.686
	DaTa-ML ( $\omega=5$ ) <sup>†</sup>	0.709 <sup>†</sup>	0.709 <sup>†</sup>	0.709 <sup>†</sup>	0.709 <sup>†</sup>
	DaTa-ML ( $\omega=7$ )	0.696	0.698	0.696	0.697

<sup>†</sup>, the best results under the 5-way, 1-shot experiment. MAML, model-agnostic meta-learning; Da-ML, difficulty-aware based meta-learning; Ta-ML, task-augmentation based on meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

data increased.

**Comparison against the state-of-the-art classification models**

The DaTa-ML model was compared with six deep learning-based classification models including AlexNet (43),

VGG16 (44), VGG19 (44), ResNet50 (45), GoogLeNet (46), and SqueezeNet (47).

Table 6 compares the performance of the DaTa-ML model and several pre-trained deep-learning models. The DaTa-ML (20-shot) model had an accuracy of 0.896 and outperformed ResNet50 by an increment of



**Table 3** Experiment results of 5-way, 5-shot

Settings	Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1
5-shot (0.25%)	MAML (baseline)	0.516	0.520	0.516	0.518
	Da-ML ( $\omega=3$ )	0.564	0.564	0.564	0.564
	Da-ML ( $\omega=5$ )	0.588	0.588	0.588	0.588
	Da-ML ( $\omega=7$ )	0.576	0.575	0.576	0.575
	Ta-ML	0.645	0.647	0.645	0.646
	DaTa-ML ( $\omega=3$ )	0.759	0.759	0.759	0.759
	DaTa-ML ( $\omega=5$ ) <sup>†</sup>	0.775 <sup>†</sup>	0.775 <sup>†</sup>	0.775 <sup>†</sup>	0.775 <sup>†</sup>
	DaTa-ML ( $\omega=7$ )	0.766	0.767	0.766	0.766

<sup>†</sup>, the best results under the 5-way, 5-shot experiment. MAML, model-agnostic meta-learning; Da-ML, difficulty-aware based meta-learning; Ta-ML, task-augmentation based on meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

**Table 4** Experiment results of 5-way, 10-shot

Settings	Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1
10-shot (0.5%)	MAML (baseline)	0.569	0.569	0.569	0.569
	Da-ML ( $\omega=3$ )	0.636	0.636	0.636	0.636
	Da-ML ( $\omega=5$ )	0.657	0.657	0.657	0.657
	Da-ML ( $\omega=7$ )	0.648	0.649	0.648	0.648
	Ta-ML	0.709	0.709	0.709	0.709
	DaTa-ML ( $\omega=3$ )	0.811	0.811	0.811	0.811
	DaTa-ML ( $\omega=5$ ) <sup>†</sup>	0.832 <sup>†</sup>	0.834 <sup>†</sup>	0.832 <sup>†</sup>	0.833 <sup>†</sup>
	DaTa-ML ( $\omega=7$ )	0.819	0.820	0.819	0.819

<sup>†</sup>, the best results under the 5-way, 10-shot experiment. MAML, model-agnostic meta-learning; Da-ML, difficulty-aware based meta-learning; Ta-ML, task-augmentation based on meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

**Table 5** Experiment results of 5-way 20-shot

Settings	Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1
20-shot (1%)	MAML (baseline)	0.605	0.605	0.605	0.605
	Da-ML ( $\omega=3$ )	0.682	0.681	0.682	0.681
	Da-ML ( $\omega=5$ )	0.701	0.701	0.701	0.701
	Da-ML ( $\omega=7$ )	0.691	0.691	0.691	0.691
	Ta-ML	0.773	0.776	0.773	0.774
	DaTa-ML ( $\omega=3$ )	0.875	0.876	0.875	0.875
	DaTa-ML ( $\omega=5$ ) <sup>†</sup>	0.896 <sup>†</sup>	0.896 <sup>†</sup>	0.896 <sup>†</sup>	0.896 <sup>†</sup>
	DaTa-ML ( $\omega=7$ )	0.883	0.883	0.883	0.883

<sup>†</sup>, the best results under the 5-way, 20-shot experiment. MAML, model-agnostic meta-learning; Da-ML, difficulty-aware based meta-learning; Ta-ML, task-augmentation based on meta-learning; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.

**Table 6** Performance comparisons between the DaTa-ML model and different pre-trained deep-learning models

Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1	Params ( $\times 10^6$ )
AlexNet	0.863	0.863	0.863	0.863	60.96
VGG16	0.868	0.868	0.868	0.868	138.36
VGG19	0.874	0.874	0.874	0.874	143.68
ResNet50 <sup>†</sup>	0.879 <sup>†</sup>	0.879 <sup>†</sup>	0.879 <sup>†</sup>	0.879 <sup>†</sup>	25.64
GoogLeNet	0.870	0.870	0.870	0.870	6.80
SqueezeNet	0.823	0.821	0.823	0.822	1.25 <sup>†</sup>
DaTa-ML (1-shot)	0.709	0.709	0.709	0.709	0.12
DaTa-ML (5-shot)	0.775	0.775	0.775	0.775	0.12
DaTa-ML (10-shot)	0.832	0.834	0.832	0.833	0.12
DaTa-ML (20-shot) <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.12 <sup>‡</sup>

<sup>†</sup>, the best results from different pre-trained deep-learning models; <sup>‡</sup>, the best results from the methods overall. DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning; VGG, Visual Geometry Group; ResNet, residual neural network.

**Table 7** Performance comparisons between the DaTa-ML model and different scratch-built deep-learning models

Models	Accuracy	Macro_Precision	Macro_Recall	Macro_F1	Params ( $\times 10^6$ )
AlexNet	0.710	0.710	0.710	0.710	60.96
VGG16	0.716	0.716	0.716	0.716	138.36
VGG19	0.723	0.724	0.723	0.723	143.68
ResNet50 <sup>†</sup>	0.728 <sup>†</sup>	0.729 <sup>†</sup>	0.728 <sup>†</sup>	0.728 <sup>†</sup>	25.64
GoogLeNet	0.720	0.720	0.720	0.720	6.80
SqueezeNet	0.674	0.677	0.674	0.675	1.25 <sup>†</sup>
DaTa-ML (1-shot)	0.709	0.709	0.709	0.709	0.12
DaTa-ML (5-shot)	0.775	0.775	0.775	0.775	0.12
DaTa-ML (10-shot)	0.832	0.834	0.832	0.833	0.12
DaTa-ML (20-shot) <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.896 <sup>‡</sup>	0.12 <sup>‡</sup>

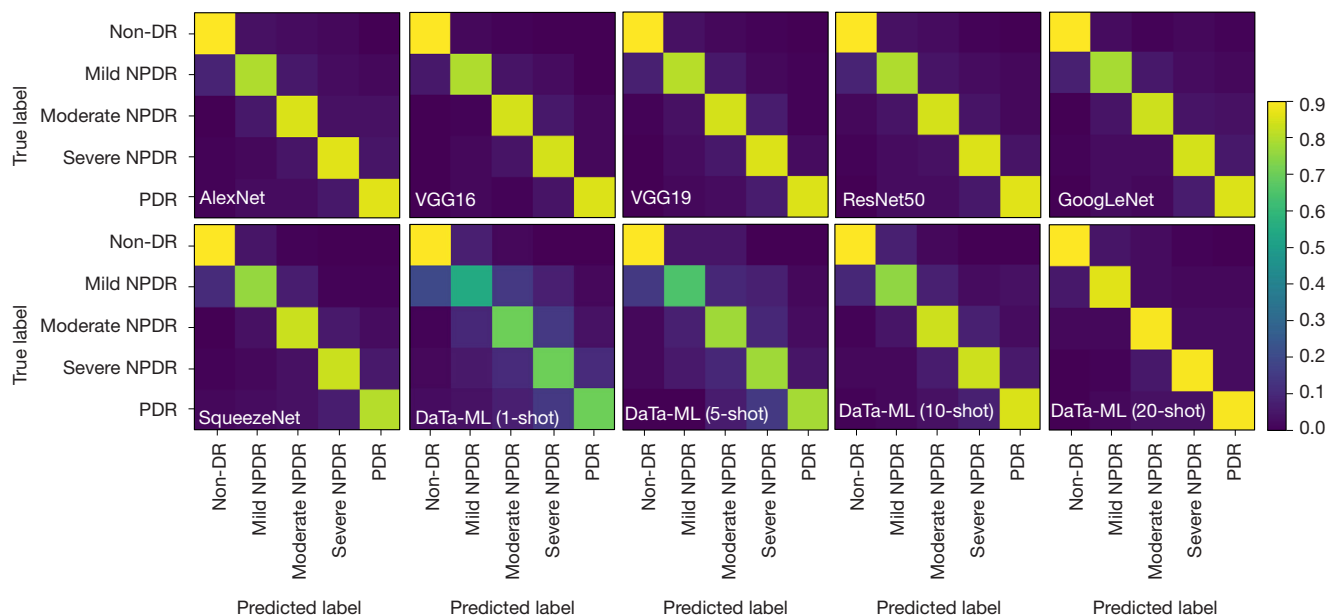
<sup>†</sup>, the best results from different scratch-built deep-learning models; <sup>‡</sup>, the best results from the methods overall. DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning; VGG, Visual Geometry Group; ResNet, residual neural network.

1.7%. More importantly, the DaTa-ML model achieved better performance with only 0.47% of the parameters of ResNet50. The DaTa-ML model also outperformed SqueezeNet with an accuracy increment of 7.3%, even though SqueezeNet had fewer parameters than the comparison methods.

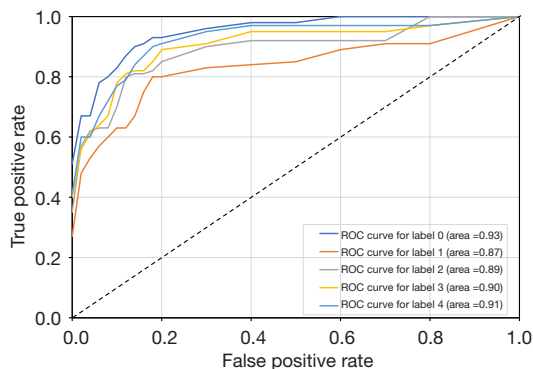
Table 7 compares the performance of the DaTa-ML model and the scratch-built deep-learning models. The DaTa-ML (5-shot) model had an accuracy of 0.775, outperforming ResNet50 by 4.7%. As the training samples increased, the accuracy of the DaTa-ML model

also increased, with the DaTa-ML (20-shot) model outperforming the ResNet50 by an increment of 16.8%.

The confusion matrix for DR classification of the different models is shown in Figure 7. Non-DR, moderate NPDR, severe NPDR, and PDR were easier to classify than mild NPDR. The mild NPDR, moderate NPDR, severe NPDR and PDR of the DaTa-ML (1-shot) model and the DaTa-ML (5-shot) model had lower classification accuracy. In addition to the DaTa-ML (20-shot) model, all the other models had poor classification accuracy for mild NPDR, and the confusion matrix of the DaTa-ML (20-shot) model



**Figure 7** A confusion matrix comparing different the classification models is presented. Higher values are indicated by colors such as yellow and yellow-green. The correctness of the predictions increases as the values increase in the confusion matrix. DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR; VGG, Visual Geometry Group; ResNet, residual neural network; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning.



**Figure 8** The ROC curve of the results of the DR classification. The value of the area indicates the AUC value. The ROC curves are labeled 0–4 (0= non-DR, 1= mild NPDR, 2= moderate NPDR, 3= severe NPDR, and 4= PDR). ROC, receiver operating characteristic; DR, diabetic retinopathy; AUC, area under the curve; NPDR, non-proliferative DR; PDR, proliferative DR.

outperformed those of the other models.

Figure 8 shows the receiver operating characteristic (ROC) curve of the DaTa-ML 20-shot model, where labels 0–4 represent the five categories of DR. The areas under the ROC curve for non-DR, mild NPDR, moderate

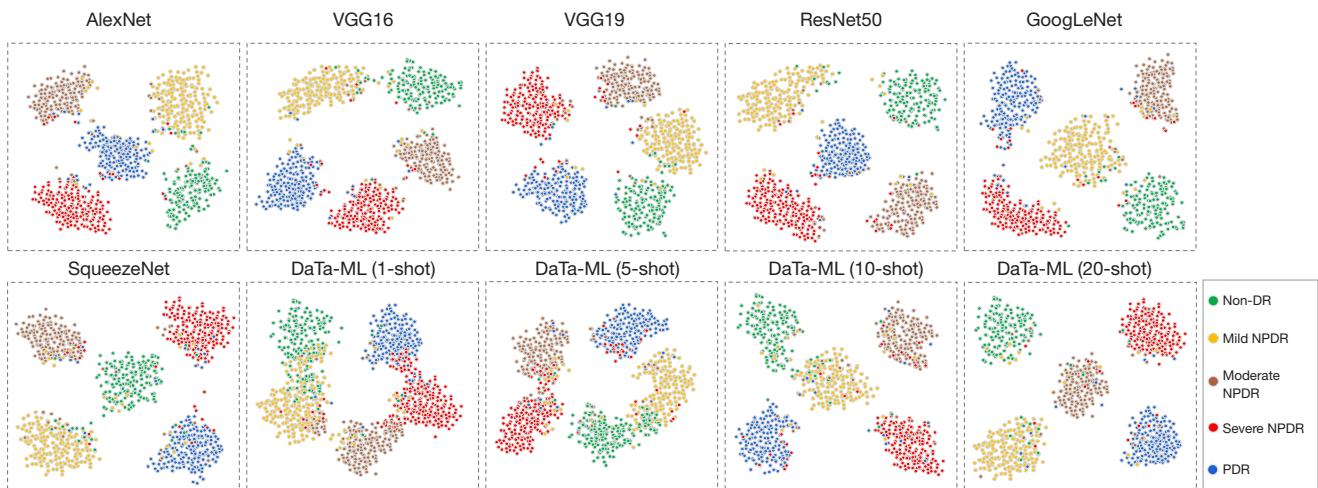
NPDR, severe NPDR, and PDR were 0.93, 0.87, 0.89, 0.90, and 0.91, respectively, demonstrating the effectiveness of the DaTa-ML model.

*Visualization using t-distributed stochastic neighbor embedding (t-SNE)*

The t-SNE was developed to reduce the dimensionality of high-dimensional data to two-dimensional and three-dimensional space for display purposes (48). Figure 9 presents the t-SNE analysis results of the DR test data and compares the DaTa-ML with other models. All the models in the t-SNE diagram aim to divide the learned features into five clusters. We observed that the t-SNE embedding generated by the DaTa-ML (20-shot) model could clearly separate the five groups. Thus, the DaTa-ML model was better able to learn identifiable features compared to other methods when there were only 20 training samples for each category, which made the classification more accurate.

**Discussion**

In this article, we presented an effective few-shot classification model for DR called the DaTa-ML model,



**Figure 9** The t-SNE visualization of the features generated from the last layer using the DaTa-ML model and the different methods. The embeddings from the DaTa-ML (20-shot) model are better clustered. VGG, Visual Geometry Group; ResNet, residual neural network; DaTa-ML, difficulty-aware and task-augmentation method based on meta-learning; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR; t-SNE, t-distributed stochastic neighbor embedding.

which outperformed other state-of-the-art classification models. The proposed DaTa-ML model has a number of advantages. First, the DaTa-ML model can acquire strong learning ability by learning multiple classification tasks and realize the classification of unseen tasks with only a few samples. Notably, both the Da and Ta methods contributed to the improvement of the model's classification performance. In addition, the DaTa-ML model effectively alleviates the negative effects caused by limited data.

The DaTa-ML model outperformed the MAML model, with an accuracy improvement of 27.8% when there were only 20 training data per category. The improved accuracy indicates the effectiveness of the Da and Ta methods for DR classification. As *Figure 5* shows, the DaTa-ML model can extract discriminative features. Ablation studies have shown that the Da and Ta methods are mutually reinforcing, and our proposed DaTa-ML model improved the performance of few-shot DR classification. As *Figure 6* shows, the accuracy of the DaTa-ML model significantly increased after a gradient update, which further demonstrated the fast adaptation ability of the DaTa-ML model.

The confusion matrix showed that the DaTa-ML model showed better classification performance with a small number of samples (20-shot), but the classification accuracy of mild NPDR was lower than that of other classes. The features are difficult to extract because microaneurysms with mild NPDR are very small and similar in color and shape to hemorrhages. As a result, mild NPDR may be misclassified

as non-DR or moderate NPDR. This phenomenon is also very common in the clinic among ophthalmologists.

The t-SNE results showed that the categories were not easily distinguishable in the case of 1-shot and that the more DR shots are taken, the better the classification of the categories. Notably, the t-SNE embedding (20-shot) clearly showed that these five components were separated.

The deep learning-based model achieved excellent generalization performance with a large amount of labeled training data. The DaTa-ML model used the Da and Ta methods to optimize the meta-learning process and obtain the optimal initialization parameters to achieve rapid adaptation on unseen tasks with only a few samples. It outperformed the other deep learning-based models.

We drew comparisons with two similar pieces of research to illustrate the unique aspects of our work. First, diabetes retinopathy network (DRNet) (41) is a prototype network for DR detection and grading. This network constructs a meta-classifier between basic classifiers using the attention mechanism. DRNet achieved a precision of 99.73% on the APTOS 2019 data set compared to our DaTa-ML model that attained an accuracy of 89.6% employing only 1% of the training data (5-way, 20-shot) with merely one updating step. While DRNet's accuracy was visibly higher, our approach still demonstrated noteworthy performance under stringent restrictions on training data. The variation in results could likely be attributed to the attention mechanisms employed by DRNet. FED1 (49), a

few-shot learning model based on Earth Mover's Distance algorithm equipped with a deep residual network for the diagnosis of DR, performed classifications on 39 categories in a 1,000-sample fundus image data set. The FEDI model achieved a precision rate of 93.5667% on the 3-way, 10-shot task. The accuracy of the FEDI model was higher than that of the DaTa-ML model; however, the FEDI model needs the construction of numerous classification tasks and training deep networks that could possibly require substantial computation resources and time. Conversely, by dynamically adjusting the emphasis on more demanding tasks and improving feature-extraction ability via task augmentation, the DaTa-ML model, significantly boosted the model training efficiency, achieving an accuracy of 89.6% with merely 0.47% of the ResNet50 training parameters.

In relation to the limitations of our study, our approach needs to explore ways to enhance model performance even with sparse data, especially when we use a training set smaller than 1% of the current one. Further validation is also needed to assess its generalization capabilities over a broader range and diverse input data conditions. The richness and distribution of the training data could have significantly affected our study results, and this is an aspect of our research that has not yet been explored. Through comparison with other methodologies, our work not only uncovered the potential of the DaTa-ML model in addressing challenging DR classification tasks but also offered avenues for further performance enhancement. This could be achieved, for instance, by adding an attention mechanism or by optimizing our Da and Ta methods. Our approach may also offer valuable insights for researchers working on DR classification research with sparse labeled data.

In the future, we plan to introduce different meta-learner structures or attention mechanisms to achieve the five classifications of DR and further improve the performance of the proposed model, especially for fine-grained classification tasks where subtle differences between classes need to be captured. Additionally, we will continue to explore the potential applications of meta-learning in the field of medical image processing.

## Conclusions

We presented a DaTa-ML model that combines Da and Ta methods within a meta-learning framework to improve few-shot DR classification. The Da method dynamically

adjusts the cross-entropy loss function to focus on difficult tasks automatically. The Ta method increases the number of meta-training tasks by rotating images and improves feature-extraction capability. Ablation studies have demonstrated that Da and Ta methods complement each other. Compared to existing deep-learning models, the DaTa-ML model achieved satisfactory results with a small number of samples. Thus, this model could be used to provide a second opinion to ophthalmologists as to the severity of DR.

## Acknowledgments

*Funding:* This work was supported by the National Natural Science Foundation of China (Nos. 61971271, 62201330, 62271294), the Shandong Province Science and Technology Small and Medium-Sized Enterprises Innovation Capacity Improvement Project (No. 2022TSGC2017), the Shandong Province Major Technological Innovation Project (Nos. 2022CXGC010502, 2022CXGC020507), and the Shandong Provincial Natural Science Foundation Major Basic Research Project (No. ZR2022ZD16).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-23-567/rc>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-567/coif>). Tuo Li and X.Z. are current employees of the Shandong Yunhai Guochuang Cloud Computing Equipment Industry Innovation Co., Ltd., Jinan, China. The other authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by Institutional Review Board of Shandong Normal University, and the requirement of individual consent for this retrospective analysis was waived.

*Open Access Statement:* This is an Open Access article

distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract* 2018;138:271-81.
2. He A, Li T, Li N, Wang K, Fu H. CABNet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. *IEEE Trans Med Imaging* 2021;40:143-53.
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016;316:2402-10.
4. Haneda S, Yamashita H. International clinical diabetic retinopathy disease severity scale. *Nihon Rinsho* 2010;68 Suppl 9:228-35.
5. Sayres R, Taly A, Rahimy E, Blumer K, Coz D, Hammel N, Krause J, Narayanaswamy A, Rastegar Z, Wu D, Xu S, Barb S, Joseph A, Shumski M, Smith J, Sood AB, Corrado GS, Peng L, Webster DR. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology* 2019;126:552-64.
6. Heo SP, Choi H. Development of a robust eye exam diagnosis platform with a deep learning model. *Technol Health Care* 2023;31:423-8.
7. Jinfeng G, Qummar S, Junming Z, Ruxian Y, Khan FG. Ensemble Framework of Deep CNNs for Diabetic Retinopathy Detection. *Comput Intell Neurosci* 2020;2020:8864698.
8. Yan J, Fang Y, Du R, Zeng Y, Zuo Y. No reference quality assessment for 3D synthesized views by local structure variation and global naturalness change. *IEEE Trans Image Process* 2020;29:7443-53.
9. Morales S, Engan K, Naranjo V, Colomer A. Retinal Disease Screening Through Local Binary Patterns. *IEEE J Biomed Health Inform* 2017;21:184-92.
10. Ramlugun GS, Nagarajan VK, Chakraborty C. Small retinal vessels extraction towards proliferative diabetic retinopathy screening. *Expert Syst Appl* 2012;39:1141-6.
11. Shahin EM, Taha TE, Al-Nuaimy W, Rabaie S El, F.Zahran O, El-Samie FEA. Automated detection of diabetic retinopathy in blurred digital fundus images. In: 2012 8th International Computer Engineering Conference (ICENCO). IEEE; 2012:20-5.
12. Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT. Application of random forests methods to diabetic retinopathy classification analyses. *PLoS One* 2014;9:e98587.
13. Huang P, Wang J, Zhang J, Shen Y, Liu C, Song W, Wu S, Zuo Y, Lu Z, Li D. Attention-Aware Residual Network Based Manifold Learning for White Blood Cells Classification. *IEEE J Biomed Health Inform* 2021;25:1206-14.
14. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput Sci* 2016;90:200-5.
15. García G, Gallardo J, Mauricio A, López J, Del Carpio C. Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images. In: Artificial Neural Networks and Machine Learning-ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26. Springer International Publishing; 2017:635-42.
16. Shanthi T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering* 2019;76:56-64.
17. Qi GJ, Luo J. Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods. *IEEE Trans Pattern Anal Mach Intell* 2022;44:2168-87.
18. Burlina P, Paul W, Mathew P, Joshi N, Pacheco KD, Bressler NM. Low-Shot Deep Learning of Diabetic Retinopathy With Potential Applications to Address Artificial Intelligence Bias in Retinal Diagnostics and Rare Ophthalmic Diseases. *JAMA Ophthalmol* 2020;138:1070-7.
19. de La Torre J, Valls A, Puig D. A deep learning interpretable classifier for diabetic retinopathy disease grading. *Neurocomputing* 2020;396:465-76.
20. Yoo TK, Choi JY, Kim HK. Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification. *Med Biol Eng Comput*

- 2021;59:401-15.
21. Gao J, Leung C, Miao C. Diabetic retinopathy classification using an efficient convolutional neural network. In: 2019 IEEE International Conference on Agents (ICA). IEEE; 2019:80-5.
  22. Zhou Y, Wang B, He X, Cui S, Shao L. DR-GAN: Conditional Generative Adversarial Network for Fine-Grained Lesion Synthesis on Diabetic Retinopathy Images. *IEEE J Biomed Health Inform* 2022;26:56-66.
  23. Pan S J, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2009;22:1345-59.
  24. Li X, Pang T, Xiong B, Liu W, Liang P, Wang T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: 2017 10th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMED). IEEE; 2017:1-11.
  25. Fan X, Zou L, Liu Z, He Y, Zou L, Chi R. CSAC-Net: Fast Adaptive sEMG Recognition through Attention Convolution Network and Model-Agnostic Meta-Learning. *Sensors (Basel)* 2022;22:3661.
  26. Gao K, Liu B, Yu X, Yu A. Unsupervised Meta Learning With Multiview Constraints for Hyperspectral Image Small Sample set Classification. *IEEE Trans Image Process* 2022;31:3449-62.
  27. Ma R, Han T, Lei W. Cross-domain meta learning fault diagnosis based on multi-scale dilated convolution and adaptive relation module. *Knowledge-Based Systems* 2023;261:110175.
  28. Yang Q, Ni Z, Ren P. Meta captioning: A meta learning based remote sensing image captioning framework. *ISPRS J Photogramm Remote Sens* 2022;186:190-200.
  29. Li X, Yu L, Jin Y, Fu CW, Xing L, Heng PA. Difficulty-aware meta-learning for rare disease diagnosis. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part I* 23. Springer International Publishing; 2020:357-66.
  30. Yuan P, Mobiny A, Jahanipour J, Li X, Cicalese PA, Roysam B, et al. Few is enough: task-augmented active meta-learning for brain cell classification. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part I* 23. Springer International Publishing; 2020:367-77.
  31. Vilalta R, Drissi Y. A perspective view and survey of meta-learning. *Artificial Intelligence Review* 2002;18:77-95.
  32. Zhang P, Bai Y, Wang D, Bai B, Li Y. Few-shot classification of aerial scene images via meta-learning. *Remote Sensing* 2020;13:108.
  33. Singh R, Bharti V, Purohit V, Kumar A, Singh AK, Singh SK. MetaMed: Few-shot medical image classification using gradient-based meta-learning. *Pattern Recognition* 2021;120:108111.
  34. Li C, Li S, Zhang A, He Q, Liao Z, Hu J. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing* 2021;439:197-211.
  35. Xu Z, Chen X, Tang W, Lai J, Cao L. Meta weight learning via model-agnostic meta-learning. *Neurocomputing* 2021;432:124-32.
  36. Gao K, Liu B, Yu X, Zhang P, Tan X, Sun Y. Small sample classification of hyperspectral image using model-agnostic meta-learning algorithm and convolutional neural network. *Int J Remote Sens* 2021;42:3090-122.
  37. Oreshkin BN, Rodriguez P, Lacoste A. TADAM: task dependent adaptive metric for improved few-shot learning. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 2018:719-29.
  38. Kaggle. APTOS 2019 Blindness Detection. APTOS 2019. Accessed November 2019. Available online: <https://www.kaggle.com/c/aptos2019-blindness-detection/overview/aptos-2019>
  39. Li J, Chiu B, Feng S, Wang H. Few-shot named entity recognition via meta-learning. *IEEE Trans Knowl Data Eng* 2020;34:4245-56.
  40. Zhou H, Ma Y, Li X. Feature selection based on term frequency deviation rate for text classification. *Applied Intelligence* 2021;51:3255-74.
  41. Murugappan M, Prakash NB, Jeya R, Mohanarathinam A, Hemalakshmi GR, Mahmud M. A novel few-shot classification framework for diabetic retinopathy detection and grading. *Measurement* 2022;200:111485.
  42. He T, Guo J, Chen N, Xu X, Wang Z, Fu K, Liu L, Yi Z. MediMLP: Using Grad-CAM to Extract Crucial Variables for Lung Cancer Postoperative Complication Prediction. *IEEE J Biomed Health Inform* 2020;24:1762-71.
  43. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.
  44. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556 [Preprint]*. 2014. Available online: <https://arxiv.org/abs/1409.1556>
  45. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

- 2016:770-8.
46. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:1-9.
  47. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv:1602.07360 [Preprint]. 2016. Available online: <https://arxiv.org/abs/1602.07360>
  48. Quellec G, Lamard M, Conze PH, Massin P, Cochener B. Automatic detection of rare pathologies in fundus photographs using few-shot learning. *Med Image Anal* 2020;61:101660.
  49. Pan L, Zhang P, Xia F, Ji B, Liu W, Wang H, et al. FEDI: Few-shot learning based on Earth Mover's Distance algorithm combined with deep residual network to identify diabetic retinopathy. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2021:1032-6.

**Cite this article as:** Liu X, Dong X, Li T, Zou X, Cheng C, Jiang Z, Gao Z, Duan S, Chen M, Liu T, Huang P, Li D, Lu H. A difficulty-aware and task-augmentation method based on meta-learning model for few-shot diabetic retinopathy classification. *Quant Imaging Med Surg* 2024;14(1):861-876. doi: 10.21037/qims-23-567



**Table S1** Parameter setting of meta learner

Type	Filter	Stride	Filter	Padding
Convolution I	3×3×32	1×1		Same
Max pooling I	2×2×32	2×2		Valid
Convolution II	3×3×32	1×1		Same
Max pooling II	2×2×32	2×2		Valid
Convolution III	3×3×32	1×1		Same
Max pooling III	2×2×32	2×2		Valid
Convolution IV	3×3×32	1×1		Same
Max pooling IV	2×2×32	2×2		Valid
Full connection	288×1	N/A		N/A
Output	2×1	N/A		N/A

N/A, not available.

**Table S2** New terms for MAML

Term	Detail
Support set	Each support set contains N categories, each of which contains K samples
Query set	Every query set contains a set number of remaining samples from the selected N categories
Meta-training set	The meta-training set ( $D_{tr}$ ) contains multiple meta-training tasks, each of which includes a support set $D_s$ and a query set $D_q$
Meta-testing set	The meta-testing set ( $D_{te}$ ) constitutes the meta-testing task, whose support sets are all unseen categories, and the query set is applied to test the performance of few-shot model

MAML, model-agnostic meta-learning.