Fine-Tuning Network in Federated Learning for Personalized Skin Diagnosis

Kyungsu Lee[1], Haeyun Lee[2], Thiago Coutinho Cavalcanti[1], Sewoong Kim[1], Georges El Fakhri[3], Dong Hun Lee[4], Jonghye Woo[3], and Jae Youn Hwang[1(✉)]

[1] Department of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology, Daegu 42988, South Korea
{ks_lee,jyhwang}@dgist.ac.kr
[2] Production Engineering Research Team, Samsung SDI, Yongin 17084, South Korea
[3] Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[4] Department of Dermatology, Seoul National University College of Medicine, Institute of Human-Environment Interface Biology, Seoul National University, Seoul 03080, South Korea

Abstract. Federated learning (FL) has emerged as a promising technique in the field of medical diagnosis. By distributing the same task through deep networks on mobile devices, FL has proven effective in diagnosing dermatitis, a common and easily recognizable skin disease. However, in skin disease diagnosis, FL poses challenges related to (1) prioritizing generalization over personalization and (2) limited utilization of mobile devices. Despite its improved comprehensive diagnostic performance, skin disease diagnosis should aim for personalized diagnosis rather than centralized and generalized diagnosis, due to personal diversities and variability, such as skin color, wrinkles, and aging. To this end, we propose a novel deep learning network for personalized diagnosis in an adaptive manner, utilizing personal characteristics in diagnosing dermatitis in a mobile- and FL-based environment. Our framework, dubbed APD-Net, achieves adaptive and personalized diagnosis using a new model design and a genetic algorithm (GA)-based fine-tuning method. APD-Net incorporates a novel architectural design that leverages personalized and centralized parameters, along with a fine-tuning method based on a modified GA to identify personal characteristics. We validated APD-Net on clinical datasets and demonstrated its superior performance, compared with state-of-the-art approaches. Our experimental results showed that APD-Net markedly improved personalized diagnostic accuracy by 9.9% in dermatitis diagnosis, making it a promising tool for clinical practice.

Keywords: Personalized Diagnosis · Federate Learning · Mobile-based Diagnosis · Skin Cancer

K. Lee and H. Lee—Contributed equally.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43898-1_37.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 378–388, 2023.
https://doi.org/10.1007/978-3-031-43898-1_37

1 Introduction

For the past several years, in skin disease diagnosis, deep learning (DL) techniques have been extensively studied, due to its effectiveness and outstanding diagnostic performance [8,16,19]. For example, Wu *et al.* used a custom dataset to develop an EfficientNet-b4-based DL model and successfully diagnosed skin diseases [19]. Srinivasu *et al.* designed an advanced DL model, by combining long short-term memory (LSTM) with MobileNet and achieved improved performance as well as fast prediction time in skin disease diagnosis [16]. For the development of DL models, a large number of datasets are needed for accurate model fitting at the training stage. Acquiring enormous skin disease datasets, however, at a single medical site is challenging. As such, the performance of DL models is often limited, due to a small number of datasets [2,16,20].

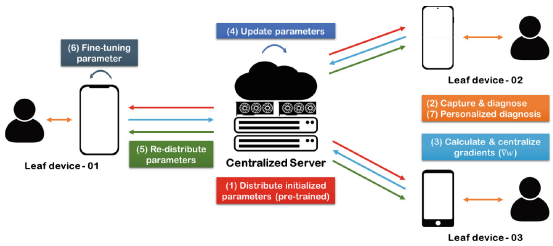


Fig. 1. Pipeline of ADP-Net framework for a personalized diagnosis.

To overcome the limitations mentioned above, federated learning (FL) can be a viable solution in the context of digital healthcare, especially in the COVID-19 pandemic era [15]. In particular, FL enables that the edge devices only share the gradient of DL models without sharing data, such that FL improves the data privacy and security. Moreover, FL allows to acquire many heterogeneous images from edge devices at multiple medical sites [1,6,7,10,15]. For example, B. McMahan *et al.* developed a protocol to average the gradients from decentralized clients, without data sharing [10]. K. Bonawitz *et al.* built high-level FL systems and architectures in a mobile environment [1]. However, DL models in the FL environment were optimized to deal with datasets from multiple clients; therefore, while the DL models yielded a generalized prediction capability across all of the domains involved, the DL models cannot efficiently perform personalized diagnosis, which is deemed a weakness of FL. To alleviate this, personalized FL methodologies have been emerged [12,14,17].

We propose to develop a personalized diagnosis network, called APD-Net in the FL framework to target personalized diagnostics. Our APD-Net comprises two novel techniques, including (1) a genetic algorithm (GA)-based fine-tuning method and (2) a dual-pipeline (DP) architecture for the DL models. The GA-based fine-tuning method improves APD-Net on each edge device by adaptively customizing the optimized DL model. We validated our framework on three public datasets, including 7pt [4], Human Against Machine (HAM) [18], and International Skin Imaging Collaboration (ISIC) [13], as well as our own datasets.

Experimental results demonstrated that the APD-Net yielded outstanding performance, compared with other comparison methods, and achieved adaptively personalized diagnosis. The contributions of this paper are three-fold:

- We developed a mobile- and FL-based learning (APD-Net) for skin disease diagnosis and achieved superior performance on skin disease diagnosis for public and custom datasets.
- We introduce a customized GA for APD-Net, combined with a corresponding network architecture, resulting in improved personalized diagnostic performance as well as faster prediction time.
- We provide a new fluorescence dataset for skin disease diagnosis containing 2,490 images for four classes, including Eczema, Dermatitis, Rosacea, and Normal. This dataset is made publicly available for future research in the field.

2 Methodology

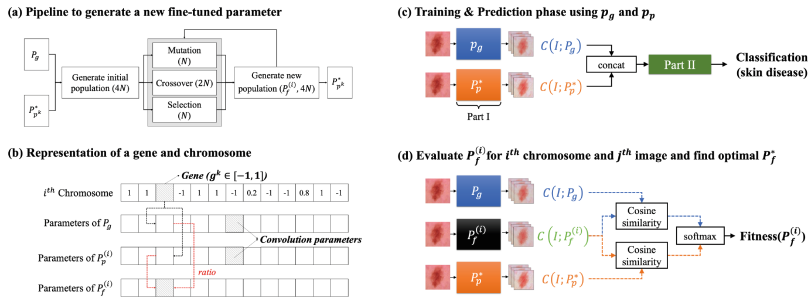


Fig. 2. Pipeline for APD-Net and representation of GA

This work aims to develop a mobile-based FL system that can provide a personalized and customized diagnosis to patients across different clients. The APD-Net framework, shown in Fig. 1, includes common procedures (1)–(5) that are typically used in a general FL system, as well as unique sequences (6) and (7) for adaptively personalized diagnosis in the proposed system. In particular, as depicted in Fig. 1 (6), the parameters transferred from the server are fine-tuned to be suitable for each domain.

The detailed procedure to fine-tune APD-Net is described in Fig. 2(a). Here, P_g represents a set of generalized parameters in the FL server, and $P^*_{p^k}$ represents a set of optimally fine-tuned parameters in the k^{th} client. Before diagnosis, a set of newly fine-tuned parameters (P_f) is generated by GA. By jointly using both P_g and $P^*_{p^k}$, the initial population of a set of personalized parameters ($P^{(i)}_{p^k}$, here chromosome) is achieved, where $i = 1, 2, 3, ..., N$, and N is the number of populations. The evolutionary operations, including crossover and mutation, then offer a new $4N$ number of chromosomes. Here, the fitness scores are compared for all

individual chromosomes, and the N number of chromosomes that achieve a high fitness score is selected to form a new population. Subsequently, the fitness score is calculated using the DP architecture, as illustrated in Fig. 2(d). After several generations of fine-tuned parameters, the chromosome, evaluated as the highest fitness score, replaces the personalized parameters, $P^*_{p^k}$. Finally, the optimally fine-tuned parameters are utilized to diagnose patients in each client.

2.1 Dual-Pipeline (DP) Architecture for APD-Net

Since the proposed FL system is implemented in a mobile-based environment, MobileNetV3 is employed as a baseline network of APD-Net. In particular, to achieve faster fine-tuning time, MobileNetV3-small is utilized. Here, the novel architecture of APD-Net is differentiated by the use of the DP architecture that allows (1) to diagnose patients adaptively, and (2) to evaluate the fitness function. Therefore, in the DP architecture, one pipeline employs the generalized parameters (P_g) from the FL server, whereas the other pipeline employs the personalized parameters ($P^*_{p^k}$) in the client.

2.2 Customized Genetic Algorithm

In the FL environment, data privacy is achieved by transferring gradients without sharing data. Therefore, since the domain of one client is not recognizable by another client, the domain gap between two clients is not computable. Therefore, in the proposed FL system, to adaptively fine-tune the parameters transferred from the FL server to be personalized concerning the domain of one client, the GA is employed. The GA is the optimal solution for adaptively personalized diagnosis in the FL environment, since it heuristically searches for another local minimum point regardless of the domain gaps. The detailed procedures of the GA are illustrated in Algorithm I (Appendix).

Gene and Chromosome. A gene and a chromosome are modeled to represent fine-tuned parameters ($P^{(i)}_f$) by jointly using P_g and P^*_f. The gene indicates the internal division between P_g and P^*_p. Figure 2(b) represents the mathematical modeling of a gene and a chromosome. In particular, let $g^{(i)}_k \in [-1, 1]$ be a k^{th} gene in the i^{th} chromosome, and then the k^{th} convolution weight ($P^{(i)}_f|_k$) is formulated as follows:

$$P^{(i)}_f|_k = 0.5(1 - g^{(i)}_k)P_g|_k + 0.5(1 + g^{(i)}_k)P^*_p|_k \tag{1}$$

where $P_g|_k$ and $P^*_p|_k$ are the k^{th} convolution parameter in P_g and P^*_f, respectively. Here, $P^{(i)}_f|_k$ is the average value of $P_g|_k$ and $P^*_p|_k$ if $g^{(i)}_k = 0$. In contrast, $P^{(i)}_f|_k = P_g|_k$ if $g^{(i)}_k = -1$, whereas $P^{(i)}_f|_k = P^*_p|_k$ if $g^{(i)}_k = 1$. In short, $P^{(i)}_f$ is calculated by the internal division between P_g and P^*_p.

Crossover. Let $crossover(P^{(i)}, P^{(j)})$ be a crossover function by jointly using two chromosomes, and then the k^{th} genes of $P^{(i)}|_k$ and $P^{(j)}|_k$ are changed in the 50% probability when the constraint of $|P^{(i)}|_k - P^{(j)}|_k| \leq l_k$ is satisfied. Here, since the convolution parameters in a deep depth are rarely fine-tuned due to a gradient vanishing problem, l_k exhibits a relatively larger value when k becomes larger. In addition, since we experimentally demonstrated that $l_k \geq 0.15$ provides a much longer time to fine-tune APD-Net, we constrained $l_k \leq 0.15$, and the fixed values of l_k are randomly determined for each experiment.

Mutation. $mutation\ (P^{(i)})$ represents a mutation function onto a chromosome, of which the k^{th} gene is $P^{(i)}|_k$, such that it is defined as follows:

$$mutation(P^{(i)}|_k) = \eta P^{(i)}|_k \quad \eta \in [1-\mu, 1+\mu], \tag{2}$$

where η is the randomly selected value in the constraint range for each individual gene. While training the DL model, we experimentally verified that the convolution weights are changed within the range of the maximum 0.2%. Therefore, here, μ is initially determined as 2e−3 (0.2%), but it depends on the variance of convolution weights in every epoch.

Selection. As illustrated in Algorithm I, the newly generated chromosomes, which yield a large value of the fitness score, are contained in a new population. In APD-Net, the fitness score is evaluated by the fitness function that is jointly utilized in the architecture of APD-Net as illustrated in Fig. 2(d). Let $C(I;P)$ be the output of Part I using the parameter P with an input image (I). As illustrated in Fig. 2(d), we can then calculate three outputs of $C(I;P_g)$, $C(I;P_p^*)$, and $C(I;P_f^{(i)})$, which use a generalized parameter, a personalized parameter, and a candidate parameter for a fine-tuned parameter, respectively. To achieve personalized diagnosis, the fine-tuned parameter should be similar to the personalized parameter (client) rather than the generalized parameter (server). Therefore, we define the fitness score as the ratio of the similarity between P_p^* and $P_f^{(i)}$ to the similarity between P_g and $P_f^{(i)}$. Here, we use a softmax function to convert the ratio into probability, as formulated below:

$$fitness(P^{(i)}) = \frac{exp(sim(P_p^*, P_f^{(i)}))}{exp(sim(P_p^*, P_f^{(i)})) + exp(sim(P_g, P_f^{(i)}))}, \tag{3}$$

where $sim(x,y)$ is cosine similarity between x and y, and $exp(x)$ is the exponential function. Note that APD-Net provides the fitness function related to its architecture, and the fitness function is more reliable than other fitness functions used in accuracy-based GA. Therefore, the APD-Net with our GA offers high accuracy in both the conventional diagnosis for overall patients and the personalized diagnosis for each patient at a specific client.

2.3 Training and Fine-Tuning APD-Net

To summarize, (1) APD-Net is initially trained in the FL server using gradients from many clients. The cross-entropy loss function is utilized for training APD-Net, and the generally optimized parameter (P_g) is achieved in the initial training. (2) P_g is then transferred to each client, and (3) the fine-tuned procedures are processed by jointly using P_g and a personalized parameter (P_p^*). By using the proposed GA, the new population of fine-tuned parameters ($P_f^{(i)}$ where $i = 1, 2, ..., 4N$) is generated, and the chromosome with the highest fitness score becomes a new personalized parameter. (4) After the diagnosis, the gradients are shared with the FL server, and P_g is newly optimized.

Table 1. Configurations of datasets (Left) and Experiment I (Right).

Datasets	Nevus	Melanoma	Others	Total
7pt	575	268	168	1011
ISIC	5193	284	27349	32826
HAM	6705	1113	2197	10015
Testset (25%)	Nevus	Melanoma	Others	Total
7pt	144	67	42	253
ISIC	1299	71	6838	8207
HAM	1677	279	550	2504

# Images (# for Testset)	Group	Nevus	Melanoma	Others
7pt	**G-00**	575 (144)	268 (67)	168 (42)
ISIC	**G-01**	180 (45)	80 (20)	17070 (4268)
	G-02	685 (172)	363 (91)	7915 (1979)
	G-03	4328 (1082)	141 (36)	2364 (591)
HAM	**G-04**	85 (22)	16 (4)	89 (23)
	G-05	828 (207)	359 (90)	673 (169)
	G-06	5792 (1448)	738 (185)	1435 (359)

Table 2. (Left) The number of skin images in the customized dataset for Experiment II. **(Right)** Summary of key features of the classification models.

	Total Samples	Each Client	Test Samples
Eczema	258	86	18
Dermatitis	294	98	20
Rosacea	738	246	50
Normal	1200	400	80
Total	2490	830	168

	Task	Multi Domain	Skin Dataset	Mobile	Runtime	StoA
Y. Gu [3]	DA	✓	✓			
M. Yu [11]	DA	✓				✓
K. Lee [5]	DA + Seg	✓			✓	✓
Y. Wu	FL + Seg					
P. Yao [21]			✓			✓
P. Srini [16]			✓	✓		
Y. Mou [12]						✓
A. Tan [17]						✓
A. Sham [14]						✓
Ours	FL + DA	✓	✓	✓		✓

3 Experimental Results

3.1 Experimental Setup

Dataset. To evaluate the performance and feasibility of APD-Net, we used three public datasets, including 7pt, ISIC, and HAM, and detailed descriptions for datasets are illustrated in Table 1. Furthermore, in this work, we collected skin images through the tertiary referral hospital under the approval of the institutional review board (IRB No. 1908-161-1059) and obtained images with the consent of the subjects according to the principles of the Declaration of Helsinki from 51 patients and subjects. The dataset included four categories,

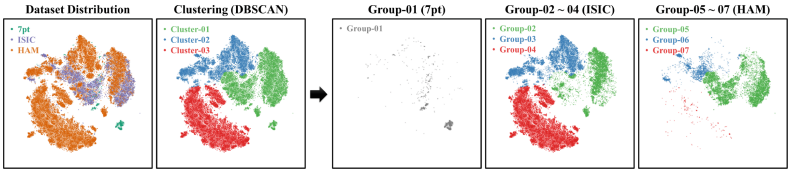


Fig. 3. Re-sampling the distribution of images in each client for Experiment I

Table 3. (Left) Results of ablation studies and **(Right)** Comparison analysis of APD-Net and the other DL models. The highest values are highlighted as **bold**, and the second values are underlined.

	APD-Net	APD-GA	APD-GA-DP
G-01	84.11%	79.26%	75.58%
G-02	96.97%	92.62%	91.07%
G-03	94.82%	91.83%	88.89%
G-04	86.79%	80.38%	76.30%
G-05	76.41%	64.08%	61.45%
G-06	84.17%	79.72%	75.38%
G-07	84.67%	78.77%	75.22%
Avg	86.85%	80.95%	77.70%

	G-01	G-02	G-03	G-04	G-05	G-06	G-07
Y.Gu	80.04%	94.94%	93.35%	84.65%	71.13%	80.61%	78.86%
M. Yu	80.22%	95.04%	92.68%	84.72%	65.92%	80.27%	78.78%
K. Lee	78.15%	95.48%	93.04%	84.15%	72.56%	79.05%	79.60%
Y. Wu	80.54%	88.41%	88.60%	84.81%	61.72%	81.64%	81.41%
P. Yao	74.81%	90.04%	88.90%	76.88%	61.64%	75.86%	76.10%
P. Srini	75.01%	92.09%	90.51%	76.77%	62.14%	76.17%	75.00%
Y. Mou	80.61%	93.64%	91.54%	83.42%	73.02%	80.74%	81.06%
A. Tan	83.57%	96.44%	94.30%	86.28%	75.88%	83.63%	84.18%
A. Sham	81.30%	94.15%	92.13%	83.90%	73.52%	81.35%	82.10%
Ours	**84.11%**	**96.97%**	**94.82%**	**86.79%**	**76.41%**	**84.17%**	**84.67%**

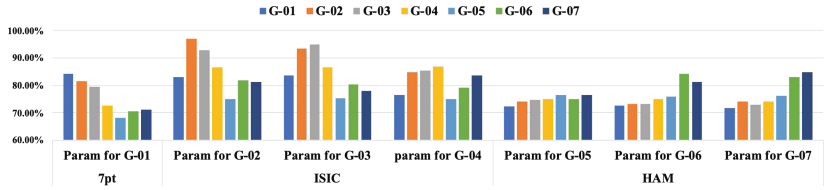


Fig. 4. Classification accuracy of APD-Nets in each FL client

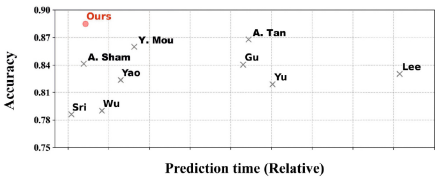
including eczema, dermatitis, rosacea, and normal skin, with 258, 294, 738, and 1,200 images, respectively, as illustrated in Table 2(Left).

To compensate for the limited number of images in the test set, a 4-fold cross-validation approach was employed. To assess the performance of the proposed network as well as compared networks, two distinct FL environments were considered: (1) an FL simulation environment to evaluate the performance of APD-Net and (2) a realistic FL environment to analyze the feasibility of APD-Net. For the FL simulation environment in Experiment I, public datasets were employed, and the distribution of samples was re-sampled using t-Distributed Stochastic Neighbor Embedding (t-SNE) [9]. The images in all skin datasets were subsequently re-grouped, as illustrated in Fig. 3. In contrast, for Experiment II, we utilized a custom dataset for a realistic FL environment. Six DL models that have shown exceptional performances in DA, FL, and skin disease diagnosis were used as comparison methods to evaluate the FL and DL performance of APD-Net. The salient characteristics of these models are summarized in Table 2(Right).

3.2 Experiment I. FL Simulation

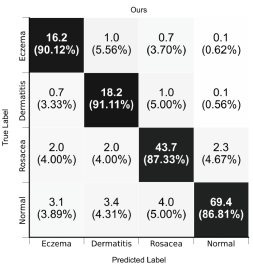
Ablation Study. An ablation study was conducted to evaluate the impact of GA and DP on diagnostic performance. APD-GA and APD-GA-DP indicate APD-Net without GA and without GA and DP, respectively. APD-DP was not evaluated since GA could not be realized without the DP architecture. As illustrated in Table 3(Left), the APD-Net with GA and DP yielded the best performance. Here, it is important to note that the DP architecture also improved the performance of the models for adaptively personalized diagnosis, similar to GA, by jointly using personalized and generalized parameters in the DP architecture.

Comparison Analysis. Performance of APD-Net was compared against those of the other DL models for adaptively personalized diagnosis. Table 3(Right) shows the performances of the DL models in every client (group). APD-Net yielded an accuracy of 9.11%, which was higher than the other DL models for adaptively personalized diagnosis. Furthermore, Fig. 4 demonstrates that the prediction with images from other clients yielded lower accuracy.



	Client-1	Client-2	Client-3	Avg
Y. Gu	83.96%	84.21%	84.04%	84.07%
M. Yu	80.99%	83.71%	81.12%	81.94%
K. Lee	83.09%	84.61%	81.42%	83.04%
Y. Wu	78.24%	79.75%	79.02%	79.01%
P. Yao	81.79%	82.88%	82.54%	82.40%
P. Srini	78.07%	79.81%	77.95%	78.61%
Y. Mou	85.68%	86.92%	85.57%	86.06%
A. Tan	86.27%	87.39%	86.41%	86.69%
A. Sham	84.30%	85.72%	84.34%	84.79%
Ours	**88.03%**	**89.28%**	**88.23%**	**88.51%**

Fig. 5. (Left) Fine-tuning time and classification accuracy by APD-Nets and other models. **(Right)** Comparison analysis of APD-Net and the other DL models in a desirable FL environment. The highest values are highlighted as **bold**, and the second values are underlined.



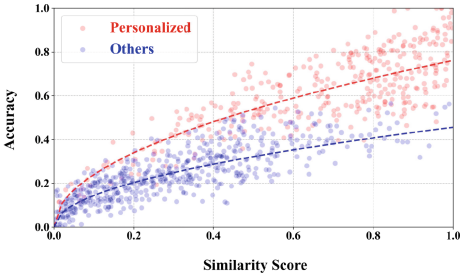


Fig. 6. (Left) Confusion matrix of APD-Net with the customized dataset in terms of four-classes classification. **(Right)** Correlation between similarity score (fitness function) and accuracy.

3.3 Experiment II. Realistic FL Environment

To verify the feasibility of APD-Net, the performance of APD-Net was evaluated using the customized datasets acquired from our devices for adaptively personalized diagnosis. The performance of APD-Nets was compared against the other DL models. Since the prediction time is critical in the mobile-based environment, the prediction times of the DL models were compared in addition to the accuracy. As illustrated in Fig. 5(Left), APD-Net (Ours) yielded an outstanding performance as well as a shorter prediction time compared with the other DL models for adaptively personalized diagnosis. In addition, Fig. 5(Right) shows the performances of APD-Nets for adaptively personalized diagnosis in every client. APD-Net achieved an improved accuracy of 9.9%, compared with the DL models in adaptively personalized diagnosis. Furthermore, Fig. 6(Left) illustrates the performance of APD-Net. Since the number of images is relatively small, the images were divided into three clients. In this work, 3-fold cross-validation was applied to evaluate our APD-Net. The results showed that our APD-Net has the potential to be used in the FL environment with an accuracy of 88.51%.

In addition, to verify the similarity as a fitness score, we examined the correlation between the similarity score and the prediction accuracy. The fitness score and accuracy were calculated corresponding to many input images and various versions of the fine-tuned parameters. Figure 6 shows that the classification accuracy was improved when the parameters with high similarity scores were used. Since the GA requires significantly longer fine-tuning time, the fine-tuning step was detached from the training and prediction steps at the application level. In particular, in the synchronization step where the generalized parameters were transferred from the FL server, the GA-based fine-tuning was realized, and a new optimal personalized parameter was generated. Therefore, the prediction time without the fine-tuning step was significantly reduced. It is important to note that, at the application level, a slightly longer synchronization time is generally considered more acceptable than a longer prediction time.

4 Conclusion

In this work, we first carried out the adaptively personalized diagnosis task in the FL system, and developed a novel DL model, termed APD-Net, with the FL system. Our APD-Net yielded outstanding performance, by jointly using a novel DP architecture and a GA-based fine-tuning technique for adaptively personalized diagnosis. The DP architecture enabled extracting feature maps using generalized and personalized parameters, thereby providing high diagnostic accuracy. In addition, GA heuristically generated the best performance for the personalized parameters. Using the public skin datasets, including 7pt, ISIC, and HAM, our APD-Net was able to achieve an improved accuracy of 9.9% compared with other state-of-the-art DL models for the adaptively personalized diagnosis. The ablation study also demonstrated that the partial fine-tuning technique achieved higher accuracy as well as a faster fine-tuning time. Furthermore, the feasibility of APD-Net in FL was demonstrated by using the customized datasets acquired

from our system, thus suggesting that the proposed system with APD-Net can be applied to the multiclass classification task of various skin diseases. However, achieving even faster prediction speeds is possible by incorporating computing performance enhancement methodologies from related fields or employing lightweight techniques like quantization, and it remains a future work.

References

1. Bonawitz, K., et al.: Towards federated learning at scale: system design. Proc. Mach. Learn. Syst. **1**, 374–388 (2019)
2. Chowdhury, M.M.U., Hammond, F., Konowicz, G., Xin, C., Wu, H., Li, J.: A few-shot deep learning approach for improved intrusion detection. In: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 456–462. IEEE (2017)
3. Gu, Y., Ge, Z., Bonnington, C.P., Zhou, J.: Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. IEEE J. Biomed. Health Inform. **24**(5), 1379–1393 (2019)
4. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE J. Biomed. Health Inform. **23**(2), 538–546 (2018)
5. Lee, K., Lee, H., Hwang, J.Y.: Self-mutating network for domain adaptive segmentation in aerial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7068–7077 (2021)
6. Li, L., Fan, Y., Tse, M., Lin, K.Y.: A review of applications in federated learning. Comput. Ind. Eng. **149**, 106854 (2020)
7. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: challenges, methods, and future directions. IEEE Sig. Process. Mag. **37**(3), 50–60 (2020)
8. Liu, Y., et al.: A deep learning system for differential diagnosis of skin diseases. Nat. Med. **26**(6), 900–908 (2020)
9. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), 2579–2605 (2008)
10. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282. PMLR (2017)
11. Mitsuzumi, Y., Irie, G., Ikami, D., Shibata, T.: Generalized domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1084–1093 (2021)
12. Mou, Y., Geng, J., Welten, S., Rong, C., Decker, S., Beyan, O.: Optimized federated learning on class-biased distributed data sources. In: Kamp, M., et al. (eds.) Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD 2021. Communications in Computer and Information Science, vol. 1524, pp. 146–158. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-93736-2_13
13. Rotemberg, V., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Sci. Data **8**(1), 1–8 (2021)
14. Shamsian, A., Navon, A., Fetaya, E., Chechik, G.: Personalized federated learning using hypernetworks. In: International Conference on Machine Learning, pp. 9489–9502. PMLR (2021)

15. Shyu, C.R., et al.: A systematic review of federated learning in the healthcare area: from the perspective of data properties and applications. Appl. Sci. **11**(23), 11191 (2021)
16. Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J.: Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors **21**(8), 2852 (2021)
17. Tan, A.Z., Yu, H., Cui, L., Yang, Q.: Towards personalized federated learning. IEEE Trans. Neural Netw. Learn. Syst. (2022)
18. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**(1), 1–9 (2018)
19. Wu, H., et al.: A deep learning, image based approach for automated diagnosis for inflammatory skin diseases. Ann. Transl. Med. **8**(9), 581 (2020)
20. Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J., Wang, F.: Federated learning for healthcare informatics. J. Healthc. Inf. Res. **5**(1), 1–19 (2021)
21. Yao, P., et al.: Single model deep learning on imbalanced small datasets for skin lesion classification. IEEE Trans. Med. Imaging **41**, 1242–1254 (2021)