WaRA: Wavelet Low-Rank Adaptation for Medical Image Classification

Moein Heidari[1], Yijin Huang[1,2], Yasamin Medghalchi[1], Alireza Rafiee[1], Roger Tam[1], and Ilker Hacihaliloglu[1]

[1] The University of British Columbia, Vancouver, BC, Canada
[2] Southern University of Science and Technology, Shenzhen, China
moein.heidari@ubc.ca

Abstract. Adapting large pretrained vision models to medical image classification is often limited by memory, computation, and task-specific specializations. Parameter-efficient fine-tuning (PEFT) methods like LoRA reduce this cost by learning low-rank updates, but operating directly in feature space can struggle to capture the localized, multi-scale features common in medical imaging. We propose WaRA, a wavelet-structured adaptation module that performs low-rank adaptation in a wavelet domain. WaRA reshapes patch tokens into a spatial grid, applies a fixed discrete wavelet transform, updates subband coefficients using a shared low-rank adapter, and reconstructs the additive update through an inverse wavelet transform. This design provides a compact trainable interface while biasing the update toward both coarse structure and fine detail. For extremely low-resource settings, we introduce Tiny-WaRA, which further reduces trainable parameters by learning only a small set of coefficients in a fixed basis derived from the pretrained weights through a truncated SVD. Experiments on medical image classification across four modalities and datasets demonstrate that WaRA consistently improves performance over strong PEFT baselines, while retaining a favorable efficiency profile. Our code is publicly available at GitHub.

Keywords: Parameter-efficient fine-tuning · Low-rank adaptation · Pretrained models.

1 Introduction

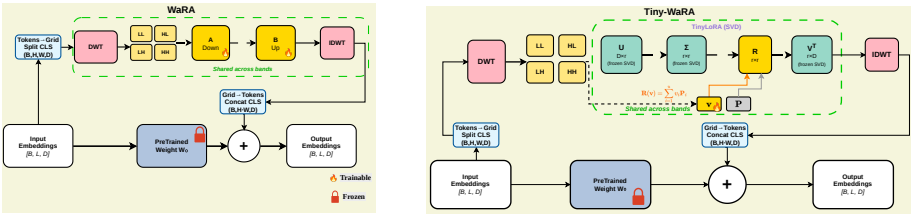
Medical image classification has advanced substantially with deep learning, yet modern foundation models are data-intensive, while curating large annotated medical datasets remains costly and constrained by privacy and class imbalance. Transfer learning and Parameter-Efficient Fine-Tuning (PEFT) have therefore become central strategies for adapting foundation models to specialized medical tasks while mitigating overfitting [20,11]. PEFT methods typically freeze the pre-trained model and adapt only a small set of parameters, making them highly relevant for medical domains with limited data and specialized institutional settings [5]. Representative directions include selective fine tuning [13], adapter based methods [14,25], and prompt based tuning [19,17].

A prominent PEFT approach is low-rank adaptation (LoRA) [15], which augments a frozen weight matrix W_0 with a low-rank update $\Delta W = BA$. This mechanism has been extended to control parameter growth in medical imaging settings, including continual segmentation and personalized federated learning [6,23,5]. In this work, we focus on medical image classification, where adapting large visual backbones efficiently is important for small datasets and decentralized clinical deployment. Despite its empirical success, standard LoRA has notable limitations. First, most low-rank adapters operate directly in feature space [21]. Moreover, in extreme resource-limited regimes—common in decentralized healthcare deployments—even conventional LoRA-style adapters can be non-trivial to store, distribute, and load, incurring storage and bandwidth overhead [12].

Recent works explore reparameterizing adaptation in a transformed domain to induce more structured updates. In the frequency domain, FouRA learns low-rank adaptations via Fourier projections [3], FourierFT explores compressing trainable updates via discrete Fourier parameterizations [12], and LoCA proposes selective frequency component learning [9]. These works suggest that frequency parameterizations can provide structured update spaces. For medical image classification, such structure is important because subtle variations and high visual similarity across categories create both intra-class variation and inter-class similarity [16]. Effective adaptation should therefore integrate multi-scale information, capturing fine local details while retaining global context. This motivates wavelet-based alternatives, which, unlike the global Fourier basis, are localized in both space and frequency. Wavelets have been integrated into scattering networks [4], transformer architectures [29], and generative modeling [22], demonstrating that multi-scale decompositions can improve both efficiency and fidelity.

In this work, we propose WaRA (Wavelet Low-Rank Adaptation), a PEFT approach designed for medical image classification that performs adaptation in a wavelet domain to exploit multi-scale structure. WaRA decomposes intermediate representations into frequency subbands that capture complementary information, performs low-rank updates in this transformed space, and reconstructs in the original domain. This wavelet-structured parameterization biases adaptation toward both coarse structure (low-frequency) and fine-grained features (high-frequency) that are prominent in medical imaging. We further develop Tiny-WaRA for extremely low-resource settings, which reduces the trainable budget by parameterizing the adapter update with a compact trainable core while keeping a fixed basis derived from pre-trained weights [1,24].

In summary, our key contributions are: (1) We present WaRA, a wavelet-domain low-rank adaptation module that leverages multi-scale subband decomposition to capture both global structure and fine-grained detail for medical image classification. (2) We develop Tiny-WaRA for extremely low-resource settings, substantially reducing trainable parameters compared to baseline frequency-domain methods by learning a compact core update in a fixed basis derived from pre-trained weights. (3) We conduct extensive experiments on medical image classification benchmarks, complemented by targeted ablations and frequency-domain analyses that elucidate when and why wavelet-structured adaptation is effective.



(a) WaRA. (b) Tiny-WaRA.

Fig. 1: **Overview of WaRA and Tiny-WaRA.** (a) WaRA performs low-rank adaptation in a wavelet domain by applying DWT on patch tokens, updating subband coefficients with shared adapter parameters, and reconstructing an additive update via IDWT. (b) Tiny-WaRA replaces the standard adapter with a compact parameterization that learns only a small coefficient vector in a fixed basis derived from pretrained weights [1,24].

2 Method

The core idea of WaRA, shown in Figure 1, is to perform parameter-efficient adaptation in a wavelet domain that exposes multi-scale structure while keeping the pretrained layer frozen. Consider a linear projection with frozen pretrained weights $W_0 \in \mathbb{R}^{d_2 \times d_1}$ applied to token embeddings $X \in \mathbb{R}^{B \times L \times d_1}$ (B: Batch Size, L: number of patch token + class token , d_1: Embedding size) which produces the base output $Y_{\text{base}} = XW_0^\top$. LoRA adds an additive update using low-rank factors, that is $\Delta W = BA$ with $A \in \mathbb{R}^{r \times d_1}$ and $B \in \mathbb{R}^{d_2 \times r}$, where $r \ll d_1, d_2$ [15]. Inspired by LoRA, WaRA applies low-rank adaptation in the multi-scale wavelet domain rather than in the feature domain. Specifically, we split X into the class token X_{cls} and patch tokens X_{tok}, reshape the patch tokens into a spatial grid $X_{\text{tok}_r} \in \mathbb{R}^{B \times H \times W \times d_1}$, and then apply a fixed 2D wavelet transform:

$$X = [X_{\text{cls}}, X_{\text{tok}}], \qquad C = \mathcal{W}(X_{\text{tok}_r}) = \{C_{\text{LL}}, C_{\text{LH}}, C_{\text{HL}}, C_{\text{HH}}\}, \tag{1}$$

where $C_\alpha \in \mathbb{R}^{B \times H' \times W' \times d_1}$ and $(H', W') = (H/2, W/2)$ for one level of decomposition, with α indexing the wavelet subbands (e.g., LL, LH, HL, HH). In the wavelet domain, we apply a shared low-rank adapter along the feature dimension for every subband:

$$\Delta C_\alpha = C_\alpha A^\top B^\top, \qquad A \in \mathbb{R}^{r \times d_1}, \quad B \in \mathbb{R}^{d_2 \times r}. \tag{2}$$

Following the standard LoRA initialization [15], we initialize A with a Gaussian distribution and B with zeros. Parameter sharing across the four subbands avoids the parameter growth that would arise from using independent adapters per subband. We reconstruct the patch token update via the inverse transform, and we adapt the class token with the same adapter:

$$\Delta Y_{\text{tok}} = \mathcal{W}^{-1}(\{\Delta C_\alpha\}), \qquad \Delta Y_{\text{cls}} = X_{\text{cls}} A^\top B^\top. \tag{3}$$

Finally, the layer output is formed by residual fusion, $Y = Y_{\text{base}} + [\Delta Y_{\text{cls}}, \Delta Y_{\text{tok}}]$.

2.1 Tiny-WaRA for extreme low resource adaptation

LoRA is efficient, but storing and training adapters can still be costly when many specializations are required or when the adaptation budget is extremely small. Tiny-WaRA reduces the trainable budget even more by parameterizing the adapter update with a compact trainable core on a fixed basis derived from W_0 [1,24]. For each layer, we compute Singular Value Decomposition (SVD) and then truncate it to rank r by only keeping the r-largest singular values of pretrained weights:

$$W_0 \approx U\,\mathrm{diag}(S)\,V^\top, \qquad U \in \mathbb{R}^{d_2\times r},\ S \in \mathbb{R}^{r},\ V \in \mathbb{R}^{d_1\times r}. \tag{4}$$

We then define a mixing matrix as a linear combination of fixed basis matrices $\{P_i\}_{i=1}^{u}$ (initialized with a Gaussian distribution) with a trainable coefficient vector $v \in \mathbb{R}^u$ (with a zero initialization):

$$R(v) = \sum_{i=1}^{u} v_i P_i, \qquad \Delta W(v) = U\,\mathrm{diag}(S)\,R(v)\,V^\top, \tag{5}$$

so only v is trainable. Tiny-WaRA uses the same wavelet pipeline as WaRA, but replaces the subband update with $\Delta W(v)$:

$$\Delta C_\alpha = C_\alpha \Delta W(v)^\top, \qquad \Delta Y_{\mathrm{cls}} = X_{\mathrm{cls}} \Delta W(v)^\top, \tag{6}$$

followed by IDWT reconstruction and residual fusion as above.

3 Results

Datasets, Training and Evaluation Setup. We evaluate on four medical image classification datasets, namely Messidor-2 [8], ISIC2018 [7], DDSM [18], and CXR COVID [26], with 1,740, 11,527, 9,104, and 1,708 images, respectively. These correspond to 5-class diabetic retinopathy grading, 7-class skin lesion classification, 3-class mammography classification, and 3-class chest radiograph classification, respectively. For each dataset, we use the provided split directory with separate train, val, and test folders. We initialize the vision backbone with the BiomedCLIP ViT-B model using 16×16 patches, and replace the classification head to match the dataset's label space. All images are resized to 224×224 and trained for 20 epochs with a batch size of 16. We use AdamW with weight decay 0.1 and a cosine learning rate schedule. We use a learning rate of 2×10^{-5} for full fine tuning, 2×10^{-4} for attention tuning, and 2×10^{-3} for the remaining parameter efficient methods, including WaRA. We select the best checkpoint using validation AUC and train all models on a 16 GB NVIDIA V100 GPU. For all baselines, we follow the training configurations from their official repositories, and for FourierFT [12], LoCA [9], and FouRA [3], for matched parameter budget comparisons in Tiny-WaRA experiments (Table 3), we shrink their adaptation capacity to match the Tiny-WaRA parameter budget. Specifically, we reduce

Table 1: Comparison results across four evaluation datasets. Per dataset columns report **F1** (%). 'Params.' denotes the ratio of learnable parameters to the total number of parameters. Entries are reported as mean±std. Among non-gray rows, the best and second-best means are colored blue and red, respectively.

Method	Params.	Fundus (Messidor-2)	Dermoscopy (ISIC2018)	Mammography (DDSM)	Chest X-ray (COVID)	Avg. F1
Full fine tuning	100	39.04±5.01	67.46±1.17	72.19±0.71	97.50±0.40	69.05±1.82
Linear probing	0.0045	28.40±3.90	62.49±1.17	61.54±1.70	97.56±0.20	62.50±1.74
Bias	0.12	34.84±4.39	64.78±1.49	65.48±0.90	97.83±0.23	65.73±1.75
Prompt tuning [17]	0.176	30.58±3.88	62.94±1.02	62.38±2.35	97.36±0.26	63.31±1.88
Attention tuning [28]	33.04	30.42±8.18	63.26±2.23	62.88±1.21	97.31±0.56	63.47±3.05
Adapter [14]	2.05	25.68±9.67	67.22±2.12	65.97±2.61	**97.62±0.38**	64.12±3.69
LoRA (rank=16) [15]	0.69	39.69±3.07	**71.81±1.10**	68.22±7.04	96.40±0.84	69.03±3.01
LoRA (rank=8)	0.34	40.50±4.55	70.61±0.92	73.96±1.77	96.96±0.38	70.51±1.90
FourierFT [12]	0.16	18.63±1.03	39.90±1.18	62.24±1.91	92.44±0.95	53.30±1.27
LoCA [9]	0.17	31.36±2.44	58.11±2.02	62.65±1.22	95.93±0.38	62.01±1.52
FouRA [3]	0.68	37.60±4.63	71.30±1.22	67.47±4.20	97.58±0.42	68.49±1.66
WaveFT [2]	0.17	17.99±1.42	32.93±3.16	44.47±4.81	90.32±0.64	46.43±2.51
WaRA (rank=8)	0.34	**45.35±2.60**	70.94±0.99	**74.68±0.94**	96.81±0.55	**71.94±1.27**
WaRA (rank=16)	0.69	38.73±7.07	71.44±1.05	70.20±2.61	96.64±0.26	69.25±2.75

FourierFT by decreasing the number of learned Fourier components from 600 to 360, LoCA by lowering its low-rank adaptation size from 600 to 262, and FouRA by setting its rank to 64. For Tiny-WaRA, we use rank $r = 64$ with basis size $u = 128$.

3.1 Comparisons with State-of-the-art

Tables 1 and 2 compare WaRA against representative parameter-efficient transfer learning baselines on four medical image classification datasets. Full fine-tuning provides a strong reference but updates all parameters, whereas linear probing is a low-cost lower bound and consistently underperforms. We report F1 scores in Table 1, where higher values indicate a better balance between sensitivity and precision at the operating threshold, a clinically relevant criterion that reflects improved decision-level performance and fewer false alarms.

According to Table 1, WaRA (rank=8) achieves the best average performance among all methods, outperforming all parameter-efficient baselines. Compared to LoRA (rank=8), our strongest competitor, WaRA, improves the average F1 by +1.43 (71.94 vs. 70.51). The gains are most pronounced on Fundus (+4.85) and remain positive on Mammography (+0.72), while WaRA is slightly lower on Dermoscopy (-0.33) and comparable on Chest X-ray with a small difference (-0.81). We further report both ranks 8 and 16 for LoRA and WaRA; in both cases, rank=8 performs better overall, indicating that increasing the rank does not necessarily improve performance.

Additionally, Table 2 reports performance averaged across the four datasets over multiple metrics: F1, accuracy (ACC), area under the ROC curve (AUC), Recall, Precision, and Cohen's κ, which corrects for chance agreement and is therefore

Table 2: Average performance across the four datasets (%). Entries are reported as mean±std, where the std term is the average of per-dataset stds. Among non-gray rows, the best and second-best means are colored blue and red, respectively.

Method	Params.	F1	ACC	AUC	Recall	Precision	Kappa
Full fine tuning	100	69.05±1.82	78.12±1.07	90.50±0.42	67.75±1.80	74.12±4.00	73.43±1.15
Linear probing	0.0045	62.50±1.74	72.61±2.72	86.71±0.35	62.54±2.02	66.77±2.24	64.89±2.89
Bias	0.12	65.73±1.75	75.78±0.86	88.76±0.36	64.99±2.05	68.99±2.33	69.50±2.63
Prompt tuning [17]	0.176	63.31±1.88	74.34±1.00	86.90±0.28	62.45±1.83	67.18±2.36	66.68±1.67
Attention tuning [28]	33.04	63.47±3.05	73.54±1.28	86.76±1.08	63.23±2.76	67.02±4.89	66.13±4.02
Adapter [14]	2.05	64.12±3.69	75.47±1.74	87.38±2.10	63.57±3.62	67.27±4.58	64.35±6.68
LoRA (rank=16) [15]	0.69	69.03±3.01	76.91±2.97	90.32±1.14	68.27±2.87	73.72±3.15	73.03±3.69
LoRA (rank=8)	0.34	70.51±1.90	78.71±1.20	91.14±0.46	70.17±2.66	74.52±2.42	75.62±2.47
FourierFT [12]	0.16	53.30±1.27	71.69±0.87	84.15±1.49	53.08±1.00	59.72±3.36	53.98±2.29
LoCA [9]	0.17	62.01±1.52	74.26±0.75	87.49±0.26	60.98±1.47	67.30±1.73	64.43±2.10
FouRA [3]	0.68	68.49±1.66	77.28±0.98	89.73±0.85	67.57±1.64	72.74±1.76	72.39±2.12
WaveFT [2]	0.17	46.43±2.51	66.36±2.59	80.42±1.04	48.04±1.40	51.25±3.50	44.30±4.75
WaRA (rank=8)	0.34	**71.94±1.27**	**79.69±0.90**	**91.62±0.37**	**71.59±1.78**	**75.02±2.27**	**77.01±1.48**
WaRA (rank=16)	0.69	69.25±2.75	77.22±1.60	90.43±0.79	69.05±2.60	72.83±3.58	72.81±3.15

more informative than accuracy under class imbalance. Consistent with Table 1, WaRA (rank=8) outperforms LoRA (rank=8) across all metrics in Table 2, while also exhibiting lower standard deviation, indicating more stable performance across runs.

We also compare against recent frequency-domain adaptation baselines, including FourierFT [12], LoCA [9], FouRA [3], and the wavelet-domain method WaveFT [2], which was originally proposed for personalized text-to-image generation. FouRA is the strongest among these methods; nevertheless, WaRA (rank=8) improves over FouRA by substantial margins (e.g., +3.45 F1, +2.41 ACC, +1.89 AUC, +4.02 Recall, +2.28 Precision, and +4.62 κ). Overall, these results suggest that wavelet-based adaptation outperforms both feature-space and Fourier-space parameterizations, which we attribute to multi-resolution parameter updates that capture both global structure and localized details more effectively than Fourier-based updates.

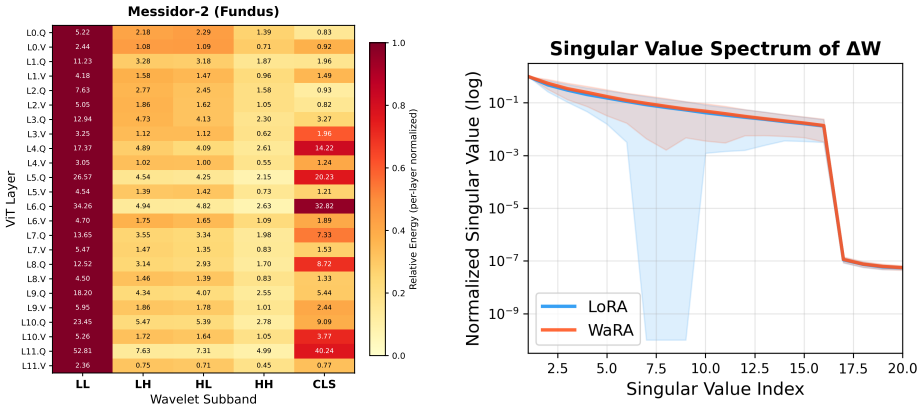
We next consider an ultra-low parameter regime in Table 3, where all methods are matched to a comparable trainable parameter budget by reducing their adaptation capacity. Tiny-WaRA trains a compact coefficient vector of size $u = 128$ over a fixed basis while using rank $r = 64$, which yields only 0.0078% trainable parameters. Despite this severe constraint, Tiny-WaRA achieves an average AUC of 88.81 with 78.54 ACC and 69.43 F1, and it remains the top performer across the four datasets. This setting is practically important for medical image classification as labeled data can be limited, and adaptation is often performed under tight memory, compute, and communication budgets, including on-premises deployment or privacy-preserving federated training, where the update size directly affects bandwidth and system cost [10].

Table 3: Comparison results of Tiny-WaRA with selected PEFT baselines across four evaluation datasets. Per dataset columns report **AUC** (%). 'Params.' denotes the ratio of learnable parameters to the total number of parameters. The best and second best results are bold and underlined, respectively.

Method	Computing cost	Fundus	Dermoscopy	Mammography	Chest X-ray	Performance (Avg.)		
	Params.	(Messidor-2)	(ISIC2018)	(DDSM)	(COVID)	AUC	ACC	F1
FourierFT [12]	0.010	53.18	61.97	60.36	63.17	59.67	46.24	24.90
LoCA [9]	0.010	79.65	92.25	79.16	99.04	87.53	74.22	62.30
FouRA [3]	0.008	78.82	93.41	80.73	99.01	87.99	75.11	65.04
Tiny-WaRA (Ours)	0.0078	**81.05**	**93.48**	**81.47**	**99.22**	**88.81**	**78.54**	**69.43**

3.2 Analysis and Ablation Study

Wavelet energy across subbands. Figure 2a reports the relative wavelet subband energy of the learned update across ViT layers on the Fundus dataset. As shown, the LL band consistently carries the majority of the update energy, while the high frequency bands (LH, HL, HH) remain smaller but non-zero. This supports our design choice to fix and share the update parameterization across subbands, since allocating separate large capacity per band would be inefficient given the observed energy imbalance. At the same time, the presence of structured energy in the higher frequency bands, together with nontrivial CLS energy, indicates that the model still benefits from a fused multi-resolution update, where coarse structure is primarily adjusted through LL, and finer cues are injected through the remaining bands via the inverse wavelet reconstruction.
Singular value spectrum of the update. Figure 2b compares the singular value spectra of the effective weight updates (ΔW) of LoRA and WaRA across all evaluated layers and datasets (solid lines: mean; shaded areas: cross-layer and cross-dataset variance). LoRA's singular values decay rapidly, concentrating update capacity in a small subset of dominant directions, consistent with existing literature describing LoRA updates as highly anisotropic [27]. WaRA, by contrast, yields a notably flatter spectrum, indicating that its update energy is distributed more evenly across directions, which is consistent with the lower run-to-run variance observed in Table 2.
Shared vs. band-specific adapters across wavelet subbands. In WaRA, we apply the same low-rank adapter parameters across all four wavelet subbands. A natural alternative is to assign independent LoRA adapters for different subbands, i.e., using separate (A_α, B_α) for each $\alpha \in \{\mathrm{LL}, \mathrm{LH}, \mathrm{HL}, \mathrm{HH}\}$, instead of sharing a single (A, B) across bands. Table 4 reports the ablation results. Sharing one adapter across subbands yields consistently better performance on Messidor-2, ISIC2018, and DDSM, while achieving comparable performance on COVID. Overall, the shared design improves the average AUC from 90.77 to 91.62 (+0.85), while also being more resource efficient due to the substantially reduced trainable footprint and optimizer overhead. These results support our design choice of subband sharing, indicating that the wavelet decomposition al-



(a) (b)

Fig. 2: Complementary analysis of WaRA. (a): Wavelet subband energy pattern on Fundus. (b): Singular value spectrum of the corresponding weight update.

ready provides complementary multi-resolution pathways, and allocating separate adapter capacity per band is unnecessary under the same training setup.

Table 4: Ablation study on wavelet-subband adapter design (AUC, %).

Method	Messidor-2	ISIC2018	DDSM	COVID
WaRA (shared adapter across bands)	83.85	95.29	87.73	99.61
WaRA (band-specific adapters)	83.35	94.21	85.84	99.68

4 Impact in Resource-Constrained Settings

Deploying foundation models in resource-constrained settings is limited less by attainable accuracy than by data scarcity and limited computational resources. WaRA addresses both: adapting only 0.34% of parameters, it outperforms every fine-tuning strategy, including full fine-tuning and Fourier-domain adaptation, and our Tiny-WaRA variant retains the best performance at a matched ultra-low budget of 0.0078%. A WaRA adapter occupies 1.2 MB and a Tiny-WaRA adapter 12 KB, against 330 MB for full fine-tuning. A single frozen backbone can therefore be deployed once and specialized to new tasks or sites by exchanging adapters small enough for intermittent or metered connections. The same reduction lowers the per-round uplink in federated training by two orders of magnitude relative to LoRA, and as only a low-dimensional update leaves the institution, the scheme remains applicable where raw data cannot be shared [10].

5 Conclusion

We introduced WaRA, a wavelet-domain low-rank adaptation module for medical image classification, together with Tiny-WaRA for extremely low-parameter regimes. Across four modalities, WaRA improves over strong parameter-efficient baselines at a matched budget, and Tiny-WaRA outperforms matched-budget frequency-domain methods while training only 0.0078% of parameters, making it well-suited to deployment under tight storage, compute, and bandwidth constraints. Future work includes extending WaRA to multiple backbones and dense prediction tasks such as segmentation and detection, and exploring multi-level decompositions and alternative wavelet families.

Acknowledgments. This work was supported by the Canadian Foundation for Innovation-John R. Evans Leaders Fund (CFI-JELF) program grant number 42816. Mitacs Accelerate program grant number AWD024298-IT33280. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2023-03575]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [RGPIN-2023-03575].

Disclosure of Interests. The authors have no competing interests in the paper.

References

1. Bałazy, K., Banaei, M., Aberer, K., Tabor, J.: Lora-xs: Low-rank adaptation with extremely small number of parameters. In: 28th European Conference on Artificial Intelligence, 25-30 October 2025, Bologna, Italy–Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025). IOS press (2025)
2. Bilican, A., Yılmaz, M.A., Tekalp, A.M., Cinbiş, R.G.: Exploring sparsity for parameter efficient fine tuning using wavelets. arXiv preprint arXiv:2505.12532 (2025)
3. Borse, S., Kadambi, S., Pandey, N.P., Bhardwaj, K., Ganapathy, V., Priyadarshi, S., Garrepalli, R., Esteves, R., Hayat, M., Porikli, F.: Foura: Fourier low-rank adaptation. Advances in Neural Information Processing Systems **37**, 71504–71539 (2024)
4. Bruna, J., Mallat, S.: Invariant scattering convolution networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(8), 1872–1886 (2013)
5. Chen, J., Ma, B., Pan, Y., Pu, B., Cui, H., Xia, Y.: Personalized federated side-tuning for medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 452–462. Springer (2025)
6. Chen, Q., Zhu, L., He, H., Zhang, X., Zeng, S., Ren, Q., Lu, Y.: Low-rank mixture-of-experts for continual medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 382–392. Springer (2024)

7. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)
8. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J.R., Massin, P., Erginay, A., et al.: Feedback on a publicly distributed image database: the messidor database. Image Analysis & Stereology pp. 231–234 (2014)
9. Du, Z., Min, Y., Li, J., Lu, K., Zou, C., Peng, L., Chu, T., Gong, M.: Loca: Location-aware cosine adaptation for parameter-efficient fine-tuning. In: The Thirteenth International Conference on Learning Representations
10. Dutt, R., Ericsson, L., Sanchez, P., Tsaftaris, S.A., Hospedales, T.: Parameter-efficient fine-tuning for medical image analysis: The missed opportunity. In: Medical Imaging with Deep Learning. pp. 406–425. PMLR (2024)
11. Fu, Z., Yang, H., So, A.M.C., Lam, W., Bing, L., Collier, N.: On the effectiveness of parameter-efficient fine-tuning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 12799–12807 (2023)
12. Gao, Z., Wang, Q., Chen, A., Liu, Z., Wu, B., Chen, L., Li, J.: Parameter-efficient fine-tuning with discrete fourier transform. In: International Conference on Machine Learning. pp. 14884–14901. PMLR (2024)
13. Guo, D., Rush, A.M., Kim, Y.: Parameter-efficient transfer learning with diff pruning. In: Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers). pp. 4884–4896 (2021)
14. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
15. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
16. Huo, X., Sun, G., Tian, S., Wang, Y., Yu, L., Long, J., Zhang, W., Li, A.: Hifuse: Hierarchical multi-scale feature fusion network for medical image classification. Biomedical signal processing and control **87**, 105534 (2024)
17. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European conference on computer vision. pp. 709–727. Springer (2022)
18. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data **4**(1), 170177 (2017)
19. Lester, B., Al-Rfou, R., Constant, N.: The power of scale for parameter-efficient prompt tuning. In: Proceedings of the 2021 conference on empirical methods in natural language processing. pp. 3045–3059 (2021)
20. Li, W.H., Liu, X., Bilen, H.: Cross-domain few-shot learning with task-specific adapters. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7161–7170 (2022)
21. Lialin, V., Deshpande, V., Yao, X., Rumshisky, A.: Scaling down to scale up: A guide to parameter-efficient fine-tuning. arXiv preprint arXiv:2303.15647 (2023)
22. Mattar, W., Levy, I., Sharon, N., Dekel, S.: Wavelets are all you need for autoregressive image generation. arXiv preprint arXiv:2406.19997 (2024)
23. Medghalchi, Y., Zakariaei, N., Rahmim, A., Hacihaliloglu, I.: Synthetic vs. classic data augmentation: Impacts on breast ultrasound image classification. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2025)

24. Morris, J.X., Mireshghallah, N., Ibrahim, M., Mahloujifar, S.: Learning to reason in 13 parameters. arXiv preprint arXiv:2602.04118 (2026)
25. Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., Gurevych, I.: Adapterfusion: Non-destructive task composition for transfer learning. In: Proceedings of the 16th conference of the European chapter of the association for computational linguistics: main volume. pp. 487–503 (2021)
26. Siddhartha, M.: Covid cxr image dataset (research) (2021)
27. Song, Y., Zhao, J., Harris, I.G., Jyothi, S.A.: Sharelora: Parameter efficient and robust large language model fine-tuning via shared low-rank adaptation. arXiv preprint arXiv:2406.10785 (2024)
28. Touvron, H., Cord, M., El-Nouby, A., Verbeek, J., Jégou, H.: Three things everyone should know about vision transformers. In: European Conference on Computer Vision. pp. 497–515. Springer (2022)
29. Yao, T., Pan, Y., Li, Y., Ngo, C.W., Mei, T.: Wave-vit: Unifying wavelet and transformers for visual representation learning. In: European conference on computer vision. pp. 328–345. Springer (2022)