ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

DeepFedNAS: Pareto Optimal Supernet Training for Improved and Predictor-Free Federated Neural Architecture Search

Bostan Khan[1] and Masoud Daneshtalab[1,2] *

1- DeepHERO Lab, Mälardalen University, Sweden

2- Tallinn Univesity of Technology, Estonia

Abstract. Federated Neural Architecture Search (FedNAS) is hindered by unguided supernet training and costly post-training search pipelines. We introduce DeepFedNAS, a two-phase framework that resolves these issues. We propose *Federated Pareto Optimal Supernet Training*, using a pre-computed path of elite architectures as an intelligent curriculum to train a superior supernet. Subsequently, our *Predictor-Free Search Method* uses a principled fitness function as a zero-cost proxy for accuracy, finding optimal subnets in seconds. DeepFedNAS achieves state-of-the-art accuracy, superior robustness to data heterogeneity, and a $\sim 61\times$ search pipeline speedup, making FedNAS practical and efficient.

1 Introduction

The proliferation of intelligent on-device applications, from mobile keyboard prediction [2] to personalized healthcare monitoring [10], highlights the need for efficient, privacy-preserving machine learning. Federated Learning (FL) [8] addresses *how* to train models on decentralized data privately. However, determining *what* model architecture to train remains a key challenge, often relying on manual design suboptimal for diverse client data and hardware [7].

To automate this, Federated Neural Architecture Search (FedNAS) discovers specialized networks in a federated setting [3]. Supernet-based methods, inspired by centralized approaches like Once-For-All (OFA) [1], represent the state-of-the-art for cost-efficient FedNAS. The SuperFedNAS framework [6] advanced the field by decoupling supernet training from a fast, training-free architecture search, reducing the cost to find specialized subnets for multiple hardware targets from $O(N)$ to $O(1)$.

However, existing FedNAS, including SuperFedNAS, suffers from two fundamental limitations. First, supernet training is *unguided*, relying on random subnet sampling (e.g., Sandwich Rule [11]) that yields noisy gradients and suboptimal models. Second, finding optimal subnets post-training requires a *costly predictor pipeline*: generating thousands of architecture-accuracy pairs to train a separate predictor model, a process that can take tens of hours or even days.

*This work was supported in part by the Swedish Research Council project GreenDL, the European Chip-JU via the project NextArc, as well as by the European Union and the Estonian Research Council via the project TEM-TA138. The computations in this work were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

In this paper, we introduce **DeepFedNAS**, a novel, two-phase framework that comprehensively addresses these limitations. Our work is grounded in the observation that a more principled approach to architecture evaluation can fundamentally transform both supernet training and subsequent search. Inspired by the mathematical design concepts of DeepMAD [9], we develop a novel, multi-objective fitness function, $\mathcal{F}(\mathcal{A})$, that synthesizes network information theory with empirical architectural heuristics. Crucially, this function allows for a holistic, single-objective optimization for desired architectural properties within a federated, weight-sharing context. This unified fitness function underpins our two primary contributions:

1. **Federated Pareto Optimal Supernet Training:** We pre-compute a "Pareto optimal path" of elite, high-fitness architectures that serves as an intelligent training curriculum. Training the supernet on these principled subnets produces a superior final model.
2. **Predictor-Free Search Method:** After the optimal path-guided training, our fitness function $\mathcal{F}(\mathcal{A})$ becomes a high-fidelity, zero-cost proxy for accuracy. This enables on-demand discovery of optimal subnets in seconds, completely eliminating the expensive predictor pipeline.

DeepFedNAS not only achieves state-of-the-art accuracy across diverse datasets and non-IID conditions but also dramatically reduces the post-training search cost by over 98%, replacing a multi-hour predictor pipeline with a near instantaneous search for specialized subnets.

2 The DeepFedNAS Framework

Our DeepFedNAS framework, depicted in Fig. 1, transforms federated supernet design through a principled, two-phase approach. Unlike traditional methods minimizing expected loss over a vast search space $\mathbb{S}$ [6], DeepFedNAS re-frames this objective by guiding supernet training with a curated architectural distribution. Specifically, DeepFedNAS aims to learn shared weights W by minimizing the expected loss of sampled subnets:

$$\min_{W} \mathbb{E}_{\mathcal{A}\sim P_{\mathcal{C}}} \left[\sum_{k=1}^{K} \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \mathcal{L}_k(\mathcal{G}(W, \mathcal{A})) \right] \tag{1}$$

Here, W denotes the shared supernet weights, $\mathcal{A}$ a sampled subnet, and $\mathcal{G}(W, \mathcal{A})$ the subnet instantiated with W. $P_{\mathcal{C}}$ is the Pareto-guided sampling distribution from our elite subnet cache $\mathcal{C}$.

Crucially, supernet weight updates from local clients adapt the MaxNet method [6], utilizing the Sandwich Rule [11] and cosine annealing (parameter weightage from largest to smaller subnets). Our core innovation replaces MaxNet's random sampling with principled sampling from $P_{\mathcal{C}}$, thus guiding supernet training with a curriculum of high-fitness architectures. This shift to $P_{\mathcal{C}}$ is central to training a superior supernet.

Principled Fitness Function $\mathcal{F}(\mathcal{A})$***:*** The foundation of our method is a novel, multi-objective fitness function, $\mathcal{F}(\mathcal{A})$, that evaluates a subnet's architectural quality without training. Inspired by DeepMAD [9], we synthesize network

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964.
Available from http://www.i6doc.com/en/.

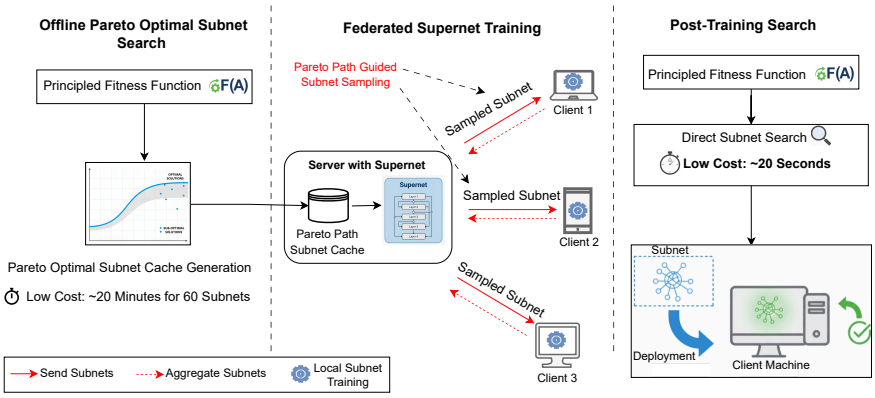

Fig. 1: **The DeepFedNAS Pipeline.** An offline search generates a Pareto-optimal cache of subnets using our fitness function $\mathcal{F}(\mathcal{A})$. This cache provides a superior curriculum ($P_{\mathcal{C}}$) for federated supernet training. Post-training, $\mathcal{F}(\mathcal{A})$ is used as a zero-cost proxy for direct, predictor-free search.

entropy (H_j, for expressive power) and **architectural effectiveness** (ρ, for design stability) into a single, optimizable objective:

$$\mathcal{F}(\mathcal{A}) = \left(\sum_{j=1}^{S} H_j(\mathcal{A}) - \beta Q(\mathcal{A}) \right) + \lambda \rho(\mathcal{A}) - \gamma V(\mathcal{A}) \tag{2}$$

This unified function maximizes entropy and effectiveness while penalizing architectural imbalances ($Q(\mathcal{A})$) and informational bottlenecks ($V(\mathcal{A})$). It is maximized during our offline search, subject to hard constraints MACs($\mathcal{A}$) $\leq$ Budget and $\rho(\mathcal{A}) \leq \rho_0$.

Phase 1- Federated Pareto Optimal Supernet Training: We replace the random subnet sampling of prior work [6] with a principled curriculum. An efficient offline search uses our fitness function to generate a *Pareto optimal path*-a cache of 60 elite architectures tracing a smooth accuracy-cost frontier (Fig. 2). During federated training, subnets are sampled from this cache, ensuring that shared weights are updated using high-fitness, mathematically sound architectures and yielding a stronger supernet.

Phase 2- Predictor-Free Search: Beyond identifying good models, the path-guided training in Phase 1 reshapes the supernet's weight landscape. Training exclusively on high-fitness architectures aligns weight-sharing patterns with $\mathcal{F}(\mathcal{A})$, making it a reliable, zero-cost proxy for accuracy. Consequently, we eliminate the multi-hour predictor pipeline-optimal subnets for new targets are found in seconds via a lightweight genetic search maximizing $\mathcal{F}(\mathcal{A})$ under hardware constraints.

3 Experiments

We validate DeepFedNAS against baselines including FedAvg [8], FedNAS [3], FedPNAS [4], and our primary baseline, SuperFedNAS [6]. We conduct experi-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

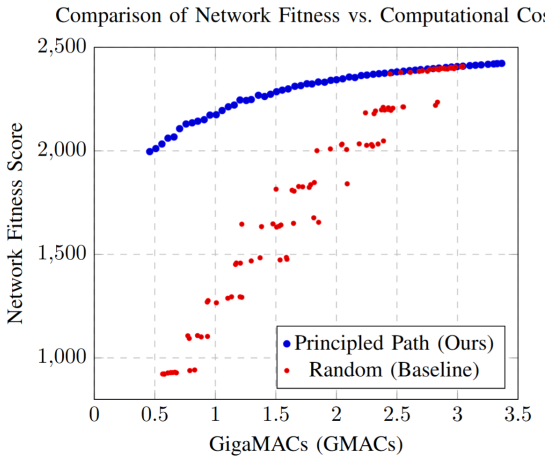

Fig. 2: Subnet fitness vs. computational cost. Our method discovers a smooth Pareto-optimal frontier (blue) that dominates the random sampling used in the baseline method [6] (red), validating its use as a training curriculum.

ments on **CIFAR-10**, **CIFAR-100**, and **CINIC-10**, replicating the federated setup of SuperFedNAS. We use a powerful 4-stage ResNet-based supernet with a search space of $\sim 10^{15}$ subnets. All federated training parameters match the baseline for a fair comparison. The Pareto optimal cache of 60 subnets was generated using a multiprocess parallel local search on a single machine taking roughly 20 minutes.

Table 1: Performance comparison on CIFAR-10, CIFAR-100, and CINIC-10.

Billion MACs	**Method**	**Test Accuracy (%)**		
		CIFAR-10	**CIFAR-100**	**CINIC-10**
0.45-0.95	FedAvg	85.25 ± 0.46	43.19 ± 0.54	61.76 ± 0.78
	FedNAS	77.33 ± 0.31	40.92 ± 2.21	58.15 ± 0.18
	FedPNAS	88.83 ± 0.5	45.77 ± 0.68	64.3 ± 0.98
	SuperFedNAS	93.47 ± 0.08	60.92 ± 0.10	75.69 ± 0.29
	DeepFedNAS (Ours)	$\mathbf{94.16 \pm 0.18}$	$\mathbf{62.60 \pm 0.16}$	$\mathbf{77.04 \pm 0.39}$
0.95-1.45	FedAvg	86.36 ± 0.22	43.92 ± 0.57	63.00 ± 0.17
	FedPNAS	89.27 ± 0.81	47.8 ± 2.6	65.74 ± 0.32
	SuperFedNAS	93.52 ± 0.16	61.66 ± 0.37	76.53 ± 0.19
	DeepFedNAS (Ours)	$\mathbf{94.51 \pm 0.02}$	$\mathbf{62.87 \pm 0.13}$	$\mathbf{77.60 \pm 0.02}$
1.45-2.45	FedAvg	87.59 ± 0.27	44.4 ± 0.56	64.00 ± 0.07
	FedNAS	86.41 ± 0.1	45.82 ± 0.29	59.97 ± 0.27
	SuperFedNAS	93.72 ± 0.01	62.06 ± 0.06	77.09 ± 0.07
	DeepFedNAS (Ours)	$\mathbf{94.50 \pm 0.02}$	$\mathbf{63.09 \pm 0.08}$	$\mathbf{77.80 \pm 0.06}$
2.45-3.75	FedAvg	89.44 ± 0.67	45.00 ± 0.27	65.02 ± 0.13
	FedNAS	89.43 ± 0.36	58.39 ± 0.23	71.93 ± 0.13
	SuperFedNAS	93.72 ± 0.02	62.30 ± 0.01	77.09 ± 0.07
	DeepFedNAS (Ours)	$\mathbf{94.51 \pm 0.00}$	$\mathbf{63.20 \pm 0.00}$	$\mathbf{77.85 \pm 0.09}$

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 2: Impact of data heterogeneity on CIFAR-10 (α value).

Billion MACs	Method	Test Accuracy (%)		
		100	**1**	**0.1**
0.95-1.45	FedAvg	86.36 ± 0.22	84.65 ± 0.11	77.99 ± 1.6
	FedPNAS	89.27 ± 0.51	87.53 ± 0.32	81.13 ± 0.4
	SuperFedNAS	93.52 ± 0.17	92.13 ± 0.12	85.56 ± 0.18
	DeepFedNAS (Ours)	$\mathbf{94.51 \pm 0.02}$	$\mathbf{93.22 \pm 0.05}$	$\mathbf{86.56 \pm 0.11}$
2.45-3.75	FedAvg	89.44 ± 0.67	87.88 ± 0.7	81.24 ± 1.99
	FedNAS	89.43 ± 0.36	85.85 ± 0.35	68.13 ± 5.04
	SuperFedNAS	93.72 ± 0.02	92.63 ± 0.02	86.00 ± 0.12
	DeepFedNAS (Ours)	$\mathbf{94.51 \pm 0.00}$	$\mathbf{93.33 \pm 0.00}$	$\mathbf{86.83 \pm 0.00}$

Main Results and Robustness: As shown in Table 1, DeepFedNAS consistently establishes a new state-of-the-art, outperforming all baselines across all datasets and computational budgets. On the more complex CIFAR-100 and CINIC-10 datasets, our method achieves up to a **1.35%** absolute accuracy improvement over SuperFedNAS, highlighting the significant benefit of our principled training curriculum.

Furthermore, we evaluate robustness to data heterogeneity, controlled by the Dirichlet distribution parameter $\alpha \in \{100, 1, 0.1\}$ [5]. Table 2 shows that while all methods degrade as data becomes more non-IID (smaller α), DeepFedNAS maintains and even widens its performance advantage. In the highly heterogeneous setting of $\alpha = 0.1$, our method outperforms the baseline by up to 1.00% on CIFAR-10, demonstrating that our principled training produces a more generalized supernet, better equipped to handle the conflicting gradients that arise from non-IID data.

Search Efficiency: Our predictor-free search provides a dramatic improvement in efficiency. As detailed in Table 3, the baseline SuperFedNAS requires over **20 hours** for predictor data generation. In contrast, DeepFedNAS requires only a **20-minute** one-time offline process to generate the Pareto optimal path cache. Subsequently, finding an optimal subnet for any new target takes just **20 seconds**. This represents an astounding **∼61x speedup** in total pipeline time over the baseline.

Table 3: Search Pipeline Cost: Predictor-Based (SuperFedNAS) vs. Predictor-Free (Ours).

Search Pipeline Stage	**SuperFedNAS**	**Ours**
Prior to SuperNet Training		
Subnet Cache Generation Time	N/A	∼20 minutes (for 60 subnets)
After SuperNet Training		
Predictor Data Generation Time	∼20.65 hours	N/A
Predictor Training Time	*few minutes*	N/A
Search Time per MAC target	∼43 seconds	∼20 seconds
Total Pipeline Time	∼20.65 hours	∼20.33 minutes
Speedup Factor	1x	**∼61x**

Baseline's data generation time is for a 10,000-sample subnet-accuracy dataset.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

4 Conclusion

We introduced DeepFedNAS, a framework that advances federated neural architecture search by replacing unguided training and costly search pipelines with a principled, two-phase approach. Our *Federated Pareto Optimal Supernet Training* uses an intelligent curriculum of elite architectures to produce a superior supernet. Subsequently, our *Predictor-Free Search Method* leverages a principled fitness function as a zero-cost proxy for accuracy, enabling on-demand subnet discovery in seconds. DeepFedNAS achieves state-of-the-art accuracy, superior robustness to data heterogeneity, and a ∼61x search pipeline speedup, marking a significant step toward practical, scalable FedNAS for resource-constrained environments. Future work will explore extending DeepFedNAS to diverse hardware targets and broader federated learning scenarios.

References

[1] H. Cai, C. Gan, T. Wang, Z. Zhang, and S. Han. Once-for-all: Train one network and specialize it for efficient deployment. In *International Conference on Learning Representations*, 2020.

[2] A. Hard, C. M. Kiddon, D. R. Ramage, F. S. Beaufays, H. Eichner, K. Rao, R. Mathews, S. Augenstein, and S. Ramaswamy. Federated learning for mobile keyboard prediction, 2019.

[3] C. He, E. Mushtaq, J. Ding, and S. Avestimehr. FedNAS: Federated deep learning via neural architecture search, 2022.

[4] M. Hoang and C. Kingsford. Personalized neural architecture search for federated learning. *1st NeurIPS Workshop on New Frontiers in Federated Learning (NFFL 2021)*, 2021.

[5] T.-M. H. Hsu, H. Qi, and M. Brown. Measuring the effects of non-identical data distribution for federated visual classification, 2019.

[6] A. Khare, A. Agrawal, A. Annavajjala, P. Behnam, M. Lee, H. Latapie, and A. Tumanov. SuperFedNAS: Cost-efficient federated neural architecture search for on-device inference. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, editors, *Computer Vision – ECCV 2024*, pages 161–179, Cham, 2024. Springer Nature Switzerland.

[7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith. Federated optimization in heterogeneous networks. In I. Dhillon, D. Papailiopoulos, and V. Sze, editors, *Proceedings of Machine Learning and Systems*, volume 2, pages 429–450, 2020.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In A. Singh and J. Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017.

[9] X. Shen, Y. Wang, M. Lin, Y. Huang, H. Tang, X. Sun, and Y. Wang. DeepMAD: Mathematical architecture design for deep convolutional neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6163–6173, June 2023.

[10] S. Silva, B. A. Gutman, E. Romero, P. M. Thompson, A. Altmann, and M. Lorenzi. Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 270–274, 2019.

[11] L. Xie, X. Chen, K. Bi, L. Wei, Y. Xu, L. Wang, Z. Chen, A. Xiao, J. Chang, X. Zhang, and Q. Tian. Weight-sharing neural architecture search: A battle to shrink the optimization gap. *ACM Comput. Surv.*, 54(9), Oct. 2021.