MedNNS: Supernet-based Medical Task-Adaptive Neural Network Search

Lotfi Abdelkrim Mecharbat✉, Ibrahim Almakky, Martin Takac, and Mohammad Yaqub

Mohammed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
firstname.lastname@mbzuai.ac.ae

Abstract. Deep learning (DL) has achieved remarkable progress in the field of medical imaging. However, adapting DL models to medical tasks remains a significant challenge, primarily due to two key factors: (1) architecture selection, as different tasks necessitate specialized model designs, and (2) weight initialization, which directly impacts the convergence speed and final performance of the models. Although transfer learning from ImageNet is a widely adopted strategy, its effectiveness is constrained by the substantial differences between natural and medical images. To address these challenges, we introduce Medical Neural Network Search (MedNNS), the first Neural Network Search framework for medical imaging applications. MedNNS jointly optimizes architecture selection and weight initialization by constructing a meta-space that encodes datasets and models based on how well they perform together. We build this space using a Supernetwork-based approach, expanding the model zoo size by **51x times** over previous state-of-the-art (SOTA) methods. Moreover, we introduce rank loss and Fréchet Inception Distance (FID) loss into the construction of the space to capture inter-model and inter-dataset relationships, thereby achieving more accurate alignment in the meta-space. Experimental results across multiple datasets demonstrate that MedNNS significantly outperforms both ImageNet pretrained DL models and SOTA Neural Architecture Search (NAS) methods, achieving an average accuracy improvement of **1.7%** across datasets while converging substantially faster. The code and the processed meta-space is available at https://github.com/BioMedIA-MBZUAI/MedNNS.

Keywords: Neural Network Search · Meta Learning · Supernetworks

1 Introduction

Deep learning (DL) is transforming medical imaging in many aspects, such as enabling tumor type identification [14], skin lesion analysis [18], and diabetic retinopathy diagnosis [20]. These advances boost diagnostic accuracy and clinical efficiency, ultimately striving to improve patient outcomes. However, adapting DL techniques to medical tasks presents two fundamental challenges. First, choosing the optimal architecture, where different tasks often require different model designs tailored to capture subtle, domain-specific features [21,7]. Second,

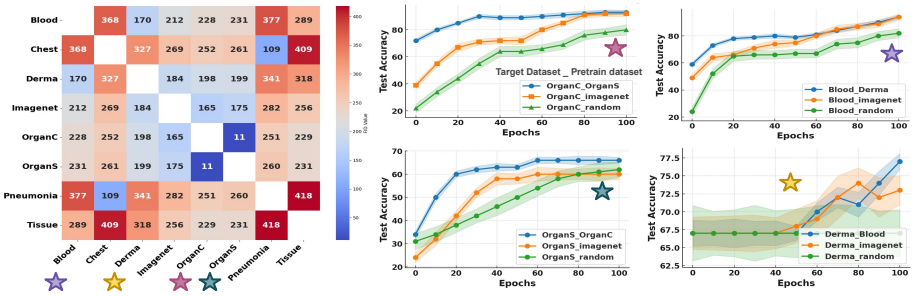

Fig. 1. (Left) Heatmap showing pairwise FID distance between MedMNIST datasets and ImageNet. (Right) Training curves for ResNet-18 using random, ImageNet, and nearest-FID dataset pertaining.

weight initialization strategies, which influence convergence during training, ultimately affecting the overall performance of the model [2,22].

Transfer learning enhances convergence by providing a better initialization from models pretrained on large datasets like ImageNet [6]. However, medical images differ substantially from natural images in format, modality, and appearance [19,21], making feature reuse from ImageNet often ineffective [11]. Investigating this, we compute the Fréchet Inception Distance (FID) [27] between several MedMNIST [26] datasets and ImageNet [6] as shown in Fig. 1 and observe that every medical dataset has at least one other medical dataset that is closer in FID compared to ImageNet. Moreover, training ResNet-18 [8] on four datasets with different initializations (Fig. 1) demonstrates that pretraining on a more similar dataset accelerates convergence and improves the final performance, suggesting more effective feature reuse.

While selecting a more relevant dataset for pretraining improves feature transfer, the architecture itself plays an important role in determining which features are extracted and reused effectively. This is because the architecture's inductive bias, which are the inherent assumptions it encodes about the data and task, dictates which features are prioritized during learning [16]. This has led to growing interest in Neural Architecture Search (NAS), an automated approach that explores a space of architectures to find configurations that best align with task and data-specific characteristics. For a given dataset D, NAS aims to discover an optimal architecture α^* that minimizes the validation loss $\mathcal{L}_{val}^{(D)}$ when evaluated on the model f_α with its optimal weights w_α^*. Formally:

$$\alpha^* = \arg\min_{\alpha \in \mathcal{A}} \mathcal{L}_{val}^{(D)} \left(f_\alpha \left(w_\alpha^* \right) \right), \quad where: \quad w_\alpha^* = \arg\min_{w} \mathcal{L}_{train}^{(D)} \left(f_\alpha(w) \right). \quad (1)$$

From an application perspective, NAS has shown promise in medical imaging, delivering notable performance improvements [1]. While improving architectural design, NAS does not inherently offer weight initialization or facilitate feature reuse for fast adaptation. Therefore, recent strategies have adopted Supernet-

works [3], which are over-parameterized neural networks that contain multiple smaller subnetworks within their structure. However, these Supernetworks are task-specific and require retraining when applied to a new target dataset [25].

Addressing both the architecture selection and weight initialization challenges, Neural Network Search (NNS) has emerged, it leverages a model zoo ($\mathcal{Z}$) of pretrained networks to jointly discover both the architecture (α) and its corresponding weights (w) based on its alignment with a given Dataset D [10]. NNS can be expressed as:

$$(\alpha^*, w^*) = \arg \min_{(\alpha,w)\in\mathcal{Z}} \mathcal{L}_{val}^{(D)} \left(f_\alpha(w)\right). \tag{2}$$

Formulating NNS as a meta-learning problem, Task-Adaptive Neural Network Search (TANS) [10] constructs a cross-modal dataset latent space using contrastive learning to align datasets with high-performing models, thereby enabling efficient retrieval of task-adaptive models. However, TANS faces notable limitations. Specifically, TANS requires training a large number of model-dataset pairs to construct its meta-space, which introduces significant computational complexity with each dataset. This scalability issue makes applying TANS for medical imaging tasks computationally expensive. Additionally, TANS's reliance on a contrastive loss function emphasizes relationships between individual datasets and models, while overlooking similarities across datasets or architectures. This limitation hinders generalization, which is critical in medical imaging where cross-dataset knowledge transfer is essential. In this work, we aim to overcome these limitations and extend the approach to the medical tasks; we introduce the first Medical Neural Network Search (MedNNS) methodology. The key contributions of MedNNS include:

- We propose MedNNS, the first NNS for medical vision tasks that jointly determines the optimal architecture and pretrained initialization for a given medical dataset by leveraging a pre-constructed meta-learning space.
- We leverage Supernetworks to construct a model zoo, enabling efficient subnetwork extraction and expanding the space of model-data pairs.
- We introduce a novel rank loss for the meta-space that incorporates inter-model information, ensuring that performance ranking is reflected in model proximity to the dataset representation.
- We introduce the FID loss to capture inter-dataset relationships, ensuring that datasets that have close feature representations are close to each other in the meta-space.

2 Methodology

We propose MedNNS to solve the NNS problem defined in Eq. (2) by employing Supernetworks to efficiently construct a large model zoo of model-dataset pairs. We also propose a novel approach to embed these pairs into a meta-space that preserves model ranks and dataset similarities. Once built, this meta-space can be queried to select a suitable model and initialization for an unseen target dataset, as shown in Fig. 2. In this section, we will describe these steps in detail.

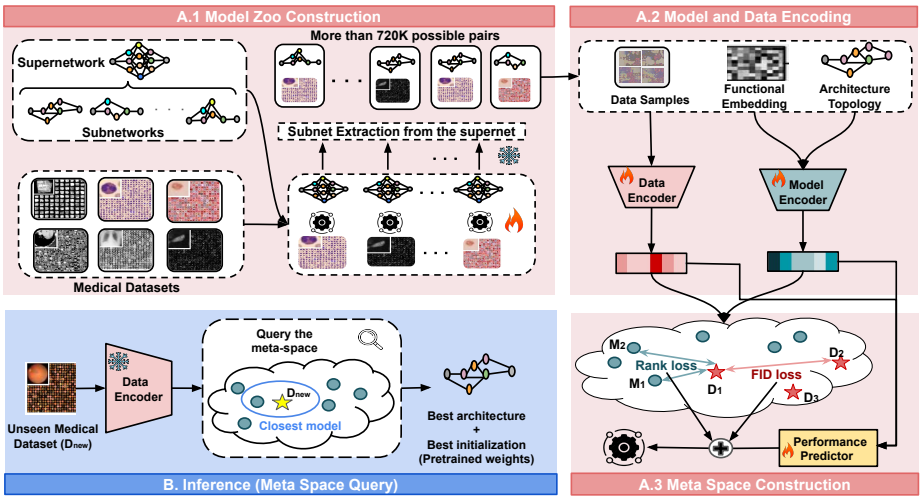

Fig. 2. Overview of our MedNNS framework. (A) During training: (A.1) A large model zoo is built by training a single Supernetwork per dataset and extracting thousands of subnetworks via weight sharing. (A.2) Models and datasets are embedded into the latent space. (A.3) The meta-space is optimized using a combination of rank loss, FID loss, and performance loss to align models and datasets according to their relative performance. (B) During inference, given an unseen dataset, its embedding is computed and used to query the meta-space, selecting the closest model embedding as the most suitable pre-trained model.

A. Creating the Meta-learning Space

A.1 Model Zoo Construction. We define a model zoo as $\mathcal{Z} = \{(D_1, M_1, P_{M_1}^{D_1}), \ldots, (D_k, M_l, P_{M_l}^{D_k})\}$, where $P_{M_l}^{D_k}$ denotes the performance of model M_l on dataset D_k. Each model M consists of an architecture α and weights θ, learned by training M on D. Constructing an extensive model zoo provides a diverse pool of pairs, enriching the meta-space and boosting overall performance. However, training thousands of models per dataset is computationally prohibitive. TANS [10] reduces this cost using accuracy estimation to selectively train model-data pairs, yet it still requires training hundreds of models per dataset. We overcome this inefficiency by training a single Supernetwork per dataset and extracting thousands of subnetworks from it without retraining via weight sharing. Through this, MedNNS enables the efficient construction of a model zoo with 720k model-dataset pairs compared to 14k pairs in TANS [10] at a significantly lower computational cost.

We define a Supernetwork $S(x; \Theta)$ with parameters Θ, while its subnetworks are derived by applying a binary mask $m \in \{0,1\}^n$ to Θ, yielding $s(x; \theta = m \odot \Theta)$ where $\odot$ denotes elementwise multiplication. Varying masks allow us to extract subnetworks with varied architectures (in depth, width, etc). Accordingly, the model zoo is represented as $\mathcal{Z} = \{(D_1, s_1, \hat{P}_{s_1}^{D_1}), \ldots, (D_k, s_l, \hat{P}_{s_l}^{D_k})\}$, where each

subnetwork s_l is characterized by an architecture α_l and weights θ_l extracted from Θ using mask m_l. While, $\hat{P}_{s_l}^{D_k}$ is the estimated performance of s_l with inherited weights θ_l, and $P_{s_l}^{D_k}$ is the true performance when trained from scratch. To ensure that the extracted weights are meaningful, the Supernetwork is trained using a two-stage scheme. Initially, the full network is trained to establish robust shared weights. In later epochs, several subnetworks are sampled each batch, and weights are updated according to the average loss. Also, to ensure rank preservation, which means if $P_{s_1} > P_{s_2}$, then $\hat{P}_{s_1} > \hat{P}_{s_2}$, we employ the FaiRNAS technique [5] that applies the strict fairness principle during sampling of subnetworks in the second stage of training. Our experiments confirm the effectiveness of this training scheme, as it is demonstrated in the experiments section.

A.2 Model and Dataset Encoding. For models, an architectural encoding of α is created by flattening configuration parameters such as depth and width. On the other hand, a functional encoding of θ is acquired at the penultimate layer when passing a fixed Gaussian noise $z \sim \mathcal{N}(0, I)$ through the model. These two encodings are then concatenated and processed through a Multi-Layer Perceptron (MLP) encoder E_m to yield the final model representation. For datasets, a random set of images is selected for each dataset and embedded using a pretrained model; their embeddings are then averaged to produce the dataset representation that serves as input to the MLP data encoder E_d.

A.3 Meta Space Construction. We use the acquired Supernet-based model zoo $\mathcal{Z}$ and construct our meta space by aligning model and dataset embeddings, obtained using the encoders E_m and E_d, so that their similarities reflect the true performance relationships. We also employ an MLP as a performance predictor (φ) which takes the model and data embeddings and output $\hat{y}$ as follows $\hat{y} = \varphi\big(E_m(s), E_d(D)\big)$. Following this, we optimize a composite loss defined as: $\mathcal{L} = \mathcal{L}_{\text{perf}} + \mathcal{L}_{\text{rank}} + \mathcal{L}_{\text{FID}}$, where $\mathcal{L}_{\text{perf}}$ is performance prediction loss computed as the mean squared error between the predicted performance and the estimated one $\hat{P}$. In addition, we introduce a novel rank loss ($\mathcal{L}_{\text{rank}}$) to preserve the true ranking of model performances across datasets in the meta-space. For a given dataset D and any pair of models s_j and s_k with a true performance difference $\Delta P_{j,k}^D = P_{s_j}^D - P_{s_k}^D$, we enforce that the corresponding difference in cosine similarity is positive when $\Delta P_{j,k}^D > 0$ via a logistic loss:

$$\mathcal{L}_{\text{rank}} = \frac{1}{|P|} \sum_{\Delta P_{j,k}^D > 0} -\log\Big(\sigma\Big(\beta \,.\big(E_d(D) \cdot E_m(s_j) - E_d(D) \cdot E_m(s_k)\big)\Big)\Big), \quad (3)$$

where σ is the sigmoid function and the scaling factor β adjusts the sensitivity.

We propose $\mathcal{L}_{\text{FID}}$ to capture inter-dataset similarities and align dataset embeddings based on the similarity of features captured by the FID metric. Although FID relies on ImageNet embeddings, which may not fully capture salient medical features, it is sufficiently expressive to quantify similarities between datasets based on image samples. This is supported by Fig. 1, where FID demonstrates a clear correlation with the visual similarity of the datasets. Furthermore, as indicated by [16], FID also correlates with feature reuse. The standard computation of FID is used, and its positive impact is demonstrated in our ablation

study. For any two datasets D_i and D_j with embeddings $E_d(D_i)$ and $E_d(D_j)$, a lower FID implies better similarity, so their embeddings should be closer. This is enforced by weighting the squared Euclidean distance between the embeddings:

$$\mathcal{L}_{\text{FID}} = \frac{1}{|P|} \sum_{i \neq j} \exp\left(-\frac{\text{FID}(D_i, D_j)}{\sigma}\right) \|E_d(D_i) - E_d(D_j)\|^2. \tag{4}$$

B. Querying the Meta Space. During inference, the objective is to select the most suitable architecture α and its initialization θ for a new unseen target dataset D_{new} leveraging the pre-constructed meta-space. The latent representation $E_d(D_{new})$ is computed using the trained dataset encoder E_d . This representation serves as a query to the meta-space, where a model M^* is selected by finding the model embedding $E_m(M^*)$ that is closest to $E_d(D_{new})$ as follows:

$$M^* = \arg\max_{M \in \mathcal{M}} \frac{E_d(D_{new}) \cdot E_m(M)}{\|E_d(D_{new})\| \, \|E_m(M)\|} \tag{5}$$

3 Experiments

We use MedMNIST [26] datasets comprising of a diverse benchmark well-suited for evaluating generalization as it is built from real-world medical datasets, including multiple modalities such as: X-ray, histopathology, and Optical Coherence Tomography (OCT). Our experimental protocol follows a cross-validation strategy. For each test on a dataset from Table 1, we construct the meta-space using all remaining datasets from the collection. This approach ensures that the model selection process is unbiased and generalizes well to unseen data. For MedNNS, we employ the OFA (Once For All) Supernetwork [3], a ResNet-like model with variable depth, width, and expansion ratio. All Supernetworks are trained on one A100 GPU following the scheme in subsection A1 (2). After that, we evaluate the Spearman rank correlation of subnetworks and achieve 90% with 0.3% margin. The meta space is trained using Adam [12] with a learning rate of 10^{-2}.

For MedNNS, we report three variants to explore the trade-off between efficiency and performance. 1) MedNNS$_{T1}$ is obtained by querying the meta-space based on the target dataset to select the best model and train it. 2) MedNNS$_{T5}$ extends this approach by selecting the top five candidate models, training each for one epoch, and then choosing the model with the highest evaluation accuracy to continue training. 3) MedNNS$_{T10}$ is obtained by extracting the top ten models and then using the same selection process. We compare our approach with SOTA NAS methodologies and DL models initialized using ImageNet pretrained weights. To avoid bias from the choice of optimization strategy, all models are trained using five different strategies, and we report the highest test accuracy achieved at epoch 10 and epoch 100 to reflect both convergence speed and final performance. However, due to the absence of their model zoo structure file, we were unable to test TANS [10] method.

Table 1. Comparison of methods across datasets reporting test accuracies at Epoch 10 (@10) and Epoch 100 (@100) to show convergence speed and final accuracy.

	Method	Pneumonia		OrganS		Tissue		Derma		Blood		Breast		Average	
		@10	@100	@10	@100	@10	@100	@10	@100	@10	@100	@10	@100	@10	@100
DL models	MobileNetV3[13]	88.5%	92.8%	71.3%	76.0%	58.8%	68.1%	66.9%	75.1%	81.0%	88.6%	73.7%	85.9%	73.4%	81.1%
	EfficientNet$_{B0}$[24]	85.6%	92.3%	78.1%	81.2%	64.4%	69.4%	70.8%	76.2%	75.7%	87.2%	77.6%	83.3%	75.4%	81.6%
	EfficientNet$_{B4}$[24]	85.6%	92.0%	74.5%	81.1%	55.4%	68.7%	68.7%	74.3%	82.8%	96.2%	73.7%	87.2%	77.8%	83.3%
	ResNet18[8]	93.1%	95.2%	66.5%	80.6%	57.6%	66.8%	72.0%	75.0%	91.9%	96.2%	85.9%	89.7%	76.5%	83.9%
	ResNet50[8]	91.3%	93.9%	68.8%	81.1%	59.9%	63.1%	72.3%	75.0%	87.3%	92.1%	79.5%	89.1%	80.7%	82.4%
	DenseNet121[9]	93.6%	95.2%	79.8%	82.1%	64.0%	68.1%	72.9%	75.9%	91.6%	96.4%	82.0%	89.1%	80.0%	84.5%
NAS methods	ProxylessNAS[4]	89.7%	94.1%	78.3%	81.1%	63.3%	68.6%	71.5%	76.1%	92.4%	96.4%	84.6%	89.7%	76.7%	84.3%
	MNASNet[23]	88.0%	93.4%	70.4%	79.8%	63.2%	**69.6%**	71.5%	75.0%	88.0%	96.2%	78.8%	87.8%	80.17%	83.6%
	NASNet[28]	92.9%	94.9%	77.6%	81.9%	64.7%	68.6%	71.0%	75.4%	91.5%	96.0%	83.3%	90.4%	80.17%	84.5%
	PNASNet[15]	93.1%	94.4%	75.8%	80.8%	63.0%	69.1%	72.8%	75.0%	93.3%	95.9%	82.7%	87.2%	80.12%	83.7%
	HardcoreNAS[17]	88.3%	92.5%	71.5%	80.9%	62.3%	68.7%	72.0%	75.1%	84.8%	96.3%	79.5%	85.3%	76.4%	83.1%
	OFA[3]	92.2%	93.8%	66.4%	80.1%	58.1%	67.2%	72.2%	75.1%	90.3%	93.4%	84.1%	89.3%	77.2%	83.2%
	MedNNS$_{T1}$	**93.6%**	94.4%	79.4%	81.8%	64.5%	69.0%	**75.5%**	**79.7%**	**94.8%**	**96.6%**	**86.5%**	**92.3%**	**82.4%**	**85.6%**
	MedNNS$_{T5}$	**93.6%**	**96.2%**	**80.6%**	**82.1%**	64.5%	69.1%	**76.2%**	**79.7%**	**95.8%**	**97.5%**	**87.2%**	**92.3%**	**83.0%**	**86.2%**
	MedNNS$_{T10}$	**93.8%**	**96.2%**	**80.6%**	**82.4%**	**65.0%**	69.2%	**77.0%**	**79.7%**	**95.8%**	**97.5%**	**90.4%**	**92.3%**	**83.8%**	**86.2%**

4 Results and Discussion

Table 1 demonstrates that our method achieves higher accuracy across all datasets except for the Tissue dataset. Analyzing the selected models, we found a strong correlation between the target dataset and the source dataset used for model initialization, as indicated by the FID scores in Fig. 1. For instance, for the OrganS dataset, most of the chosen models are initialized from source datasets such as OrganC and OrganA, which exhibit a good alignment in feature distribution with OrganS. This alignment facilitates feature reuse, which is the key to the observed performance gains. In contrast, the Tissue dataset shows the highest FID difference from the other datasets, indicating a distinct feature distribution that undermines the benefit of reusing features from the meta-space datasets. Moreover, our approach demonstrates remarkably fast convergence: the accuracy at epoch 10 not only surpasses that of competing models at the same stage but, in many cases, matches or even exceeds their final accuracy at epoch 100. This rapid adaptation is a sign of effective feature reuse and it is valuable in medical imaging applications, where data is often limited and efficient learning is crucial.

Meta-space Analysis. We analyze the structure of our created meta-space by visualizing the dataset and model embeddings using t-SNE, as shown in Fig. 3. The central plot reveals that models are clustered based on their performance on datasets, with each group positioned near a corresponding dataset embedding. Also, datasets with similar FID value such as OrganC and OrganA are found in close proximity within the meta-space, validating the efficacy of using the FID loss. To further explore the space, we zoom in on selected regions, as depicted in the four side plots. These visualizations demonstrate a clear correlation between proximity to the dataset embedding and model accuracy: as the distance from the

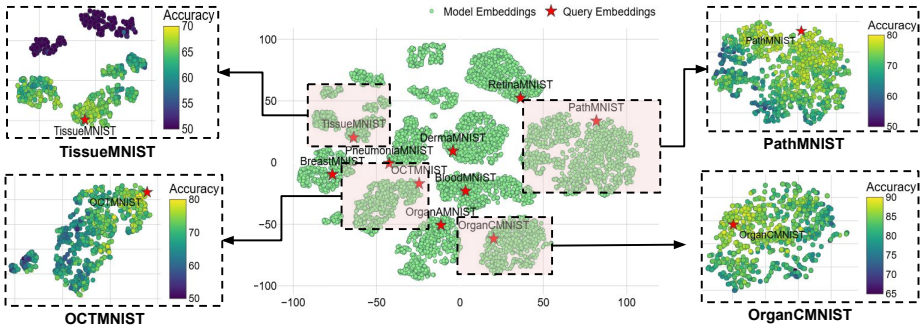

Fig. 3. T-SNE visualization of dataset (query) and model embeddings in the meta-space. The central plot shows the MedNNS meta-space. The zoomed-in side plots provide a closer view of specific regions, illustrating the spatial arrangement of models, colored according to their true accuracy, around their corresponding datasets.

Table 2. Comparison of MedNNS accuracy results using different loss combinations.

Rank	FID	Contrastive	BreastMNIST			PneumoniaMNIST		
			T1	T5	T10	T1	T5	T10
•	○	○	89.8%	89.8%	91.7%	93.0%	94.4%	94.6%
○	•	○	88.5%	90.4%	91.0%	92.6%	93.1%	95.0%
○	○	•	88.5%	89.8%	90.4%	91.9%	93.2%	95.4%
○	•	•	89.7%	91.7%	91.7%	94.4%	95.5%	95.5%
•	•	○	**92.3%**	**92.3%**	**92.3%**	**94.4%**	**96.2%**	**96.2%**

dataset embedding increases, the accuracy decreases. This finding underscores the impact of the rank loss in effectively ordering models by performance.

Computational Cost. Constructing the Supernet-based model zoo requires moderate resources, with each Supernetwork taking approximately three hours to train on a single A100 GPU. However, this is a one-time, upfront investment. Once the meta-space is built, the framework enables rapid model selection for any new medical dataset by encoding only a few samples, a process that takes about one second. Most importantly, MedNNS significantly reduces subsequent training costs for any given target dataset. As shown in table 1, our approach allows models to closely match or even surpass the final accuracy of competing methods trained for 100 epochs in as few as 10 epochs, demonstrating a substantial acceleration in convergence.

Ablation. Table 2 shows the effect of different loss combinations on constructing the meta-space. Single loss comparison reveals that the rank loss outperforms both contrastive and FID losses. This is because the rank loss enforces the correct order of the models based on performance, while contrastive loss operates at model-dataset pair level without model-to-model interaction. The model ordering attribute of the rank loss can be further observed in Fig. 3, where model-to-model interaction leads to a continuous meta-space. Pairing losses yields better performance, with the pairing of rank and FID losses delivering the best results.

5 Conclusion

In this work, we introduce MedNNS, the first medical neural network search framework, which simultaneously identifies the optimal architecture and its corresponding pretrained weights for a new dataset by querying a pre-constructed meta-space. Our approach efficiently increases the number of models represented in the meta-space by leveraging a Supernet-based model zoo, while refining the alignment of embeddings by integrating inter-model and inter-dataset performance information through rank and FID losses. This leads to enhanced performance and faster convergence across various medical datasets. Despite these advancements, the challenge of generalization persists, particularly when confronted with datasets that exhibit significant dissimilarity to those within the meta-learning space (e.g., TissueMNIST). In the future, we intend to incorporate a more extensive and diverse array of medical datasets into our framework to expand its coverage. Additionally, we plan to incorporate hardware constraints as an additional dimension in the search process, with the ultimate goal of augmenting the overall applicability and versatility of MedNNS across a broader spectrum of medical imaging scenarios.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Benmeziane, H., Hamzaoui, I., Cherif, Z., El Maghraoui, K.: Medical neural architecture search: Survey and taxonomy. In: International Joint Conference on Artificial Intelligence (2024)
2. Boulila, W., Alshanqiti, E., Alzahem, A., Koubaa, A., Mlaiki, N.: An effective weight initialization method for deep learning: Application to satellite image classification. Expert Systems with Applications p. 124344 (2024)
3. Cai, H., Gan, C., Wang, T., Zhang, Z., Han, S.: Once-for-all: Train one network and specialize it for efficient deployment. arXiv preprint arXiv:1908.09791 (2019)
4. Cai, H., Zhu, L., Han, S.: ProxylessNAS: Direct neural architecture search on target task and hardware. In: International Conference on Learning Representations (2019), https://arxiv.org/pdf/1812.00332.pdf
5. Chu, X., Zhang, B., Xu, R.: Fairnas: Rethinking evaluation fairness of weight sharing neural architecture search. In: Proceedings of the IEEE/CVF International Conference on computer vision. pp. 12239–12248 (2021)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. Godasu, R., Zeng, D., Sutrave, K.: Transfer learning in medical image classification: Challenges and opportunities. Transfer **5**, 28–2020 (2020)
8. 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. pp. 770–778 (2016)
9. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)

10. Jeong, W., Lee, H., Park, G., Hyung, E., Baek, J., Hwang, S.J.: Task-adaptive neural network search with meta-contrastive learning. Advances in Neural Information Processing Systems **34**, 21310–21324 (2021)
11. Kim, H.E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M.E., Ganslandt, T.: Transfer learning for medical image classification: a literature review. BMC medical imaging **22**(1), 69 (2022)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Koonce, B., Koonce, B.: Mobilenetv3. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization pp. 125–144 (2021)
14. Kouli, O., Hassane, A., Badran, D., Kouli, T., Hossain-Ibrahim, K., Steele, J.D.: Automated brain tumor identification using magnetic resonance imaging: A systematic review and meta-analysis. Neuro-oncology advances **4**(1), vdac081 (2022)
15. Liu, C., Zoph, B., Neumann, M., Shlens, J., Hua, W., Li, L.J., Fei-Fei, L., Yuille, A., Huang, J., Murphy, K.: Progressive neural architecture search. In: Proceedings of the European conference on computer vision (ECCV). pp. 19–34 (2018)
16. Matsoukas, C., Haslum, J.F., Sorkhei, M., Söderberg, M., Smith, K.: What makes transfer learning work for medical images: Feature reuse & other factors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9225–9234 (2022)
17. Nayman, N., Aflalo, Y., Noy, A., Zelnik, L.: Hardcore-nas: Hard constrained differentiable neural architecture search. In: International Conference on Machine Learning. pp. 7979–7990. PMLR (2021)
18. Pattnayak, P., Patnaik, S., Patra, S.S., Gourisaria, M.K., Singh, S., Barik, L.: Automated skin lesion analysis and classification using ai. In: 2024 4th International Conference on Sustainable Expert Systems (ICSES). pp. 1491–1498. IEEE (2024)
19. Prevedello, L.M., Halabi, S.S., Shih, G., Wu, C.C., Kohli, M.D., Chokshi, F.H., Erickson, B.J., Kalpathy-Cramer, J., Andriole, K.P., Flanders, A.E.: Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. Radiology: Artificial Intelligence **1**(1), e180031 (2019)
20. Qureshi, I., Ma, J., Abbas, Q.: Recent development on detection methods for the diagnosis of diabetic retinopathy. Symmetry **11**(6), 749 (2019)
21. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems **32** (2019)
22. Skorski, M., Temperoni, A., Theobald, M.: Revisiting weight initialization of deep neural networks. In: Asian Conference on Machine Learning. pp. 1192–1207. PMLR (2021)
23. Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., Le, Q.V.: Mnasnet: Platform-aware neural architecture search for mobile. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2820–2828 (2019)
24. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
25. White, C., Safari, M., Sukthanker, R., Ru, B., Elsken, T., Zela, A., Dey, D., Hutter, F.: Neural architecture search: Insights from 1000 papers. arXiv preprint arXiv:2301.08727 (2023)
26. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**(1), 41 (2023)

27. Yu, Y., Zhang, W., Deng, Y.: Frechet inception distance (fid) for evaluating gans. China University of Mining Technology Beijing Graduate School **3** (2021)
28. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8697–8710 (2018)