NEURON-LEVEL ARCHITECTURE GROWTH: A CONTROLLED EVALUATION FOR EEG TIME-SERIES DECODING

Adam Mounir[1,2] *Stella Douka*[1] *Arnault H. Caillet*[2,4] *Bruno Aristimunha*[2,3,*] *Sylvain Chevallier*[1,*]

[1]Inria TAU – LISN, Université Paris-Saclay, France [2]Yneuro, Paris, France
[3]University of California San Diego, CA, USA [4]Imperial College London, UK

ABSTRACT

Convolutional EEG decoders are trained at a fixed width, usually set by their authors on other data. Growing methods add neurons during training where the loss could decrease the most, but whether they improve compared to a reference width is untested on EEG. Here, we grow three convolutional backbones on 12 motor-imagery datasets under three protocols and compare each with its reference model per subject. The growing ShallowFBCSPNet scores 2.9 points above its reference model with only half the parameters ($0.57\times$), SCCNet changes by at most 1.2 points. Deep4Net growing models show decreased accuracy, but they require adaptation that prevent to compare faithfully the results. These differences follow the selection step, which keeps a candidate neuron relying on a dynamic threshold from singular values decomposition. Overall, these results suggest that growth helps when its criterion can rank the candidate neurons, and that the abstention rate tells where a decoder can be grown small from scratch.

***Index Terms*—** EEG, motor imagery, brain–computer interfaces, neural architecture growth, deep learning

1. INTRODUCTION

Convolutional networks are the standard end-to-end decoders for EEG (ShallowFBCSPNet and Deep4Net [1], SCCNet [2], EEGNet [3]). The large brainwave foundation models proposed since gain about one point over these small architectures on BCI benchmarks, for a thousand times more parameters [4, 5, 6]. The width of these decoders, the number of filters per layer, is tuned by their authors on one dataset and reused on every other, including in cross-dataset benchmarks [7, 8, 9]. That choice is domain sensitive [10, 11]. EEG datasets differ in subjects, trials and classes, and architecture searches on EEG are tailored to a task or a subject [12, 13].

Indeed, one could adapt the width per dataset [14], at the cost of multiply the training cost, as BCI applications call for decoders that are compact and effective from the start [3, 15, 16]. Growing methods offer a alternative approach: the network starts narrow, and training decides where and how much to widen it [17, 18, 19]. Within this approach, functional methods [20, 21] implements growth steps that reduce the loss most at first order (Fig. 1a), and a line search selects the number of neuron to add. It builds on Net2Net [22] and extends to directed acyclic graphs [23], with evidence mostly from image benchmarks.

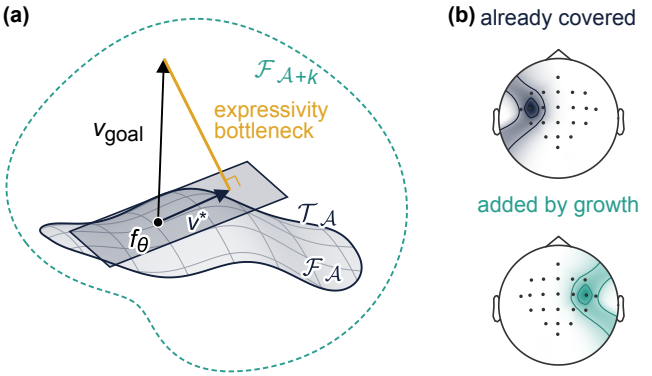

Fig. 1. (a) The desired update v_{goal} splits into a reachable part and a residual, the expressivity bottleneck, which growth covers. (b) On EEG, both read as scalp patterns (schematic).

On EEG, it has so far been applied to a classification head [24] or triggered when learning stalls [25]. Whether growing the convolutional stages improves on the reference width, and for which backbones, remains unknown.

To address this question, we grow three convolutional backbones from a narrow start on twelve motor-imagery datasets under three protocols, 34,596 fits in all, and compare each with its reference model on the same subjects. Our results show that:

1. A decoder grown from scratch can beat its reference width with fewer parameters. ShallowFBCSPNet gains 2.9 points within-session with $0.57\times$ the parameters and SCCNet stays within 1.2 points of its reference. Differences in architectures and ca- pacity prevent reaching any strong conclusion regarding Deep4Net and its growing surrogate.
2. Whether growth will help can be read during training. Growing ShallowFBCSPNet stabilizes at three quar-

*Equal supervision.

ters of the reference model width, the deep backbone reaches the same width than reference Deep4Net.
3. Growth thus works best when its criterion can rank the candidate neurons. Compared to the final width, the abstention rate tells where a decoder can be grown small from scratch.

2. METHODS

2.1. The growth step

We consider a convolutional decoder trained on the trials of one fold. A *growable layer* is a pair of consecutive convolutions with at most a normalisation between them. Channels can then be added to the output of the first and to the input of the second without touching any other layer. The unknown is the width of this layer, which starts narrow and is widened during training with the method of Verbockhaven et al. [20], as implemented in the gromo library. Every five epochs, the layer is offered a growth opportunity, which runs in four steps.

Table 1. Where each backbone grows and what it is compared with. Only the first two rows are width-matched (Section 2.2).

Backbone	Growable layer	Width	Reference
Shallow	temporal $\rightarrow$ spatial	$8 \rightarrow 40$	ShallowFBCSPNet, 40
SCCNet	spatial $\rightarrow$ spatio-temp.	$4 \rightarrow 22$	SCCNet, 22
Deep	conv2a $\rightarrow$ conv2b	$8 \rightarrow 32$	Deep4Net, 4 stages

Propose. Let v_{goal} be the change in the layer's output that would most decrease the loss at first order, computed per sample. Its projection v^* onto the tangent space of the current parameters is the best that updating the existing weights can achieve, and the residual $v_{\text{goal}} - v^*$ is the expressivity bottleneck of the layer (Fig. 1a). Second-order moments of the layer's inputs and their cross-covariance with the residual are accumulated over the training set. Fitting new neurons to the residual is then a low-rank approximation problem: a singular value decomposition returns candidate neurons in decreasing order of first-order contribution, with singular values $\lambda_1 \geq \lambda_2 \geq \dots$

Select. We keep the candidates whose singular value is at least $10\,\%$ of λ_1. The default threshold of gromo is absolute, and on EEG it fell either above a backbone's whole spectrum or below all of it. The relative floor applies the same criterion as a fraction of the spectrum. A width cap stops the layer at its target width (Table 1). A growing backbone therefore never ends wider than its reference model.

Amplify. The selected block is added with a scaling factor s chosen by line search over $\{0,\ 0.1,\ 0.5,\ 1.0\}$. Since the candidates were fitted to the training signal, the search is scored on the last fifth of the epoch's training batches, withheld from the statistics, using the chosen loss function (cross-entropy in this study). When an epoch has fewer than four batches, as in most within-session fits, it falls back to the training batches.

Decide. If the line search returns $s = 0$, the update is deleted and the network is unchanged, which we call an *abstention*. Growth continues, with statistics re-estimated at the next opportunity, and a backbone can abstain at all 39 opportunities of a 200-epoch fold and end at its starting width.

2.2. Three backbones

In ShallowFBCSPNet, the temporal and spatial convolutions are chained with no nonlinearity between them, and the temporal filters grow from 8 towards the reference width of 40. SCCNet is spatial-first, with only a BatchNorm between its two convolutions, which grows with the layer. Its spatial components grow from 4 towards the reference width of 22. The downstream layers (normalisation, readout, classifier) are those of the reference models. They are re-implemented together with the layer, so the comparison is one between two codebases as well as two procedures.

Deep4Net cannot be grown faithfully, because a max-pooling layer sits between every pair of its convolutions and changes the spatial dimension, which breaks the layer. Our deep backbone is therefore a VGG-style surrogate with two stages, pooling only at the end of a stage and the layer inside stage two. We compare it with the reference four-stage Deep4Net, which has an order of magnitude more parameters. This contrast mixes growth with a change of geometry, so we call the backbone *Deep* rather than Deep4Net.

2.3. Datasets and protocols

We evaluate growth on twelve motor-imagery datasets from MOABB [7]: AlexMI, BNCI2014-001, BNCI2014-002, BNCI2014-004, BNCI2015-001, Cho2017, Lee2019-MI, PhysionetMI, Schirrmeister2017, Shin2017A, Weibo2014 and Zhou2016. They have two to four classes and between four and 109 subjects each. The task is to classify the imagined movement of each trial, band-pass filtered between 8 and 32 Hz and resampled to 250 Hz. Following MOABB, the score is ROC-AUC for the six two-class datasets and accuracy for the six others, both in percent, so differences between models are in points.

We evaluate three protocols: within-session (5-fold cross-validation within a session), cross-session (leave-one-session-out) and cross-subject (leave-one-subject-out). Each backbone contributes two models, six in all: the reference model, at the width its authors published, and its growing counterpart, which starts narrow. All models share the same 200-epoch budget and optimiser (AdamW, learning rate 6.25×10^{-4}, batch size 64), with $20\,\%$ of the training trials held out for validation, no early stopping and three seeds per fold. Each fit is scored at the epoch of best validation accuracy.

We report raw scores, without alignment. Euclidean alignment [26], run as a second complete condition, lowers

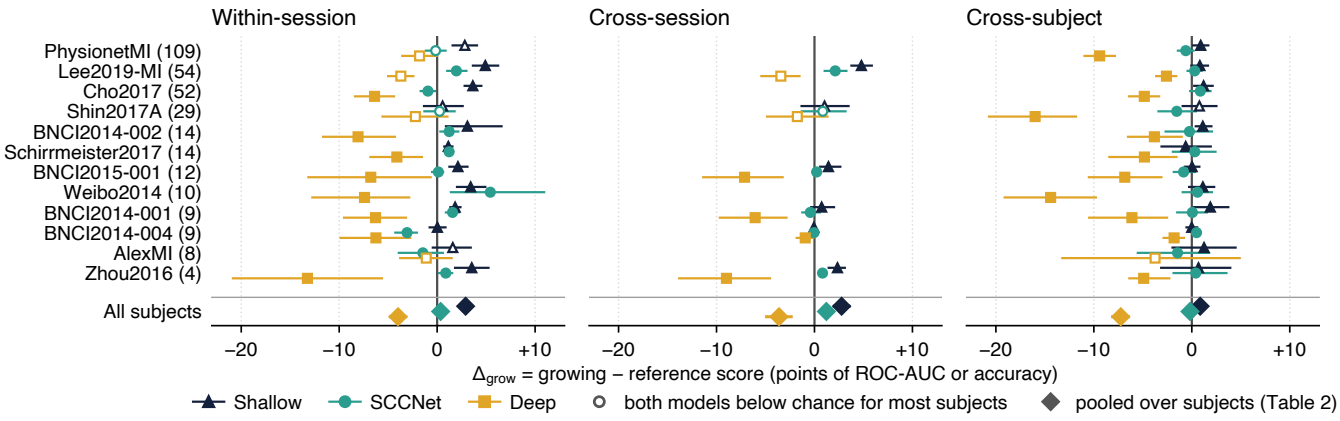

Fig. 2. The growing Shallow implementation generally outperformed its reference, whereas the growing Deep surrogate underperformed the architecturally different Deep4Net. Growing minus reference score, in points, per dataset (rows, subjects in parentheses) and protocol (panels), with 95 % bootstrap intervals over subjects. Bottom row: all subjects pooled (Table 2). Hollow markers: both models below chance for most subjects.

the scores of all six models by 2.4 to 5.1 points within-session and raises them by 1 to 2 points only cross-subject. Taking it as the default would have favoured the growing backbones cross-subject and hidden a loss on every model within-session.

2.4. Statistical analysis

Sessions and seeds are averaged before any test, so the unit of analysis is the subject, and the 34,596 scored folds reduce to 9,168 subject-level units across protocols, models and alignment conditions. Pooling them would inflate every p-value, since they are not independent. For each protocol and backbone, Δ_{grow} is the subject-paired difference between the growing backbone and the reference model. We report its mean with a 95 % percentile bootstrap interval over subjects (20,000 resamples) and a Wilcoxon signed-rank test, Holm-corrected over the nine protocol-by-backbone cells as one family, together with the minimum detectable effect (MDE) of each cell at 80 % power.

3. RESULTS

3.1. The growing backbone is ahead for Shallow and behind for Deep

To test whether growth improves on the reference width, we compare each growing backbone with its reference model on every dataset (Fig. 2). The growing Shallow is above zero on all twelve datasets within-session and on ten of twelve cross-subject, so its gain does not come from a single dataset. The growing SCCNet changes sign from one dataset to the next. The growing Deep is behind on every dataset in all three protocols.

Table 2. Growth is ahead for Shallow and behind for Deep in every protocol. Δ_{grow}: growing minus reference score, in points. Brackets: 95 % bootstrap intervals over subjects. *Ahead*: subjects on which the growing backbone scores higher. MDE: minimum detectable effect at 80 % power. † marks the two cells whose effect falls below it.

Protocol	Net	Δ_{grow} [95 % CI]	MDE	ahead
Within-sess.	Shallow	$+2.9\,[+2.3, +3.5]$	0.9	240/324
	Deep	$-4.0\,[-4.9, -3.1]$	1.3	107/324
	SCCNet	$+0.4^{\dagger}\,[-0.2, +0.9]$	0.8	172/324
Cross-sess.	Shallow	$+2.8\,[+1.9, +3.7]$	1.3	85/116
	Deep	$-3.6\,[-5.1, -2.2]$	2.0	28/116
	SCCNet	$+1.2\,[+0.4, +2.1]$	1.2	72/116
Cross-subj.	Shallow	$+0.9\,[+0.4, +1.3]$	0.6	195/324
	Deep	$-7.3\,[-8.2, -6.3]$	1.4	46/324
	SCCNet	$-0.2^{\dagger}\,[-0.7, +0.3]$	0.7	168/324

Table 2 pools the subjects and gives Δ_{grow} for the nine cells, next to their MDE. Two cells, both for SCCNet, fall below the MDE with intervals that cross zero, and we draw no conclusion from them. Of the other seven, four favour the growing backbone and three the reference model, and all seven survive Holm correction. The effects are small next to the spread between runs. The median standard deviation across seeds is 4.4 points, so the largest gain (+2.9) is 0.7 times the spread of a single configuration. It shows up only when averaged over 324 subjects.

3.2. The gain comes with fewer parameters

To see what the gain costs in size, we place the six models on the accuracy–size plane (Fig. 3). The growing Shallow scores

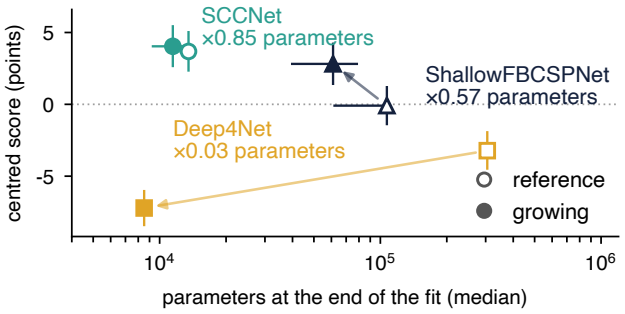

Fig. 3. The growing Shallow scores higher than its reference with fewer parameters. Score against size, within-session: scores centred within dataset and averaged over subjects, against the median parameter count. Filled markers: growing backbone, open markers: reference model. Horizontal bars: inter-quartile range of the parameters, vertical bars: 95 % bootstrap interval over subjects.

higher with fewer parameters. It reaches $+2.8$ with about 6×10^4 parameters, against -0.1 at 1.05×10^5 for the reference model, which is $0.57\times$ the parameters. For SCCNet, the two models nearly coincide ($+4.0$ against $+3.7$).

The network is smaller because growth stops before the target width. Only 29 % of Shallow folds reach width 40, and the median fold is scored at width 26. How far it grows follows the amount of data. By the end of the fit, the share of Shallow fits that reach width 40 rises from none on Shin2017A, with 16 training trials per fold, to 94 % on BNCI2014-001, with 230 (Spearman $\rho = 0.77$ over the twelve datasets). Cross-subject, where the training set pools all other subjects, it reaches 78 %. SCCNet reaches its full width of 22 on 78 % of folds, so it reproduces the reference width and its small contrast follows. Growth adds training time, 22 s per fold for the growing Shallow against 14 s for the reference model (medians, same 200-epoch budget).

3.3. Growth abstains most where its candidates can be ranked

What distinguishes the backbone that gains from the two that do not? The abstention rate does. We define it as the share of growth opportunities at which the line search returns $s = 0$. It orders the three backbones as Table 2 does. The growing Shallow abstains at 74.9 % of its opportunities, the growing SCCNet at 3.5 % and the growing Deep at 0.7 %. Over the 542,187 opportunities of the campaign, Deep adds almost every update it is offered.

We find the same order in the selection step, which runs before the line search. Of the neurons proposed at a growth opportunity, Shallow keeps a median of 15 % and SCCNet 85 %, whereas Deep keeps all of them (100 %). The candidate spectra explain this order. They span 3.1 to 4.0 decades for Shallow and 0.9 to 3.3 for SCCNet, but only 0.3 to 1.4 for Deep. On a spectrum this flat, no threshold separates the candidates and every proposal goes in.

A single growth step changes little. Its predicted first-order gain is a median of $10^{-3.6}$ of the gain of one ordinary epoch for Shallow, and $10^{-1.0}$ for Deep. The held-out loss one epoch before and after a step is centred on zero for all three backbones. Any effect of growth must therefore come through the training trajectory it induces [27].

A simpler explanation, that a width chosen on a large corpus is too large for one session, does not hold: binned by training-set size, the growing backbone has no advantage in the smallest bins and is behind in every bin cross-subject.

Overall, these results suggest that the abstention decision, more than the added capacity, drives the difference between backbones.

4. DISCUSSION

The present study asked whether growing the convolutional stages of an EEG decoder improves on its reference width, and for which backbones. It makes two contributions. First, we compared three growing backbones with their reference models, subject by subject, on twelve motor-imagery datasets (Fig. 2). Growth helps ShallowFBCSPNet, which ends ahead with $0.57\times$ the parameters and changes SCCNet by at most 1.2 points. Growing deep surrogate underperformed Deep4Net, but differences in architectures and capacity prevent attributing this deficit to growth only. Second, we showed that these outcomes follow how often the growth procedure abstains, a rate that needs no test data. ShallowFBCSPNet abstained at three quarters of its opportunities, the deep backbone at almost none.

These results complement the evidence for neuron-level growth on image benchmarks [20, 21], and are consistent with the growth of a c-VEP classification head [24] and with the pruning of a motor-imagery CNN at nearly equal accuracy [28]. On EEG, the benefit of adapting the width appears as a smaller network more than as a higher score, which is what efficient BCI decoders aim for [29]. At the other end of the scale, brainwave foundation models buy about one point with a thousand times more parameters [4]. Effects of a few points are usual on this data, where a nine-subject dataset cannot separate two architectures [30].

The mechanism we propose is selection. Where the candidate spectrum spans several decades, the line search can rank the neurons and rejects most of them, and the network stays small and improves. Where it is flat, every update is accepted and the network grows to its cap without benefit.

The growth literature asks when and where to add capacity. Layer-growing policies trigger on a fitting risk [31], and a recent criterion detects under-expressive layers before growing them [32]. The abstention rate and the span of the candidate spectrum answer both from the training signal of a single run.

Overall, growth is usually presented as a way to find the right capacity. Here, it helped where it abstained most often. Reported with the final width, the abstention rate would tell where a decoder can be grown small from scratch, and paves the way to decoders sized by their data, not their authors.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study used human EEG data released in open access by the original studies and accessed through MOABB [7]. Ethical approval was not required, as confirmed by the licences of the open access data.

6. REFERENCES

[1] Robin Tibor Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[2] Chun-Shu Wei, Toshiaki Koike-Akino, and Ye Wang, "Spatial component-wise convolutional network (SCCNet) for motor-imagery EEG classification," in *International IEEE/EMBS Conference on Neural Engineering (NER)*, 2019, pp. 328–331.

[3] Vernon J Lawhern et al., "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056013, 2018.

[4] Na Lee et al., "Are large brainwave foundation models capable yet? Insights from fine-tuning," 2025, arXiv:2507.01196.

[5] Gayal Kuruppu et al., "EEG foundation models: a critical review of current progress and future directions," *Journal of Neural Engineering*, vol. 23, no. 2, pp. 021001, 2026.

[6] Pierre Guetschel et al., "Open EEG bench: benchmarking parameter-efficient fine-tuning of EEG foundation models," Zenodo, 2026.

[7] Vinay Jayaram and Alexandre Barachant, "MOABB: trustworthy algorithm benchmarking for BCIs," *Journal of Neural Engineering*, vol. 15, no. 6, pp. 066011, 2018.

[8] Sylvain Chevallier et al., "The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark," 2024, arXiv:2404.15319.

[9] Manuel Eder, Jiachen Xu, and Moritz Grosse-Wentrup, "Benchmarking brain–computer interface algorithms: Riemannian approaches vs convolutional neural networks," *Journal of Neural Engineering*, vol. 21, no. 4, pp. 044002, 2024.

[10] Lichao Xu et al., "Cross-dataset variability problem in EEG decoding with deep learning," *Frontiers in Human Neuroscience*, vol. 14, pp. 103, 2020.

[11] Bruno Aristimunha et al., "Evaluating the structure of cognitive tasks with transfer learning," 2023, arXiv:2308.02408.

[12] Yiqun Duan et al., "Cross task neural architecture search for EEG signal classifications," 2022, arXiv:2210.06298.

[13] Chong Wang et al., "Subject-adaptive EEG decoding via filterbank neural architecture search for BCI applications," *IEEE Journal of Biomedical and Health Informatics*, 2026.

[14] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter, "Neural architecture search: A survey," *Journal of Machine Learning Research*, vol. 20, no. 55, pp. 1–21, 2019.

[15] Salim Khazem et al., "Minimizing subject-dependent calibration for BCI with Riemannian transfer learning," in *International IEEE/EMBS Conference on Neural Engineering (NER)*, 2021, pp. 523–526.

[16] Martin Wimpff et al., "Fine-tuning strategies for continual online EEG motor imagery decoding: insights from a large-scale longitudinal study," in *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2025, pp. 1–7.

[17] Qiang Liu, Lemeng Wu, and Dilin Wang, "Splitting steepest descent for growing neural architectures," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[18] Lemeng Wu et al., "Firefly neural architecture descent: a general approach for growing neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[19] Federico Errica et al., "Adaptive width neural networks," in *International Conference on Learning Representations (ICLR)*, 2026.

[20] Manon Verbockhaven et al., "Growing tiny networks: Spotting expressivity bottlenecks and fixing them optimally," *Transactions on Machine Learning Research*, 2024.

[21] Utku Evci et al., "GradMax: Growing neural networks using gradient information," in *International Conference on Learning Representations (ICLR)*, 2022.

[22] Tianqi Chen, Ian Goodfellow, and Jonathon Shlens, "Net2Net: Accelerating learning via knowledge transfer," in *International Conference on Learning Representations (ICLR)*, 2016.

[23] Stella Douka et al., "Growth strategies for arbitrary DAG neural architectures," in *European Symposium on Artificial Neural Networks (ESANN)*, 2025.

[24] Sébastien Velut et al., "Tackling brain signal inter-subject variability with adaptive neural architectures," in *International Joint Conference on Neural Networks (IJCNN)*, 2026.

[25] Byeong-Hoo Lee and Kang Yin, "Towards a network expansion approach for reliable brain-computer interface," in *13th International Winter Conference on Brain-Computer Interface (BCI)*, 2025, pp. 1–4.

[26] Bruna Junqueira et al., "A systematic evaluation of Euclidean alignment with deep learning for EEG decoding," *Journal of Neural Engineering*, vol. 21, no. 3, pp. 036038, 2024.

[27] Paul Caillon and Christophe Cerisara, "Growing neural networks have flat optima and generalize better," HAL preprint hal-04697428, 2024.

[28] Vishnupriya R et al., "Performance evaluation of compressed deep CNN for motor imagery classification using EEG," in *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 795–799.

[29] Xiaying Wang et al., "MI-BMInet: an efficient convolutional neural network for motor imagery brain–machine interfaces with EEG channel selection," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8835–8847, 2024.

[30] Csaba Márton Köllőd et al., "Deep comparisons of neural networks from the EEGNet family," 2023, arXiv:2302.08797.

[31] Haihang Wu et al., "When to grow? A fitting risk-aware policy for layer growing in deep neural networks," 2024, arXiv:2401.03104.

[32] Yifan Wang, Julien Mille, and Moncef Hidane, "Where to grow: a surprisingly straightforward criterion to detect underexpressive layers," in *European Symposium on Artificial Neural Networks (ESANN)*, 2026, pp. 715–720.