

Evolutionary Architecture Search for Chlorophyll-*a* Prediction in Lakes using Sentinel-2

AutoML4EO Workshop

Kürşat Kömürcü Linas Petkevičius

Vilnius University, Institute of Computer Science
IRISA, Université Bretagne Sud
European Commission Joint Research Centre

AutoML4EO Workshop, 1 October 2026
AutoML 2026 · Ljubljana, Slovenia

AIML at Vilnius University

Artificial Intelligence Methods Lab



Trustworthy, reproducible AI methods

Explainable AI, multimodal learning
and agentic systems

Earth observation research

Climate monitoring and environmental change
Peatland moisture and urban heat
Satellite and in-situ data fusion

Supported by VU high-performance computing



Open to collaboration in AI for Earth observation

Joint research and EU proposals, shared datasets, and applied pilots

Valentas Gružas - valentas.gruzas@mif.vu.lt - aiml.lt

AutoML for Earth observation

Two proposed research projects



AWARE-EO

MSCA Postdoctoral Fellowship proposal

Automated data preparation and model development

Build validated multisensor data cubes
Select and adapt foundation models
Generate and test model architectures

Use cases

Peatland moisture, water quality
and solar-panel mapping

VERITAS

ERC Starting Grant proposal

Automated construction and verification of simulations

Build EO-grounded digital twins
Check physical and economic consistency
Benchmark reliability with VERITAS-Bench

Use cases

Peatlands and urban climate risk,
with food-risk stress tests

We are looking for research and application partners

EO data and ground truth, AutoML methods, verification, and joint pilots

Valentas Gružas valentas.gruzas@mif.vu.lt aiml.it

Operational Earth observation is often small and tabular

- In-situ labels are expensive: $\sim 10^3$ rows of expert spectral features, not millions of images
- Networks for such data are usually hand-designed
- NAS is mature, but mostly studied on large-scale vision; it is unclear it helps with so little signal

Why model size matters

- Downlink bandwidth, not sensing, is often the binding constraint
- Onboard inference has flown [GFM⁺22]: send down only the informative scenes
- A tiny bloom classifier could act as an *onboard screening trigger*

Research question

Revisit a published, hand-designed Sentinel-2 algal bloom classifier [GP24]: **what does architecture search add?**

Held fixed

- Task: binary bloom / no bloom
- The same 14 spectral features
- The lake-level train/test split and its seed
- Training code shared by all models

Searched

- Architecture (MLP family)
- Training recipe: optimiser, schedule, regularisation, weight averaging

Selected data set and specifics

- 1151 observations from 429 sites in Lithuania, Latvia and Estonia, 2017–2021 [GP24]
- Sentinel-2 surface reflectances + radiometric indices: **14 features**
- Near-balanced: 624 no bloom / 527 bloom (majority floor 52.9%)
- Split **by lake**: 801 train / 350 test rows, 129 held-out lakes
- **Why by lake?** One lake gives up to 17 observations; a row-level split lets a model memorise water bodies

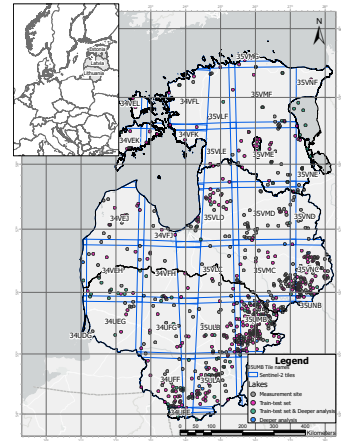


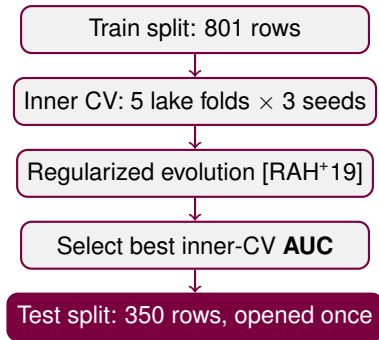
Fig. 1 of [GP24], CC BY. Pink/green: train/test sites.

Search space: $\sim 10^{12}$ MLP configurations



Dimension	Values	Dimension	Values
Blocks	1–6	Optimiser	AdamW, Adam, SGD, RMSprop
Base width	16–256 (9 values)	Learning rate	10^{-4} – 3×10^{-2}
Width profile	constant, taper, expand, hourglass	Schedule	constant, cosine, step
Block type	plain, residual, multi-head, GLU, GRN	Weight decay	0 – 10^{-2}
Normalisation	batch, layer, RMS, none	Batch size	8–256
Activation	ReLU, GELU, ELU, SiLU, Mish, tanh	Epochs	50–600
Dropout	0.0–0.5	Label smoothing	0.0–0.15
Input dropout	0.0–0.3	Class weighting	off, on
Heads	2, 4, 8	Weight EMA	off, 0.99, 0.995, 0.999

- All three reference networks are points in this space $\Rightarrow$ like-for-like comparison
- Candidates above 200k parameters are rejected



Two choices that matter at $n = 801$

- **Select on AUC:** seed-to-seed sd ≤ 0.008 vs. 0.019 for accuracy; selection variance drives overfitting [CT10]
- **Retire by age**, not score: a lucky candidate must re-earn its score every generation

Budget

- Population 64, 20 children/round, tournament 10
- 2720 candidates, ~ 5 h on 20 CPU cores

Held-out results



Model	CV AUC	Test AUC	Test acc.	F ₁	Params
Baseline-MLP	0.806	0.790 $\pm$.006	0.731 $\pm$.009	0.724	10 625
Baseline-Res	0.759	0.754 $\pm$.010	0.699 $\pm$.010	0.675	32 129
Baseline-MHA	0.791	0.779 $\pm$.008	0.733 $\pm$.011	0.727	19 073
NAS-1	0.824	0.818 $\pm$.001	0.746 $\pm$.004	0.744	409
NAS-6	0.824	0.820 $\pm$.001	0.745 $\pm$.003	0.745	433
NAS-8	0.824	0.819 $\pm$.001	0.748 $\pm$.003	0.747	409

Mean $\pm$ sd over 10 weight seeds. Baselines retrained in the same code with their original optimiser settings.

+3.0 pts

test AUC over best baseline

26 $\times$ fewer

parameters (47 $\times$ vs. MHA)

10 / 10

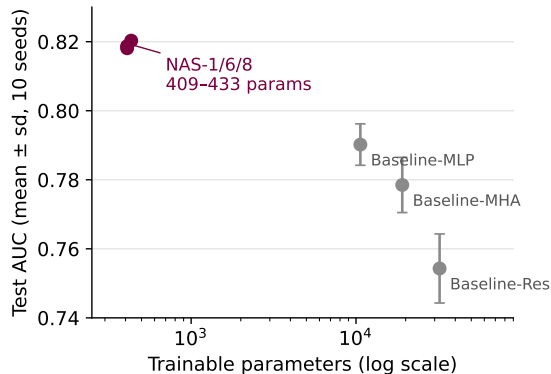
top configs beat every baseline
on AUC

What the search converged on



Top 100 configurations (unanimous except depth & width):

- **One hidden layer** of width 24 (median), plain dense block
- **tanh**: bounded, limits over-confident fits on 801 rows
- **RMS normalisation**
- **RMSprop** with step decay
- **Weight EMA 0.999 [IPG⁺18]**: main source of seed stability
- **Class weighting**



- 1 Searched MLPs beat all three hand-designed networks on test AUC, accuracy and F_1
- 2 With **409 parameters**: 26–47 $\times$ smaller, and far more stable across seeds
- 3 A consistent recipe: one narrow layer, tanh, RMS norm, RMSprop + step decay, weight EMA
- 4 **1.6 kB**, ~ 400 multiply-accumulates per lake: the decision layer of an onboard bloom trigger is essentially free

Limits: one dataset and one split (~ 2.4 pts standard error on accuracy); baselines reimplemented on the public data; features still need ground-side L2A processing and water masking.

Code: <https://github.com/VU-AIML/automl4eo-bloom-nas>

Thank you for your attention!

- [CT10] Gavin C. Cawley and Nicola L. C. Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010.
- [GFM⁺22] Gianluca Giuffrida, Luca Fanucci, Gabriele Meoni, Matej Batič, et al. The ϕ -Sat-1 mission: the first on-board deep neural network demonstrator for satellite Earth observation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.
- [GP24] Dalia Grendaitė and Linas Petkevičius. Identification of algal blooms in lakes in the Baltic states using Sentinel-2 data and artificial neural networks. *IEEE Access*, 12:27973–27988, 2024.

- [IPG⁺18] Pavel Izmailov, Dmitrii Podoprikin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 876–885, 2018.
- [RAH⁺19] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V. Le. Regularized evolution for image classifier architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 4780–4789, 2019.