

AI4Life Open Calls and Public Challenges: why, how, and what we have learned.

Vera Galinova^{1,*}, Mehdi Seifi^{1,*}, Beatriz Serrano Solano², Kristína Lidayová³, Damian Dalle Nogare⁴, Agustín Andrés Corbat^{3,5}, Joshua Talks⁶, Edoardo Giacomello¹, Estibaliz Gómez-de-Mariscal^{7,8,9,10}, Mariana G. Ferreira^{7,8}, Caterina Fuster-Barceló^{11,12}, Juan Manuel Battagliotti⁴, Carlos García-López-de-Haro^{11,13}, Benjamin Salmon^{1,14}, Melisande Croft⁴, Si Young Yie¹⁵, Guillermo Rey-Paniagua¹¹, Xiaotian Hu¹⁶, Sungjun Cho¹⁷, Aagam Sheth¹⁸, Chhayansh Porwal¹⁹, Xiaomeng Li¹⁷, AI4Life Consortium[†], Ricardo Henriques^{7,8}, Xinyang Li²⁰, Alexander Krull¹⁴, Anna Klemm³, Arrate Muñoz Barrutia¹¹, Anna Kreshuk⁶, Wei Ouyang²¹, Florian Jug^{1,+}, and Joran Deschamps^{4,+}

¹Computational Biology Research Center, Human Technopole, Milan, Italy.

²Euro-Bioimaging ERIC Bio-Hub, European Molecular Biology Laboratory, Heidelberg, Germany.

³BioImage Informatics Unit, Science for Life Laboratory, National Bioinformatics Infrastructure Sweden, Uppsala University, Uppsala, Sweden.

⁴BioImage Analysis Unit, National Facility for Data Handling and Analysis, Human Technopole, Milan, Italy.

⁵Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

⁶Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

⁷AI-driven Optical Biology Laboratory, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal.

⁸Optical Cell Biology Laboratory, Instituto Gulbenkian de Ciência, Oeiras, Portugal.

⁹NOVA Institute for Medical Systems Biology, Universidade Nova de Lisboa, Lisboa, Portugal.

¹⁰iNOVA4Health, Programme in Translational Medicine.

¹¹Neuroscience and Life Sciences Department, Universidad Carlos III de Madrid, Getafe, Spain.

¹²Department of Molecular Life Sciences, Universität Zürich, Zürich, Switzerland.

¹³BioImage Analysis Unit, Pasteur Institute, Paris, France.

¹⁴School of Computer Science, University of Birmingham, Birmingham, UK.

¹⁵Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, South Korea.

¹⁶School of Biological Science and Medical Engineering, Beihang University, Beijing, China.

¹⁷Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

¹⁸School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, India.

¹⁹Rajasthan Technical University, Kota, India.

²⁰College of AI, Tsinghua University, Beijing, China.

²¹Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

[†]A list of members and their affiliations appears at the end of the paper.

⁺florian.jug@fht.org, joran.deschamps@fht.org

*these authors contributed equally to this work

ABSTRACT

Within AI4Life, we ran three Open Calls and three Public Challenges (2023-2025), supporting 22 bioimage analysis projects from 151 applications and engaging 225 challenge participants, with the aim of applying FAIR deep learning in the life sciences. Our experience offers a view of the current state of bioimage analysis, the landscape of available tools, as well as the existing gaps between method developers, tool producers and potential users. It highlights that even after careful selection for AI-ready projects, most still require substantial effort to apply deep learning, and that the field still relies heavily on established, well-rounded methods to solve common problems. We come to the conclusion that for scientific AI in biology, the rate-limiting step is not methods and models but data, annotations, and shared infrastructure underneath them.

Introduction

AI4Life was a Horizon Europe-funded project (2022-2025) aimed at bringing state-of-the-art AI-based image analysis to life scientists by establishing and supporting services that promote deep learning FAIRness^{1,2}. With these overarching goals in mind, the AI4Life consortium conducted various activities, from developing a new format for deep learning models and an accompanying public repository³ to organizing workshops promoting tool adoption. Notably, the consortium sought to foster interactions between computational and life scientists, bridging the gap between method developers and those applying these methods. To that aim, we organized three open calls for bioimage analysis projects and three computational challenges.

The Open Calls (OCs) provided life scientists who were facing image analysis challenges with adequate deep learning workflows developed by a team of AI and image analysis experts from our consortium. Selected projects focused on problems that were of interest to the larger community and would therefore impact more than a single research group through re-usability of a solution and open-source tooling. The OCs provided technical support to researchers for a duration of six months, at the end of which the developed pipeline was documented, transferred to the users, and released online.

Technical solutions developed during the OCs were the result of a specific team of analysts' work and therefore could benefit from further investigation by other experts in the field, either by proposing alternative or superior analysis pipelines. Public competitions have previously proven effective at translating biological questions into well-scoped machine learning problems and driving methodological advances⁴⁻¹¹. Building on these examples, we organized three Public Challenges (PCs) centered around image restoration, each targeting a different aspect of microscopy image denoising, and running for two months.

In this manuscript, we provide an analysis of the OCs and PCs, describe lessons learned from organizing them, and discuss the pain points encountered along the way. In particular, we believe that our experience reflects the state of the bioimage analysis field and of the interface between method developers and potential users. By summarizing the best practices that we would recommend to avoid the pitfalls we faced, we hope to not only help similar initiatives in the future, but also incentivize the community to continue working towards improved user training, better tool interoperability and FAIRness. A full report and analysis of both the Open Calls and the Public Challenges are provided in the Supplementary Information (SI) for the interested readers. In the following sections, we discuss the main lessons and recommendations that we draw from our experience.

Open Calls - Lessons learned

The Open Calls were an opportunity to survey the diversity of image analysis problems faced by life scientists, their typical workflows and the current state of the existing solutions offered by the community.

The first challenge was to establish a selection process that would narrow support to projects that were ready to start, solvable within the time frame, and whose solution could be deep-learning-based and transferable to other projects. Our process is shown in Fig. 1, and described in SI section S1.

Across the three OCs, we progressively moved from pure remote scoring to interactive expert dialogue as OC 1 revealed a high variance in both data readiness and the actual relevance of deep learning to the proposed problem. Multiple projects were ultimately solved with classical image analysis or existing tools, including pre-trained deep learning models, while others were delayed by data transfer or the need to re-acquire data.

To reduce this mismatch, OC 2 and 3 introduced three decisive changes: mandatory upload of representative data, a template-based application, and a consultation round with experts. The consultations required substantial additional work, but were the most reliable way to identify ready-to-start projects, re-frame over-ambitious ones, and give useful feedback even to non-selected applicants. Finally, the reviewing scores were simplified to only 3 categories to increase the transparency of the scoring: readiness, applicability of deep learning, and scientific excitement (SI section S1.2).

Challenges faced during the OCs are familiar to image analysis facilities and highlight the need to build better community resources to educate users about the availability of user-friendly tools. As described in SI section S2.3, we observed that many widely used tools still failed to get into the hands of users (e.g. Cellpose¹², CellProfiler¹³, QuPath¹⁴, ilastik¹⁵, Labkit¹⁶, or the napari¹⁷ ecosystem), while users rely heavily on proprietary software, hindering the spread of reproducible analysis and FAIRness. The nature of microscopy images and scientific goals often leads to complex pipelines, chaining several existing tools and requiring custom scripting, as exemplified by the pipelines developed for the OC (see SI section S2.5).

The recurrence of specific tools in the developed pipelines demonstrates that these are cornerstones of bioimage analysis. Notably, this was the case of Cellpose^{12,18} and StarDist¹⁹ for instance segmentation, and Noise2Void for denoising²⁰. Similarly, while the field is moving heavily towards Python under the impulse of scientific Python libraries and deep learning, ImageJ2/Fiji^{21,22} remains an important ecosystem. Several ImageJ2/Fiji plugins provide solutions to widely encountered challenges and are accessible to non-specialists (BigWarp²³, Labkit¹⁶, TrackMate²⁴). Well-established tools often owe their success to both effectiveness and user-friendliness²⁵. State-of-the-art methods, on the other hand, may suffer from complex installations and use. Newer architectures (transformers, foundation models, diffusion-based methods) are absorbed by the

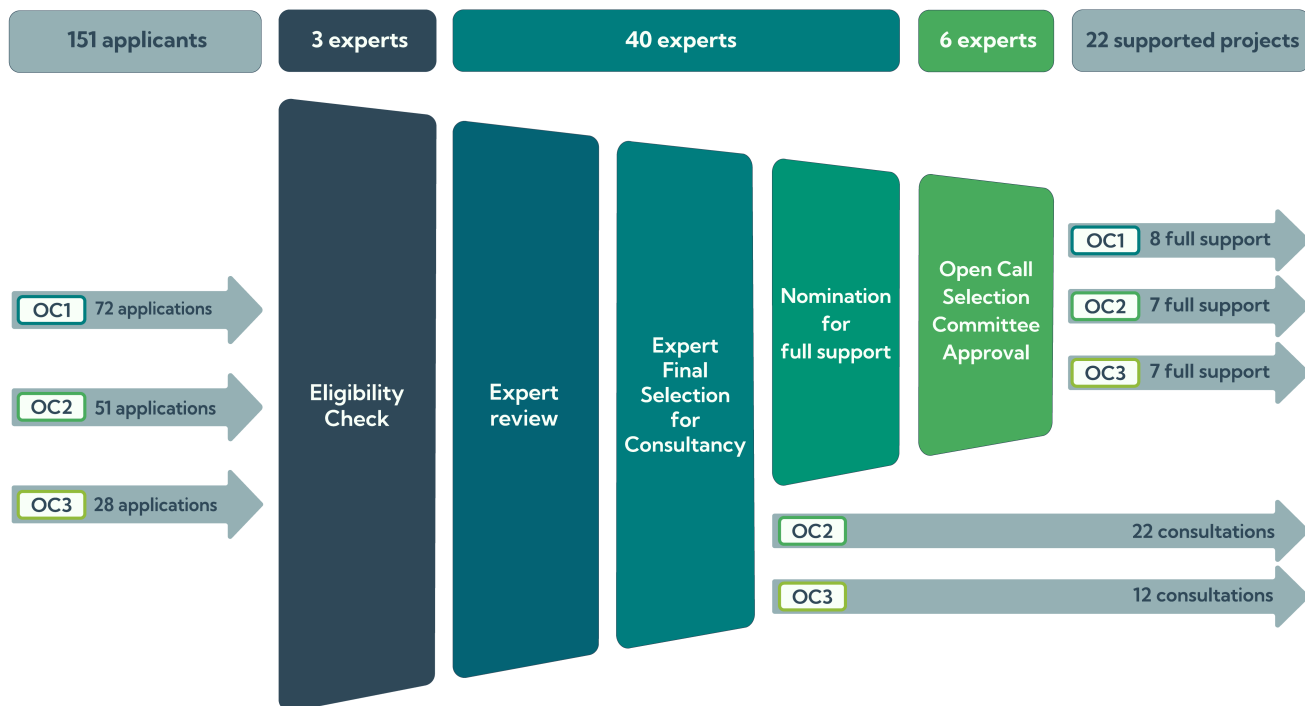


Figure 1. Selection procedure and analysis support workflow for the three Open Calls. The number of experts involved in eligibility checks, reviewing applications, consultation calls, and members of the selection committee is indicated in the top row. OC 1 did not include a consultation step.

field selectively, and the OC pipelines suggest the absorption criterion is whether the new method reduces the data and annotation burden on users rather than whether it improves a benchmark by a few points. This is exemplified by the success of SAM/SAM2^{26,27} in segmentation, which has led to a sharp increase in segmentation tools relying on it as a pre-trained model. Such tools have been used during the OCs, or developed for an OC project²⁸, hinting at a new direction for tool development in the era of large foundation models developed outside of the field of bioimage analysis.

Another issue apparent from both applications and consultations is the need to spread realistic expectations about the capabilities and requirements of deep learning algorithms^{29,30}. Two different aspects of deep learning in particular appear to be generally misunderstood: the quality of annotations required to train a deep learning model, and the current capacities of deep learning methods. For the latter, the consultation round was instrumental in managing user expectations and re-framing the scope of the projects based on more realistic goals. Data labeling, on the other hand, is a time-consuming activity and labels quality is often overestimated by users.

Only 8 out of 22 selected projects were fully ready at the start of the support period, with complete, usable data and sufficient annotations available to start working on the project. Even more striking, only 2 of them had sufficient ground-truth data to directly envision using deep-learning as a first step (see SI section S2.6). In projects that required substantial data labeling, the analysis pipelines often made use of alternative methods or attempted to leverage sparse labeling (see SI section S2.5).

While some projects turned out to be more challenging than expected, most of the incurred delays were a result of difficulty in exchanging data, in particular large datasets. Often, data exchanged relied on private services, external to the scientific institutions (see SI section S2.6). Not only does this present a security risk, but it also highlights the lack of shared data infrastructure across European public research institutions and the dissemination of local solutions.

Furthermore, while most applicants agreed in principle to publicly release their data, the immediate benefit of doing so is rarely visible to the user when sharing has to happen, typically near the end of a project, when attention has already shifted elsewhere. The consequence is that data release is consistently deprioritized: of 22 supported projects, only 16 had data deposited in a public repository by the time of writing, with 2 still pending and 4 unable to contribute data at all (see SI section S2.6). Furthermore, most of the projects incurred significant delays in releasing data publicly, with 10 projects having data uploaded more than 6 months after the official end of the project and an overall average of 8 months delay between end of

support and public data release. FAIR data sharing is not a one-time act at publication but a sustained effort that extends well beyond the visible end of a project, and current incentive structures do not reward that effort. Concretely, we found it more effective to identify shareable subsets early, begin the upload process before the analysis pipeline was finalized, and treat the data deposit as a parallel deliverable rather than a closing step. Even with this strategy, attrition was substantial.

The deeper issue underlying low data-sharing rates is not unwillingness but invisibility of return. Researchers who share image data rarely see direct credit for it: citations accrue to method papers and biological findings, not to the underlying datasets that made either possible. As long as this asymmetry holds, FAIR data sharing will remain something individuals do despite their incentives rather than because of them. Echoing others³¹, we believe that in order to meaningfully shift this trend, two main avenues should be pursued. First, training: most life scientists encounter FAIR principles as compliance language rather than as a working description of how reusable data is structured, annotated, and deposited³², and the gap between willingness to share and data usable by others is wider than most applicants realize at submission time. Second, recognition: dataset publications, repository deposits, and reusable annotations should count toward hiring, promotion, and funding decisions on a comparable footing with method publications. Neither change is within the power of any single project, but both are within the control of the community that funds, hires, and evaluates its members.

Taken together, the lessons from three Open Calls describe a community that is technically capable, intellectually engaged, and culturally aligned with open science, but is operating under a certain lack of infrastructure, training resources, and shared norms that AI methods quietly assume.

Public Challenges - Lessons learned

The Public Challenges offered a complementary perspective, producing fixed benchmarks for method developers to explore diverse solutions.

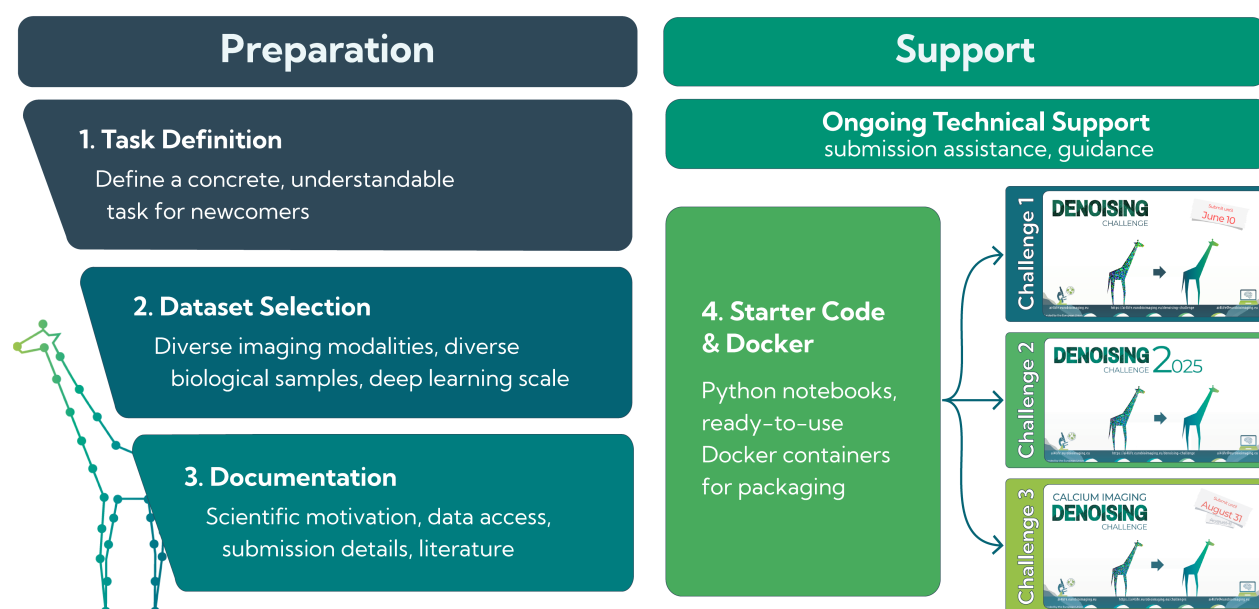


Figure 2. Challenge organization process.

The central lesson from the Public Challenges is that the bottleneck is not community interest in biological imaging, but the rarity of problems that are immediately benchmark-ready. Our initial vision was to draw challenge tasks directly from the Open Calls, but in practice this proved considerably more difficult than anticipated. Many datasets lacked the quality or quantity that deep learning benchmarking requires. Beyond data readiness, many proposals were either too narrowly scoped within a specific biological subfield to attract broad computational interest, or described multi-step experimental workflows that were difficult to abstract into a single, well-defined task, which is a prerequisite for any accessible and fairly evaluated challenge.

This gap motivated us to shift toward publicly available datasets and synthetically generated data as the foundation for challenge design (see SI section S3). While no dataset perfectly captures the diversity of real-world biological imaging, curated public resources offer a practical and transparent starting point: data provenance is known, baselines can be established, and participants can access training data openly. This approach also naturally promotes reproducibility. Future challenge organizers

in the biological imaging space should treat data readiness as a central criterion, and the availability of suitable public datasets as a precondition rather than an afterthought.

The three challenges focused on image denoising in three different settings (see Fig. 2): unsupervised (PC 1), supervised (PC 2), and domain-specific calcium imaging (PC 3). Together, they provided a useful snapshot of which methods participants currently reach for under different training regimes (see SI section S4). In all three settings, participants largely relied on established pipelines: Noise2Void and its variants were the most frequent submissions, reflecting the method's wide adoption and ease of deployment. In the unsupervised setting (PC 1), however, COSDD³³ placed first on three of the four leaderboards and a narrow second on the fourth, where an N2V submission edged it by 0.01 dB. This highlights the potential of methods that have not yet reached mainstream adoption. The supervised challenge (PC 2) attracted more methodological diversity, with participants adapting U-Net architectures and also trying generative adversarial networks (NAFNet-GAN³⁴), the latter topping three out of four leaderboards, suggesting that supervised denoising is both more familiar and more open to experimentation. PC 3, focused on the more demanding calcium imaging setting, showed the slowest submission rate. Here, a transformer-based model (Restormer3D³⁵) topped all four leaderboards, while a well-tuned N2V 3D baseline also remained highly competitive and consistently placed in the top three. Taken together, these results suggest that established methods remain the default choice for participants, but that alternative architectures can become clearly advantageous in settings that are technically more demanding.

Running these challenges also made clear that community benchmarking reveals performance differences that are hard to see from individual method papers alone: a method may be well-known, yet its relative strength across diverse datasets and noise regimes only becomes apparent through direct comparison³⁰.

The geographic diversity of registrants was an encouraging sign of broad community interest (see SI section S4). However, the overall number of successful submissions remained modest compared to established challenges in computer vision or medical imaging. We attribute this primarily to the absence of monetary or conference-based incentives: many prominent challenges offer cash prizes, hardware, or co-location with major venues, sometimes accompanied by paper authorship for top participants. Without such incentives, participation is driven almost entirely by intrinsic scientific interest, reaching a narrower audience. This suggests that future challenges in biological imaging would benefit from strong incentives, such as prize money, hardware, paper authorship for top participants who release their code, broader outreach through community channels such as the [image.sc](#) forum and social media, and closer alignment with conferences or journals to make participation more accessible and attractive to a wider community.

The three Public Challenges revealed that we are constrained by the scarcity of benchmark-ready biological problems, openly reusable datasets, and the incentives needed to sustain broad participation.

Discussion

What emerged from both initiatives was not two separate sets of issues, but the same underlying bottleneck viewed from two different angles. Three years of structured support for 22 bioimage-analysis projects, alongside three public denoising challenges, gave us an empirical view of the gap between the methods the field has published and the conditions under which those methods can actually be applied in practice. That gap is the central finding of this work. The algorithms exist, in many cases as community-supported, sometimes as well-engineered tools, but data are too small, too coarsely annotated, too poorly standardized, too costly to share, and too fragmented across institutional infrastructures.

The Open Calls showed that this gap can be narrowed project-by-project when expert support is available. 18 of 22 projects produced reproducible, openly released analysis pipelines, suggesting that targeted bioimage-analysis support can generate durable value even when starting conditions are uneven. The Public Challenges, by contrast, showed how difficult it remains to convert real biological needs into benchmarks that are sufficiently well scoped, well resourced, and broadly accessible to invite community-wide comparison. Together, these two observations sharpen the same point: methodological progress alone does not translate into practical impact unless the surrounding data and infrastructure are mature enough to carry it.

These observations are easier to dismiss in isolation than together. We observed the dominance of TIFF and proprietary microscopy formats over next-generation file formats (8 of 151 applications), the persistence of proprietary analysis software despite mature open alternatives, the fragmentation of data transfer across 14 different services (for 22 projects), and the fact that many well-adopted open tools remain off the radar of the researchers who would most benefit from them. Each of these is, on its own, a manageable inconvenience. Together, they describe a field whose data infrastructure, tool ecosystem, and cultural norms remain some distance from the level of maturity that current AI discourse often assumes.

We do not see this as cause for pessimism. Frontier AI systems trained on web-scale text were possible because the training data was already there. Our field has none of those conditions fully developed and is now building them deliberately with the examples of the Image Data Resource³⁶, the BioImage Archive³⁷, the BioImage Model Zoo³, OME-NGFF³⁸, community standards^{2,32}, and projects such as AI4Life itself. What this work shows is how far that building still has to go, and how concretely it can be done. The community resources required are not exotic: shared data-transfer infrastructure with institutional

support, sustained annotation efforts on biologically meaningful datasets, training material that calibrates user expectations of what deep learning can and cannot do, and incentive structures that reward dataset publication on par with method publication. None of this is a research problem in the conventional sense. It is infrastructure work, dispersed and undramatic. It is also the rate-limiting step on what scientific AI in biology will deliver in the coming decade³⁰.

Recommendations

If these constraints are largely infrastructural, the response must also be practical. The lessons we draw from three Open Calls and three Public Challenges therefore fall into actionable recommendations, grouped here by the audience best placed to act on them.

For life scientists submitting projects

Deep learning generally benefits from very high quality labeling. Before entering any AI collaboration, it is helpful to keep in mind that labels likely need substantial revision or extension. Plan annotation effort early in the discussion, and treat data and annotations as one of the slowest-moving parts of any AI collaboration.

When possible, adopt next-generation file formats³⁸ for projects involving volumetric, multichannel, or large-scale imaging. TIFF and proprietary formats remain workable for small data, but can make downstream reuse and scaling more difficult.

Explore the open-source bioimage-analysis ecosystem (ImageJ2/Fiji^{21,22}, CellProfiler¹³, QuPath¹⁴, Cellpose¹², ilastik¹⁵, Labkit¹⁶, napari¹⁷, etc.) before commissioning custom development. In many cases, a method that addresses much of the problem already exists and may be within reach with some guidance or training. The bioimaging community will be happy to help you find a good solution on the Image.sc forum³⁹, and the answers will also help others.

Finally, it helps to approach deep learning with realistic expectations: it is not a generic intelligence applied to images, but a class of methods with concrete data and annotation requirements^{29,30}. In our experience, projects are much more likely to progress smoothly when those requirements are made explicit from the outset.

For image analysts carrying out projects

Verify data and annotation readiness before scoping methods. The Open Call experience showed that remote scoring alone is often not enough to distinguish ready projects from those that still need some preparation. A structured consultation with sample data in hand was the most cost-effective filter we found.

Expect to spend a meaningful fraction of any project on data transfer, format conversion, and annotation triage. Budget for it explicitly.

Default to FAIR pipelines³² from the first commit, not as a publication afterthought. Data uploads after the fact are the most common point of attrition. Where possible, prefer the tools that are the best fit for the project, including when they are not the ones you developed yourself. Conversely, recognize that some tools and methods (Cellpose¹², StarDist¹⁹, Noise2Void²⁰, ilastik¹⁵, Labkit¹⁶) are cornerstones for good reason and are often worth considering before custom development.

Finally, team up with others. There are many opportunities, and meetings such as "I2K: From Images to Knowledge" or image analysis networks such as GloBIAS⁴⁰ are a good way to strengthen the international analysts community.

For bioimage-analysis facilities and community organizers

Many widely-used tools still do not reliably reach the users who would benefit from them most, suggesting that the main gap is one of discoverability and training rather than methods alone.

Maintain community-driven, regularly updated documentation of what current tools can do, what data they require, and where they fit into typical pipelines.

For open calls or similar support programmes, include sample data in applications, use a fixed presentation template for project descriptions, and run at least one consultation round before final selection.

Keep scoring schemes simple and interpretable, ideally with three or four categories that can be discussed transparently. Identify two or three preferred data-transfer channels by data size (with institutional support where possible, e.g. Globus^{41,42}), and start the data-release process before the pipeline is finalized rather than after.

In our experience, these changes substantially improved selection accuracy and reduced project attrition between OC1 and OC3.

For method developers, research software engineers and challenge organizers

State-of-the-art methods are most useful to biology when they are also installable, runnable on real data, and accompanied by clear guidance on when they apply.

The success of Cellpose¹² and Noise2Void²⁰ owes as much to their accessibility as to their algorithmic merits. Conversely, the rapid uptake of SAM2²⁷ derivatives in our pipelines shows that the field will absorb foundation models quickly when they reduce the data burden for users.

When releasing a method, consider what other researchers would need in order to apply it to their own data, and take the extra steps needed to make that process as easy as possible^{29,43}. If developing a tool, ensure accessibility, documentation and interoperability with existing ecosystems^{25,44}.

For benchmarking, treat data readiness as a precondition rather than an afterthought: many real biological problems do not yet naturally fit the shape of a well-posed challenge, and curated public datasets therefore remain the most viable foundation.

Future challenges in this space will benefit from stronger participation incentives. This includes prize money, hardware, paper authorship for top entries, and closer alignment with conferences and journals to reach beyond the existing community.

For funding bodies and institutions

The most consistent finding across both Open Calls and Public Challenges is that the limiting resource is shared infrastructure. Recognize fragmented data transfer as an infrastructure gap rather than an individual failure.

Invest in cross-institutional data-transfer infrastructure, in public bioimage repositories^{36,37}, in annotation efforts on community-prioritized datasets, and in the open-source tool maintainers who underpin the entire ecosystem.

Treat this work as strategic capacity building for scientific AI in biology, not as a peripheral service layer. The community is already building these resources⁴⁵, but the question is whether funding structures will recognize them as the frontier work that they are.

Acknowledgements

We would like to sincerely thank the bioimage analysis community for its continuous development of open tools and support via the Image.sc forum. We are grateful to all applicants and participants involved in the Open Calls and Public Challenges initiatives, and to the reviewers involved in the Open Calls: Aitor Gonzalez-Marfil, Anil Kumar Mysore Badarinarayana, Anirban Ray, Anna Klemm, Ashesh, Ben Woodhams, Callum Tromans Coia, Daniel Sage, Dominik Kutra, Elena Buglakova, Gisele Miranda, Kate Gill, Mario Del Rosario, Martim Gamboa, Martin Schätz, Marvin Albert, Pablo Rodríguez Pérez, Thibault Lagache, Varun Kapoor, and Weize Xu, and the participants to the Solvathon Hackathon organized at EMBL: Fynn Beuttenmüller, Jonas Hellgoth, Qin Yu, Craig Russell, Constantin Pape, Anwai Archit, Fariha Mahzabin Annesha, Ashesh, Igor Zubarev, Ankit Roy, and Sushmita Nair. We also thank the Bioimage Analysis Unit, National Facility of Data Analysis and Handling, Human Technopole, and the BioImage Informatics Unit, SciLifeLab, for support, feedback and constructive discussions. We also thank Matthew Hartley, Aybuke Kupcu Yoldas, Teresa Zulueta-Coarasa and the whole BioImage Archive team for their support in uploading datasets. We are grateful to all participants in the Public Challenges, and in particular to Kino Sun (user a897574323) and Yuanchen Wang (user wangyuanchen) for placing in the top 3 and making their training code publicly available.

Author contributions

The Open Calls were organized by VG, MS, BSS, DDN, FJ and JD. OC projects were supported by VG, MS, KL, DDN, AAC, JT, EGdM, MGF, CFB, JMB, CGLdH, and JD, with supervision from EGdM, RH, AK, AK, AMB, AnK, WO, FJ, and JD. Challenges were organized by VG, EG, XinL, AlK and FJ, with support from BSS, AMB, and WO. BS, MC, SYY, GRP, XH, SC, AS, and CP participated in the public challenges, achieved top-3 placements on the leaderboards, and contributed their training code for reproducibility, with BS', MC's, GRP's, and SC's participations supervised by AlK, JD, AMB, and XiaL, respectively. Analysis of OCs and PCs meta-statistics were carried out by VG, MS, BSS and JD. The manuscript was written by VG, MS, BSS, FJ and JD with inputs from co-authors.

Competing Interests

The authors declare no competing interest.

Funding statement

AI4Life has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101057970. E.G.M. acknowledges funding by Fundação para a Ciência e Tecnologia, Portugal (2023.09182.CEECIND/CP2854/CT0004) and the European Union under Horizon Europe (101060346). G.R.P. acknowledges funding from the Ministerio de Ciencia, Innovación y Universidades and the Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) under grants PID2023-152631OB-I00 and AIA2025-164165-C41, and from

the European Regional Development Fund (ERDF), "A way of making Europe". R.H. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Self-Driving4DSR, grant agreement No. 101001332); from the European Union through Horizon Europe (RT-SuperES, grant agreement No. 101099654); from FCT – Fundação para a Ciência e a Tecnologia, I.P., through the MOSTMICRO-ITQB R&D Unit (DOI 10.54499/UID/04612/2025, UID/PRR/4612/2025) and the LS4FUTURE Associated Laboratory (DOI 10.54499/LA/P/0087/2020); from a European Molecular Biology Organization (EMBO) Installation Grant (EMBO-2020-IG-4734); from a joint Wellcome, Chan Zuckerberg Initiative, and Kavli Foundation Essential Open Source Software for Science Cycle 6 award (Wellcome 313383/Z/24/Z; CZI EOSS6-0000000260); and from the La Caixa Foundation (CaixaResearch Health 2025, VirusAwareScopes, HR25-00453). The Chan Zuckerberg Initiative award is made through the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

AI4Life Horizon Europe Programme Consortium

Fynn Beuttenmueller¹, Joran Deschamps², Dorothea Dörr³, Mariana G. Ferreira^{4,5}, Caterina Fuster-Barcelo^{6,7}, Vera Galinova⁸, Carlos Garcia-Lopez-de-Haro^{6,9}, Estibaliz Gómez-de-Mariscal^{4,5,10,11}, Matthew Hartley¹², Ricardo Henriques^{4,5}, Iván Hidalgo-Cenalmor^{4,5}, Florian Jug⁸, Anna Kreshuk¹, Emma Lundberg^{13,14}, Nils Mechtle¹⁴, Arrate Muñoz-Barrutia⁶, Wei Ouyang¹⁴, Constantin Pape^{15,16}, Craig T. Russell¹², Mehdi Seif⁸, Beatriz Serrano-Solano¹⁷, Tomaz Vieira¹, Teresa Zulueta-Coarasa¹².

¹Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

²Bioimage Analysis Unit, National Facility for Data Handling and Analysis, Human Technopole, Milan, Italy.

³Euro-BioImaging ERIC Statutory Seat, Turku, Finland.

⁴AI-driven Optical Biology Laboratory, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal.

⁵Optical Cell Biology Laboratory, Instituto Gulbenkian de Ciência, Oeiras, Portugal.

⁶Neuroscience and Life Sciences Department, Universidad Carlos III de Madrid, Getafe, Spain.

⁷Department of Molecular Life Sciences, Universität Zürich, Zürich, Switzerland.

⁸Computational Biology Research Center, Human Technopole, Milan, Italy

⁹Bioimage Analysis Unit, Pasteur Institute, Paris, France.

¹⁰NOVA Institute for Medical Systems Biology, Universidade Nova de Lisboa, Lisboa, Portugal.

¹¹iNOVA4Health, Programme in Translational Medicine.

¹²European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom.

¹³Department of Pathology, Stanford University, Stanford, CA, United States of America.

¹⁴Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

¹⁵Institute of Computer Science, University of Göttingen, Göttingen, Germany.

¹⁶Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells", University of Göttingen, Göttingen, Germany.

¹⁷Euro-Bioimaging ERIC Bio-Hub, European Molecular Biology Laboratory, Heidelberg, Germany.

References

1. Wilkinson, M. D. *et al.* The fair guiding principles for scientific data management and stewardship. *Sci. data* **3**, 1–9 (2016).
2. Zulueta-Coarasa, T. *et al.* Mifa: Metadata, incentives, formats and accessibility guidelines to improve the reuse of ai datasets for bioimage analysis. *Nat. Methods* **22**, 2245–2252 (2025).
3. Ouyang, W. *et al.* Bioimage model zoo: a community-driven resource for accessible deep learning in bioimage analysis. *BioRxiv* 2022–06 (2022).
4. Goodman, A. *et al.* 2018 data science bowl. <https://kaggle.com/competitions/data-science-bowl-2018> (2018). Kaggle.
5. Caicedo, J. C. *et al.* Nucleus segmentation across imaging experiments: The 2018 Data Science Bowl. *Nat. Methods* **16**, 1247–1253, DOI: [10.1038/s41592-019-0612-7](https://doi.org/10.1038/s41592-019-0612-7) (2019).
6. Winsnes, C. *et al.* Human protein atlas image classification. <https://kaggle.com/competitions/human-protein-atlas-image-classification> (2018). Kaggle.

7. Ouyang, W. *et al.* Analysis of the Human Protein Atlas Image Classification competition. *Nat. Methods* **16**, 1254–1261, DOI: [10.1038/s41592-019-0658-6](https://doi.org/10.1038/s41592-019-0658-6) (2019).
8. Maška, M. *et al.* The Cell Tracking Challenge: 10 years of objective benchmarking. *Nat. Methods* **20**, 1010–1020, DOI: [10.1038/s41592-023-01879-y](https://doi.org/10.1038/s41592-023-01879-y) (2023).
9. Team, C. P. *et al.* Cellmap segmentation challenge, DOI: [10.25378/janelia.28034561.v1](https://doi.org/10.25378/janelia.28034561.v1) (2024).
10. Kauffmann, D. *et al.* Light my cells: Bright field to fluorescence imaging challenge 2024. International Symposium on Biomedical Imaging (ISBI 2024), Zenodo, DOI: [10.5281/zenodo.10687569](https://doi.org/10.5281/zenodo.10687569) (2024).
11. Graham, S. *et al.* Conic challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting. *Med. Image Analysis* **92**, 103047, DOI: [10.1016/j.media.2023.103047](https://doi.org/10.1016/j.media.2023.103047) (2024).
12. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. methods* **19**, 1634–1641 (2022).
13. Stirling, D. R. *et al.* Cellprofiler 4: improvements in speed, utility and usability. *BMC Bioinforma.* **22**, 433 (2021).
14. Bankhead, P. *et al.* Qupath: Open source software for digital pathology image analysis. *Sci. Reports* **7**, 16878 (2017).
15. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* (2019).
16. Arzt, M. *et al.* Labkit: Labeling and segmentation toolkit for big image data. *Front. Comput. Sci.* **4**, DOI: [10.3389/fcomp.2022.777728](https://doi.org/10.3389/fcomp.2022.777728) (2022).
17. napari contributors (2019). napari: a multi-dimensional image viewer for python. <https://napari.org/stable/>, DOI: [10.5281/zenodo.3555620](https://doi.org/10.5281/zenodo.3555620).
18. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
19. Weigert, M. & Schmidt, U. Nuclei instance segmentation and classification in histopathology images with stardist. In *The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, DOI: [10.1109/ISBIC56247.2022.9854534](https://doi.org/10.1109/ISBIC56247.2022.9854534) (2022).
20. Krull, A., Buchholz, T.-O. & Jug, F. Noise2void - learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).
21. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics* **18**, 529 (2017).
22. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
23. Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of calcium images by learned contrast synthesis. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, 1123–1126 (IEEE, 2016).
24. Tinevez, J.-Y. *et al.* Trackmate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90, DOI: [10.1016/j.jymeth.2016.09.016](https://doi.org/10.1016/j.jymeth.2016.09.016) (2017). Image Processing for Biologists.
25. Levet, F. *et al.* Developing open-source software for bioimage analysis: opportunities and challenges. *F1000Research* **10**, 302 (2021).
26. Kirillov, A. *et al.* Segment anything. *arXiv:2304.02643* (2023).
27. Ravi, N. *et al.* Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024).
28. Seifi, M. *et al.* Featureforest: the power of foundation models, the usability of random forests. *npj Imaging* **3**, 32 (2025).
29. Laine, R. F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nat. methods* **18**, 1136–1144 (2021).
30. Nogare, D. D., Hartley, M., Deschamps, J., Ellenberg, J. & Jug, F. Using ai in bioimage analysis to elevate the rate of scientific discovery as a community. *Nat. methods* **20**, 973–975 (2023).
31. Bajcsy, P. *et al.* Enabling global image data sharing in the life sciences. *Nat. methods* **22**, 672–676 (2025).
32. Schmied, C. *et al.* Community-developed checklists for publishing images and image analyses. *Nat. Methods* **21**, 170–181 (2024).
33. Salmon, B. & Krull, A. Unsupervised denoising for signal-dependent and row-correlated imaging noise. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, 2379–2389 (2025).
34. Rey-Paniagua, G., Lachowski, D. & Muñoz-Barrutia, A. Nafnet-gan: An adversarial framework for supervised fluorescence microscopy denoising. In *2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI)*, 1–5, DOI: [10.1109/ISBI61048.2026.11515733](https://doi.org/10.1109/ISBI61048.2026.11515733) (2026).

35. Zamir, S. W. *et al.* Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739 (2022).
36. Williams, E. *et al.* Image data resource: a bioimage data integration and publication platform. *Nat. methods* **14**, 775–781 (2017).
37. Hartley, M. *et al.* The bioimage archive—building a home for life-sciences microscopy data. *J. Mol. Biol.* **434**, 167505 (2022).
38. Moore, J. *et al.* Ome-ngff: a next-generation file format for expanding bioimaging data-access strategies. *Nat. methods* **18**, 1496–1498 (2021).
39. Image.sc, scientific community image forum is a discussion forum for scientific image software sponsored by the center for open bioimage analysis (coba). <https://forum.image.sc/>.
40. Corbat, A. A. *et al.* Globias: strengthening the foundations of bioimage analysis. *Nat. Methods* 1–2 (2026).
41. Foster, I. Globus online: Accelerating and democratizing science through cloud-based services. *IEEE Internet Comput.* **15**, 70–73 (2011).
42. Allen, B. *et al.* Software as a service for data scientists. *Commun. ACM* **55**, 81–88 (2012).
43. Lambert, T. & Waters, J. Towards effective adoption of novel image analysis methods. *nature methods* **20**, 971–972 (2023).
44. Deschamps, J., Dalle Nogare, D. & Jug, F. Better research software tools to elevate the rate of scientific discovery or why we need to invest in research software engineering. *Front. Bioinforma.* **3**, 1255159 (2023).
45. Kemmer, I. *et al.* Building a fair image data ecosystem for microscopy communities. *Histochem. Cell Biol.* **160**, 199–209 (2023).

Supplementary Information

Vera Galinova^{1,*}, Mehdi Seifi^{1,*}, Beatriz Serrano Solano², Kristína Lidayová³, Damian Dalle Nogare⁴, Agustín Andrés Corbat^{3,5}, Joshua Talks⁶, Edoardo Giacomello¹, Estibaliz Gómez-de-Mariscal^{7,8,9,10}, Mariana G. Ferreira^{7,8}, Caterina Fuster-Barceló^{11,12}, Juan Manuel Battagliotti⁴, Carlos García-López-de-Haro^{11,13}, Benjamin Salmon^{1,14}, Melisande Croft⁴, Si Young Yie¹⁵, Guillermo Rey-Paniagua¹¹, Xiaotian Hu¹⁶, Sungjun Cho¹⁷, Aagam Sheth¹⁸, Chhayansh Porwal¹⁹, Xiaomeng Li¹⁷, AI4Life Consortium[†], Ricardo Henriques^{7,8}, Xinyang Li²⁰, Alexander Krull¹⁴, Anna Klemm³, Arrate Muñoz Barrutia¹¹, Anna Kreshuk⁶, Wei Ouyang²¹, Florian Jug^{1,+}, and Joran Deschamps^{4,+}

¹Computational Biology Research Center, Human Technopole, Milan, Italy.

²Euro-Bioimaging ERIC Bio-Hub, European Molecular Biology Laboratory, Heidelberg, Germany.

³Biolmage Informatics Unit, Science for Life Laboratory, National Bioinformatics Infrastructure Sweden, Uppsala University, Uppsala, Sweden.

⁴Bioimage Analysis Unit, National Facility for Data Handling and Analysis, Human Technopole, Milan, Italy.

⁵Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

⁶Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

⁷AI-driven Optical Biology Laboratory, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal.

⁸Optical Cell Biology Laboratory, Instituto Gulbenkian de Ciência, Oeiras, Portugal.

⁹NOVA Institute for Medical Systems Biology, Universidade Nova de Lisboa, Lisboa, Portugal.

¹⁰iNOVA4Health, Programme in Translational Medicine.

¹¹Neuroscience and Life Sciences Department, Universidad Carlos III de Madrid, Getafe, Spain.

¹²Department of Molecular Life Sciences, Universität Zürich, Zürich, Switzerland.

¹³Bioimage Analysis Unit, Pasteur Institute, Paris, France.

¹⁴School of Computer Science, University of Birmingham, Birmingham, UK.

¹⁵Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, South Korea.

¹⁶School of Biological Science and Medical Engineering, Beihang University, Beijing, China.

¹⁷Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

¹⁸School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, India.

¹⁹Rajasthan Technical University, Kota, India.

²⁰College of AI, Tsinghua University, Beijing, China.

²¹Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

[†]A list of members and their affiliations appears at the end of the paper.

⁺florian.jug@fht.org, joran.deschamps@fht.org

^{*}these authors contributed equally to this work

S1 Open Calls organization

S1.1 Process description

In this section, we describe the process organization towards which we converged after three OCs. The OCs ran for three consecutive years (2023, 2024, and 2025), following similar structures: (1) application step, (2) reviewer recruitment, (3) reviewing process, (4) project ranking, (5) consultation and project selection, and (6) project support. This workflow aimed at identifying, evaluating, and supporting bioimage analysis projects suitable for AI-based solutions. The process was designed to balance accessibility for life scientists with rigorous technical assessment, while ensuring efficient use of expert resources. It combined structured applications, anonymised review, expert consultation, and hands-on project support, with an emphasis on project readiness, feasibility, and reusability of outcomes. The structure was continuously refined and adjusted throughout the calls based on the previous year experience, leading to a process organization that can serve as a reference for future open calls. For each selected project, a team of deep learning and image analysis experts built a suitable pipeline to solve the challenges faced by the researchers, pipeline which was then released online according to the FAIR principles together with the data used in the project.

Application. In the first step of the process, applicants submitted a standardized application consisting of a brief project description and representative image data, including annotations when available. See Appendix A1 for OC3 application form. A common presentation template of four slides and an example project were used to ensure consistency across applications and to facilitate rapid assessment of project goals, data characteristics, and existing workflows by the reviewers (see Fig. S1). Guidance was provided on terminology and ground truth quality with an illustrative figure in the application process, helping applicants describe their data and challenges accurately, and improving the reliability of reviewer assessments. All applications underwent an initial pre-screening by the main organizers to exclude submissions that were incomplete or insufficiently documented. This step focused on the clarity and completeness of the application, ignoring scientific merit, and ensuring that only projects suitable for external review progressed further.

The figure displays a four-slide presentation template for project applications. Each slide has a header bar with the AI4Life logo and contact information.

- Slide 1:** Project title
- Slide 2:** Project aim
 - Background:
 - Goal of the study:
 - Current challenges:
- Slide 3:** Data example
 - Relevant image(s):
 - Useful information:
- Slide 4:** Current analysis workflow
 - Steps:
 - 1.
 - 2.
 - 3.
 - 4.
 - 5.
 - Desired outcome:

Figure S1. Project template for applications in OC 2 and 3. Each application was requested to upload a presentation in the format of their choice (.pdf, .pptx, .key, etc.) following this template to standardize the information reviewed by reviewers.

Review. During the application window, we recruited reviewers both internally to AI4Life and in the community of bioimage analysts. Advertisement for registration as reviewers were sent by emails to image analysis facilities, posted at conferences (e.g. I2K) and advertised on the image.sc forum¹. Once the pool of reviewers was assembled, eligible project applications were evaluated through anonymized peer review using predefined scoring criteria. Reviewers were guided to assess project readiness, applicability of deep learning, feasibility within available effort, and scientific excitement, and suitability for the aim of the OCs (see Section S1.2). Measures were taken to prevent reviewing errors, including controlled project assignment and consistency checks, and to identify systematic reviewer bias. Review scores were aggregated using a simplified

and automated scoring scheme to generate a ranked list of applications. A score cut-off was applied to shortlist projects moving to the next selection phase.

Consultation phase. In OC 2 and 3, shortlisted projects were invited to a consultation phase consisting of discussions with a panel of image analysts and AI experts. These consultations were used to verify data readiness, clarify technical challenges, identify appropriate tools, and assess feasibility. When necessary, a second consultation was organized to allow the panel of experts to come to a shared understanding of the project's current state. The consultation phase panels were composed of members of AI4Life partners' institutes and external image analysts. We ensured that typically three or four experts were present for each session, with the aim of pairing expert domain knowledge with relevant projects.

Project selection. After the consultation phase, all experts who had participated in the various meetings with the applicants assembled and discussed the projects. Written feedback from the experts informed the final selection, which was performed by adjusting the technical feasibility and readiness scores from reviewers. Top-scoring projects were selected in order up to the maximum capacity for support, and in compliance with the requirements of transnational access imposed by the Horizon Europe funding.

Project support. Selected projects entered a support phase formalized through signed agreements. Data transfer strategies were defined early to avoid delays, and open data release plans were developed in alignment with FAIR principles. AI4Life experts provided hands-on support to develop reproducible, open-source analysis pipelines. Progress was monitored through regular meetings that addressed data acquisition, labeling needs, pipeline usability, and coordination between experts, with early user involvement to facilitate adoption beyond the support period.

Project conclusion. At the end of each project, a description of the analysis pipeline was uploaded to GitHub and Zenodo, alongside any custom code or tool developed for the need of the project. The GitHub repositories served as basis for transferring the pipeline to users. When possible, data used in the project was also released on the BioImage Archive². Finally, all compatible fine-tuned or de-novo trained deep learning model were saved in the BioImage Model Zoo format and uploaded to the BioImage Model Zoo repository³.

S1.2 OC3 project ranking

We appointed an international team of expert reviewers who helped us to evaluate project applications according to the following criteria: readiness of the project, applicability of deep-learning and excitement about the project. Reviews were sent via a form (see Appendix A2).

All OCs had different criteria, here we only describe OC3 ranking process as these were the criteria to which we converged. The following criteria were aggregated across all reviewers for each project.

1. Average readiness: $\mu_r \in [0, 5]$
2. Average DL applicability: $\mu_{dl} \in [0, 5]$
3. Average Excitement: $\mu_{ex} \in [0, 5]$
4. Sum of AI4Life relevance: $s_{ai} \in \{0, 1, 2\}$
5. Sum of Consultation relevance: $s_{co} \in \{0, 1, 2\}$

A general score was computed as the sum of the reviewing criteria weighted by the reviewer's belief that the project was a good AI4Life project or would greatly benefit from consultation:

$$S_i = (\mu_r(i) + \mu_{dl}(i) + \mu_{ex}(i)) * (s_{ai}(i) + s_{co}(i))$$

The projects were ranked by their S score and projects qualifying for trans-national access (as per the Horizon Europe rules) were selected for consultation. Projects not qualifying for trans-national access were selected up to the quote allowed by the funding constraints.

S1.3 Methods

The entire Open Calls process was carried out using Google Forms for the project applications, applications to be a reviewer, and review submission. Applications for supports were saved with a unique ID in a Google Drive folder accessible only to the main organizers. The data was automatically anonymized with an Apps Script and copied to a reviewer-only accessible folder. Each reviewer had only access to the relevant project applications. Scores were aggregated in a Google Sheet, in which the processing was performed.

S2 Open Calls outcomes

S2.1 Applications statistics

The OCs received 151 applications throughout the three calls (OC 1: 72, OC 2: 51, and OC 3: 28) from 28 countries worldwide, spanning all inhabited continents. Most submissions originated from European institutions, particularly from Germany (27 applications), Portugal (20), Italy (14), France (10), and the United Kingdom (10) (see Fig. S2a), while substantial participation was observed from the United States (16). Of these 151 applications, 32 projects underwent consultations (OC2: 22, OC3: 12) and 22 were ultimately selected for support (OC 1: 8, OC 2: 7, and OC 3: 7). Due to transnational access constraints from the funding body, the number of consultations and project supports was capped for non-EU countries. The distribution of home base countries for the selected projects was coherent with the overall applications (see Fig. S2a, light bars). EU countries with the most selected projects were Germany (6), France (3), Portugal (2), Italy (2) and the Netherlands (2). Three projects from non-EU countries were selected, originating from the United States (2), Australia (1) and Canada (1).

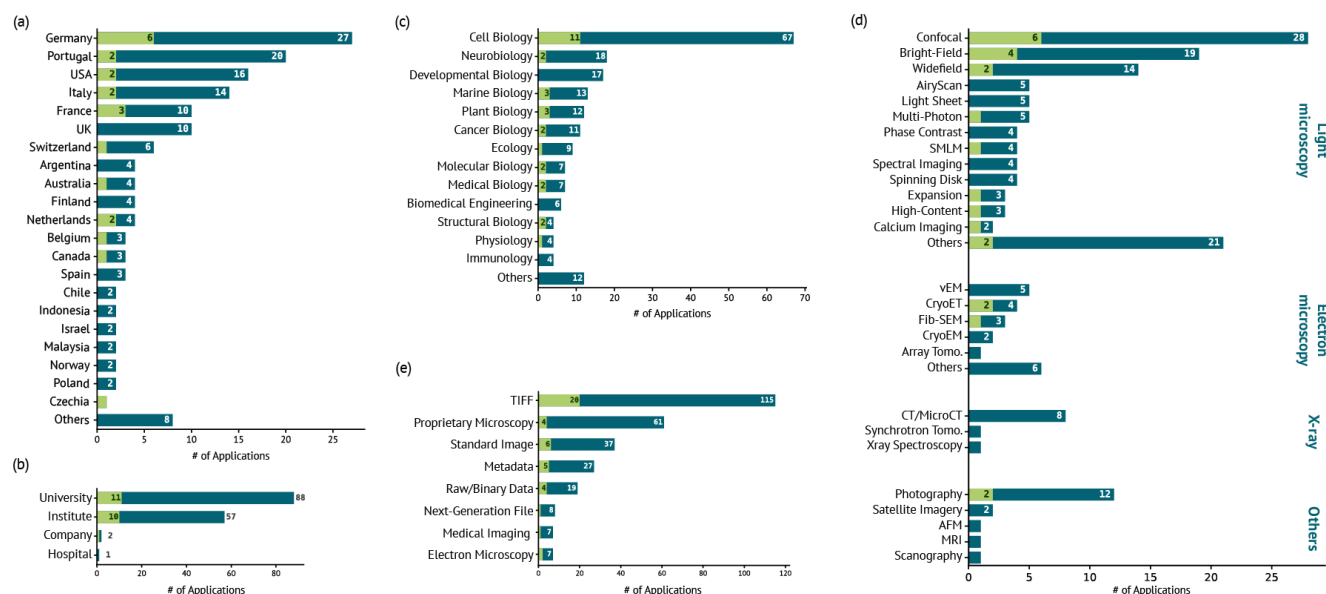


Figure S2. (a) Countries in which applications (dark) and selected projects (light) home institutions are based. Others correspond to single applications from Austria, Denmark, Hungary, Mexico, Serbia, Slovakia, South Africa, and Sweden. (b) Type of organization from which applications (dark) and selected projects (light) originated. (c) Scientific domains of applications (dark) and selected projects (light). Note that each application could declare multiple scientific domains. Others correspond to Agronomy, Biophysics, Genetics, Histology, Morphology, Pathology, Remote Sensing, and Virology. (d) Imaging methods of the applications (dark) and selected projects (light) by categories: light microscopy (LM), electron microscopy (EM), X-ray, and others. "Others" category in EM and LM includes projects for which the particular method was not explicit (EM and LM), or was mentioned by a single application only (LM). "LM Others" includes STED, OCT, ODT, and flow cytometry (selected project). Selected CLEM project was counted in both LM (wide-field) and EM (Others). (e) Data formats, as declared during applications (dark) and for the selected projects (light).

Applicants showed a balanced gender distribution (statistics only collected during OC2 and OC3), with a slight predominance of female applicants (47 females, 30 males), while a small number reported non-binary gender or preferred not to disclose (2). Applicants represented a wide range of nationalities (see Fig. S3), predominantly European (53 out of 79 applicants). Note that nationality and gender information were not collected for Open Call 1.

Applicants home institutions were overwhelmingly universities (88 applications, of which 11 were selected) and research institutes (57 applications, 10 selected), as shown in Fig. S2b, and few applications came from companies (2 applications, 1 selected) or hospitals (1 application).

Submitted projects covered a broad range of scientific domains, with a strong concentration in cell biology and adjacent fields. The most frequent domains declared during applications (see Fig. S2c, dark bars) included cell biology (67 applications), neurobiology (18), developmental biology (17), marine biology (13), plant biology (12), and cancer biology (11), reflecting both core bioimaging communities and applied life-science use cases. Less frequent submissions spanned specialized areas such as structural biology, physiology, or immunology. Note that applications could declare multiple scientific domains. Once again, selected projects were sampled appropriately from these domains (see Fig. S2c, light bars), with a majority of them

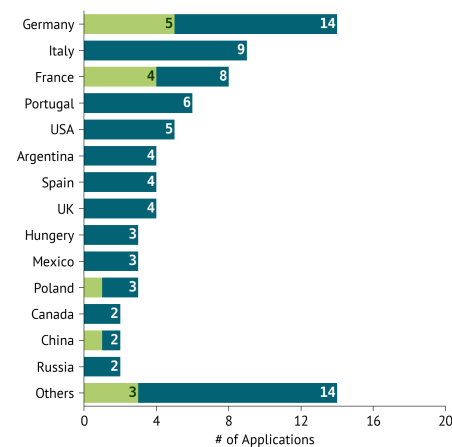


Figure S3. Self-declared applicant nationalities collected during OC 2 and 3. "Others" category consists of Chile, Croatia, Czech, Greece, Indonesia, Ireland, Israel, Japan, Malaysia, Netherlands, New Zealand, Sweden, Switzerland, and Turkey. EU nationals accounted for 53 out of 79 applicants (some applicants were multinationals).

belonging to cell biology (11 projects), and spanning most declared scientific domains. Structural biology, while being declared by only 4 applications, was well represented with 2 selected projects.

S2.2 Data statistics

We extracted the source imaging approaches from the applications, with applications potentially declaring multiple ones. Applications were predominantly using light microscopy (121 applications, see Fig. S2d), far more than electron microscopy (21) and X-ray (10). Other types of techniques (17 applications) included natural images and photography (12), mostly by projects from ecology, plant biology, and agronomy domains, and few applications with satellite images (2), Scanography (1), AFM (1), or MRI (1). Within light imaging, widespread imaging methods dominated the applications, with confocal (28 out of 121 applications), bright-field (19), or widefield (14). More recent and specialized techniques, such as Airyscan, light sheet, multi-photon, or single-molecule localization, were used by a few applications (4 to 5 applications). The rest of the methods covered a wide range of the microscopist toolbox. For many applications, the particular imaging method was not explicit and was therefore grouped in the "Others" category in light microscopy, together with single occurrences (STED, OCT, ODT, flow cytometry). Electron microscopy methods covered the most common methods available to scientists (e.g. vEM, cryoET, cryoEM, FIB-SEM). Several applications only mentioned scanning or transmission electron microscopy, and were grouped together with undefined EM methods ("Others", see Fig. S2d). X-ray imaging or spectroscopy was largely dominated by computed tomography, with 8 applications out of 10. None of the X-ray applications was selected.

For all applications, applicants could declare image dimensionalities and data formats. There were slightly more applications declaring 2D (99 out of 151 applications) than 3D (83). Multi-channel imaging concerned 63 applications, and time-series were limited to only 40 applications out of the total. Note that, here again, each application could select multiple categories.

Finally, TIFF format was by far the most common declared data type involved in the application (115 out of 151 applications, see Fig. S2e), before proprietary microscopy formats (61), or standard non-scientific image formats (37). Next-generation file formats, such as HDF5 or Zarr, were declared in only 8 applications in total. Here again, multiple options could be selected by each application.

S2.3 Tool statistics

Applicants had the opportunity to mention the tools that they tried or use for image analysis in the context of their project. While in OC 1 tools were mentioned as free text, we added tools and tools categories suggestions to OC 2 and 3 (see section A1). In Fig. S4a, we aggregated the answers from all OCs into the categories used in OC 2 and 3. As expected from a call dominated by life sciences applicants, ImageJ/Fiji^{4,5} was the most cited tool (116 out of 151 applications). Scripts ("ImageJ macro, Python, R, Matlab, etc.") were cited in 60 applications. Note that ImageJ/Fiji macros were categorized in scripts, thus underestimating the use of ImageJ/Fiji by the applicants. The next category was commercial analysis software ("Zeiss, Arivis, Imaris, Amira, Huygens, etc.") with 41 applicants declaring using them. Additional commercial tools used for statistical analysis and plotting, namely GraphPad Prism and Excel, were mentioned in 25 applications. Specialized libraries for machine learning (PyTorch, TensorFlow, scikit-learn, etc.) were cited 19 times. Interestingly, popular tools in the image analysis community, such as ilastik⁶ (15), Cellpose^{7,8} (14), napari⁹ (14), CellProfiler¹⁰ (12), QPath¹¹ (8), Labkit¹² (4), or Icy¹³ (2), were cited less than 15 times

each. Tools allowing training a deep learning network and specifically focused on bioimages were scarce, mostly Cellpose (14) and StarDist¹⁴ (7). Finally, 29 projects declared using other tools. A complete list of the individual tools cited in OC 1 can be found in (see Fig. S5a). The categorized answers from OC 2 and 3 are shown in Fig. S5b, including all individual tools from the "Others" category.

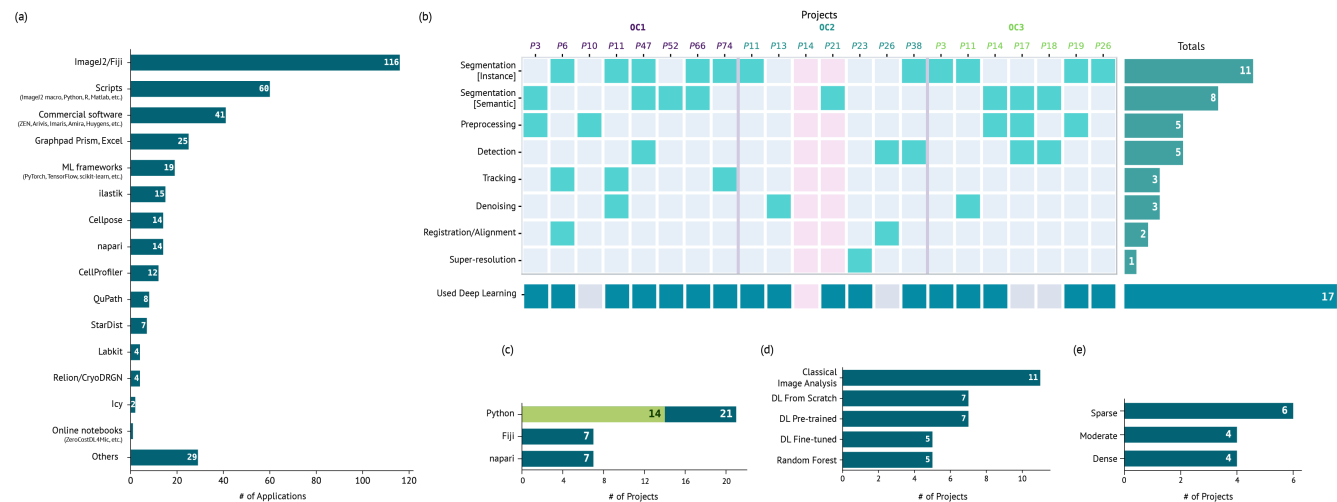


Figure S4. (a) Tools and tool categories declared in the applications as used or tested for the purpose of the project. The "Others" category was available at application time for users to specify additional tools. (b) Discrete analysis steps and their occurrences in every project selected for the three OCs. Red columns indicate failed projects. The right-hand side plot shows the distribution. We indicate in the bottom row whether the project used deep learning. (c) Number of projects using one of three software ecosystems (Python, Fiji, napari). Dark bars indicate to the number of project pipelines partially carried out on the corresponding ecosystem, while green bars denote the number of project pipelines entirely performed in that ecosystem. Project pipelines could make use of multiple software ecosystems. (d) Number of projects using classical image analysis, random forest classification (shallow learning), and various deep-learning flavours (training a model from scratch, using a pre-trained model, and fine-tuning a pre-trained model). Project pipelines could use a mix of these categories. (e) Number of projects requiring various degrees of labeling effort (sparse, moderate, and dense labeling). Sparse labeling corresponds to squiggles typically drawn for training random forests or to point annotations, while moderate labeling to adding a few masks for Cellpose fine-tuning, and dense to complete annotation of a dataset for de-novo training.

S2.4 Selected projects

Once projects were selected, expert image analysts were tasked with developing a FAIR pipeline to solve the challenges encountered by the applicants, while ensuring that they could use it in the future. Summary of selected projects can be found in Table S8. In the next sections, we provide an analysis of the pipelines developed for these projects. A description of the resulting pipelines, and links to code and data repositories can be found in Table S9.

S2.5 Project Pipelines

The pipelines converged to a limited set of recurring image analysis tasks (see Fig. S4b), dominated by segmentation (19 out of 22 pipelines), pre-processing (5), and detection (5), sometimes combined with denoising (3), tracking (3), or registration/alignment (2). One project focused on super-resolution only. We excluded final quantification as it was the end goal of most projects. Each pipeline relied on different tools and scripts, highlighting the heterogeneity of the analysis goals and the specialization of image analysis tools, as well as preferences of the analysts themselves. The implemented solutions relied on a mix of widely adopted community tools, domain-specific tools and project-specific developments. See Fig. S6 for a breakdown of the various tools and programming/scripting languages used in each image analysis task. All tools used in the OCs are listed in Fig. S7.

All projects used either Python or ImageJ/Fiji (see Fig. S4c), via custom scripts or tools, or both together in heterogeneous pipelines. Out of the 22 projects, 21 used Python-based tools, of which 14 solutions were Python only (including Python-based tools, light green bars in Fig. S4c). ImageJ/Fiji was used in 7 projects, but none of the pipelines solely consisted of steps in the ImageJ/Fiji ecosystem. In one project, the core of the pipeline was carried out in ImageJ/Fiji, and the Python component was an optional labeling step. napari was also used in 7 projects, 5 times for deep learning related plugins (FeatureForest¹⁵, nnInteractive¹⁶, Noise2Void¹⁷ plugin), one time to train a random forest classifier (APOC¹⁸), and once for image visualization

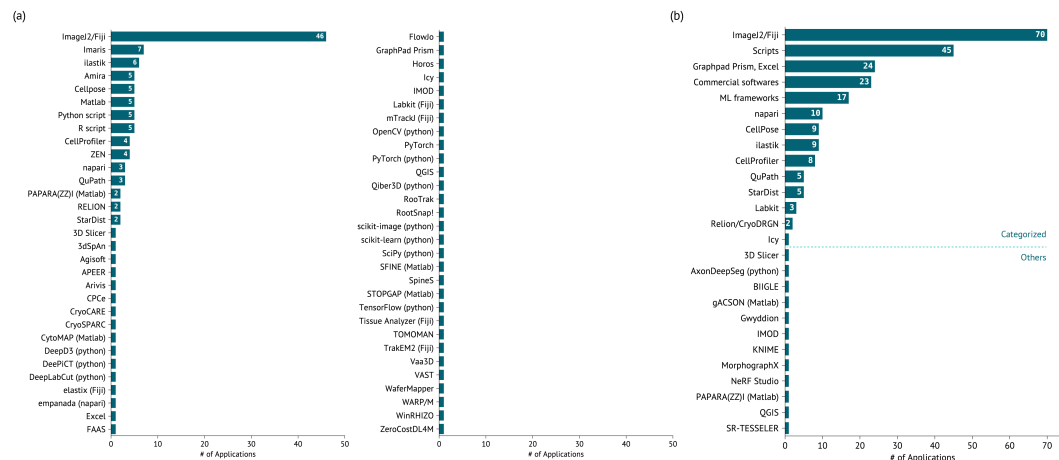


Figure S5. (a) Tools mentioned by users in their OC 1 application as free text. (b) Individual tools and categories of tools selected by users in their OC 2 and 3 application. Additional tools could be mentioned by users as free text (see line separation for "Others").

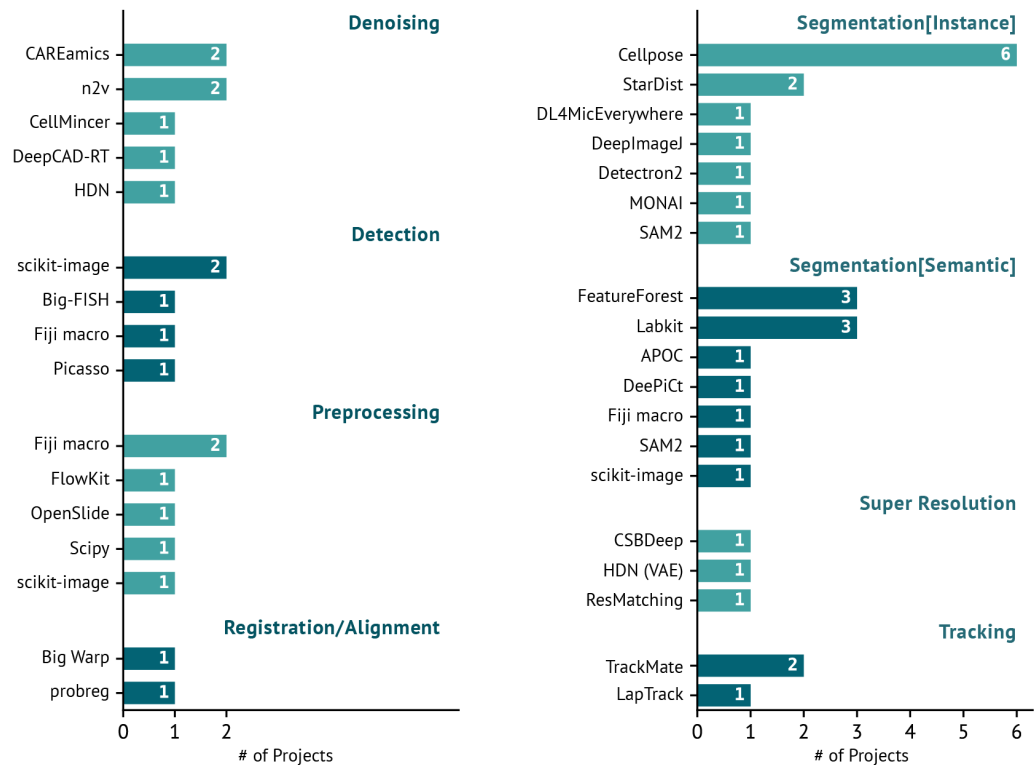


Figure S6. The various tools and libraries used in the project pipelines by analysis steps, to the exclusion of labeling, post-processing and quantification. Project pipelines could make use of several tools in a single analysis step. A project compared denoising algorithms in CAREamics (Noise2Void, Noise2Noise), DeepCAD-RT, CellMincer and HDN. The two instance segmentation steps using StarDist were carried out in DeepImageJ and DL4MicEverywhere. All super-resolution tools were used in a single project.

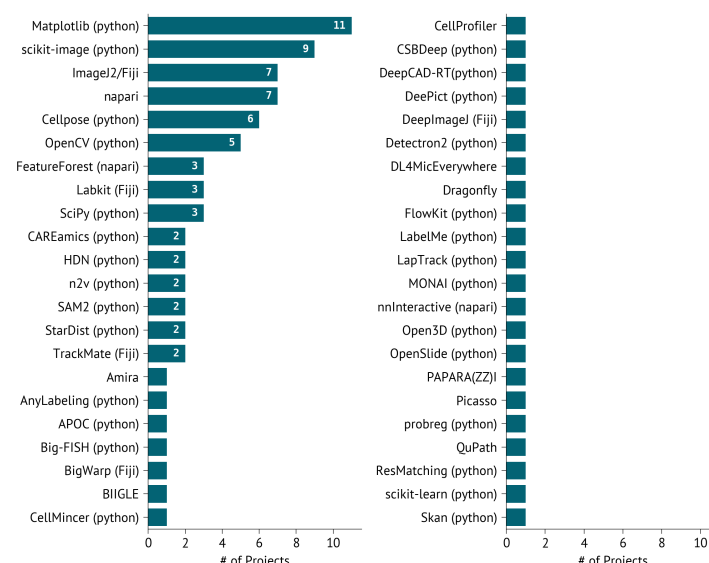


Figure S7. Summary of tools used in the project pipelines.

(OC2-38). A majority of projects required custom scripts to process images or call specialized tools and libraries. The pipelines delivered Jupyter notebooks (13 projects), Python scripts (7 projects), and ImageJ2/Fiji macros (4 projects). Finally, 2 projects delivered containers to the users.

The application of deep learning to bioimages was the focus of the OCs. In the majority of projects, some flavour of deep learning was applied (17 out of 22, see Fig. S4b), while 5 projects also used shallow machine learning (random forests) and 11 projects additionally performed some classical image analysis explicitly (filtering, watershed, morphological operations, etc.). A total of 6 projects employed solely deep learning, without shallow learning or classical image analysis. Random forests pixel classification, without any deep-learning generated features, was mostly carried out in Labkit¹² (4 projects) and one time with APOC¹⁸. Classical image analysis was essentially performed in Python (10 projects out of 11), with a mix of scikit-image¹⁹ (9), OpenCV²⁰ (3), SciPy¹⁹ (2), and Skan²¹ (1), and one time purely in ImageJ2/Fiji. In 3 projects (OC2-26, OC3-17, and OC3-18), the entire pipeline consisted of classical image analysis only, without shallow or deep learning steps.

Some pipeline steps did not involve deep-learning. Pre-processing was performed with ImageJ2/Fiji macros in 2 projects, and with various Python libraries (FlowKit²², OpenSlide²³, SciPy²⁴, scikit-image) in three others. Detection was performed using classical image analysis with domain-specific tools (Big-FISH²⁵, Picasso²⁶), Python scripts using scikit-image (2 projects), or ImageJ2/Fiji macros (1). Registration and alignment were also performed using BigWarp²⁷ and probreg²⁸. Finally, the last step not performed using deep learning was tracking, although the generation of the instances used for tracking was obtained using deep learning methods. Tracking was present in 3 pipelines and was carried out using TrackMate²⁹ (2 projects) or LapTrack³⁰ (1).

As expected, deep learning was the approach of choice in denoising, segmentation (instance and semantic) and super-resolution. Among the various pipelines that involved deep learning, pre-trained models were used in 7 projects (Cellpose and SAM2³¹, either directly or via its use within FeatureForest), while 5 projects fine-tuned models (Cellpose) and 7 pipelines trained models entirely from scratch. De-novo trained models were either trained using an existing tool (CAREamics³², CellMincer³³, DeepCAD-RT³⁴, DeePiCt³⁵, HDN³⁶, Noise2Void¹⁷, StarDist¹⁴) or custom scripts (3 projects). Denoising was the main focus of project OC2-13 that compared various approaches for the denoising of calcium imaging, leading to the application of CellMincer, DeepCAD-RT, HDN, Noise2Noise³⁷ and Noise2Void¹⁷, where the latter two were trained within CAREamics. Two other projects applied denoising using Noise2Void (using the original n2v tool and the more recent CAREamics), demonstrating that it remains the most accessible denoising deep learning-based algorithm. Instance segmentation was dominated by Cellpose (6 out of 11 projects), specifically Cellpose2⁸. StarDist was used twice, once from within DeepImageJ³⁸ via CellProfiler¹⁰ and the other using DL4MicEverywhere³⁹. The other instance segmentation approaches were a custom training of Detectron2, of a U-Net⁴⁰ from Monai⁴¹, and an application of SAM2³¹. Semantic segmentation with deep learning was either carried out using random forests trained from the embeddings of SAM2 (FeatureForest¹⁵, 3 projects), or a de-novo trained model (DeePiCt, 1 project). Other semantic segmentation steps were done using random forests directly (Labkit¹² in 3 projects, APOC¹⁸ jointly in one of these projects). Finally, the super-resolution project compared several methods: CARE⁴² (using the CSBDeep library), ResMatching⁴³ and the variational auto-encoder (VAE) from HDN.

While the deep learning methods used in the OC pipelines often differ from each other by their purpose and features, the majority of the methods were UNet-based (13 out of 17, see Fig. S8), with 4 visual transformers-based approaches (SAM2-derived) and two other architectures (Mask R-CNN, variational auto-encoder).

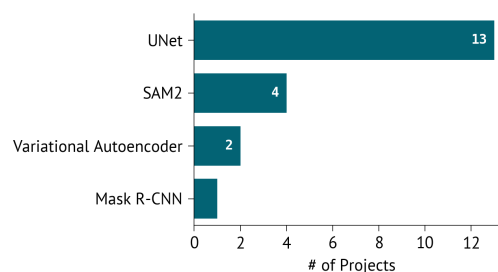


Figure S8. Deep learning architectures or framework used in the OC pipelines. Note that in one project, SAM2 encoder was used both in FeatureForest and in a hybrid SAM2+UNet model.

As a consequence of the prevalence of segmentation in the analysis steps, 13 of the 22 projects required annotations (see Fig. S4e). Since gold-standard annotations were scarcely available (see next section), methods requiring a sparse or moderate amount of annotations were often adopted (6 out of 22). That was specifically the random forests-based pipelines, Labkit and FeatureForest, which required sparse squiggles-like annotations, and the point labels used for SAM2. Cellpose fine-tuning only necessitated a few additional masks, corresponding to a moderate labeling effort (4 projects). De-novo training of a deep learning network for segmentation purposes, on the other hand, required dense annotations (5 projects). A single project had enough gold standard annotation at the start of the project to train a deep neural network (OC3-14), albeit only for one class out of the four originally expected to be segmented. Annotations were either performed in the segmentation tools directly (Labkit, APOC, FeatureForest, Cellpose), in software with annotation tools (napari⁹, BIIGLE⁴⁴, Amira⁴⁵, Dragonfly⁴⁶, LabelMe⁴⁷, AnyLabeling⁴⁸), or with deep learning tools (nnInteractive¹⁶). Of the aforementioned tools, Labkit is a ImageJ2/Fiji plugin, while APOC, FeatureForest, and nnInteractive are napari plugins. In two cases, user annotations had already been generated in PAPARA(ZZ)I⁴⁹ and QuPath¹¹, and these tools were used to export the annotations.

Finally, image visualization was often performed with ImageJ2/Fiji (5 projects) or napari (6). Many Python scripts and notebooks used Matplotlib⁵⁰ to visualize results (11 projects), and one notebook employed Bokeh⁵¹. Domain-specific tools were also used to check results, with the example of BIIGLE⁴⁴ (1 project), CellProfiler¹⁰ and Picasso²⁶ (1).

S2.6 Project readiness and outcome analysis

Only 8 out of 22 selected projects were fully ready at the start of the support period, with complete, usable data and sufficient annotations available to start with. The remaining projects required some form of intervention before or during execution: additional data acquisition, re-imaging, or further ground-truth annotation.

Data shareability was assessed at the application stage across all 151 applicants: 93 indicated their data was fully shareable, 53 partially shareable, and only 5 not shareable (see Fig S9). In OC 2 and 3, the option to declare the data as non-shareable was not available as it was a mandatory outcome of the OC projects.

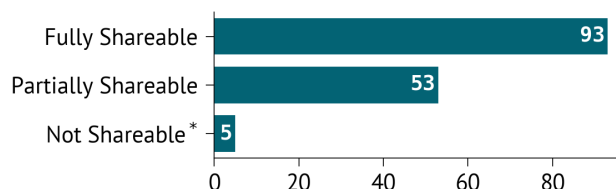


Figure S9. Portion of their data that is shareable as declared by the applicants. "Not Shareable" was only available as an answer in OC 1.

Data transfer to the image analysis team was a recurring bottleneck, with 3 out of 22 projects experiencing significant delays. Delays arose both from applicant-side constraints, including slow uploads and late data submission, and from technical limitations, such as the absence of a unified data-transfer solution across institutions. Across 22 projects, 14 different data-

transfer methods were used, ranging from commercial cloud services to institutional platforms and one-time physical hard drives sent via postal services (see [S10](#)), highlighting the fragmentation of current data-sharing practices.

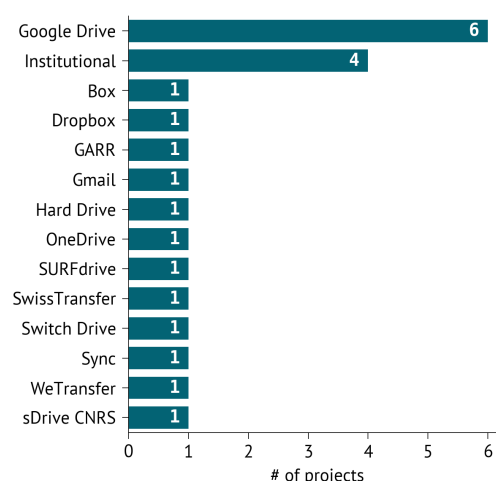


Figure S10. Data transfer methods used during project support as declared by the image analysts. Institutional refers to any service provided by the home institution of the applicants or the analysts. "Hard Drive" refers to a physical hard drive sent by the post.

All projects were assessed at the end of the project by the support team experts for data readiness for deep learning, taking into account annotation quantity, quality, and coverage. Only 2 out of 22 projects were considered to have sufficient ground-truth data for deep learning. Among the others, 3 projects had no ground-truth annotations available at the start of the support period, and 3 further projects required substantially more annotation than originally provided during project runtime to reach the intended analysis objectives. In 4 out of 22 projects, data quality was lower than initially expected, causing delays and requiring corrective steps due to corruption, saturation, imaging noise, or other acquisition issues. All readiness and quality assessments reported here were filled out by the experts directly involved in supporting each project.

A small number of projects (2 out of 22) did not achieve the initially defined objectives by the end of the support period. In these cases, outcomes were limited by factors such as insufficient data readiness, overly ambitious or evolving project scope, or technical issues encountered during execution.

All projects pipelines, except for two failed projects (OC2-14 and OC2-21), one under publication embargo (OC2-23), and one currently in its final steps (OC3-26), were uploaded to GitHub with scripts, notebooks, or detailed descriptions on how to reproduce the pipeline. All GitHub repositories were also archived to Zenodo. For 12 of the projects, data was uploaded to the BioImage Archive, while 4 were uploaded to Zenodo (see [Fig. S12](#)). Illustrating the difficulty in transferring data, track metadata and incentivize users in publicly sharing it, 2 projects are still awaiting data upload. 4 projects had no useful data to upload: 2 failed projects, 1 software-only project for which data was not applicable, and 1 case where the applicant stopped responding after the project ended. Among the 16 projects with publicly released data, we observed significant delays in sharing the data publicly (see [Fig. S11](#)). We used two indicators: the official end of support project and the last update to the public pipeline. Indeed, project support may have been extended beyond the official support time frame due to multiple factors: incurred delays in transferring or acquiring data, final steps requiring more polishing, issues in the pipeline surfacing after official support has ended, etc.. Data was publicly available within a month of the end of the official support for 1 project only (see [Fig S11a](#)), while the data for another 10 projects (out of 16) took more than 6 months to upload, of which 3 took more than a year. The average time from the end of official support to public data release was 8 months. The last changes to the pipelines were consistently done after the end of the official release. Consequently, the distribution of estimated time between the last pipeline modification and data release leads to shorter time frames (see [Fig S11b](#)). Nonetheless, the average time to public data release remains sizable in that case: 4.5 months. Only 2 projects saw data being released at the same time as the final touches to the public pipeline (OC1-6 and OC1-10). 8 datasets were uploaded online more than 3 months after the pipeline was finalized, of which 2 took 7 months and 2 more than a year. Note that 2 projects, OC2-23 and OC3-26, do not have yet a public pipeline as one is currently under embargo (OC2-23) and the other one is not yet finalized (OC3-26).

Finally, 5 trained models were uploaded to the BioImage Model Zoo, 2 projects' models were uploaded to Zenodo (OC2-13, OC3-19), and 1 additional model was deposited alongside its dataset on the BioImage Archive. Pre-trained models that were not fine-tuned were not uploaded to the BioImage Model Zoo. Links to the pipelines, data and models can be found in [Table S9](#).

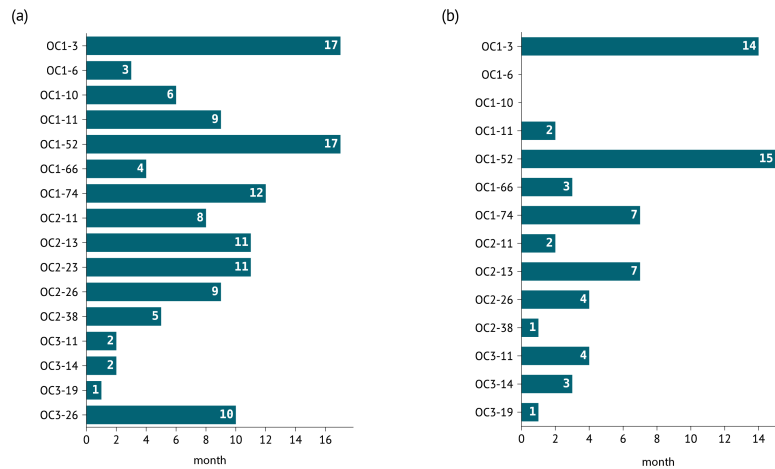


Figure S11. (a) Time in months between end of official support and public data release for all projects with released data. (b) Time in months between last update of the public pipeline and public data release for all projects with released data. OC1-6 and OC1-10 coincidentally released data and carried out the last update to the pipeline, while OC2-23 (embargo) and OC2-26 (ongoing) have released data but no public pipeline.

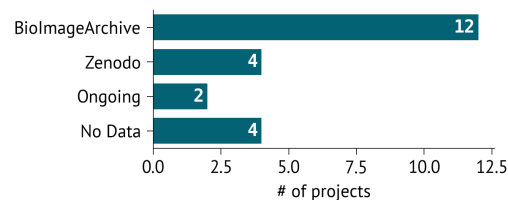


Figure S12. Repository used for openly released data. "Ongoing" refers to selected project whose data is not yet released at the time of writing. "No Data" corresponds to the two failed projects, one lost contact, and one project whose aim was to write a software plugin.

S2.7 Methods

All the application stats were extracted from the application forms filled by applicants across the three open calls. All the responses were concatenated together in an Excel sheet and mined directly or otherwise curated. Some categories resulted in lists of entries (e.g., data formats, imaging methods, image analysis tools), which were then individually split. We then manually curated the answers to standardize the terms (e.g. "Lighsheet" vs "light-sheet"). Nationalities (only OC 2 and 3) and countries of home institution (Fig. S2a and S3) were compiled directly from user answers. Since the scientific domains (Fig. S2c) were free text, users could enter answers that were not strictly speaking scientific domains (e.g. confocal imaging). In a second-pass curation, we removed any such entry. The list of available data formats was a detailed list in OC 1 (specific formats and a free text field), while it corresponded to the general categories shown in Fig. S2e in OC 2 and 3. We therefore automatically allocated the answers of OC 1 into the categories of OC 2 and 3. The type of home organization (Fig. S2b) was manually investigated based on web searches.

Imaging methods (Fig. S2d) were not a field in the application, and were manually derived from specific mentions in the project description field, data description field, project summary presentation, and investigation of uploaded data. Wherever ambiguity could not be lifted, the applications were categorized as "Others". CLEM application was split into "Widefield" and electron microscopy "Others" (transmission electron microscopy). One successful application mentioned "SHG" among other imaging methods but the project did not make use of that particular data, it was therefore excluded from the statistics. While there are overlaps between categories (e.g., SMLM and Widefield, Fib-SEM and vEM) or techniques that are strictly imaging methods (e.g., Expansion), we reported the denomination used by the applicants. TEM and SEM added ambiguity in the classification of EM methods and were aggregated into "Others" as a result. Natural images (e.g. underwater photography, seed pictures) were aggregated into a single "Photography" category. Scanography refers to images generated by using a scanner.

Tools declared by the applicants required manual curation and web search to confirm the existence of the tools/software/library and their correct spelling. In OC 1, tools were declared as free text. Based on the statistics from OC 1, we introduced individual and categories of tools in the application form OC 2 and 3. In order to plot Fig. S4a, we categorized the answers from OC 1 using the same categories as in OC 2 and 3.

To compile the statistics of the selected projects (Fig. S4b, c, d and e), we investigated each GitHub repository to extract analysis tasks, programming language, tools, library and software usage, labeling effort, deep learning usage, training mode (pre-trained model, fine-tuning, de-novo training).

Project outcomes and data readiness were assessed by the support team member assigned to each project and recorded at the end of the support period. Each assessment included explicitly structured fields: a readiness score on a 1–5 scale, the support team's judgment of whether sufficient ground-truth data was available for deep learning, the self-reported ground-truth status provided by the applicant at project initiation, the data transfer method used, and the overall project outcome. In addition, each assessor provided a free-text comment describing any issues encountered during the project. The statistics on data transfer delays, re-imaging events, annotation insufficiency, and data quality problems reported in this section were derived from systematic analysis of these free-text comments.

To estimate the delay between the end of project and public data release for the 16 projects that did release data, we used two different end of project dates. The first one is the official end of project support time frame: December 31st 2023 (OC 1), October 31st 2024 (OC 2), and August 31st (OC 3). However, projects may have incurred delays and actual project support be extended in time. We therefore added an estimated end of project support using the latest git commit modifying the public pipeline. Public data release was readily available from BioImage Archive and Zenodo releases. The two projects with no public pipelines were excluded from the average calculation in the case of the end of support date estimated from the pipeline commits.

S3 Public Challenges

Problem selection. During the Open Calls we observed that biological image analysis pipelines typically involved pre-processing, often denoising, followed by segmentation and downstream analysis. While segmentation remains a key step, state-of-the-art tools such as Cellpose⁷ and StarDist¹⁴ already perform robustly on many common biological imaging tasks. Downstream analysis, by contrast, tends to be highly project-specific, making it difficult to formulate it as a general-purpose challenge.

Image restoration, particularly denoising, stood out as a more broadly applicable and impactful task. Improvements in denoising directly influence the performance of subsequent steps, such as segmentation and quantification, and enable the effective use of low-signal or low-quality data, which is common in biological imaging. We therefore identified a clear need for improved denoising methods and for systematic benchmarking of existing approaches, and more broadly, for objective comparison of methods under controlled, reproducible conditions, which is a prerequisite for understanding the true state of the field.

Depending on the data at hand, image denoising may require fundamentally different approaches. We therefore designed three challenges to cover complementary denoising settings, each targeting a distinct data regime.

Hosting platform selection. Hosting a deep learning challenge requires robust infrastructure for dataset management, solution submission, automated evaluation, and public leaderboards. Several established platforms (Kaggle, Grand Challenge) already provide these capabilities. In addition to technical considerations, established platforms offer built-in visibility and access to an engaged user base, which is critical for reaching the broader machine learning community.

After a comparative evaluation, we selected [Grand Challenge](#), a European platform specializing in medical imaging challenges. Although no existing platform focuses exclusively on biological imaging, Grand Challenge offered the closest alignment with our requirements. Its emphasis on medical imaging, support for reproducible Docker-based submissions and compliance with European data regulations made it a strong fit.

Organization process. Each challenge began with the definition of a concrete and achievable task, which was formulated to be understandable to participants without prior experience in bioimage analysis. Appropriate datasets were then selected based on multiple criteria, including sufficient size for training deep learning models, alignment with the challenge objective, coverage of different imaging modalities, and representation of diverse biological sample types.

For each challenge, we prepared comprehensive documentation describing the scientific motivation, task definition, dataset characteristics, relevant background literature, and detailed instructions for accessing the data and submitting solutions via the hosting platform. To lower the barrier to participation, we also provided starter code and clear instructions for packaging submissions using Docker. Throughout the challenge periods, ongoing technical support was offered, including assistance with submission issues, clarification of task requirements, and guidance related to the challenge infrastructure.

Top-performing participants were required to release their training code publicly as a condition of recognition and were invited as co-authors of this summary paper. Participants who did not publish their code or who could not be reached for follow-up are acknowledged instead.

Challenges organized. The three challenges were unsupervised denoising (PC 1), supervised denoising (PC 2), and domain-specific denoising for calcium imaging (PC 3). In PC1, the data consisted of noisy data only, while PC2 focused on methods trained on noisy-clean image pairs. PC3 required methods to preserve temporal signal structure in addition to spatial image quality, and was inspired by an Open Call project (OC2-13).

Each challenge included four leaderboards corresponding to different datasets or task variants (full details are provided in Table S1). Datasets were selected from openly available sources with ground truth, or, where new data were introduced, training data were made publicly available on Zenodo. To ensure fair evaluation, test sets for leaderboards with ground-truth labels were privately hosted on the platform and not released to participants.

Table S1. Datasets used in the Public Challenges.

Challenge	Leaderboard	Dataset	Imaging modality	Dim.	Training data
Unsupervised Microscopy Denoising Challenge	Hagen et al.: Structured noise	Hagen et al. ⁵² [data]	Confocal microscopy	2D	79 images
	SUPPORT: Structured noise	SUPPORT ⁵³ [data]	Confocal microscopy	3D	1 volume (1001 z-slices)
	W2S: Unstructured noise	W2S ⁵⁴ [data]	Widefield microscopy	2D	120 images
	JUMP: Unstructured noise	JUMP ⁵⁵ [data]	Fluorescence microscopy	2D	517 images
Supervised Microscopy Denoising Challenge	Planaria: 3D denoising	Planaria ^{42,56} [data]	Confocal microscopy	3D	17,945 paired 3D crops
	Tribolium: 3D denoising	Tribolium ^{42,56} [data]	Multiphoton laser-scanning	3D	15,500 paired 3D crops
	Nuclei: 2D denoising	Nuclei ^{57,58} [data]	Widefield/spinning disk confocal	2D	2,477 image pairs
	FMD: 2D denoising	FMD ⁵⁹ [data]	Confocal/two-photon/wide-field	2D	63,669 image pairs
Calcium Imaging Denoising Challenge	All leaderboards	AI4Life-CIDC25 [train data, val data]	Synthetic calcium imaging	3D	5 time-series volumes

All submissions were evaluated automatically: PC 1 and PC 2 were scored using scale-invariant PSNR, while PC 3 employed a spatiotemporal PSNR metric to account for the temporal component of the signal (see methods).

S4 Public Challenges outcomes

S4.1 Challenge statistics

In total, the three challenges attracted 225 registered participants from 41 countries and resulted in 145 successful submissions (with 289 total submission attempts). Participation was geographically diverse, with the largest contingents from China (44), India (28), the United States (27), the United Kingdom (15), and Italy (12), alongside participants from 36 additional countries. These statistics are represented in Table S2.

Table S2. Countries by registrant count.

Other countries: Germany (5), Spain (5), Switzerland (5), Bangladesh (4), Iran (4), Pakistan (4), Belgium (3), Canada (3), Czechia (3), Sweden (3), Taiwan (3), Brazil (2), Colombia (2), Ecuador (2), Israel (2), Armenia (1), Australia (1), Denmark (1), Egypt (1), Greece (1), Indonesia (1), Ireland (1), Malaysia (1), Morocco (1), Nepal (1), Nigeria (1), Portugal (1), Romania (1), Slovakia (1), Sri Lanka (1), and Tanzania (1).

Country	Registrants
China	44
India	28
United States of America	27
United Kingdom	15
Italy	12
Hong Kong	7
South Korea	7
France	6
Türkiye	6
Vietnam	5
Other	66

Engagement varied across challenges. The Unsupervised Microscopy Denoising Challenge (PC 1) drew the largest audience, with 148 registrants and 68 successful submissions. The Supervised challenge (PC 2) attracted far fewer registrants (19) but produced 34 successful submissions, the highest submission-to-registrant ratio of the three, suggesting a highly engaged participant base. The Calcium Imaging Denoising Challenge (PC 3) attracted 58 registrants and had 43 successful submissions out of 72 total attempts, which we attribute to the increased task difficulty: calcium imaging requires handling volumetric data under complex noise conditions, likely demanding substantially more effort from participants. Reflecting this, PC 3 ran for considerably longer than the other two challenges, with the submission window extended multiple times in response to slow initial uptake. Overall, the number of successful submissions remained modest relative to established challenges in computer vision or medical imaging. The statistics are presented in Figure S13. Full submission counts per challenge and per leaderboard are also provided in Table S3 and S4.

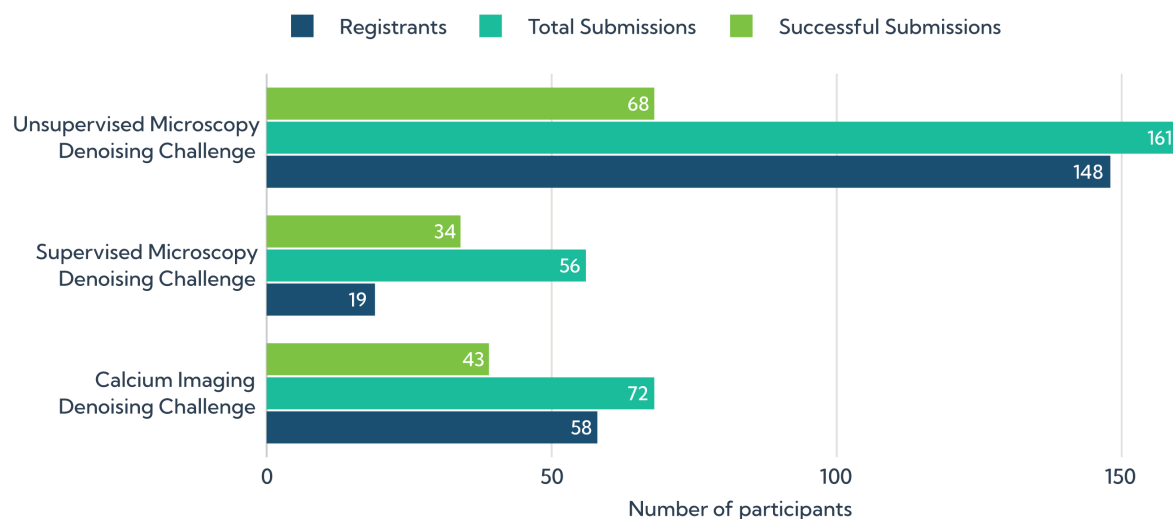


Figure S13. Participation statistics across the three AI4Life denoising challenges. Each Challenge is broken down in three categories: number of registrants, total number of submissions, and number of successful submissions.

Table S3. Registrants and submissions per challenge.

Challenge	Period	Registrants	Total Submissions	Successful Submissions
Unsupervised Microscopy Denoising Challenge	Jun-Jul 2024	148	161	68
Supervised Microscopy Denoising Challenge	Jun-Aug 2025	19	56	34
Calcium Imaging Denoising Challenge	Jun 2025-May 2026	58	72	43
Total		225	289	145

Table S4. Challenge leaderboard submissions

Challenge	Leaderboard	Total Submissions	Successful Submissions
Unsupervised Microscopy Denoising Challenge	Hagen et al: Structured noise	55	23
	SUPPORT: Structured noise	23	9
	JUMP: Unstructured Noise	48	19
	W2S: Unstructured Noise	35	17
Supervised Microscopy Denoising Challenge	Planaria: 3D denoising	15	5
	Tribolium: 3D denoising	11	7
	Nuclei: 2D denoising	17	14
	FMD: 2D denoising	13	8
Calcium Imaging Denoising Challenge	Content Generalization: Preliminary	12	8
	Content Generalization: Final	37	16
	Noise Level Generalization: Preliminary	6	6
	Noise Level Generalization: Final	17	13

Within each challenge, 2D denoising tasks consistently attracted more submissions than their 3D counterparts. In the Supervised challenge (PC 2), for example, the Nuclei (2D) leaderboard received 14 successful submissions versus 5 for Planaria (3D).

S4.2 Submitted solutions

Across all three challenges, submissions were dominated by a small number of established method families. Noise2Void^{17,32} and its variants were the most widely adopted approach, appearing across all three challenges and encompassing a range of extensions including probabilistic, structured, and 3D variants. U-Net-based architectures, including a ResUNet⁶⁰ and a method based on CARE⁴², appeared exclusively in the supervised challenge (PC 2). Noise2Fast⁶¹ and Noise2Noise³⁷ variants each contributed two submissions. Noise2Noise-based approaches appeared in the unsupervised and calcium imaging settings. A small number of submissions used more modern and specialized architectures, including COSDD⁶², HDN³⁶, a transformer-based model, and a classical method (a correlation-weighted spatial filter).

The algorithm types used are mentioned in Table S5.

Leaderboard results revealed several notable patterns. In the unsupervised challenge, COSDD (University of Birmingham) placed at or near the top across all four leaderboards, consistently outperforming the dominant Noise2Void-based submissions. In the supervised challenge, NAFNet-GAN⁶³ (Universidad Carlos III de Madrid) ranked first on both 3D and one of the 2D leaderboards, while placing third in the last leaderboard. Notably, a small number of participants placed consistently across multiple leaderboards within the same challenge, suggesting that general-purpose implementations can perform competitively across diverse datasets. The top 3 participants are shown in Table S6.

Many leaderboard entries corresponded to iterative variants of the same underlying approach, reflecting parameter tuning and dataset-specific adaptations rather than fundamentally distinct methods. Overall, the submitted solutions provide a

Table S5. Algorithm types used across challenges.

Algorithm type	Users
Noise2Void and variants	13
UNet and variants	3
Noise2Fast	2
Noise2Noise and variants	2
COSDD	1
Classical	1
HDN	1
Transformer	1
Other/Unknown	3
Total	27

representative snapshot of current denoising practices across different supervision regimes and data modalities.

Among the top-3 participants across all three challenges, 13 open-source repositories were published containing training code for reproducibility of the results. All are archived on Zenodo to ensure long-term availability, and links are compiled in Table S7.

S4.3 Methods

Submission and evaluation. Participants packaged their trained models and inference code as Docker containers. To lower the barrier to entry, we provided a template container with example code and detailed submission instructions for each challenge. Submissions were executed on the Grand Challenge platform against privately held test sets not released to participants. Predicted outputs were then evaluated using a dedicated metrics container developed by the organizers.

Metrics. PC 1 and PC 2 were scored using scale-invariant PSNR (SI-PSNR), a variant of the standard peak signal-to-noise ratio that is invariant to the scale of the predicted signal.

PC 3 was ranked using a spatiotemporal SNR metric (stSNR), combining a spatial SNR computed per frame with a temporal SNR computed per spatial location across time. This composite metric captures both image quality and the preservation of temporal signal dynamics, which is critical for calcium imaging where the temporal trace carries the biological signal of interest.

Full details and metric formulas are available on the respective challenge pages.

Challenge statistics. Raw submission and participation data were downloaded directly from grand-challenge.org. Country and institutional affiliation were taken from user profile information provided on the platform. A submission was counted as successful if both the inference container and the metrics evaluation container completed without error and produced an output in the expected format.

Algorithm types were manually inferred from the metadata provided by participants at submission time (algorithm name and comment fields) and, where available, from the linked code repositories. Submissions were assigned to algorithm families (e.g., Noise2Void and variants, U-Net and variants) based on this information. For the algorithm family summary Table S5, submissions were de-duplicated such that each unique combination of challenge, participant, and algorithm family was counted once.

The top-3 Table S6 reports the single best-scoring submission per participant per leaderboard. Where a challenge administrator placed in the top three (PC 3), their result is included as a reference baseline and noted as such in the table caption.

Table S6. Top 3 participants per leaderboard, showing only the best-scoring submission per participant. For Challenge 3, the challenge administrator (edoardogiacomello) placed in the top 3 on all leaderboards; his results are included as a baseline reference, allowing readers to contextualise participant performance relative to an expert submission.

Challenge	Leaderboard	User	Affiliation	Solution name	Metric
Unsupervised Microscopy Denoising Challenge					<i>SI-PSNR</i> ↑
	Hagen et al: Structured noise	bensalmon	University of Birmingham	COSDD	31.4847
		mcraft	Human Technopole	N2V	30.8962
		shanetoy	Seoul National University	Classical	30.8406
	SUPPORT: Structured noise	bensalmon	University of Birmingham	COSDD	30.4415
		a897574323	University of Shanghai for Science and Technology	N2V	29.7744
		edoardogiacomello	Human Technopole	N2V	28.8547
	JUMP: Unstructured Noise	kalexis	National and Kapodistrian University of Athens	N2V	35.6424
		bensalmon	University of Birmingham	COSDD	35.6282
		mcraft	Human Technopole	N2V	35.4957
	W2S: Unstructured Noise	bensalmon	University of Birmingham	COSDD	35.6855
		edoardogiacomello	Human Technopole	N2V2	35.0505
		mcraft	Human Technopole	N2V	34.3323
Supervised Microscopy Denoising Challenge					<i>SI-PSNR</i> ↑
	Planaria: 3D denoising	Guillermo_Rey	Universidad Carlos III de Madrid	NAFNet-GAN	32.7359
		sungjunc01	Hong Kong University of Science and Technology	UNet Care Hybrid	31.4879
	Tribolium: 3D denoising	Guillermo_Rey	Universidad Carlos III de Madrid	NAFNet-GAN	27.9314
		hu2274898	Beihang University	ResUNet	25.7342
		sungjunc01	Hong Kong University of Science and Technology	UNet Care Hybrid	24.8724
	Nuclei: 2D denoising	wangyuanchen	Harbin Institute of Technology	UNet	31.7287
		sungjunc01	Hong Kong University of Science and Technology	UNet Care Hybrid	30.8930
		Guillermo_Rey	Universidad Carlos III de Madrid	NAFNet-GAN	30.1082
	FMD: 2D denoising	Guillermo_Rey	Universidad Carlos III de Madrid	NAFNet-GAN	38.1628
		wangyuanchen	Harbin Institute of Technology	UNet	37.1843
		sungjunc01	Hong Kong University of Science and Technology	UNet Care Hybrid	30.4427
Calcium Imaging Denoising Challenge					<i>stSNR</i> ↑
	Content Generalization: Preliminary	aagamsheth	Ahmedabad University	Restormer3D	23.92
		edoardogiacomello	Human Technopole	N2V 3D	21.69
		mcraft	Human Technopole	Noise2Noise	20.94
	Content Generalization: Final	aagamsheth	Ahmedabad University	Restormer3D	23.72
		edoardogiacomello	Human Technopole	N2V 3D	22.14
		chhayanshpurwal	Rajasthan Technical University Kota	N2V 3D	21.18
	Noise Level Generalization: Preliminary	aagamsheth	Ahmedabad University	Restormer3D	18.46
		mcraft	Human Technopole	Noise2Noise	15.79
		edoardogiacomello	Human Technopole	Baseline	9.18
	Noise Level Generalization: Final	aagamsheth	Ahmedabad University	Restormer3D	18.08
		edoardogiacomello	Human Technopole	N2V 3D	16.75
		chhayanshpurwal	Rajasthan Technical University Kota	N2V 3D	15.59

Table S7. Code repositories for all top-3 participants for every leaderboard, grouped by challenge. “—” indicates no code was submitted.

Challenge	Username	Affiliation	Solution
Unsupervised Microscopy Denoising	bensalmon	University of Birmingham	GitHub
	mcroft	Human Technopole	GitHub
	shanetoy	Seoul National University	GitHub
	a897574323	Univ. of Shanghai for Sci. & Tech.	GitHub
	edoardogiacomello	Human Technopole	GitHub
	kalexis	Nat. & Kapodistrian Univ. Athens	—
Supervised Microscopy Denoising	guillermo_rey	Univ. Carlos III de Madrid	GitHub
	sungjunc01	Hong Kong Univ. of Sci. & Tech.	GitHub
	hu2274898	Beihang University	GitHub
	wangyuanchen	Harbin Institute of Technology	—
Calcium Imaging Denoising	aagamsheth	Ahmedabad University	GitHub
	edoardogiacomello	Human Technopole	GitHub
	mcroft	Human Technopole	GitHub
	chhayanshporwal	Rajasthan Technical Univ. Kota	GitHub
	guillermo_rey	Univ. Carlos III de Madrid	GitHub

A1 Appendix: OC3 application form

The form had five short sections in which we tried to understand the needs of the potential project:

- Applicant personal information
- General information about the project
- The existing image data
- The current image analysis workflow (if any)
- Questions about potential impact and dissemination

Data protection policy

The data you enter below will be used for the dissemination of information pertaining to the execution of the Horizon Europe Funded AI4Life project (Grant number: 101057970). In accordance with the Grant Agreement, your data will be retained during the project and deleted when it has ended as soon as the retention period established by the EC is over. If you would like to update or delete your data during the course of the project, please contact us using ai4life@eurobioimaging.eu.

Confidentiality

The scientific data you provide will be treated confidentially and the reviewers will sign a confidentiality and non-disclosure agreement.

Funding

AI4Life has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101057970.

Applicant Details

- Please re-enter your e-mail address:
 - Given Name:
 - Surname:
 - Institution or employer:
 - Are you working in an academic institution*? (Yes / No)
 - Gender* (Male / Female / Non-Binary / Prefer not to say):
 - Nationality*:
- * Answers to these question will be used for statistics, as part of our reporting to the European Commission.

Project Details

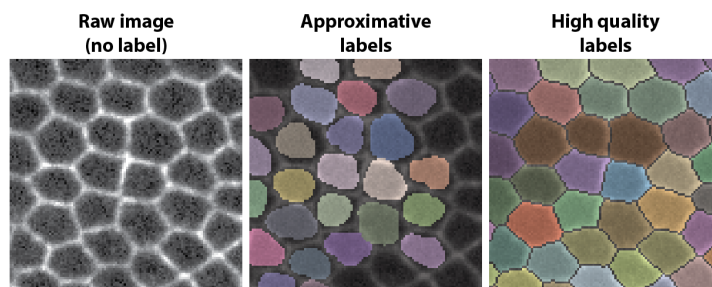
- Title of the project:
- What is/are the scientific domain(s) of the project?
- Describe the aim of the project:
Describe the type of data, modality and desired end result of the analysis, as well as the scientific relevance.
- Share a presentation (.pdf) using the AI4Life Open Call template or similar.
The template contains the following:
 - Slide 1: Project title
 - Slide 2: Aim of the study in bullet points
 - Slide 3: Example of the data and any relevant descriptions (legend, arrows, etc.)
 - Slide 4: Current analysis workflow (e.g., the various steps), if applicable, and desired outcome.
Feel free to reorganize the text in the template, but make sure that you do not use information that could identify you (names, university, own article reference, etc.), in order to guarantee blind reviews. Keep the technical details to the minimum necessary to judge the image analysis task; remember that the reviewers will not be specialists from your particular field of study!

- What is the limitation in analyzing your data we can help you overcome? Select the most fitting examples or describe your use-case below.
 - ✓ Improving your image analysis workflow
 - ✓ Advice on training data creation
 - ✓ Consultancy on available tools and solutions
 - ✓ Catering your problem to other experts
 - ✓ Other:

Your Existing Data

We would like to know more about the data you already have (type, size, format...).

- Please provide a small sample of your data (max 100 MB size, e.g. you can crop your images, only give a few slices, etc.).
- What does your image data show?
We are interested in a top-level description, e.g. the model organism(s), the type of labelled structures (without too much details), and the relationship between the structures in the data and the aim of your study.
- What type of image data do you have?
 - ✓ 2D images
 - ✓ 3D images
 - ✓ Time series
 - ✓ Multi-channel images
 - ✓ Other:
- How much project relevant data do you have (in Gigabytes)?
- Which format do you typically store your data in (tiff, jpg, png, czi, nd2, hdf5, zarr, ...)?



- Do you have any labeled data ("ground truth")? (None / Approximative labels / High-quality labels)
- Which percentage of your data is labeled/annotated?
- Is your data shareable or partly shareable?
If your project is selected, part of the data analyzed (in agreement with you) will be shared publicly at the end of the Open Calls, alongside the resulting analysis workflow.
 - ☐ I can only share parts of it (e.g. controls).
 - ☐ I can share all data publicly.

If you can only share part of your data...

- What fraction of your data is shareable? Tell us all we need to know to understand better your situation.

Your Existing Analysis

In this section, we'd like to know if you have already analyzed your data, and which approach or workflow you used.

- Do you already have an analysis pipeline?

- ☐ Yes, and the outcome is pretty good.
- ☐ Yes, but the outcome is not satisfying my minimum requirements.
- ☐ Not yet.

- Which software tools do you typically use to analyze image data?

- ✓ ImageJ/Fiji
- ✓ Scripts (ImageJ macro, Python, R, Matlab, etc.)
- ✓ Commercial software (Zeiss, Arivis, Imaris, Amira, Huygens, etc.)
- ✓ Machine learning frameworks (PyTorch, TensorFlow, scikit-learn, etc.)
- ✓ Cellpose
- ✓ CellProfiler
- ✓ DeepImageJ
- ✓ Icy
- ✓ ilastik
- ✓ Labkit
- ✓ napari
- ✓ QuPath
- ✓ Relion/CryoDRGN
- ✓ StarDist
- ✓ Webknossos
- ✓ Online notebooks (ZeroCostDL4Mic, etc.)
- ✓ Graphpad Prism, Excel, etc.
- ✓ Other:

- Which software tools are typically used in your field?

(The same options as above)

- Is there any part of your image analysis workflow that is currently done manually?

Impact and Dissemination

Our aim is to help you to make the best out of your image analysis and we would love to share the outcomes with the whole community.

- How impactful would be any improvement to your analysis on the rest of your community?

- ☐ Highly impactful, most would use the analysis pipeline
- ☐ Moderately impactful, some other labs would reuse the analysis pipeline
- ☐ Not directly impactful, our analysis is very specialised for now

- Would you be interested in helping us promote the outcome of the Open Call? (Yes / No)

- If your application is unsuccessful, would you be interested in posting your problem publicly to our bartering corner? (Yes / No)

Open-science principle

As part of AI4Life open-science and FAIR mandate, a subset of the data (in agreement with the data owners) and the whole analysis pipeline developed during the Open Call will be **publicly shared** with the community for each selected project. Raw data and annotations will be deposited on the BioImage Archive, while the analysis workflows will be shared on GitHub, and any trained model uploaded to the BioImage Model Zoo.

Data protection policy

The data you enter below will be used for the dissemination of information pertaining to the execution of the Horizon Europe Funded AI4Life project (Grant number: 101057970). In accordance with the Grant Agreement, your data will be retained during the project and deleted when it has ended as soon as the retention period established by the EC is over. If you would like to update or delete your data during the course of the project, please contact us using ai4life@eurobioimaging.eu.

✓ I accept this policy.

A2 Appendix: OC 3 review form

1. From 0 (not at all) to 5 (hell yes!):

- (a) Readiness of the project
- (b) Applicability of deep-learning
- (c) Excitement about the project

2. Is this a good AI4Life project? (yes/no)

3. Could this project be solved through consultation without taking it as an AI4Life project? (yes/no)

4. Additional notes:

A3 Appendix: Open Calls projects

Table S8: Selected projects in OC 1, 2 and 3.

Open Calls	[Project ID] - Title	Description
OC 1	[3] - Treat CKD	The aim of the studying is to characterize the effect of Chronic Kidney Disease (CKD) on the density of collagen in specific tissues, such as the heart and kidney, using mouse and rat tissue slices.
	[6] - Murine skeletal muscle analysis	The goal is to study the regulation of the metabolic properties in murine skeletal muscles, and the specific challenge is to segment, align and identify cells across individually acquired sections.
	[10] - Image-guided gating strategy for image-enabled cell sorting of phytoplankton	The goal of this project iss to create an automated gating strategy for a cutting-edge commercial flow cytometer to sort phytoplankton from lab-grown cultures or field samples.
	[11] - Segmentation and tracking of epithelial cells	The aim of this project is to develop a modular pipeline to segment and track epithelial cell nuclei in time-lapse fluorescence microscopy datasets, focusing on live imaging of dynamic epithelial monolayers.
	[47] - Automated detection and quantification of senescent cells	Here, we aim at providing a semi-automated workflow to identify and quantify senescent cells in embryonic mouse neural tube tissue imaged by confocal microscopy.
	[52] - Atlas of Symbiotic partnerships in plankton	The aim is to develop a segmentation workflow for identifying key organelles in FIB–SEM image volumes of microalgae engaged in symbiosis using limited annotations.
	[66] - Automated and integrated cilia profiling	This project involves developing an automated detection and analysis of cilia length in cultured cells overexpressing motor proteins of the kif family. The goal is to correlate motor protein expression levels to cilia length.
	[74] - Leaf tracker plant species proof	Developing a deep learning–based pipeline for instance segmentation and time-resolved tracking of plant leaves in RGB images across developmental stages.

Continued on next page

Table S8: Selected projects in OC 1, 2 and 3. (Continued)

Open Calls	[Project ID] - Title	Description
OC 2	[11] - Automated species annotation in turbid waters	The project aim is to create an automated image-annotation pipeline to monitor the ecological status of vulnerable marine habitats by detecting various species in underwater photographic images.
	[13] - Optimizing calcium image acquisition with machine learning denoising algorithms	The goal of the project is to benchmark various deep learning methods to denoise calcium imaging data acquired from in vivo brain recordings.
	[14] - Diatom symbioses at planetary scale	The aim of the project is to identify symbiotic interactions between plankton species sampled from many ocean sites as part of the Tara Ocean project.
	[21] - Automated segmentation of actin filaments in intact cells (EM)	Here, the aim is to segment actin filaments in cryo-ET tomograms, building a model that can generalize to future tomograms to produce quality segmentations.
	[23] - Enhancing the effective resolution of intra-vital microscopy with digital expansion microscopy	Developing a deep learning pipeline capable of enhancing low-resolution in-vivo two-photon microscopy images to match the 4x resolution improvement obtained by expansion microscopy.
	[26] - Ultrastructural protein mapping through Correlation of Light and Electron Microscopy	The project focuses on accurately aligning widefield fluorescence microscopy and transmission electron microscopy images to combine molecular and ultrastructural information for high-resolution analysis of endo-lysosomal alterations.
	[38] - NeuroScan - A 3D human motor neuron disease platform for high throughput drug screening	The goal of the project is to segment nuclei and Islet1 positive nuclei in 3D neurospheres from their respective channels.
OC 3	[3] - Improving nuclei segmentation using CellProfiler and StarDist	Enhancing DeepImageJ to run StarDist and other BioImage.io-compatible models directly within CellProfiler, ensuring seamless execution of advanced segmentation and feature-extraction routines.
	[11] - 3D Matrix Motility Map (3DM ³)	This project aims to analyze cancer cell dissemination in 3D extracellular matrices.
	[14] - The speed of life in trees	The goal of this study is to automatically measure the fractional areas of vessels, axial and ray parenchyma, and fibers across the entire wood tissue slices and multiple species.
	[17] - Imprints of Wind Disturbances on Wood Anatomy	To analyze differences in wood quality among trees experiencing varying degrees of wind disturbances. This requires an automatic segmentation of tree lumen and membrane in confocal microscopy.
	[18] - Detection of Nuclear Pore Complexes with shape variability imaged by DNA PAINT	To identify nuclear pore complexes in DNA-PAINT images, including those that deviate from the canonical structure in terms of diameter, stoichiometry, and circularity.
	[19] - Segmentation of sparse bacteria in human tissue	To segment and analyze bacteria clusters in human biopsies and characterize the biological microenvironment of niches (intra/extracellular, cell type, cluster size).
	[26] - Automatic microtubule doublet picking in tomograms	To automatically detect and pick doublets with the correct polarity (i.e., all pointing the same) in Cryo-EM tomograms of cilia.

Table S9: Project pipeline descriptions and outcomes

OC	ID	Pipeline	Outcomes
OC 1	3	In order to estimate collagen density in tissue sections, we used Labkit to classify pixels into background, cells, tissue and collagen classes. Fiji scripts were developed for image normalization, masking and quantification, while region exclusion masks were generated either manually in Fiji or using FeatureForest.	repo , data
	6	With the aim to analyze skeletal muscle fiber organization, we aligned different muscle slices to each other using BigWarp. We then segmented the cells with Cellpose. Finally, to link the resulting cell segmentation across multiple slices, we used the Fiji Trackmate plugin.	repo , data , model
	10	To generate gates for selecting cells in a cell sorting experiments, we then applied linear discriminant analysis to select relevant features. We used these features to generate polygonal gates on the 2D histograms with a similarity threshold between samples. Finally, we exported the gates in a FlowKit compatible format for gating strategy upload to the instrument.	repo , data
	11	To segment and track epithelial cells and nuclei in a wound-healing fluorescence time series, we first denoised the images with Noise2Void (n2v library). We then used Cellpose to perform cell bodies and nuclei segmentation, and finally used TrackMate to track the cells over time.	repo , data , model
	47	In order to quantify marker intensity in senescent and non-senescent nuclei in mouse neural-tube cross-sections, the pipeline first sum-projects stitched 3D stacks along Z, segments nuclei with Cellpose, and performs coarse first-pass senescent-cell-body pixel classification with Labkit in Fiji. Then, we detected senescent cell bodies using a threshold on their size and extracted a 3D volume around them. We trained a new pixel classifier with APOC for precise segmentation, matched the resulting 3D senescent-cell mask to the closest 2D nucleus, and exported nuclear marker intensities and senescence labels for downstream analysis.	repo
	52	To segment organelles in 3D FIB-SEM plankton datasets, we developed FeatureForest, a napari plugin that combines Segment Anything Model 2 (SAM2) embeddings with a random forest classifier trained from sparse scribble annotations. This allowed to segment various organelles throughout large volumes.	repo , data
	66	So as to correlate motor-protein expression levels with cilia length in fluorescence microscopy images, we segmented expressing cell bodies and nuclei using fine-tuned Cellpose models and segmented cilia using Labkit. We then measured perinuclear fluorescence intensity and cilia length to compare cilia morphology between expressing and non-expressing cells across the image datasets.	repo , data , model
	74	To automatically track leaf growth and movement in time-lapse plant imaging datasets, we first labeled individual leaves using LabelMe and AnyLabeling. Finally, we fine-tuned Detectron2 and segmented the entire dataset. Finally, we tracked the resulting instances over time using LapTrack and exported morphological features from individual leaves generated using OpenCV.	repo , data , model

Continued on next page

Table S9: Project pipeline descriptions and outcomes (Continued)

OC	ID	Pipeline	Outcomes
OC 2	11	To transform point annotations from an earlier study of underwater species into instances, we developed a pipeline using these point annotations as prompts for SAM2 to generate instance masks. The results were exported to the COCO format, compatible with BIIGLE.	repo , data
	13	In order to evaluate existing denoising deep-learning algorithms on time-resolved calcium imaging, we applied Noise2Noise (CAREamics), Noise2Void (CAREamics), HDN, DeepCAD-RT, and CellMincer on the applicants' and public data.	repo , data , models
	14	The project uncovered an alignment issue between data (from a large survey) and annotations and could not progress further.	repo
	21	In order to segment actin filaments in cryo-electron tomography data, we generated training annotations from in a semi-automated fashion with Amira and manually in Dragonfly. We then trained a 3D U-Net model using DeePiCt.	repo
	23	To perform superresolution from in-vivo two-photon microscopy images, we trained several deep-learning approaches using matching pairs of low resolution two-photon microscopy images and high resolution expanded samples. We trained CARE (CSB-Deep), a variational auto-encoder (adapted from HDN), and a ResMatching model.	repo , data
	26	So as to align high-resolution TEM images with fluorescence microscopy channels, we detected fiducial markers in both modalities using Big-FISH spot detection (fluorescence) and template matching (electron microscopy). Then, the probreg package was used to perform point-cloud registration between the detections. Finally, the fluorescence image was warped to overlay the markers onto the electron micrograph.	repo , data
	38	To quantify segment nuclei in two different channels, we used Cellpose on images pre-processed with scikit-image and visualized in napari. Segmentation was then further improved by automatically processing resulting masks and fine-tuning the model.	repo , data

Continued on next page

Table S9: Project pipeline descriptions and outcomes (Continued)

OC	ID	Pipeline	Outcomes
OC 3	3	In order to enable StarDist models to generate features in CellProfiler, we have adapted DeepImageJ to run directly in the CellProfiler plugin for ImageJ2/Fiji and load models from BioimageModel.io.	repo
	11	To analyze cancer cell interaction with the extracellular matrix in live-cell imaging of breast tumor spheroids embedded in collagen, we first denoised the images with Noise2Void (either in n2v or CAREamics) and then applied Cellpose to segment the cell nuclei.	repo , data , model
	14	With the aim to segment different cell types in wood tissue slices, we designed two different approaches. For one cell type, we used existing labels to train the decoder of a hybrid SAM2-UNet model. For the other cell types, we used FeatureForest to perform segmentation from iterative user-provided squiggles.	repo , data
	17	To investigate how wind disturbance affects wood anatomy, we performed cell lumen and wall segmentation using classical image analysis in Fiji macros.	repo , data upload in progress
	18	To classify irregular nuclear pore structures in localization microscopy, we performed drift correction, image rendering and detection of individual nuclear pores in Picasso. We then separated, segmented and performed morphological analysis of each individual nuclear pore using scikit-image, Scipy and OpenCV.	repo , data upload in progress
	19	In order to automatically segment in 3D <i>Staphylococcus aureus</i> bacteria in tissue samples, we trained a 3D StarDist model in DL4MicEverywhere.	repo , data , model
	26	To detect microtubules doublet in cryo electron microscopy tomograms, we first generated high-quality 3D ground-truth annotations using nnInteractive. Then, we used the generated masks together with manual particle annotations as input to a custom UNet for instance segmentation for training.	repo , data

References

1. Image.sc, scientific community image forum is a discussion forum for scientific image software sponsored by the center for open bioimage analysis (coba). <https://forum.image.sc/>.
2. Hartley, M. *et al.* The bioimage archive - building a home for life-sciences microscopy data. *J Mol Biol* **434**, 167505 (2022).
3. Ouyang, W. *et al.* Bioimage model zoo: a community-driven resource for accessible deep learning in bioimage analysis. *BioRxiv* 2022–06 (2022).
4. Rueden, C. T. *et al.* ImageJ2: Imagej for the next generation of scientific image data. *BMC bioinformatics* **18**, 529 (2017).
5. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
6. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* (2019).
7. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
8. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. methods* **19**, 1634–1641 (2022).
9. napari contributors (2019). napari: a multi-dimensional image viewer for python. <https://napari.org/stable/>, DOI: [10.5281/zenodo.3555620](https://doi.org/10.5281/zenodo.3555620).
10. Stirling, D. R. *et al.* Cellprofiler 4: improvements in speed, utility and usability. *BMC Bioinforma.* **22**, 433 (2021).
11. Bankhead, P. *et al.* Qupath: Open source software for digital pathology image analysis. *Sci. Reports* **7**, 16878 (2017).
12. Arzt, M. *et al.* Labkit: Labeling and segmentation toolkit for big image data. *Front. Comput. Sci.* **4**, DOI: [10.3389/fcomp.2022.777728](https://doi.org/10.3389/fcomp.2022.777728) (2022).
13. De Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nat. methods* **9**, 690–696 (2012).
14. Weigert, M. & Schmidt, U. Nuclei instance segmentation and classification in histopathology images with stardist. In *The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, DOI: [10.1109/ISBIC56247.2022.9854534](https://doi.org/10.1109/ISBIC56247.2022.9854534) (2022).
15. Seifi, M. *et al.* Featureforest: the power of foundation models, the usability of random forests. *npj Imaging* **3**, 32 (2025).
16. Isensee, F. *et al.* nninteractive: Redefining 3d promptable segmentation. *arXiv preprint arXiv:2503.08373* (2025).
17. Krull, A., Buchholz, T.-O. & Jug, F. Noise2void - learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).
18. Haase, R., Yamauchi, K., Müller, J. & Fernando, I. haesleinhuepf/apoc: 0.12.0, DOI: [10.5281/zenodo.7489264](https://doi.org/10.5281/zenodo.7489264) (2022).
19. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
20. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* (2000).
21. Nunez-Iglesias, J., Blanch, A. J., Looker, O., Dixon, M. W. & Tilley, L. A new python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton. *PeerJ* **6**, e4312 (2018).
22. White, S. *et al.* Flowkit: a python toolkit for integrated manual and automated cytometry analysis workflows. *Front. immunology* **12**, 768541 (2021).
23. Goode, A., Gilbert, B., Harkes, J., Jukic, D. & Satyanarayanan, M. Openslide: A vendor-neutral software foundation for digital pathology. *J. pathology informatics* **4**, 27 (2013).
24. Virtanen, P. *et al.* Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. methods* **17**, 261–272 (2020).
25. Imbert, A. *et al.* Fish-quant v2: a scalable and modular tool for smfish image analysis. *Rna* **28**, 786–795 (2022).
26. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with dna-paint. *Nat. Protoc.* **12**, 1198–1228 (2017).
27. Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of calcium images by learned contrast synthesis. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, 1123–1126 (IEEE, 2016).
28. Kenta-Tanaka *et al.* probreg: <https://probreg.readthedocs.io/en/latest/>.
29. Tinevez, J.-Y. *et al.* Trackmate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90, DOI: [10.1016/j.ymeth.2016.09.016](https://doi.org/10.1016/j.ymeth.2016.09.016) (2017). Image Processing for Biologists.

30. Fukai, Y. T. & Kawaguchi, K. Laptrack: linear assignment particle tracking with tunable metrics. *Bioinformatics* **39**, btac799 (2023).
31. Ravi, N. *et al.* Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024).
32. Careamics is a library aimed at simplifying the use of state of the art image restoration deep-learning algorithms, from well-established algorithms (care, noise2void) to cutting-edge methods (microsplit). <https://careamics.github.io/latest/>.
33. Wang, B. *et al.* Robust self-supervised denoising of voltage imaging data using cellmincer. *bioRxiv* DOI: [10.1101/2024.04.12.589298](https://doi.org/10.1101/2024.04.12.589298) (2024).
34. Li, X. *et al.* Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nat. Biotechnol.* **41**, 282–292 (2023).
35. de Teresa-Trueba, I. *et al.* Convolutional networks for supervised mining of molecular patterns within cellular context. *Nat. Methods* **20**, 284–294 (2023).
36. Prakash, M., Delbracio, M., Milanfar, P. & Jug, F. Interpretable unsupervised diversity denoising and artefact removal. In *International Conference on Learning Representations* (2022).
37. Lehtinen, J. *et al.* Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).
38. Gómez-de Mariscal, E. *et al.* Deepimagej: A user-friendly environment to run deep learning models in imagej. *Nat. Methods* **18**, 1192–1195 (2021).
39. Hidalgo-Cenalmor, I. *et al.* DL4miceverywhere: deep learning for microscopy made flexible, shareable and reproducible. *Nat. Methods* **21**, 925–927 (2024).
40. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).
41. Cardoso, M. J. *et al.* Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701* (2022).
42. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
43. Ray, A., Galinova, V. & Jug, F. Resmatching: Noise-resilient computational super-resolution via guided conditional flow matching. *arXiv preprint arXiv:2510.26601* (2025).
44. Langenkämper, D., Zurowietz, M., Schoening, T. & Nattkemper, T. W. Biigle 2.0-browsing and annotating large marine image collections. *Front. Mar. Sci.* **4**, 83 (2017).
45. Stalling, D., Westerhoff, M., Hege, H.-C. *et al.* Amira: A highly interactive system for visual data analysis. *The visualization handbook* **38**, 749–767 (2005).
46. Makovetsky, R., Piche, N. & Marsh, M. Dragonfly as a platform for easy image-based deep learning applications. *Microsc. microanalysis* **24**, 532–533 (2018).
47. Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. Labelme: a database and web-based tool for image annotation. *Int. journal computer vision* **77**, 157–173 (2008).
48. Anylabeling. <https://github.com/vietanhdev/anylabeling>.
49. Marcon, Y. & Purser, A. Papara (zz) i: An open-source software interface for annotating photographs of the deep-sea. *SoftwareX* **6**, 69–80 (2017).
50. Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. & Eng.* **9**, 90–95, DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55) (2007).
51. Bokeh Development Team. *Bokeh: Python library for interactive visualization* (2018).
52. Hagen, G. M. *et al.* Fluorescence microscopy datasets for training deep neural networks. *GigaScience* **10**, giab032, DOI: [10.1093/gigascience/giab032](https://doi.org/10.1093/gigascience/giab032) (2021).
53. Eom, M. *et al.* Statistically unbiased prediction enables accurate denoising of voltage imaging data. *Nat. Methods* **20**, 1581–1592, DOI: [10.1038/s41592-023-02005-8](https://doi.org/10.1038/s41592-023-02005-8) (2023).
54. Zhou, R. *et al.* W2s: Microscopy data with joint denoising and super-resolution for widefield to sim mapping. In *Computer Vision – ECCV 2020 Workshops*, 474–491 (Springer International Publishing, Cham, 2020).
55. Chandrasekaran, S. N. *et al.* Jump cell painting dataset: morphological impact of 136,000 chemical and genetic perturbations (2023).

56. Weigert, M. *et al.* Content Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy data, DOI: [10.17617/3.FDFZOF](https://doi.org/10.17617/3.FDFZOF) (2025).
57. Jacquemet, G. Noisy nuclei dataset for testing deep learning-based denoising tools, DOI: [10.5281/zenodo.5750174](https://doi.org/10.5281/zenodo.5750174) (2021).
58. Ljosa, V., Sokolnicki, K. L. & Carpenter, A. E. Annotated high-throughput microscopy image sets for validation. *Nat. Methods* **9**, 637–637, DOI: [10.1038/nmeth.2083](https://doi.org/10.1038/nmeth.2083) (2012).
59. Zhang, Y. *et al.* A poisson-gaussian denoising dataset with real fluorescence microscopy images. 11702–11710, DOI: [10.1109/CVPR.2019.01198](https://doi.org/10.1109/CVPR.2019.01198) (2019).
60. Diakogiannis, F. I., Waldner, F., Caccetta, P. & Wu, C. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote. Sens.* **162**, 94–114 (2020).
61. Lequyer, J., Philip, R., Sharma, A., Hsu, W.-H. & Pelletier, L. A fast blind zero-shot denoiser. *Nat. Mach. Intell.* **4**, 953–963 (2022).
62. Salmon, B. & Krull, A. Unsupervised denoising for signal-dependent and row-correlated imaging noise. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, 2379–2389 (2025).
63. Rey-Paniagua, G., Lachowski, D. & Muñoz-Barrutia, A. Nafnet-gan: An adversarial framework for supervised fluorescence microscopy denoising. In *2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI)*, 1–5, DOI: [10.1109/ISBI61048.2026.11515733](https://doi.org/10.1109/ISBI61048.2026.11515733) (2026).