

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/162817/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Sivagurunathan, Suganya, Marcotti, Stefania, Nelson, Carl J., Jones, Martin L., Barry, David J., Slater, Thomas J. A. , Eliceiri, Kevin W. and Cimini, Beth A. 2024. Bridging imaging users to imaging analysis - A community survey. *Journal of Microscopy* 296 (3) , pp. 199-213. 10.1111/jmi.13229

Publishers page: <http://dx.doi.org/10.1111/jmi.13229>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Bridging Imaging Users to Imaging Analysis - A community survey

Suganya Sivagurunathan¹, Stefania Marcotti², Carl J Nelson³, Martin L Jones⁴, David J Barry⁴, Thomas J A Slater⁵, Kevin W Eliceiri⁶, Beth A Cimini^{1*}

¹=Broad Institute of MIT and Harvard, Cambridge, MA, USA

²=King's College London, London, UK

³=University of Glasgow, Glasgow, UK

⁴=Francis Crick Institute, London, UK

⁵=Cardiff University, Cardiff, UK

⁶=University of Wisconsin-Madison, Madison, WI, USA

* = To whom correspondence should be addressed; Contact details:

Dr. Beth Cimini,
Imaging Platform,
Broad Institute,
415 Main St,
Cambridge, MA 02142
Email: bcimini@broadinstitute.org
Phone: 617-714-7000

ABSTRACT

The “Bridging Imaging Users to Imaging Analysis” survey was conducted in 2022 by the Center for Open Bioimage Analysis (COBA), Bioimaging North America (BINA), and the Royal Microscopical Society Data Analysis in Imaging Section (RMS DAIM) to understand the needs of the imaging community. Through multi-choice and open-ended questions, the survey inquired about demographics, image analysis experiences, future needs, and suggestions on the role of tool developers and users. Participants of the survey were from diverse roles and domains of the life and physical sciences. To our

knowledge, this is the first attempt to survey cross-community to bridge knowledge gaps between physical and life sciences imaging. Survey results indicate that respondents' overarching needs are documentation, detailed tutorials on the usage of image analysis tools, user-friendly intuitive software, and better solutions for segmentation, ideally in a format tailored to their specific use cases. The tool creators suggested the users familiarize themselves with the fundamentals of image analysis, provide constant feedback, and report the issues faced during image analysis while the users would like more documentation and an emphasis on tool friendliness. Regardless of the computational experience, there is a strong preference for 'written tutorials' to acquire knowledge on image analysis. We also observed that the interest in having 'office hours' to get an expert opinion on their image analysis methods has increased over the years. The results also showed less-than-expected usage of online discussion forums in the imaging community for solving image analysis problems. Surprisingly, we also observed a decreased interest among the survey respondents in deep/machine learning despite the increasing adoption of artificial intelligence in biology. In addition, the community suggests the need for a common repository for the available image analysis tools and their applications. The opinions and suggestions of the community, released here in full, will help the image analysis tool creation and education communities to design and deliver the resources accordingly.

INTRODUCTION

Microscopy has grown tremendously in the last few decades as a discipline ranging from simple light microscopes to super-resolution and electron microscopes which can image specimens beyond the diffraction limit. In parallel, quantitative image analysis has become an integral part of microscopy, and results that were previously communicated by 'representative images' now are typically required to be quantified. Automated microscopes now generate a large amount of data (up to TBs per day), which increasingly requires automated analysis to handle this ever-increasing data load. New modalities and submodalities of microscopy are now frequently invented, many of which require a diverse set of tools to analyze them. The requirements of the imaging community in terms of image analysis are therefore highly diverse and ever-changing.

With the aim of improving the understanding of the imaging community's needs, the Center for Open Bioimage Analysis (COBA) along with Bioimaging North America (BINA) and the Royal Microscopical Society Data Analysis in Imaging Section (RMS DAIM) has conducted the 'Bridging Imaging Users to Imaging Analysis' survey in 2022. The survey consisted of 32 questions directed towards the imaging community in both life science and physical sciences on different topics such as the demographics of the participants, usage of image analysis tools, preferences for learning materials, topics of interest for future workshops, image analysis experiences, suggestions for image analysis tool creators and users. The responses received were compared with the results of the 2020 Bioimage analysis survey conducted by COBA to understand the

preferences and needs of the imaging community and thereby develop and disseminate the resources accordingly.

RESULTS

Participants, their work type, and computational skills

The survey was conducted from May 2022 to July 2022; the survey was open to the general public and promoted in the imaging community through the Images2Knowledge (I2K) and Electron Microscopy and Analysis Group (EMAG) conferences, the image.sc forum¹, Microforum, Twitter, Confocal, ImageJ, and BioImaging North America (BINA) listservs. In contrast to a previous survey from this team² and other community surveys³, which was limited to bioimage analysis only, questions were added around physical science analysis as well. The final results contain 493 participants from a variety of roles and scientific experiences (Fig 1A&C). While our experience and data such as CellProfiler website analytics (Fig S1B) indicate global interest in accessing image analysis, most survey participants were from Europe and North America (Fig 1B) possibly because of the medium of distribution of the survey. The geographic results, along with the fact that students and postdoctoral fellows together only make up less than 40% of our sample, mean our results if left unfiltered do not create an unbiased sample of the imaging community as a whole; nevertheless, with cautious examination and subsampling, trends and conclusions can be drawn.

Five major descriptors were used to break down other survey respondents: discipline (life sciences vs physical sciences), self-reported computational skill level, self-reported computational comfort level, primary work classification (between Imaging, Analyst, or Balanced), and trainee status (categorized as students and postdocs vs other roles) (Fig S2A-C); further details can be found in the methods section. Among the 493 participants, 459 belonged to the life sciences discipline and 34 belonged to physical science; trainee status classification showed that 316 non-trainees and 177 trainees took part in the survey. A majority of the respondents were under the 'Balanced' work type irrespective of the domain and trainee status (Fig 2, S2D-G); specialists in either imaging or analysis were more common in non-trainee roles, possibly due to the >25% of respondents who describe themselves as facility directors or staff. Most of the participants had moderate computational skills except for the 'Analysts' in life sciences, whose self-reported computational skills were higher than the other work types. Respondents were also asked to describe their 'Comfort in developing new computational skills'; as one might expect, more computational job roles and self-reported skill levels are associated with higher comfort (Fig 2). Overall, most of the participants of the survey including the trainees and non-trainees had moderate computational skills in both life and physical sciences (Fig 2, S2E-G).

Images and image analysis tools

Understanding the kinds of images most commonly analyzed by the imaging community is important to help developers to design their tools accordingly. When surveyed, the majority of life science participants wanted to analyze fluorescent images

that were manually acquired followed by the ones that were acquired in an automated manner (such as by a plate scanning microscope) (Fig 3A). Fluorescence microscopic images were also the most analyzed images in the physical sciences, but physical science respondents were far more likely to be doing electron microscopy than life scientists. (Fig 3B). 2D images were the most commonly generated images, followed by 2D +time, 3D, 3D+ time, and large-volume 3D images (Fig 3A&B).

Participants were next asked about the image analysis tools that they use; as in our previous survey², we observed open-source point-and-click software are the most used image analysis tools (Fig 4), which we hypothesize may be due to any or all of the following - ease of use, ease of access, availability of tutorials, perceived need for human intervention, or the ability to perform analyses without prior programming knowledge. Participants may also feel more comfortable with such software as they facilitate ease of visual confirmation of a processing or analysis step. No-code-required tools are not necessarily needed by every member of the community with approximately $\frac{3}{4}$ of participants reporting using computational libraries and scripts for analyzing their images at least sometimes; while sampling bias no doubt partially contributes to this number, it does indicate a somewhat higher community competency with scripting than self-reported comfort levels might suggest (Fig 4 A&D). The frequency at which the scripts are used is higher among the physical science participants (Fig 4 C&F, S3), which likely relates to the higher-reported computational skills of physical science vs life science participants (Fig S2 F&G). The higher-reported computational skills of physical science participants might be reflective of a higher tendency of undergraduate programs

in these disciplines to include programming. The techniques available in physical sciences imaging (electron microscopy, AFM, spectrum imaging etc.) may also have fewer established point-and-click software packages for analysis, particularly in some specialized techniques.

Solving image analysis problems

Given the complexity of images often generated, it is common for participants to move beyond simple analysis using single methodologies. When the participants were asked about the approaches they use, it was clear that they prefer to use the tools that they have already used and are comfortable with. Participants rely on the internet, scientific literature, and their colleagues' protocols for any problems they face when coming up with a solution for analyzing their images. Despite the creation of the Scientific Community image forum¹ as a central hub for questions around image analysis and with answers provided by the experts in the field, usage remains comparable to 2020 levels² (Fig 5A) and low (28%) even among analysts (Fig S4A).

We also asked open-ended questions about what the participants thought were the well-solved image analysis problems and also about the problems that need better solutions. Respondents considered 'segmentation' to be an image-analysis problem that is both well-solved and needs a better solution, reflective of both segmentation's centrality to many image analysis problems as well as how wide the variety of segmentation problems are (Fig 5B-E). Three-dimensional image analysis and tracking are listed as major needs, similar to the answers that we received from the 2020 image analysis survey participants (Fig S4B&C). A notable difference between life and physical sciences were major topics in analysis needs, with life science respondents

clearly highlighting tracking as a major issue (with the assumption this relates to cell tracking) and physical science respondents highlighting reconstruction (although it's unclear whether this corresponds to 3D reconstruction in tomography or reconstruction of object phase in electron microscopy techniques).

Experience in image analysis

To get a general idea of the participants' experience in imaging and image analysis, we asked whether they have attended any workshops/conference sessions/conferences specific to these areas. The responses indicate that the participants have had considerable exposure to the field of imaging and image analysis as most of the participants have attended a workshop/tutorial on these topics; specifically for analysis, approximately $\frac{3}{4}$ of respondents had attended or considered attending at least one image analysis event, though only about $\frac{1}{4}$ described attending "some" or "many" (Fig 6 A&B). Workshops like NEUBIAS have been consistently quoted by the community members as one of the most useful workshops showing the interest of the imaging community in learning about the advancements in the field (Fig 6C)². The participants found these experiences useful for the following reasons - the workshops/conferences provided hands-on experience in working with many tools, interaction with the experts in the field, detailed theory, easy-to-follow video tutorials, and Q&A sessions.

Topics of interest and preferable methods for learning image analysis

Given a choice, the community prefers to learn image analysis practices that are more specific to a certain sub-discipline, the methods to analyze large images, and tools to visualize the results, with more than half of participants describing themselves as "very interested" (Fig 7A). The strong preference for sub-discipline-specific learning is consistent with previous results²; while it is unsurprising that most users want the tools only most relevant to them, the extra time required to tailor a generalist tool (and/or its training materials) to a specific audience often must be balanced against other aspects such as bug fixes and feature additions.

Based on the comments given by the participants on what made the conferences/workshops on image analysis very beneficial, it is clear that modes of instruction have played a significant role. The participants were asked about the modes of delivering knowledge on image analysis to be aware of their preferences. Written tutorials have always been the highly favored method to acquire knowledge on image analysis (Fig 7B)². The main advantages could be self-paced learning, step-by-step instructions which the users could follow more easily, and flexibility in usage. To know if this is true across people with different computational skills and work types, the preferable instructional methods were cross-matched with specific categories. Regardless of the computational knowledge and work type, 'written tutorials' are highly preferred by the imaging community (Fig S6). Participants who fell under the category of 'low computational skills' and 'imaging' work type prefer video tutorials and office hours (i.e., dedicated time for consultation on a specific problem) along with written tutorials (Fig S6). The imaging users may not possess comprehensive awareness of the

availability of resources that are required for their image analysis. In such instances, expert guidance and video tutorials would be more efficient than written tutorials alone. It is noteworthy that in contrast to all other topics or instructional methods, interest in office hours increased since 2022 (Fig 7C)², which could be because of the community's general interest in learning about customized methods that would work for their own images. It also indicates the need for image analysis as a service where users could consult with experts for guidance. Interactive webinars had the largest decrease in interest between 2020 and 2022, possibly because they were so heavily leaned on as primary instructional methods in 2020 and 2021 when COVID restrictions made other instructional methods far less common; if interest continues to decline in future years, instructors may need to reassess such formats in terms of desirability and effectiveness.

Suggestions for future workshops

Having asked about the topics and preferable methods for delivering knowledge on image analysis, the participants were next asked about the conferences that would benefit from such image analysis sessions. The respondents proposed that including image analysis workshops/sessions would be helpful for the attendees in conferences such as the American Society for Cell Biology (ASCB), Microscience Microscopy Congress (MMC), European Light Microscopy Initiative (ELMI), Association for Biomolecular Resource Facilities (ABRF), Focus on Microscopy (FOM), Biophysical Society and developmental biology conferences (Fig 8A). Some of these meetings have started introducing such sessions in recent years (ASCB, MMC, ELMI, FOM) and

offering satellite workshops on image analysis topics; however, some of the sessions were more oriented towards tool highlighting than workflow construction.

In regards to the content of the workshops/tutorials on image analysis, the participants suggested including the methods of using Python (coding/scripting), and deep or machine learning for image analysis, the comparisons of different image analysis tools, segmentation methods, ImageJ/Fiji plugins and macros, open-source software, certain tools of interest, CellProfiler, and also the best practices in image analysis (Fig 8B). Advancements in imaging technologies have made image analysis a multi-disciplinary field and the imaging community's curiosity to learn coding/scripting/machine learning in image analysis is reasonable. Getting comfortable with computational skills gives the end-user an opportunity to automate the image processing steps and helps in analyzing difficult-to-analyze images.

General suggestions

Participants were asked to provide suggestions on the roles of tool creators as well as tool users in improving image analysis; the responses were then categorized based on the work type to understand the opinions of each group, since each group may have particular insight into their own role as well as the roles of others. In response to how tool creators could improve image analysis, regardless of work type, the common suggestions were the need for documentation, open-source software, video tutorials on how to use the software, user-friendly intuitive GUI, help with installation, and example data to practice the software. The need for video tutorials is highly quoted by the 'imaging' participants when compared to the other two groups. There was also a

suggestion on recognizing the contribution of the developer through awards or incentivizing the developers who are creating user-friendly code (Fig 9A).

On the suggestion of users' role to make image analysis better, 'Imaging' participants feel that the users should practice the image analysis workflows with the examples, discuss the problems with developers and learn the basics by attending workshops. 'Balanced' and 'Analyst' work types suggest the users to understand fundamental image analysis by attending workshops/conferences/basic courses, learning the basics of coding/scripting, reporting image analysis problems with the developer through forums, sharing videos of how-to tutorials from users' perspective, discussing the microscopy experiments with an image analyst before image acquisition, and providing feedback to the developers (Fig 9B). In general, frequent, and constant interaction between the tool user and developer is a common suggestion.

Any other thoughts

A final, open-ended question invited respondents to express "any other thoughts" that they had. The respondents expressed their need for more user-friendly tools/plugins, a common repository for the available image analysis tools with their applications, and guides/checklists for both the users and the developers. Repeated requests for 'user-friendly' tools in the surveys indicate the need for computational knowledge to analyze biological images. There were also requests for detailed image analysis pipelines, training the microscope core facility staff, and also to include more techniques/tools for analyzing histology images and for the images generated using other modalities such as multi-photon microscopy and scanning probe microscopy.

Many of these requests such as a common repository for image analysis tools, detailed image analysis pipelines, guides for users and tools for analyzing images of specific modalities were also received from the 2020 survey participants, indicating the continued importance of these issues².

DISCUSSION

We have reported the responses provided by 493 participants regarding their needs and suggestions for image analysis, which were collected through a survey questionnaire containing 32 questions. Combining input from physical and biological sciences was added to the survey this year, in order to determine consistent image analysis needs across the two communities. Our data from the survey will be helpful for the tool creators to design the tools based on the suggested needs and also help COBA in designing and providing the resources accordingly.

Our data suggest a significant undersampling of trainees, physical scientists, users exclusively involved in imaging, and respondents outside of North America and Europe, possibly due to distributing the survey via existing networks which might have caused the partial exclusion of newer communities. The undersampling of physical sciences data makes drawing firm comparisons difficult, and future surveys will aim to target this community further to strengthen the bridging of life science and physical science communities that was initiated in this survey. Involving imaging core facilities at various universities and imaging consortiums that are focused on other continents such as the African Bioimaging Consortiums would broaden the representation of the imaging community in future such surveys. Incorporating input and feedback from a broader

community will be helpful in better understanding the users' needs allowing the tool developers to prioritize the efforts accordingly.

When approached as a whole, some inherent contradictions can be found within our survey results: a major one is that users feel more central repositories are needed (as a response to a question '*Any other thoughts?*') and are overwhelmed with the variety of new tools being created, but they also simultaneously request more emphasis on field-specific rather than general approaches and developers in response create new tools rather than adding functionalities to existing ones. These contradictions are understandable, especially in the light of competing incentive structures: users want to focus on image analysis more as a tool than as a discipline when they are trying to publish scientific results in a time and/or funding-constrained manner, so naturally want something targeted; developers may find it easier to publish and/or receive funding for a new tool than for maintenance and feature addition for existing tools, and there are fewer still incentives for making tools and solutions FAIR (findable, accessible, interoperable, and reproducible)([Wilkinson et al. 2016](#)). While solving this contradiction is beyond this group's power, it should be noted that some repositories such as the BioImage Informatics Index (BIII, <https://biii.eu/>)⁴, the NEUBIAS F1000 Gateway⁵, and the Workflow Hub⁶ (<https://workflowhub.eu/>) do already exist, even if their adoption still seems somewhat limited, possibly due to users and developers lacking awareness of such tools. Moreover, resources collating conference and training materials are available, such as MicroscopyDB (<https://microscopydb.io/>), maintained by Global Bioimaging and BINA, and community-led efforts such as FocalPlane

(<https://focalplane.biologists.com/>). To better serve community needs, forums can include 'resources' pages with pointers to these centralized sources alongside specific updates on the latest resources and their applications. Some centralized repositories such as BIII allow for field and/or function-specific tagging, which could be adopted by other repositories and which end users wishing to help their communities could prioritize adding for their sub-discipline. These tags can be used for tool searching, similar to what was done by the Correlated Multimodal Imaging in Life Sciences (comulis) on their resource pages (<https://www.comulis.eu/correlation-software>).

Another point that stands out when analyzing the differences between end-users' and tool developers' needs and requests relates to the need for more communication between the two communities. The authors feel that this gap is often bridged by the figure of the image analyst, who can offer dedicated support to the end users, while also being able to provide detailed feedback to the developers. Another hint to the need for this figure might be represented by the request for dedicated office hours for user support (Fig S6). Image analysts are often associated with imaging facilities, however, more examples of dedicated centers and consortia are reported⁷. One example is AI4Life (<https://ai4life.eurobioimaging.eu/>), a Horizon Europe-funded consortium offering research services and infrastructure to support life scientists in the adoption of machine learning solutions for image data. In Germany, a centralized initiative (NFDI4BIOIMAGE, <https://nfdi4bioimage.de/en/start/>) has recently kicked-off to provide scientists from all natural science and biomedical research fields with workable and trusted solutions to handle image data. Moreover, research infrastructures offering open access to imaging technologies, such as Euro-Bioimaging (<https://www.eurobioimaging.eu/>), have started

offering image and data analysis services for biological and medical data. It is likely that such projects will have a positive impact in improving the communication between end-users and developers, by providing the role of the image analyst who can communicate with both and facilitate each group's interactions.

The Scientific Community image forum¹ was introduced as a central platform to discuss, share, and provide help on the issues faced when acquiring or analyzing images. While overall traffic levels imply that utilization of the Image.sc forum has steadily increased over the years, the number of survey participants saying it is among the approaches they use the most did not increase here, nor was it >30% in any subgroup including analysts. To address this, future efforts can be made to bring awareness about the existence and benefits of the forum through image analysis workshops/tutorials, specifically, the ones attached to discipline-specific conferences, and also through targeted education opportunities for imaging facility staff, who can then relay information about best practices and about where to ask further questions to their users. Use of the forum is one of the most direct ways for an imaging tool user to gain access to expert help in their image analysis, as it is essentially a 24/7 virtual office hour. Beyond providing direct help, the forum creates a public record of problems encountered and solved making it easier for the next user with the same issue to overcome the challenges they are facing. Questions of all difficulty levels but especially from beginners are warmly encouraged; users who are uncomfortable attaching their real names to their posts are welcome to create pseudonymous accounts.

Beyond documentation of individual issues encountered, broader training materials on how to solve particular problems, specially tailored to more individual

communities such as multi-photon microscopy and scanning probe microscopy, and also created in such a way that people with any/no computational background can follow, are needed in greater numbers. The results of the survey suggest that most users work with 2D, 2D+time and 3D images (Figure 3), as opposed to more complex data, possibly suggesting that foundation training might be a way forward to support the community. Broader still were the requests from the imaging community, irrespective of the work type, to have a checklist or guidelines for imaging and image analysis; a number of such works are under construction in a variety of consortia to help users in planning their microscopy experiments, analyzing, and publishing images^{8,9}. As discussed above, many users may also not be aware that public workflows for many common tasks exist; spreading knowledge of such solutions is a critical first step to adoption.

One of the most surprising results in our opinion was the decrease in interest in deep learning. There was an approximately 2-fold increase in the peer-reviewed papers published with the terms 'deep/machine learning/artificial intelligence' (Fig S8) during the period of 2020 to 2022, reflecting the overall general increasing interest in creating machine learning models or applying them in scientific discovery. This is reflected in several new (or "newly-deep") deep learning/machine-learning based tools designed to be easy for end users, such as Cellpose¹⁰, StarDist¹¹, ilastik¹², DeepImageJ¹³ and Piximi¹⁴, that were introduced for image analysis during the last few years. While overall trends were towards fewer "very interested" responses in 2022 vs 2020, deep learning represented the second largest decrease in topic interest. We cannot rule out that people were less "very interested" in learning about deep learning because increasing

fractions of users mastered it in that two year window, or that they are using well established “newly-deep” tools without fully acknowledging their DL/ML component. However, based on our experiences with end users we think a more likely explanation is that greater education and outreach, as well as continued improvements in installation and access, will be needed before end users trust these tools enough to become excited about applying them to their own work. Deep learning tools are of course not always the best tool for a job, especially when large amounts of training data are unavailable and/or painful to make; improved human-in-the-loop¹⁵ approaches to reduce the annotation burden and improved self-supervised training approaches may reduce this burden in the future, together with more data available in open repositories such as the BioImage Archive¹⁶ (<https://www.ebi.ac.uk/bioimage-archive/>). Finally, it is possible that the reduced uptake of such methods might relate to the lack of suitable local infrastructure and training data. Some efforts in democratizing access via Google Colab (ZeroCostDL4Mic¹⁷, <https://henriqueslab.github.io/resources/ZeroCostDL4Mic/>) and creating hubs for bringing AI models to researchers (Bioimage Model Zoo¹⁸, <https://bioimage.io/>) have been made in recent years, and it will be interesting to see if they will have an impact in the near future.

CONCLUSION

While the last several years have brought significant advances in many subfields of microscopy analysis, these results suggest that we as a community still have much work to do in order to empower and/or encourage imaging users to fully embrace analytical approaches and tools. We hope that these results, as well as the analysis of

what has and has not changed since our initial report, can help further motivate researchers to "cross the divide" as well as motivate journals, universities, and funding agencies to assess how they can help in this process, such as by increasing access to statistical and computational training for students and early career researchers or by funding developers to create more friendly and more tailored instructional material. Unlike in our past survey, this paper provides the underlying response data in fully intact form, which we hope will encourage stakeholders to perform their own subsetting and analysis of their own section of the community. Further, conducting periodic surveys will help developers and the educators keep tabs on changes in needs and populations for the populations they wish to serve and deliver the resources accordingly.

Ultimately, the development of image analysis tools, workflows, and approaches is an iterative process requiring input from all stakeholders. No single group or community in isolation can accomplish these goals on their own; all must work together in order to fully overcome the past history of microscopy as a purely quantitative science. We hope this data adds to previous work in order to help light the way.

KEYWORDS

image analysis, survey, deep learning, life science, physical science.

PRACTITIONER POINTS

1. Gaps remain between image analysis tools developers and tool users; image analysts as a professional class may help close this gap. The Scientific Community Image Forum (forum.image.sc) is proposed as a central organizing platform.

2. Major requests for developers included increased video tutorials and office hours.
3. Tailoring of tools to particular subfields as well as including image analysis into topic-specific conferences may help increase image analysis tool adoption.

METHODS

“Bridging Imaging Users to Imaging Analysis” survey conducted by the Center for Open Bioimage Analysis (COBA), Bioimaging North America (BINA), and the Royal Microscopical Society (RMS) was made available in Google forms to the imaging community through the Images2Knowledge (I2K) and Electron Microscopy and Analysis Group (EMAG) conferences, the image.sc forum¹, Microforum, Twitter, Confocal, ImageJ, and BioImaging North America (BINA) listservs. The responses of the survey were exported to tables, analyzed, and graphed in Jupyter Notebook (6.4.12)¹⁹ using Python3(3.9.13)²⁰, pandas(1.4.4)²¹, plotly(5.9.0)²², matplotlib(3.5.2)²³, numpy(1.24.2)²⁴, wordcloud(1.8.2.2)²⁵, kaleido (0.1.0.post1) and sci-kit learn(1.0.2)²⁶ libraries. Interactive versions were made with Streamlit²⁷.

Participants who answered as ‘Undergraduate/Graduate student’ and ‘Postdoctoral fellow’ to the question ‘Which of the following roles best describes you?’ were categorized as ‘Trainees’ and the rest of the roles were categorized as ‘Nontrainees’. Answers to the questions on ‘work description’, ‘computational skills’, and ‘comfort in developing new computational skills’ were grouped into three categories. Scale values of 1 & 2 were grouped as ‘Imaging/Low skill/Low comfort’, values of 3 to 5 as ‘Balanced/Medium skill/Medium comfort’, and values 6&7 as ‘Analysts/High skill/High

comfort' based on the relevant questions. The graphs show the responses from all respondents unless specified otherwise in the figure legend.

For the fold change in the interest level or the preferable methods, percentage fold change was calculated with responses received under the category 'Very interested'/'Very preferable' for the questions - 'In regards to learning more about image analysis, how preferable do you find each of these instructional methods?' and 'How interested are you in learning more about the following topics?' from the 2020 and 2022 surveys.

The open-ended questions on solving image analysis problems, creators and users role were analyzed using a word cloud library where the answers were split into single words and the frequency of occurrence of the words decided the font size of the words in the word cloud plot. Words that appeared in the related questions and the words that were not providing useful information were removed from the wordcloud plot along with the standard 'English' stopwords available in the wordcloud library. Unigrams, bigrams, and trigrams were generated from the responses to the open-ended questions 'Are there any conferences you've attended in the past that you think would particularly benefit from the addition/expansion of image analysis offerings?' and 'What specific topics (i.e. overviews of a particular tool, comparisons between pieces of software, or how to use a certain tool for a certain kind of experiment) would you like to see prioritized for future image analysis workshop and tutorial offerings?' using CountVectorizer library. The words/ conferences that were making it to the top of the list and also the meaningful words were counted and plotted as a graph.

ACKNOWLEDGMENTS

The authors gratefully acknowledge members of the Cimini and Eliceiri labs for feedback on the survey questions. The authors thank Erin Weisbart, Mario Cruz, Ellen Dobson, Rebecca Senft for their suggestions and Rebecca L. Ledford for help on the data statement and publicizing the survey. They also acknowledge the strong support of Vanessa Orr, Nikki Bialy and the Image Informatics working group of BioImaging North America (BINA) in giving useful feedback and promoting the survey. They are additionally grateful to the 493 unnamed participants, without whom this work would have of course not been possible.

FUNDING STATEMENT

The work was supported by the Center for Open Bioimage Analysis (COBA) funded by National Institute of General Medical Sciences P41 GM135019 awarded to BAC and KWE. The work was also supported by grant number 2020-225720 to BAC from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. This work was supported by The Francis Crick Institute which receives its core funding from Cancer Research UK (CC001999), the UK Medical Research Council (CC001999), and the Wellcome Trust (CC001999). This work was supported by grant number BB/V006169/1 to SM by the Biotechnology and Biological Sciences Research Council (BBSRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

COMPETING INTERESTS

The authors declare that there are no competing interests associated with the manuscript.

DATA AVAILABILITY STATEMENT

The survey questions, the CSV files, the Jupyter notebooks, and the code that was used to generate the plots are available at https://github.com/COBA-NIH/2023_ImageAnalysisSurvey/tree/main. Interactive versions of most figures are available via Streamlit²⁷ at <https://broad.io/2022SurveyApp>.

REFERENCES

1. Rueden, C. T. *et al.* Scientific Community Image Forum: A discussion forum for scientific image software. *PLoS Biol.* **17**, e3000340 (2019).
2. Jamali, N., Dobson, E. T., Eliceiri, K. W., Carpenter, A. E. & Cimini, B. A. 2020 BioImage Analysis Survey: Community experiences and needs for the future. *Biol Imaging* **1**, (2022).
3. Miura, K. A Survey on Bioimage Analysis Needs, 2015. doi:10.5281/zenodo.4648076.
4. Paul-Gilloteaux, P. *et al.* Bioimage analysis workflows: community resources to navigate through a complex ecosystem. *F1000Res.* **10**, 320 (2021).
5. Cimini, B. A. *et al.* The NEUBIAS Gateway: a hub for bioimage analysis methods and materials. *F1000Res.* **9**, 613 (2020).

6. Goble, C. *et al.* *Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory*. <https://zenodo.org/record/4605654> (2021)
doi:10.5281/zenodo.4605654.
7. Soltwedel, J. R. & Haase, R. Challenges and opportunities for bioimage analysis core-facilities. *J. Microsc.* (2023) doi:10.1111/jmi.13192.
8. Senft, R. A. *et al.* *A biologist's guide to the field of quantitative bioimaging*. (2023).
doi:10.5281/zenodo.7810631.
9. Schmied, C. *et al.* Community-developed checklists for publishing images and image analysis. *ArXiv* (2023) doi:10.1242/jcs.254151.
10. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
11. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* 265–273 (Springer International Publishing, 2018).
12. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
13. Gómez-de-Mariscal, E. *et al.* DeepImageJ: A user-friendly environment to run deep learning models in ImageJ. *Nat. Methods* **18**, 1192–1195 (2021).
14. Goodman, A. *et al.* *Piximi - an images to discovery web platform*. (2021).
15. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
16. Hartley, M. *et al.* The BioImage Archive - Building a Home for Life-Sciences Microscopy Data. *J. Mol. Biol.* **434**, 167505 (2022).

17. von Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276 (2021).
18. Ouyang, W. *et al.* BioImage Model Zoo: A Community-Driven Resource for Accessible Deep Learning in BioImage Analysis. *bioRxiv* 2022.06.07.495102 (2022) doi:10.1101/2022.06.07.495102.
19. Kluyver, T. *et al.* Jupyter Notebooks -- a publishing format for reproducible computational workflows. in *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (eds. Loizides, F. & Schmidt, B.) 87–90 (IOS Press, 2016).
20. Van Rossum, G. & Drake, F. L. Python reference manual. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/gwydion-1/OldFiles/OldFiles/python/Doc/ref.ps> (1995).
21. Reback, J. *et al.* *pandas-dev/pandas: Pandas 1.0.3.* (2020). doi:10.5281/zenodo.3715232.
22. Inc, P. T. Collaborative data science. *Montréal, QC* (2015).
23. Hunter. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
24. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
25. Mueller, A. *et al.* *amueller/word_cloud: 1.9.1.1.* (2023). doi:10.5281/zenodo.7901644.
26. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*
27. Streamlit. <https://streamlit.io/>.

FIGURES

Figure 1

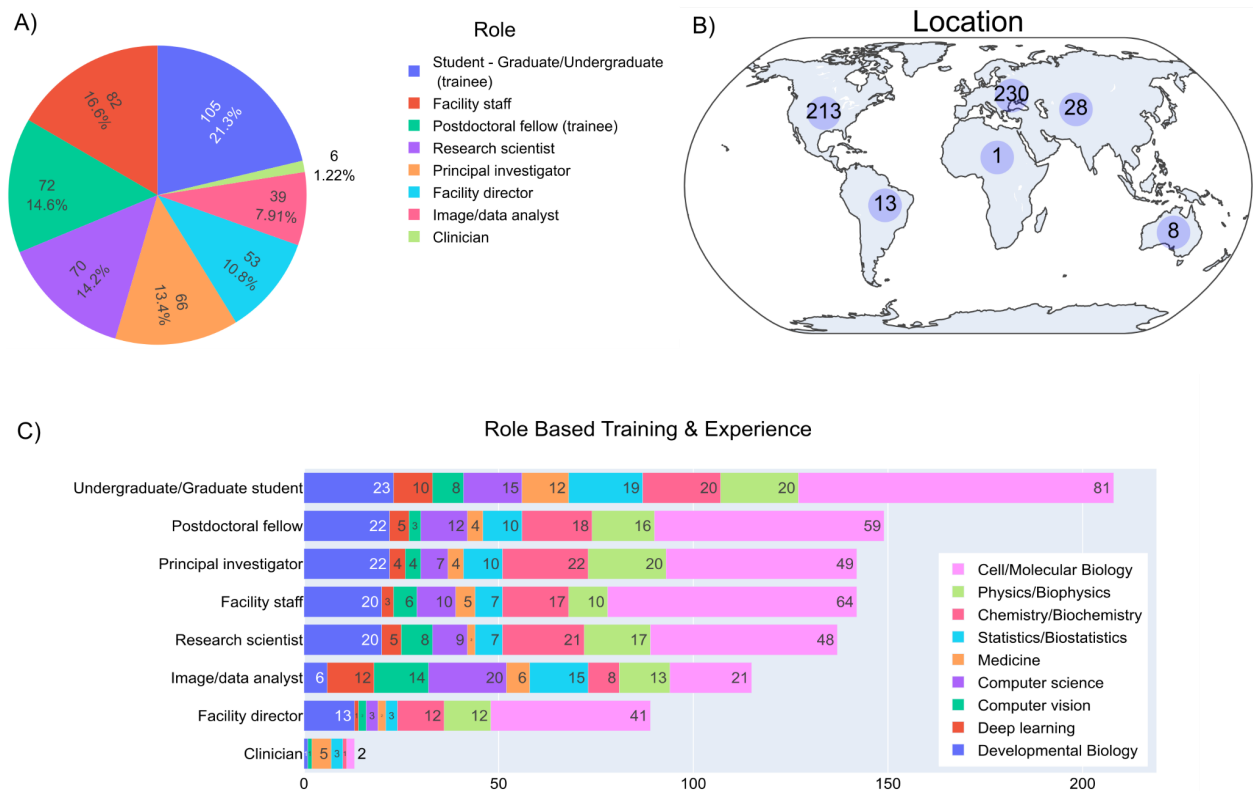


Figure 1 - Survey respondents' roles training histories and location vary across the sampled responses

- A) Answers to the multiple-choice question “Which of the following roles best describes you?”.
- B) Answers to the multiple-choice question “Where do you currently primarily work?”
- C) Answers to the check-box question “Which of the following do you have significant formal training in or experience with? Select all that apply.” Responses were categorized based on the answers provided for part A.

Figure 2

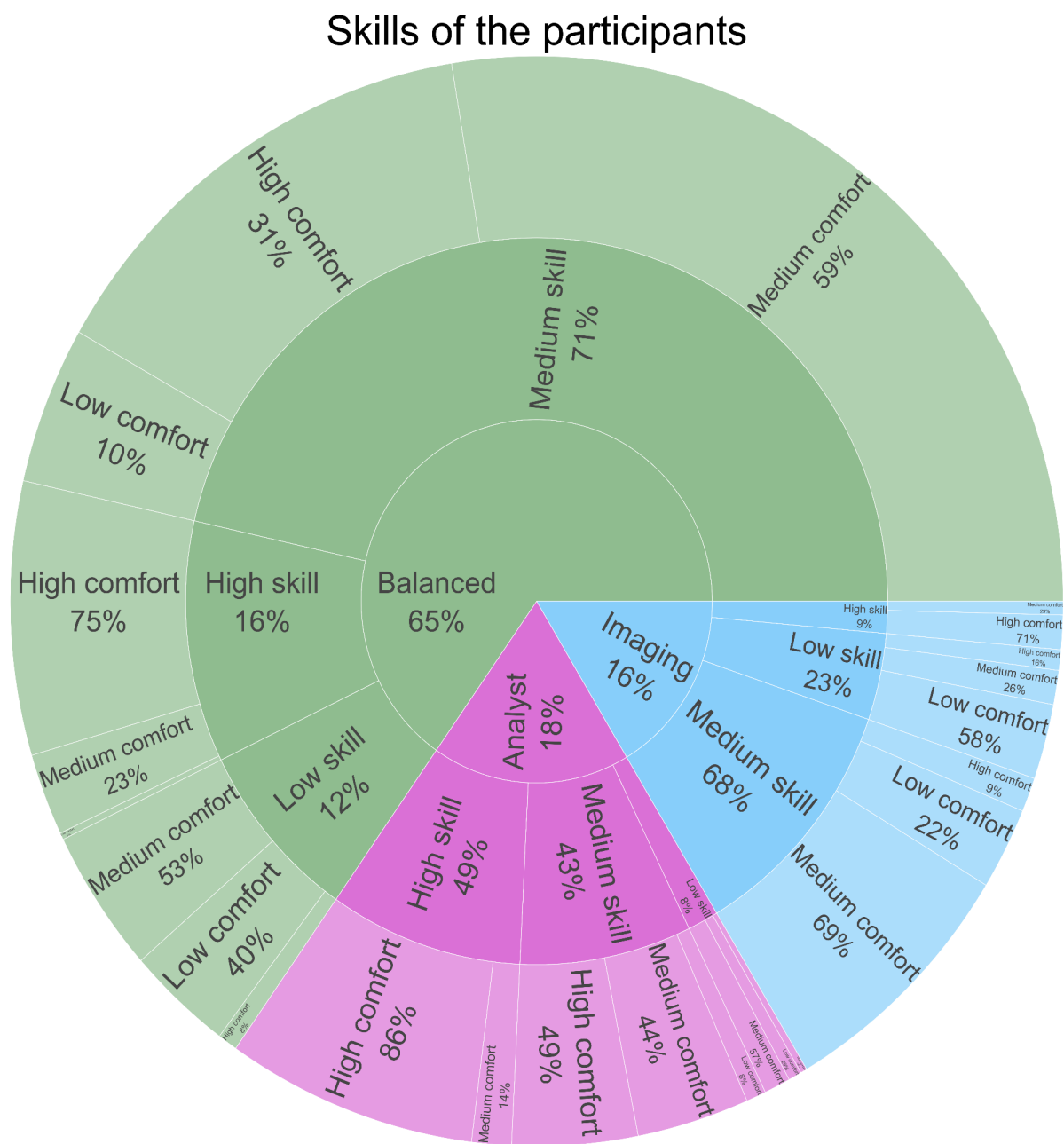


Figure 2- Skills of the participants

Breakdown of answers to the multiple-choice questions “How would you describe your work?”, “How would you rate your computational skills?” and “How would you rate your comfort in developing new computational skills?” Percentages were rounded to the nearest percent; in outer wedges, percentages are of the adjacent inner wedge population. See methods and Supplemental Figure 2 for fuller descriptions of each

category; interactive versions of these plots are available at <https://broad.io/2022SurveyApp>.

Figure 3

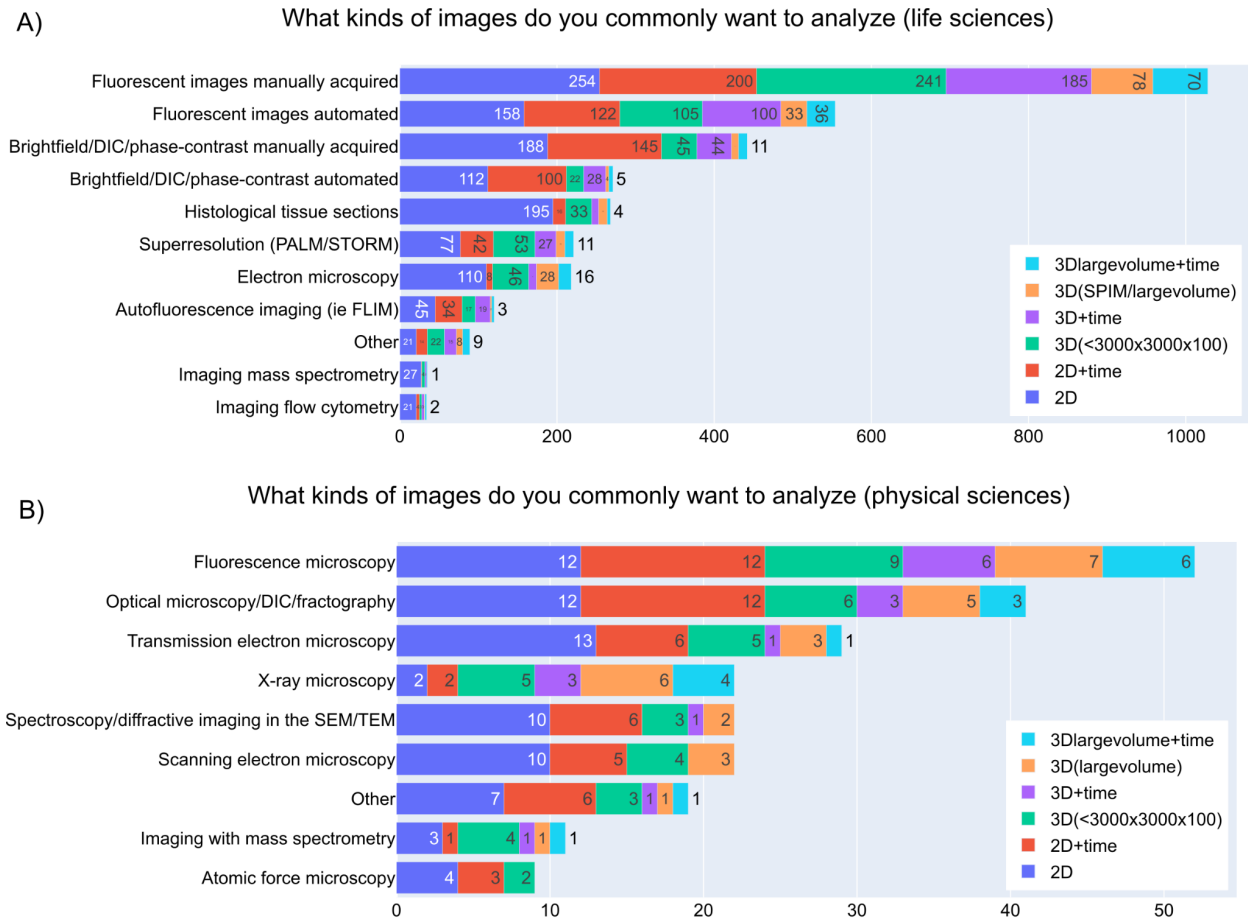


Figure 3-Kinds of images analyzed

- A) Answers to the checkbox grid question “What kinds of images do you commonly want to analyze (select all that apply)?” under the “Life Sciences Image Analysis” section.
- B) Answers to the checkbox grid question “What kinds of images do you commonly want to analyze (select all that apply)?” under the “Physical Sciences Image Analysis” section.

Figure 4

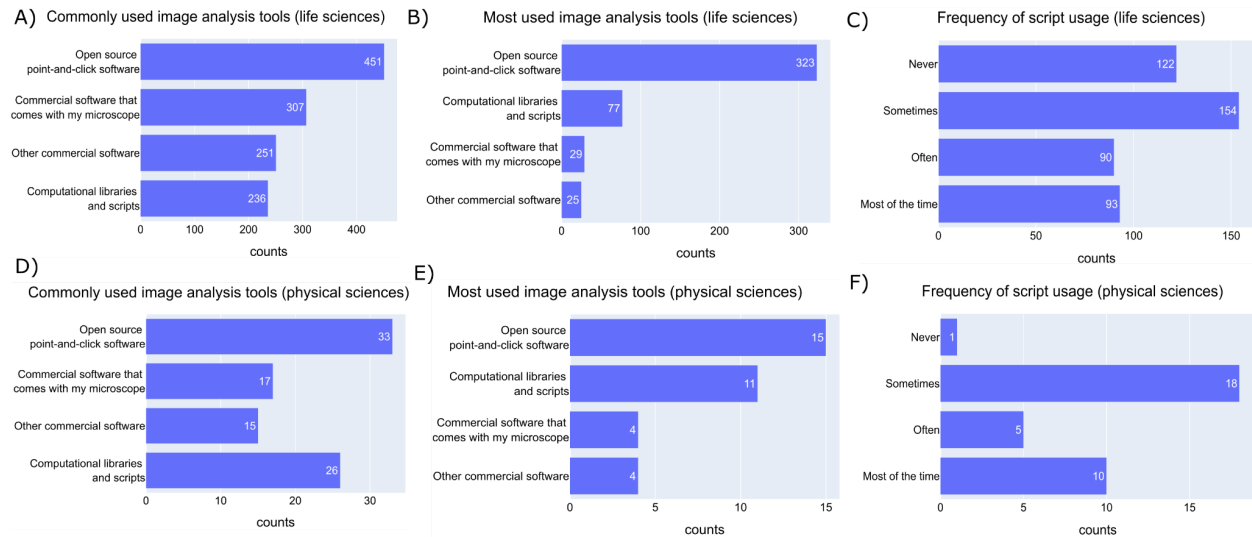


Figure 4- The community prefers open-source point-and-click software

A) Answers to the multiple-choice question “What image analysis tools have you used before? (check all that apply)” under the “Life Sciences Image Analysis” section.

B) Answers to the checkbox question “What image analysis tools do you use the most?” under the “Life Sciences Image Analysis” section.

C) Answers to the question “How frequently do you use scripting to solve image analysis problems?” by ‘Life Science’ participants.

D) Answers to the multiple-choice question “What image analysis tools have you used before? (check all that apply)” under the “Physical Sciences Image Analysis” section.

E) Answers to the checkbox question “What image analysis tools do you use the most?” under the “Physical Sciences Image Analysis” section

F) Answers to the question “How frequently do you use scripting to solve image analysis problems?” by ‘Physical Science’ participants.

Figure 5

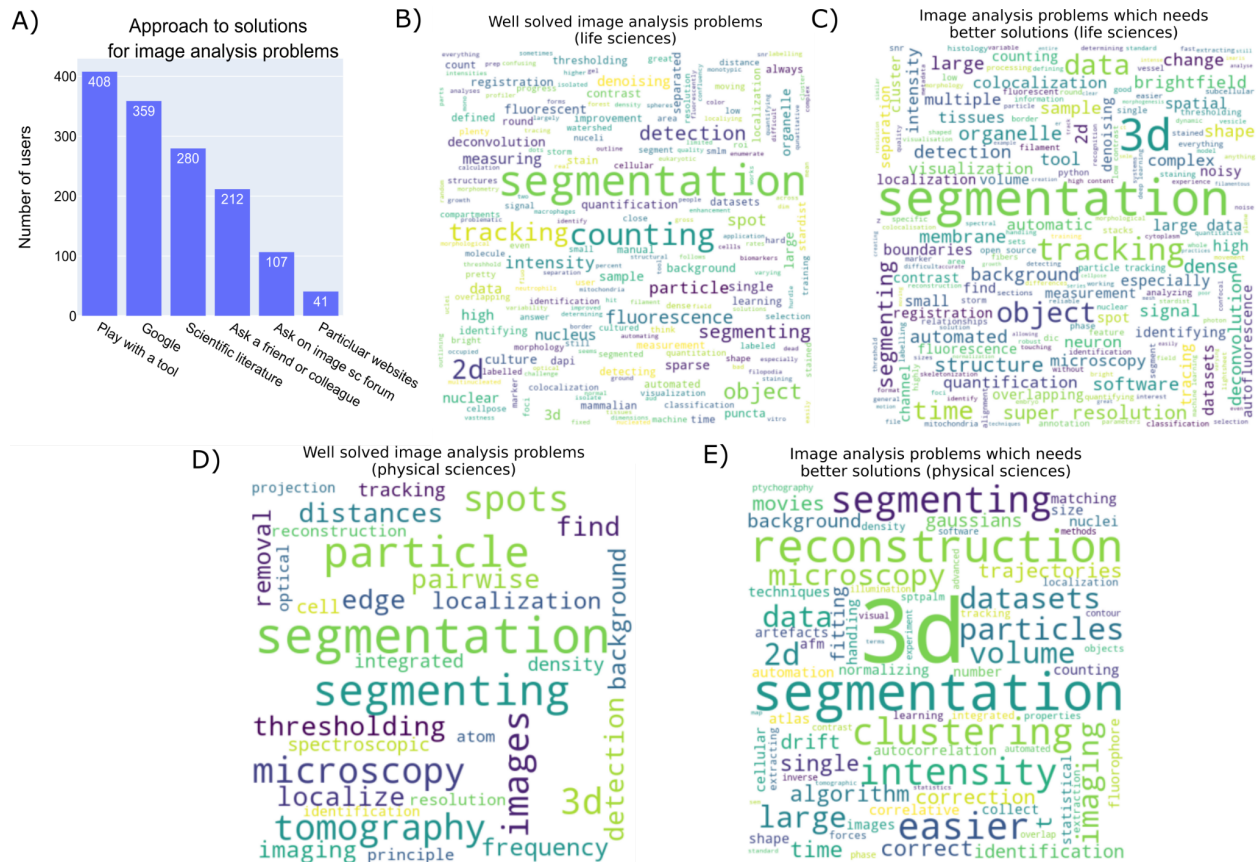


Figure 5- Solving image analysis problems

- Answers to the checkbox question “How do you generally go about solving an image analysis problem? Check the approach(es) you use the most.”
- Word Cloud representation of the unigrams of the answers by ‘life science’ participants to an open-ended question “What image analysis problems (i.e. finding nuclei, tissue analysis, analysis of super-resolution data, etc) do you think are generally well-solved?”
- Word Cloud representation of the unigrams of the answers by ‘life science’ participants to an open-ended question “What image analysis problems (i.e. finding nuclei, tissue analysis, analysis of super-resolution data, etc) do you wish had easier/better solutions?”
- Word Cloud representation of the unigrams of the answers by ‘physical science’ participants to an open-ended question “What image analysis problems (i.e. finding nuclei, tissue analysis, analysis of super-resolution data, etc) do you think are generally well-solved?”

E) Word Cloud representation of the unigrams of the answers by 'physical science' participants to an open-ended question "What image analysis problems (i.e. finding nuclei, tissue analysis, analysis of super-resolution data, etc) do you wish had easier/better solutions?"

Figure 6

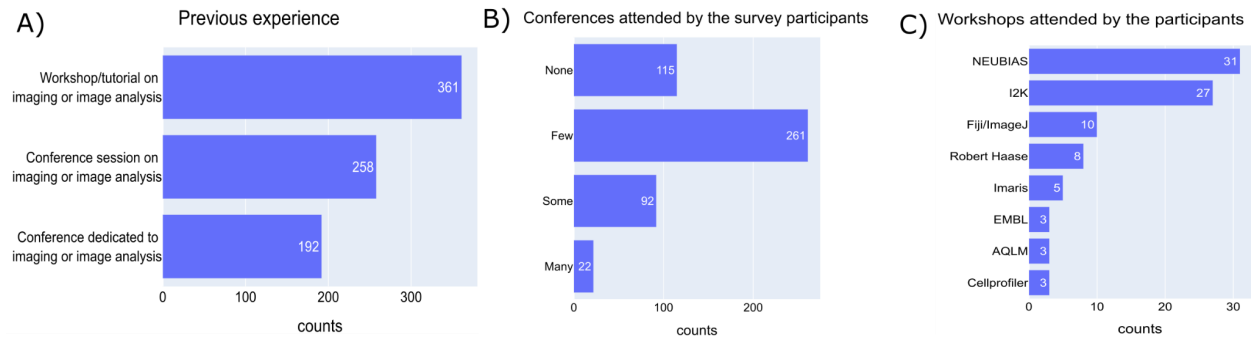


Figure 6 - Experience in image analysis

- A) Answers to a multiple-choice question “Please select any of the following you have attended in the past”
- B) Answers to the checkbox question “Are there any image analysis workshops, tutorials, or conferences you are aware of and attended or considered attending? If so, how many?”
- C) Answers to an open-ended question “Are there any image analysis workshops, tutorials, or conferences that you have participated in and found particularly helpful? If yes, what made them beneficial?”

Figure 7

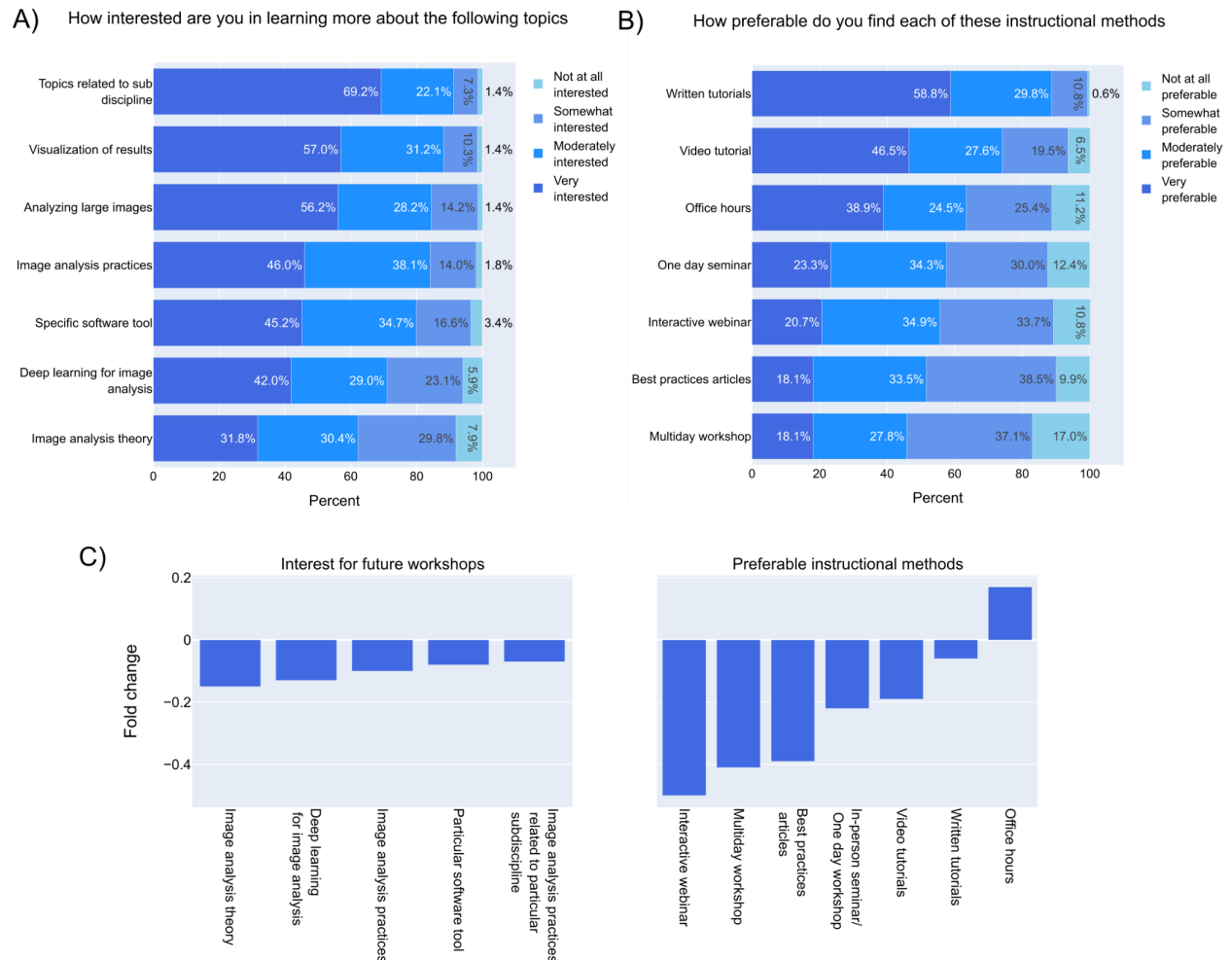


Figure 7- The imaging community prefers to learn about customized image analysis tools at their own pace

- A) Answers to a multiple-choice grid question “How interested are you in learning more about the following topics?”
- B) Answers to a multiple-choice grid question “In regards to learning more about image analysis, how preferable do you find each of these instructional methods?”
- C) Percent change in the “Very interested/Very preferable” category of part A and B compared to the results from the 2020 bioimage analysis survey.

Figure 8

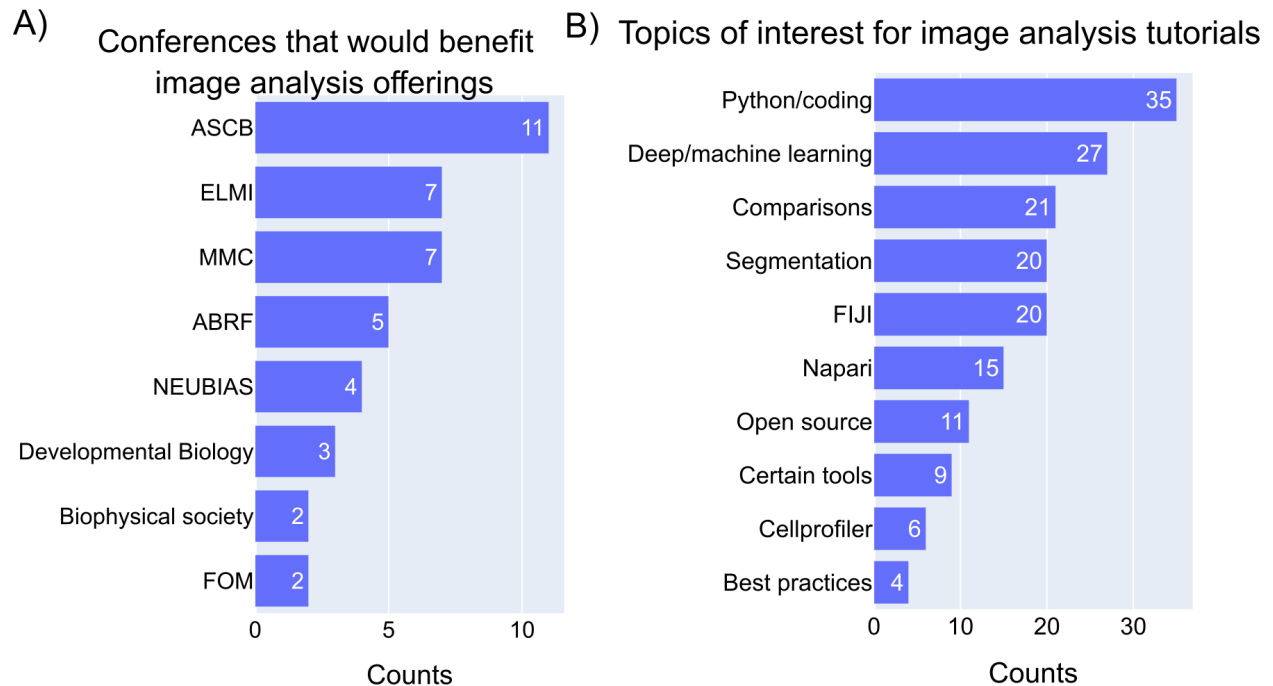


Figure 8- Topics of interest for the image analysis sessions in a conference

- A) Answers to an open-ended question “Are there any image analysis workshops, tutorials, or conferences that you have participated in and found particularly helpful? If yes, what made them beneficial?” Unigrams and bigrams were counted from the answers and the meaningful words were plotted.
- B) Answers to an open-ended question “What specific topics (i.e. overviews of a particular tool, comparisons between pieces of software, or how to use a certain tool for a certain kind of experiment) would you like to see prioritized for future image analysis workshop and tutorial offerings?” Unigrams and bigrams were counted from the answers and the meaningful words were plotted.

Figure 1: Word clouds illustrating the role of the user in the learning process. The figure is divided into four quadrants: A) Creators role-Imaging, B) Users role-Imaging, C) Creators role-Balanced, and D) Users role-Analyst. Each quadrant contains a word cloud of terms related to the learning process, with colors corresponding to the four roles: Imaging (blue), Balanced (green), Analyst (red), and Creator (yellow).

A) Creators role-Imaging

B) Users role-Imaging

C) Creators role-Balanced

D) Users role-Analyst

A) Answers to an open-ended question “What do you think analysis tool CREATORS (such as software developers) could/should do to make image analysis better and more successful? How best could we encourage them to do it?” was categorized based on the “work type” as described in supplementary S2A and the unigrams of the answers are represented as word clouds. Word clouds were made with a cut-off of 50 words.

B) Answers to an open-ended question “What do you think analysis tool USERS (such as microscopists) could/should do to make image analysis better and more successful? How best could we encourage them to do it?” was categorized based on the “work type” as described in supplementary S2A and the unigrams of the answers are represented as word clouds. Word Clouds were made with a micut-off of 50 words.

Supplementary Figures

☰ Supplementary file - Survey paper J Mic

Associated links

📄 Cover letter - Survey Paper J Mic.pdf

☰ 2024(?)ImageAnalysisSurvey

https://github.com/COBA-NIH/2023_ImageAnalysisSurvey/tree/main

Notes for revision

- Kevin requested comparison to 2020 survey results for fraction using the forum by job category; pretty sure we didn't do that. Would be a nice to add
- Beth had these papers open in her tab group for her paper on improving education - I think the idea was just to be like "let's learn from these people", but should revisit in revision in case past-her had a good idea that current-her has forgotten
<https://arxiv.org/abs/2305.06213> <https://www.bikeprinciples.org/>
<https://www.biorxiv.org/content/10.1101/2023.03.10.531570v2.full>
<https://osf.io/preprints/socarxiv/7d2yx/>