MACHINE LEARNING
Science and Technology

PURPOSE-LED PUBLISHING™

Part of the MACHINE LEARNING Series

CHALLENGES • OPEN ACCESS

Mic-hackathon 2024: hackathon on machine learning for electron and scanning probe microscopy

To cite this article: Utkarsh Pratiush *et al* 2025 *Mach. Learn.: Sci. Technol.* **6** 040701

IOP Publishing Mach. Learn.: Sci. Technol. 6 (2025) 040701 https://doi.org/10.1088/2632-2153/ae1f5d

MACHINE LEARNING Science and Technology

OPEN ACCESS

RECEIVED
29 June 2025

REVISED
16 October 2025

ACCEPTED FOR PUBLICATION
13 November 2025

PUBLISHED
2 December 2025

CHALLENGES

Mic-hackathon 2024: hackathon on machine learning for electron and scanning probe microscopy

Utkarsh Pratiush[1,41,*], Austin Houston[1], Kamyar Barakati[1,41], Aditya Raghavan[1,41], Ralph Bulanadi[16,41], Xiangyu Yin[26,41], Samuel S Welborn[12,41], Dasol Yoon[10], Harikrishnan K P[11], Zhaslan Baraissov[11], Desheng Ma[11], Mikolaj Jakowski[15], Shawn-Patrick Barhorst[15], Alexander J Pattison[13], Panayotis Manganaris[39], Sita Sirisha Madugula[2], Sai Venkata Gayathri Ayyagari[14], Vishal Kennedy[15], Michelle Wang[17], Kieran J Pang[18], Ian Addison-Smith[19,21], Willy Menacho[20,21], Horacio V Guzman[20,21], Alexander Kiefer[15], Nicholas Furth[15], Nikola L Kolev[35], Mikhail Petrov[38], Viktoriia Liu[40], Sergey Ilyev[36], Srikar Rairao[15], Tommaso Rodani[37], Ivan Pinto-Huguet[22], Xuli Chen[22], Josep Cruañes[22], Marta Torrens[22], Jovan Pomar[22], Fanzhi Su[22], Pawan Vedanti[23], Zhiheng Lyu[24], Xingzhi Wang[24], Lehan Yao[3], Amir Taqieddin[25], Forrest Laskowski[25], Yu-Tsun Shao[27], Benjamin Fein-Ashley[28], Yi Jiang[26], Vineet Kumar[29], Himanshu Mishra[29], Yogesh Paul[30], Adib Bazgir[32], Rama Chandra Praneeth Madugula[31], Yuwen Zhang[32], Pravan Omprakash[33], Jian Huang[33], Eric Montufar-Morales[33], Vivek Chawla[1], Harshit Sethi[34], Jie Huang[34], Lauri Kurki[34], Grace Guinan[4], Addison Salvador[4,5], Arman Ter-Petrosyan[6], Madeline Van Winkle[4], Steven R Spurgeon[4,7,8], Ganesh Narasimha[2], Zijie Wu[2], Richard Liu[1], Yongtao Liu[2,41], Boris Slautin[9], Andrew R Lupini[2], Rama Vasudevan[2], Gerd Duscher[1] and Sergei V Kalinin[1,3,41,*]

1 Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, United States of America
2 Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
3 Pacific Northwest National Laboratory, Richland, WA 99354, United States of America
4 National Renewable Energy Laboratory, Golden, CO 80401, United States of America
5 University of Cincinnati, Cincinnati, OH 45221, United States of America
6 University of California Irvine, Irvine, CA 92697, United States of America
7 Metallurgical and Materials Engineering Department, Colorado School of Mines, Golden, CO 80401, United States of America
8 Renewable and Sustainable Energy Institute (RASEI), University of Colorado, Boulder, Boulder, CO 80309, United States of America
9 Institute for Materials Science and Center for Nanointegration, Duisburg-Essen (CENIDE), University of Duisburg-Essen, Essen 45141, Germany
10 Department of Materials Science and Engineering, Cornell University, Ithaca, NY, United States of America
11 School of Applied and Engineering Physics, Cornell University, Ithaca, NY, United States of America
12 National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, United States of America
13 Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America
14 Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, United States of America
15 University of Tennessee, Knoxville, TN, United States of America
16 Department of Quantum Matter Physics, University of Geneva, 1211 Geneva, Switzerland
17 Department of Electrical and Photonics Engineering, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
18 Department of Experimental Psychology, Justus Liebig University Giessen, 35394 Giessen, Germany
19 Department of Mechanical Engineering, Universidad de Chile, Beauchef, 851 Santiago, Chile
20 Institut de Ciència de Materials de Barcelona, CSIC, 08193 Barcelona, Spain
21 Biophysics and Intelligent Matter Lab, E-08193 Barcelona, Spain
22 Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, United Kingdom
23 Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, United States of America
24 Department of Materials Science and Engineering, University of Illinois at Urbana—Champaign, Urbana, IL 61801, United States of America
25 Technology Integration—Materials Informatics and Modeling Department, Solid Power Operating Inc, Louisville, CO 80027, United States of America
26 Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, United States of America
27 Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089, United States of America
28 Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, United States of America
29 Department of Surface and Plasma Science, Faculty of Mathematics and Physics, Charles University, 18000 Prague 8, Czech Republic
30 Institute for Neuromodulation and Neurotechnology, University Hospital and University of Tuebingen, Tübingen, Germany
31 Department of Mechanical Engineering, New York University, New York, NY 10012, United States of America
32 Department of Mechanical and Aerospace Engineering, University of Missouri-Columbia, Columbia, MO 65211, United States of America

[33] Institute of Material Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, United States of America
[34] Department of Applied Physics, Aalto University, Helsinki, Otaniemi, Espoo FI-02150, Finland
[35] University College London, Gower St, London WC1E 6BT, United Kingdom
[36] Moscow Institute of Physics and Technology, Institutskiy Pereulok, 9, Dolgoprudny, Moscow Oblast 141701, Russia
[37] AREA Science Park, Università degli Studi di Trieste, Località Padriciano, 99, 34149 Trieste, TS, Italy
[38] Tufts University, 419 Boston Ave, Medford, MA 02155, United States of America
[39] Department of Nuclear Engineering, North Carolina State University, Raleigh, NC 27695-7213, United States of America
[40] Aspiring Scholars Directed Research Program, Fremont, CA, United States of America
[41] Contributors who helped in responding to reviewers.
[*] Authors to whom any correspondence should be addressed.

E-mail: upratius@vols.utk.edu and sergei2@utk.edu

Keywords: hackathon, machine learning, scanning probe microscopy, electron microscopy

Supplementary material for this article is available online

Abstract

Microscopy is one of the primary sources of information on materials structure and functionality at the nanometer and atomic scales. The data generated through microscopy is often contained in well-structured datasets, enriched with extensive metadata and sample histories, although not always with the same level of detail or storage format. The broad incorporation of data management plans by major funding agencies ensures the preservation and accessibility of this data. However, deriving insights from these rich datasets remains challenging due to the lack of established code ecosystems, standardized benchmarks, and integration strategies. Correspondingly, the efficiency of data usage is very low, and time expenditures at the analysis stage are enormous. In addition to post-acquisition data analysis, the emergence of application programming interfaces by major microscope manufacturers now creates opportunities for real-time ML-based data analytics to enable automated decision making, and particularly ML-agent controlled real-time microscope operation. Despite these opportunities, there is a significant gap in integrating the ML community with the broader microscopy community, limiting the value that these methods bring to physics and materials discovery and materials optimization. Hackathons address these challenges by fostering collaboration between ML experts and microscopy professionals, encouraging the development of innovative solutions that leverage ML for microscopy and preparing the workforce of the future both for microscopy-intensive domains areas, instrument manufacturers, and ML scientists interested in real world applications for fundamental research, materials optimization, and manufacturing. The hackathon generated benchmark datasets and digital twins of microscopes that further contribute to the development of the field and establish data analysis ecosystems. All the codes can be found at GitHub(https://github.com/KalininGroup/Mic-hackathon-2024-codes-publication/tree/1.0.0.1) and Zenodo (https://zenodo.org/records/15579940).

1. Opportunity

Microscopy and more generally physical and chemical imaging form one of the primary toolsets for exploring matter in fields ranging from materials science and condensed matter physics to chemistry, biology, electrochemistry, and their extensive subfields [1]. Many of the imaging tools such as electron microscopy (EM) e-beam diffraction [2, 3], certain modalities of scanning probe microscopy (SPM) [4, 5] and time-of-flight secondary ion mass spectrometry [6, 7] offer high-fidelity information on local properties and compositions. Structural imaging methods such as scanning transmission EM (STEM) [8, 9] allow visualization of the structure of matter atom by atom with few picometer precision [10],sufficient for direct mapping of the chemical expansion in electrochemically active materials [11–13] and ferroelectric [14, 15] and ferroelastic [16, 17] order parameter fields. Overall, microscopy datasets contain a wealth of information on materials structure and functionalities, processing histories, fundamental physical laws governing materials formation and property emergence, and hence are the keys for materials and physics discovery and materials optimization.

Yet whereas scientific domains such as astronomy, genomics, mass-spectrometry, x-ray scattering [18, 19], thermodynamics, and many others have by now a well-established tradition of data analysis and integration for downstream applications, this is largely absent in microscopy. While simulation frameworks [20–25] are ample in scanning probe and particularly EM fields, data analysis ecosystems are scarce and disjointed. We identify the two primary factors behind this as (a) extreme heterogeneity of experimental microscopy workflows spanning fields from condensed matter and quantum physics to biology and (b) resultant dearth of downstream applications. In other words, no subfield of microscopy (outside of cryo-EM [26–28] and electron diffraction) has critical mass of researchers and applications within a single group required to build robust data analysis and software ecosystems. We pose that ML/AI methods are now positioned to change this paradigm and organized the first Hackathon on ML for Microscopy on 16–17 December 2024 (with the second planned for December 2025) to spark this transition by bringing together microscopists from multiple communities and organizations together with ML experts and enthusiasts to learn and build the ML-based solutions for electron and probe microscopy.

1.1. Microscopy matters

Electron and scanning probe Microscopies are foundational tools in materials science, condensed matter physics, chemistry, catalysis, electrochemistry, and other fields. STEM allows imaging and characterization of materials at the nanoscale extending to atomic resolution, providing insights into the atomic and molecular structures of materials. By now, STEM has become one of the primary tools in multiple academic and industrial research labs worldwide [9, 29–33]. The versatility of STEM is further enhanced by its integration with techniques such as electron energy loss spectroscopy (EELS) [34–36], which allows for precise analysis of the chemical composition [37],electronic structure [38] of materials, and low energy quasiparticles [39]. The ability of STEM to image materials at nanometer and atomic levels makes it a crucial tool for advancing understanding of the structure property relationship in wide range of material systems. This combination is particularly beneficial in the development of new materials for technological applications, including semiconductors, solar cells [40],catalysts [41],and battery materials [42].

Similarly, the extensive application of SPM has opened the doors to explore and modify the nanoworld. Compared to other materials characterization tools, SPM offers a desktop footprint, low cost, and versatility in operating in multiple environments [43, 44]. It provides a wide range of functional imaging capabilities, extending from basic topographic imaging [45–49] to probing of electronic [46, 50], magnetic [45, 51, 52], mechanical [45, 46], biological [49, 53–57], and chemical [58, 59] properties. Furthermore, SPM supports multiple spectroscopy techniques in a variety of imaging modes, enabling comprehensive understanding and manipulation of materials at the nanoscale [60–62] and exploring phenomena such as single-molecule chemical transformation in biomolecules, polarization switching phenomena in ferroelectrics, and local electrochemical activity in a broad range of energy materials from batteries to fuel cells.

1.1.1. ML for microscopy: single point data analysis

Microscopy tools now offer almost unbounded sources of information on the structure and property of matter in the form of images, spectra, and hyperspectral imaging data. Even for simple imaging modes, modern cameras and detectors offer capabilities up to 32k $\times$ 32k pixel images, which are already above the capability of a human operator to usefully examine. Similarly, high-dimensional data sets that are common across scanning probe and electron microscopies are generally outside the human ability to comprehend and interpret, necessitating lengthy iterative analyses. Currently, the efficiency of data utilization in microscopy is extremely low, often with 2–3 best images from a day of work being analyzed for publications of downstream applications. The data analysis itself often relies on custom multistep workflows developed based on operator intuition and best practices and often take weeks to perfect. These factors lead to enormous hidden inefficiencies and strong operator biases in microscopy use. Compared to success stories such as molecular structure discovery by CryoEM, they also suggest the tremendous potential to increase efficiency and impact of imaging tool on broad range of scientific disciplines if data analysis and potentially data acquisition methods are brought to the intrinsic physical limits of instruments.

Like other domains, the growth of data science and machine learning (ML) have stimulated interest into big data and ML methods in EM. A number of early works have been reported in 80ies and 90ies, including that of (co-author) Duscher [63] and the especially visionary work of Noel Bonnett [64]. However, at that time, the computational capabilities and hence potential to work with large data volumes was limited, as were the libraries of available data analysis tools.

In the general computer science community, the ML ecosystem has been growing exponentially for the last two decades, with new network architectures, methodologies, etc. becoming available almost monthly. The development of deep learning in 2012 has become an inflection point that brought these developments to the attention of the broader scientific community. As a natural sequence, disparate microscopy communities started to adopt these methods for applications such as semantic segmentation of images, unsupervised learning over spectral and imaging data, and multiple other applications. The first sustained effort in ML for EM can be dated to the work of Watanabe and Williams [65] in ∼2005. At that time, grid-based spectroscopic measurements had become common on many tools, and the computation power and codes sufficient for simple data analytics were starting to become available outside of the computer science community. From that moment, the combination of hyperspectral imaging and ML for physics extraction and dimensionality reduction has become the new paradigm in STEM-EELS [35, 66, 67], 4DSTEM [68, 69], EDS [70, 71], and spectroscopic imaging techniques in SPM. The field has been steadily developing ever since, and several comprehensive reviews on ML based analysis of STEM—EELS data have become available recently [72, 73]. It is important to note that the range of data analysis tasks in microscopy goes well beyond simple dimensionality reduction and image segmentation, and requires active learning methods, integration of microscopy data for downstream physics discovery as well as upstream experiment planning. As mentioned above, despite close similarity between analysis tasks performed on dissimilar microscopies, currently the scientific landscape is dominated by software developed in individual groups, typically lacking benchmarks or community-wide development. The few exceptions reporting benchmarking for microscopy tasks have become available only recently [74].

1.1.2. ML for microscopy: active experiments

Any SPM and STEM operator is well familiar with the classical scan paradigm, and at some point, asked the question whether rectangular scanning and grid-based hyperspectral imaging orchestrated by a human operator are indeed the only or the best way to explore new materials. The progress of big data methods in areas such as robotic vision [75] and autonomous driving [76] brings forth the question as to whether similar methods [77–80] can be useful for building automated microscopes. So far, these were preponderantly realized in the form of workflows in which execution of the codes is driven by immediately available targets via fixed policies. For example, this can include the use of the deep convolutional networks or simpler image analysis tools for the identification of the *a priori* known object of interest such as atoms in scanning tunneling microscopy [81],identification of single DNA molecules [82],spectroscopy of grain boundaries [83],and ferroelectric domain walls [84–88]. These developments are paralleled by the development of sampling methods such as compressed sensing [89]. More complex examples entail inverse workflows, in which the goal is to discover the structural features that maximize the desired aspect of the spectral response [90]. The number of examples of ML integration into active microscopy workflows has been growing rapidly over the last 2–3 years [60, 91–95]. However, despite several early successes in active learning in SPM and STEM [68, 96], the amount of work on active learning is so far very limited, due to relatively early stages of the control application programming interfaces (APIs), challenges of real time (as compared to post-acquisition) data analysis, and particularly lack of workflow design tools and experience across microscopy communities.

1.2. Data and software ecosystems

Complementary to post-acquisition data analysis and implementation of on-the-fly data analytics and active learning on individual instruments in support of single operator work, data integration across multiple data generation facilities and teams brings additional opportunities. Below, we summarize these in the light of the current open data mandates.

1.2.1. Data management plans (DMPs)

For almost a decade, major funding agencies, including the National Science Foundation and the Department of Energy (DOE), have mandated the inclusion of DMPs in research proposals. A DMP is a detailed document outlining how data generated during the research project will be handled, preserved, and shared. It ensures that data is stored in accessible formats, with appropriate metadata to facilitate future use by other researchers. This requirement aims to establish a culture of data sharing and transparency in scientific research, implementing FAIR principles across multiple domains.

By now, there is broad acceptance within the scientific community that data should be shared online. This practice enhances reproducibility, enables further discoveries, and maximizes the return on investment in research. However, while data sharing is slowly becoming the norm, there remains a significant need for methods that enable downstream use of this data. For fields such as x-ray crystallography, genomics, and thermodynamics, well-established software ecosystems exist to facilitate the analysis and utilization of shared data. Projects like the materials project and nomad (by FAIRmat) have further improved the ease of storing, accessing and applying data and connecting it across domains, and by now some are sufficiently advanced to be used as a part of graduate and undergraduate level education.

For example, one of the authors regularly uses the materials project API as a part of his *Intro to ML in Materials Science* course.

In contrast, the field of microscopy has yet to develop a comparable software ecosystem for the downstream use of its data. Despite the extensive datasets generated through advanced microscopy techniques, researchers often face challenges in extracting meaningful insights due to the lack of standardized tools and benchmarks. The lack of shared data precludes development of the downstream use applications, and the lack of the latter in turn disincentivizes data sharing (note that simply sharing the data and sharing the data in the form that encourages downstream use are very different). This gap highlights the need for concerted efforts to develop and integrate ML and artificial intelligence (AI) tools specifically tailored for microscopy data to establish a virtuous circle of downstream applications and data sharing.

1.2.2. Software ecosystem

Progress in ML in different scientific applications has been largely driven by the availability of software frameworks, many of which were initially developed by industry (the main examples being TensorFlow and PyTorch). Within the microscopy domain, given the strong heterogeneity of the space, packages exist for certain sub-domains but are virtually all utilizing classical/traditional analysis methods. These packages include efforts in electron microcopy, such as hyperspy [97],and in atomic force microscopy, such as with TopoStats [98]. However, packages to develop ML workflows for microscopy datasets are limited (notable examples include AtomAI and GPax [99]).

In parallel to development of the specialized ecosystems for microscopy data analysis, certain successes have been achieved with the broad use tools developed by ML community. For example, GUIs that can be easily used to train models for image segmentation and object detection abound, such as iLastik [100] and RoboFlow [101],are available. The recent push towards foundation models has also seen the deployment of universal segmentation models, such as the 'SAM' and 'SAM2' model from META [102]. The limitation of these is that they are optimized for real-space imaging from ordinary cameras, and do not always function as expected on microscopy datasets (e.g. atomically resolved images will be identified as chainmail or cloth).

Perhaps more than software codebases, a fundamental challenge is that benchmarking a new ML workflow or architecture remains difficult in the current ecosystem. We require better digital twins and 'ground truth' datasets, on which typical applications such as segmentation, reconstruction, and more advanced human-in-the-loop workflows can be reliably tested to gauge the performance of newly developed algorithms.

1.2.3. APIs

APIs are the key bridges between microscopy and codes that enable the application of ML for automated microscopy. Only a few years ago these were generally lacking across microscopy manufacturers. While there were some exceptions, such as Nion [103] in EM and Nanonis [104] in SPM, what APIs existed were often not well known or well documented, and third party tools such as SerialEM(https://bio3d.colorado.edu/SerialEM/) or DigitalMicrograph(www.dmscripting.com/publications.html) were frequently used. Fortunately, this trend is now changing drastically, with both availability and functionality being improved.

For SPM, the programming interface API from SpecsGroup [105] is built for LabView [106] and requires another translation layer to work with Python and ML. The Python API from NanoSurf [107] offers direct communication to Python codes, but still does not provide full control and data access. On the user side, there have also been attempts to interface microscopy with ML with APIs. The AEcroscopy API [108] developed at ORNL enables the full control of SPM manufactured by Asylum Research [109] but requires custom hardware. The DeepSPM API [110] is based on the programming interface by SpecsGroup and thus only works Nanonis controllers. While the official APIs for SPMs are still scarce and come with limited control and data I/O for their microscopes, there is a rapid trend to their emergence and operationalization, and corresponding pressure from the customer community to build and share the in-house versions.

The APIs have similarly become more advanced in the EM community over the last several years. Companies such as Thermo Fisher Scientific [111], Nion [103], and JEOL [112] have recently introduced APIs that allow for hardware control through code. However, there is a need to increase community awareness of these tools and their potential benefits for ML. Similarly, the APIs built with different custom hardware and built for different brands prevent the share of ML workflows across the microscopy community. This hackathon will play a pivotal role in fostering this understanding and adoption and stimulate the development of pure-software generic API that works universally on different microscopy has been long awaited and can promote the development of automated microscopy.

1.3. Summary and needs

Overall, the microscopy community has high anticipations for ML and has demonstrated multiple use cases for the applications of ML, both at the stage of post-acquisition data analysis, real time data analytics, and even ML/AI driven decision making during the experiment. However, these efforts are largely fragmented across the community, prototype workflows are developed within individual groups, and generally do not form large projects and software ecosystems required for benchmarking, cross-task and cross-domain use, and generally mature downstream applications.

The outstanding need in the field now is to integrate the emerging efforts in ML for scanning probe and EM. This is possible since the basic principles of operation, instrument control hyper languages, and many data analysis workflows are similar between electron and scanning robe techniques. Similarly, there is a clear opportunity to bridge the ML and microscopy communities, where the availability of APIs offers an advantage for real-world testing of the ML algorithms and microscope serving a prototype model for much more complex decision making and real-world operation environments such as robotics. The proposed hackathon aims to address this need by engaging students to develop a new ML-conversant workforce in microscopy, start building consensus on benchmark datasets, standard implementations, and approaches, developing multi-institution partnerships and a nexus of talent to seed future collaboration and development, and make a value proposition for manufacturers, community and funding agencies toward broader adoption.

Note based on the discussion with the reviewers': sections corresponding to value proposition of such events in the community, organization details and winning teams/prizes has been moved to the supplementary part of the manuscript.

2. Project summaries

The participants submitted ∼20 projects, with a total of 80 active participants. The full summaries, codes, and YouTube presentations are available on the hackathon website. Here, we provide general overview of the submitted projects. The Appendix has individual writeups of each project. Figure 1 shows some example workflows realized in the hackathon projects.

2.1. Image segmentation and shape analysis (Teams 3, 8, 10, 11)

General discussion on projects in this section

Microscopy images allow us to visualize structures at micro, nano, and atomic scales, such as grain structures, nanoparticles, and defects in materials. These structures can affect the mechanical, electrical, and chemical properties of materials, as such, understanding these structural characteristics are crucial for optimizing and improving materials. However, traditionally it often relies on manual identification of interesting structures and manual analysis of structural correlations, which are time-consuming and prone to human bias. To overcome these limitations, four projects in this hackathon have integrated ML into microscopy analysis workflows to enable automated detection, classification, and analysis of structural features with enhanced efficiency and accuracy. Welborn *et al* applied cell segmentation model to automate the detection of grain structures in AFM images of halide perovskite thin films. The cell segmentation model cellSAM [113] was originally developed for biological cells but proved effective for grain structure analysis in functional materials. Additional efforts were made to sorting grains based on size and nearest neighbors to discover regions of interest for further zoom-in imaging measurement. Kiefer *et al* developed a python package, named micrography, to analyze EM images in order to find regions of interest. By constructing a graph representation of atom positions and their nearest neighbors, the project aimed to learn more meaningful relationships about materials and predict unseen information to guide the microscope focus on regions of interest. Pinto *et al* applied ML models to automatically identify the locations of nanoparticles and integrated a microscope navigator to actively search nanoparticles in the sample. They also developed an intuitive interface allowing users to view the nanoparticle search results in real time. By automating these tasks, the system significantly reduced the manual effort required in nanoparticle research. In another project, Wang *et al* developed a workflow for real-time TEM image analysis of nanoparticles. A U-Net segmentation model was applied to detect nanoparticles. The workflow enabled real-time analysis of nanoparticle area, aspect ratio, convexity, etc., followed by unsupervised classification of contour shapes. A common theme among these projects is the emphasis on streamlining features of interest identification in microscopy images. The integration of ML not only

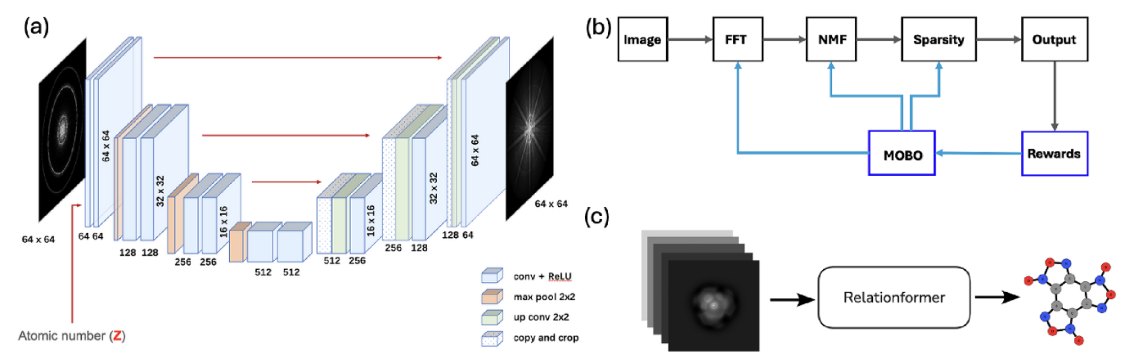

Figure 1. Example workflows. (a) Project 2: conditional U-Net architecture for predicting diffraction patterns incorporating thermal diffuse scattering. (b) Project 1: reward-driven image analysis workflow for automated phase mapping in polycrystalline Pd–Se. (c) Project 19: image-to-graph model for structure discovery from AFM images.

accelerates the identification of these features but also enables more sophisticated analysis, such as shape classification, defect prediction, and automated search experiment.

Highlight project in this section (Honorable Mention by judges) Automating AFM through model-driven image segmentation and classification:
Conventionally, microscopists rely on their expertise and intuition to steer a session—manually adjusting the acquisition region when interesting features appear on the monitor to capture higher-resolution data. In *Automating AFM through model-driven image segmentation and classification* (appendix No. 3), the team combined a foundation model and modern workflow tooling to reduce manual intervention and human bias during data collection. Leveraging *cellSAM* [114]—a segmentation model originally developed for identifying cells in optical microscopy—the team delineated grain structures in AFM images of halide perovskite thin films, whose granular morphology resembles clusters of cells. Post-processing routines sorted segmented grains by size, identified regions with high interface density, and extracted statistical descriptors of grain populations, enabling targeted follow-up measurements. These capabilities—spanning AFM data acquisition through pycroscopy's DTMicroscope(https://github.com/pycroscopy/DTMicroscope) package, segmentation with cellSAM, analysis, and automated region-of-interest selection—were deployed as containerized operators within the *interactEM*(https://github.com/NERSC/interactEM) workflow platform. As the microscope streams data between operators, the results were visualized in *interactEM*'s web-based interface, enabling both facile monitoring of the autonomous experiment and adjusting post-processing parameters in real time.

2.2. Optimization of image analysis workflows (Team 1, 4, 14)

General discussion on projects in this section
Sub-images of microscopy data are frequently used as inputs for ML analysis, where data augmentation (e.g. image cropping) is crucial for ML performance. Appropriate sub-image parameters, such as window size and step size, are essential for ensuring that relevant features are preserved while minimizing noise and redundancy. Three projects in this hackathon systematically investigated and optimized sub-image parameters to improve analysis accuracy. Barakati *et al* utilized a reward-driven optimization method to balancing compactness and separation of user defined objectives for classification. By analyzing the Pareto front, the authors can identify a window size providing best results of phase separation in a 2D polycrystalline Pd-Se film. Similarly, Guinan *et al* also optimized window size and step size to ensure effective classification of ferroelectric domain structures in STEM images of BiFeO3. Additionally, Wen *et al* analyzed the effects of window size and step increment on the predictive accuracy of im2spec to enhance the structure-property prediction.

Highlight project in this section (Student Council Award) Reward based segmentation: phase mapping of 2d polycrystalline pd-se phases:
The rapid progress of electron and scanning probe microscopies has created an urgent need for image analysis methods that can operate in real time, keeping pace with high-throughput and atomically resolved experiments. As an illustrative case, phase separation in polycrystalline thin films can be studied with DCNN-based supervised learning or unsupervised models: the former provides strong segmentation but depends on labeled datasets and is sensitive to distribution shifts, while the latter avoids labeling yet relies on complex, hyperparameter-heavy pipelines such as FFT decomposition or matrix factorization.

To establish a reliable basis for real-time image analysis, a reward-driven optimization framework was developed and validated using atomic-resolution STEM data of 2D polycrystalline Pd–Se thin films as a model system. The workflow employs sparse linear unmixing of FFT-transformed patches, followed by non-negative matrix factorization with progressively imposed sparsity to isolate dominant phases. Crucial hyperparameters including window size, sparsity level, and number of phases are dynamically tuned through physics-informed reward functions that balance objectives such as phase compactness and minimal overlap. By evaluating solutions along the Pareto front, the method achieves robust discrimination of structural phases and accurate identification of boundaries and interfaces. The framework developed in this project therefore shows strong potential for advancing real-time image analysis in microscopy, offering an explainable alternative to conventional supervised and unsupervised models, and providing a transferable strategy for automated decision-making across diverse imaging modalities.

2.3. Artefacts in SPM (Team 8 and 16)

General discussion on projects in this section

Unlike EM that uses an electron beam to scan samples, SPM uses a physical tip to scan samples. In this case, tip artifacts significantly impact the accuracy of SPM characterization, as these artifacts arise due to imperfections of probe tip (e.g. asymmetric or double tips) can obscure fine structural details, causing convolution effects that blur, duplicate, or alter the perceived surface structures and leading to misinterpretations of results. To address these problems, two projects in this hackathon leveraged ML approaches to reverse tip effect and reconstruct original surface structures. Nick *et al* trained three ML models with a mean absolute error loss and/or a structural similarity index loss to learn and reverse the distortions introduced by imperfect tips. They explored the feasibility of ML-based approach on tip shape estimation and image reconstruction. Narasimha *et al* tested multiple ML architectures for tip artifacts removal, including a U-Net model, a hybrid ResNet-UNet autoencoder, as well as point spread functions (PSFs) to predict tip conditions. The U-Net model performs well for moderate distortions. However, it over-sharpens the image or recovers wrong features under conditions with extreme bluntness or strong double-tip effects. The hybrid autoencoder struggles for reconstructing specific structures (e.g. spiral patterns). Both hybrid autoencoder and PSF approaches struggle for double tip effects. While both projects demonstrate success in reducing distortions, challenges remain in ensuring generalizability to reverse tip artifacts under diverse experimental conditions.

Highlight project in this section-Removal of tip-shape induced artifacts in AFM images using deep learning:

AFM provides critical nanoscale structural information, but the fidelity of its topographic images is often compromised by probe-tip artifacts such as bluntness, asymmetry, and double-tip effects. These distortions obscure fine surface features and limit quantitative interpretation, motivating the need for robust deconvolution approaches. In this project, we develop a deep learning–based workflow for tip-shape artifact removal, combining UNet and ResNet–UNet architectures with a novel PSF prediction step. Using a large dataset of simulated tip distortions and experimental AFM topographies, our models learn to recover ground-truth surfaces with high fidelity. Incorporation of PSF prediction into the deconvolution pipeline significantly improves reconstruction quality under severe tip degradation, outperforming conventional image-only models. This framework enables reproducible, automated correction of AFM images, enhancing data quality and enabling more reliable nanoscale analysis across a wide range of materials systems, and can be readily extended to other scanning probe techniques such as STM or KPFM for improved image reconstruction.

2.4. Physics-based analytics (Team 2, 5, 6)

General discussion on projects in this section

Like the tip conditions that are unseen in real-time experiments, many scientific problems struggle with challenges of analyzing crucial properties that are unmeasurable due to practical limitations or resource-intensive to assess using traditional methods. In microscopy experiments, for example, inelastic scattering in diffraction requires computationally expensive models to simulate and ferroelastic–ferroelectric correlations are too subtle to measure directly. To address this challenge, two projects in this hackathon leverage ML in combination with measurable data to infer these otherwise inaccessible properties. Yoon *et al* applied a U-Net model to predict inelastic scattering patterns in TEM using elastic diffraction patterns and atomic number distributions. Although the method requires further testing across diverse conditions, the results demonstrate the potential of ML to enable cost-effective prediction for inelastic scattering by learning the relationship between elastic and inelastic diffraction patterns. Bulanadi *et al* applies a generative adversarial network (GAN) to infer ferroelastic–ferroelectric domain structures from standard

AFM topography. Conventional piezoresponse force microscopy (PFM) measurements of ferroelectric domain structures require the application of a voltage to the sample, which can introduce unwanted modifications. GAN bypasses this issue by learning structural correlations between AFM topography and PFM imaging channels. Additionally, Addison-Smith *et al* applied a dynamic mode decomposition customized model with VampNets3 in high-speed AFM to map molecules below the substrates. These projects demonstrate how ML can offer new insights into material properties that are too complex or impractical to measure directly in traditional ways.

Highlight project in this section(1st Prize winner)GANder: Ferroelastic–Ferroelectric Domains Observed by Image-to-Image Translation:
Ferroelectric and ferroelastic materials are used extensively in transducers and sensors, but PFM, which is used in their study, is susceptible to noise, crosstalk, and irreversible sample damage through ferroelectric switching or charge injection. Methods that could classify ferroelectric domains or polarizations with neither external bias, nor tip–sample contact, could therefore be useful for research in ferroelectric materials. In ferroelectric–ferroelastic domains, where the ferroelectric polarization is directly linked to material structure, purely topographic information has the potential to encode information about ferroelectric properties. A GAN [115], which creates synthetic outputs through alternating competition between a discriminator model and a generator model, is thus trained on PFM data from a series of lead titanate samples [116] showing both purely ferroelectric c-domains (with an out-of-plane polarization and out-of-plane tetragonal distortions), and correlated ferroelectric–ferroelastic a-domains (with an in-plane polarization and similarly in-plane tetragonal distortions). When the GAN is used to convert experimental topography data to a synthetic piezoresponse amplitude output, orthogonal striations appear corresponding to the ferroelectric–ferroelastic a-domains observed in experimental piezoresponse amplitude data. However, oppositely polarized c-domains, which express identical tetragonal distortions to one another, and are prominent in experimental piezoresponse amplitude data but not experimental topography data, are not typically visible in the synthetic piezoresponse amplitude output. The GAN developed in this project therefore shows promise in extracting specifically ferroelectric–ferroelastic correlations in lead titanate, without requiring external bias or charge injection.

2.5. Automated experiment and LLM integration (Team 12, 15, 18)

General discussion on projects in this section
Language models have shown the potential to enhance microscopy experimentation [117]. Three projects in this hackathon also showcase the application of LLMs in microscopy. Yin *et al* developed an AutoScriptCopilot leveraging LLMs and vision language models for automation of TEM workflows. The AutoScriptCopilot allows users to dynamically adapt experiment parameters based on image quality assessment with the assistance of foundational AI models, balancing automation and expert intervention. Bazgir *et al* developed MicroscopyLLM-Bench that leverages state-of-the-art vision-language models to automate key microscopic imaging tasks, such as object detection, classification, and feature analysis. This approach reduces manual annotation efforts and accelerates analysis workflows of microscopy-based research. In addition, Kalinin *et al* used LLM to suggest possible descriptors of image patches for feature extraction.

Highlight project in this section (2nd Prize winner) AutoScriptCopilot: Agentic Workflow for TEM Experiment Automation:
AutoScriptCopilot implements a state machine based agentic workflow that couples Thermo Fisher Scientific's AutoScript instrument API with an orchestration layer [118] to coordinate instrument control, human inputs, foundation model assisted parameter recommendation, and image quality assessment. The system initializes the microscope and verifies readiness, executes acquisition under recommended settings, and then evaluates focus and astigmatism to detect drift or misalignment. When suboptimal quality is detected, it returns to alignment and retunes parameters. Decision nodes enable expert confirmation and error handling without interrupting automation, with states, decisions, and metadata logged to increase visibility and reproducibility. This modular graph-node design facilitates integration of AI with experimentation and has the potential to automate TEM experiment control and reduce ad hoc trial-and-error.

3. Summary

The hackathon served as a dynamic platform for the enhancement and improvement of scientific, engineering, and educational activities in several key ways. By bringing together experts in ML, EM, and SPM, the event fostered the interdisciplinary collaboration that is essential for addressing complex scientific challenges. Participants built teams and worked on real-world problems, such as robust segmentation, deionizing, and reconstruction of microscopy data, which are critical for advancing the field of nanoscience. The hackathon catalyzed the development of new ML tools and techniques specifically tailored for microscopy, enabling more accurate and efficient data analysis, which is fundamental to scientific discovery and innovation.

The hackathon pushed the boundaries of engineering by encouraging participants to develop and implement ML algorithms that can be integrated into microscopy hardware. This integration will lead to the creation of more advanced, autonomous instruments capable of real-time data processing and decision-making. Such innovations will not only enhance the capabilities of existing microscopy technologies but also pave the way for the development of next-generation instruments that can operate with greater precision and efficiency.

Finally, the hackathon contributed to educational activities by providing a hands-on, immersive learning experience for participants. It served as a unique training ground for students, early-career researchers, and professionals, offering them the opportunity to apply theoretical knowledge in a practical setting. The mentoring sessions, workshops, and tutorials helped participants build valuable skills in ML, data science, and microscopy. Additionally, the hackathon produced open-source resources, such as code libraries and educational materials, that can be widely disseminated and used for teaching and training in academic and research institutions. Most importantly, the hackathon laid the foundation for the development and use of digital twins, transitioning AI/ML from a post-acquisition data-analysis tool to an active participant in the research process.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/records/15579940.

Acknowledgments

The organizers gratefully acknowledge the support of Office of Naval Research, Thermo Fisher Scientific, and AI Tennessee initiative that made the hackathon possible. The PyCroscopy coding research received support from the Center for Nanophase Materials Sciences (CNMS), which is a US Department of Energy, Office of Science User Facility at Oak Ridge National Laboratory. This activity is (in part) sponsored by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This work was authored in part by the National Renewable Energy Laboratory (NREL) for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. M V, A S, and S R S were supported by the Operando to Operation (O2O) Laboratory Directed Research and Development (LDRD) program at NREL. G G was supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program. The views expressed in the presentation do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Organizers List

Sergei Kalinin (**advisor lead**)—University of Tennessee/Pacific Northwest National Laboratory.

Utkarsh Pratiush (**student lead**)—University of Tennessee.

Austin Houston—University of Tennessee.

Gerd Duscher—University of Tennessee.

Rama K. Vasudevan—Oak Ridge National Laboratory.

Boris Slautin—University of Duisburg-Essen.

Steven R. Spurgeon—National Renewable Energy Laboratory.

Yongtao Liu—Oak Ridge National Laboratory.

Yu (Richard) Liu—University of Tennessee.

Andy Lupini—Oak Ridge National Laboratory.

Author contributions

Utkarsh Pratiush 0009-0003-7249-103X
Conceptualization (equal), Data curation (lead), Investigation (equal), Software (lead), Writing – original draft (equal), Writing – review & editing (lead)

Sergei V Kalinin
Conceptualization (equal), Funding acquisition (lead), Project administration (equal), Resources (lead), Supervision (lead), Writing – original draft (equal), Writing – review & editing (equal)

Reference

[1] Franken L E, Grünewald K, Boekema E J and Stuart M C 2020 A technical introduction to transmission electron microscopy for soft-matter: imaging, possibilities, choices, and technical developments *Small* **16** 1906198
[2] Zou X, Hovmöller S and Oleynikov P 2011 *Electron Crystallography: Electron Microscopy and Electron Diffraction* (Oxford University Press)
[3] Bendersky L A and Gayle F W 2001 Electron diffraction using transmission electron microscopy *J. Res. Natl Inst. Stand. Technol.* **106** 997
[4] Loos J 2005 The art of SPM: scanning probe microscopy in materials science *Adv. Mater.* **17** 1821–33
[5] Kalinin S V, Vasudevan R, Liu Y, Ghosh A, Roccapriore K and Ziatdinov M 2023 Probe microscopy is all you need *Mach. Learn.: Sci. Technol.* **4** 023001
[6] Sodhi R N 2004 Time-of-flight secondary ion mass spectrometry (TOF-SIMS):—versatility in chemical and imaging surface analysis *Analyst* **129** 483–7
[7] Fearn S 2015 *An Introduction to Time-of-flight Secondary Ion Mass Spectrometry (Tof-sims) and Its Application to Materials Science* (Morgan & Claypool Publishers)
[8] Shibata N, Seki T, Sánchez-Santolino G, Findlay S D, Kohno Y, Matsumoto T, Ishikawa R and Ikuhara Y 2017 Electric field imaging of single atoms *Nat. Commun.* **8** 15631
[9] Krivanek O L *et al* 2010 Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy *Nature* **464** 571–4
[10] Yankovich A B, Berkels B, Dahmen W, Binev P, Sanchez S I, Bradley S A, Li A, Szlufarska I and Voyles P M 2014 Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts *Nat. Commun.* **5** 4155
[11] Kim Y M, Morozovska A, Eliseev E, Oxley M P, Mishra R, Selbach S M, Grande T, Pantelides S T, Kalinin S V and Borisevich A Y 2014 Direct observation of ferroelectric field effect and vacancy-controlled screening at the $BiFeO_3/LaxSr_{1-x}MnO_3$ interface *Nat. Mater* **13** 1019–25
[12] Kim Y M, He J, Biegalski M D, Ambaye H, Lauter V, Christen H M, Pantelides S T, Pennycook S J, Kalinin S V and Borisevich A Y 2012 Probing oxygen vacancy concentration and homogeneity in solid-oxide fuel-cell cathode materials on the subunit-cell level *Nat. Mater.* **11** 888–94
[13] Borisevich A Y, Morozovska A N, Kim Y M, Leonard D, Oxley M P, Biegalski M D, Eliseev E A and Kalinin S V 2012 Exploring mesoscopic physics of vacancy-ordered systems through atomic scale observations of topological defects *Phys. Rev. Lett.* **109** 065702
[14] Jia C L, Mi S B, Urban K, Vrejoiu I, Alexe M and Hesse D 2008 Atomic-scale study of electric dipoles near charged and uncharged domain walls in ferroelectric films *Nat. Mater.* **7** 57–61
[15] Jia C L, Nagarajan V, He J Q, Houben L, Zhao T, Ramesh R, Urban K and Waser R 2007 Unit-cell scale mapping of ferroelectricity and tetragonality in epitaxial ultrathin ferroelectric films *Nat. Mater.* **6** 64–69
[16] Borisevich A *et al* 2010 Mapping octahedral tilts and polarization across a domain wall in $BiFeO_3$ from Z-contrast scanning transmission electron microscopy image atomic column shape analysis *ACS Nano* **4** 6071–9
[17] Borisevich A Y *et al* 2010 Suppression of octahedral tilts and associated changes in electronic properties at epitaxial oxide heterostructure interfaces *Phys. Rev. Lett.* **105** 087204
[18] Li T, Senesi A J and Lee B 2016 Small angle x-ray scattering for nanoparticle research *Chem. Rev.* **116** 11128–80
[19] Ament L J, Van Veenendaal M, Devereaux T P, Hill J P and Van Den Brink J 2011 Resonant inelastic x-ray scattering studies of elementary excitations *Rev. Modern Phys.* **83** 705–67
[20] Oelerich J O, Duschek L, Belz J, Beyer A, Baranovskii S D and Volz K 2017 STEMsalabim: a high-performance computing cluster friendly code for scanning transmission electron microscopy image simulations of thin specimens *Ultramicroscopy* **177** 91–96
[21] Ophus C 2017 A fast image simulation algorithm for scanning transmission electron microscopy *Adv. Struct. Chem. Imaging* **3** 1–11
[22] Pryor A, Ophus C and Miao J 2017 A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy *Adv. Struct. Chem. Imaging* **3** 1–14
[23] Pelz P M, Rakowski A, DaCosta L R, Savitzky B H, Scott M C and Ophus C 2021 A fast algorithm for scanning transmission electron microscopy imaging and 4D-STEM diffraction simulations *Microsc. Microanal.* **27** 835–48
[24] Savitzky B H *et al* 2021 py4DSTEM: a software package for four-dimensional scanning transmission electron microscopy data analysis *Microsc. Microanal.* **27** 712–43
[25] Qian X and Villarrubia J S 2007 General three-dimensional image simulation and surface reconstruction in scanning probe microscopy using a dexel representation *Ultramicroscopy* **108** 29–42
[26] Li Y, Huang W, Li Y, Chiu W and Cui Y 2020 Opportunities for cryogenic electron microscopy in materials science and nanoscience *ACS Nano* **14** 9263–76
[27] Taylor K A and Glaeser R M 1976 Electron microscopy of frozen hydrated biological specimens *J. Ultrastruct. Res.* **55** 448–56
[28] Nogales E 2016 The development of cryo-EM into a mainstream structural biology technique *Nat. Methods* **13** 24–27
[29] Pennycook S J and Nellist P D 2011 *Scanning Transmission Electron Microscopy: Imaging and Analysis* (Springer Science & Business Media)
[30] Mkhoyan K, Babinec T, Maccagnano S, Kirkland E and Silcox J 2007 Separation of bulk and surface-losses in low-loss EELS measurements in STEM *Ultramicroscopy* **107** 345–55

[31] Williams D B, Carter C B, Williams D B and Carter C B 1996 *The Transmission Electron Microscope* (Springer)
[32] Muller D A 2009 Structure and bonding at the atomic scale by scanning transmission electron microscopy *Nat. Mater.* **8** 263–70
[33] Crewe A V 1974 Scanning transmission electron microscopy *J. Microsc.* **100** 247–59
[34] Lovejoy T, Corbin G, Dellby N, Hoffman M and Krivanek O 2018 Advances in ultra-high energy resolution STEM-EELS *Microsc. Microanal.* **24** 446–7
[35] Gázquez J, Sánchez-Santolino G, Biškup N, Roldán M A, Cabero M, Pennycook S J and Varela M 2017 Applications of STEM-EELS to complex oxides *Mater. Sci. Semicond. Process.* **65** 49–63
[36] Browning N, Wallis D, Nellist P and Pennycook S 1997 EELS in the STEM: determination of materials properties on the atomic scale *Micron* **28** 333–48
[37] Schneider R, Woltersdorf J and Röder A 1997 EELS nanoanalysis for investigating both chemical composition and bonding of interlayers in composites *Microchim. Acta* **125** 361–5
[38] Gloter A, Ewels C, Umek P, Arcon D and Colliex C 2009 Electronic structure of titania-based nanotubes investigated by EELS spectroscopy *Phys. Rev. B* **80** 035413
[39] Wu Y, Li G and Camden J P 2017 Probing nanoparticle plasmons with electron energy loss spectroscopy *Chem. Rev.* **118** 2994–3031
[40] Noircler G, Lebreton F, Drahi E, de Coux P and Warot-Fonrose B 2021 STEM-EELS investigation of c-Si/a-AlOx interface for solar cell applications *Micron* **145** 103032
[41] Qu J, Sui M and Li R 2023 Recent advances in *in-situ* transmission electron microscopy techniques for heterogeneous catalysis *IScience* **26** 7
[42] Yu L, Li M, Wen J, Amine K and Lu J 2021 TEM-EELS as an advanced characterization technique for lithium-ion batteries *Mater. Chem. Front.* **5** 5186–93
[43] Bian K, Gerber C, Heinrich A J, Müller D J, Scheuring S and Jiang Y 2021 Scanning probe microscopy *Nat. Rev. Methods Primers* **1** 36
[44] Hansma P K *et al* 1994 Tapping mode atomic force microscopy in liquids *Appl. Phys. Lett.* **64** 1738–40
[45] Binnig G, Quate C F and Gerber C 1986 Atomic force microscope *Phys. Rev. Lett.* **56** 930–3
[46] Binnig G, Rohrer H, Gerber C and Weibel E 1982 Surface studies by scanning tunneling microscopy *Phys. Rev. Lett.* **49** 57–61
[47] Ohnesorge F and Binnig G 1993 True atomic resolution by atomic force microscopy through repulsive and attractive forces *Science* **260** 1451–6
[48] Meyer G and Amer N M 1988 Novel optical approach to atomic force microscopy *Appl. Phys. Lett.* **53** 1045–7
[49] Giessibl F J 2019 The qPlus sensor, a powerful core for the atomic force microscope *Rev. Sci. Instrum.* **90** 011101
[50] Crommie M F, Lutz C P and Eigler D M 1993 Confinement of electrons to quantum corrals on a metal surface *Science* **262** 218–20
[51] Martin Y, Williams C C and Wickramasinghe H K 1987 Atomic force microscope–force mapping and profiling on a sub 100-Å scale *J. Appl. Phys.* **61** 4723–9
[52] Heinrich A J, Gupta J A, Lutz C P and Eigler D M 2004 Single-atom spin-flip spectroscopy *Science* **306** 466–9
[53] Ando T, Uchihashi T and Kodera N 2013 High-speed AFM and applications to biomolecular systems *Annu. Rev. Biophys.* **42** 393–414
[54] Drake B, Prater C B, Weisenhorn A L, Gould S A C, Albrecht T R, Quate C F, Cannell D S, Hansma H G and Hansma P K 1989 Imaging crystals, polymers, and processes in water with the atomic force microscope *Science* **243** 1586–9
[55] Dufrêne Y F, Martínez-Martín D, Medalsy I, Alsteens D and Müller D J 2013 Multiparametric imaging of biological systems by force-distance curve–based AFM *Nat. Methods* **10** 847–54
[56] Herruzo E T, Perrino A P and Garcia R 2014 Fast nanomechanical spectroscopy of soft matter *Nat. Commun.* **5** 3126
[57] Huber F, Lang H P, Backmann N, Rimoldi D and Gerber C 2013 Direct detection of a BRAF mutation in total RNA from melanoma cells using cantilever arrays *Nat. Nanotechnol.* **8** 125–9
[58] Florin E-L, Moy V T and Gaub H E 1994 Adhesion forces between individual ligand-receptor Pairs *Science* **264** 415–7
[59] Zhang R *et al* 2013 Chemical mapping of a single molecule by plasmon-enhanced Raman scattering *Nature* **498** 82–86
[60] Chen I J, Aapro M, Kipnis A, Ilin A, Liljeroth P and Foster A S 2022 Precise atom manipulation through deep reinforcement learning *Nat. Commun.* **13** 7499
[61] Leinen P, Esders M, Schütt K T, Wagner C, Müller K-R and Tautz F S 2020 Autonomous robotic nanofabrication with reinforcement learning *Sci. Adv.* **6** eabb6987
[62] Stroscio J A and Eigler D M 1991 Atomic and molecular manipulation with the scanning tunneling microscope *Science* **254** 1319–26
[63] Gatts C, Duscher G, Müllejans H and Rühle M 1995 Analyzing line scan EELS data with neural pattern recognition *Ultramicroscopy* **59** 229–39
[64] Trebbia P and Bonnet N 1990 EELS elemental mapping with unconventional methods I. Theoretical basis: image analysis with multivariate statistics and entropy concepts *Ultramicroscopy* **34** 165–78
[65] Watanabe M and Williams D 2004 Improvements of elemental mapping via x-ray spectrum imaging combined with principal component analysis and zero-peak deconvolution *J. Clin. Microbiol.* **10** 1040–1
[66] Jarausch K, Thomas P, Leonard D N, Twesten R and Booth C R 2009 Four-dimensional STEM-EELS: enabling nano-scale chemical tomography *Ultramicroscopy* **109** 326–37
[67] Gloter A, Badjeck V, Bocher L, Brun N, March K, Marinova M, Tencé M, Walls M, Zobelli A and Stéphan O 2017 Atomically resolved mapping of EELS fine structures *Mater. Sci. Semicond. Process.* **65** 2–17
[68] Roccapriore K M, Dyck O, Oxley M P, Ziatdinov M and Kalinin S V 2022 Automated experiment in 4D-STEM: exploring emergent physics and structural behaviors *ACS Nano* **16** 7605–14
[69] Bustillo K C, Zeltmann S E, Chen M, Donohue J, Ciston J, Ophus C and Minor A M 2021 4D-STEM of beam-sensitive materials *Acc. Chem. Res.* **54** 2543–51
[70] Shindo D, Oikawa T, Shindo D and Oikawa T 2002 *Energy Dispersive x-Ray Spectroscopy* (Springer) pp 81–102
[71] Hodoroaba V-D 2020 Energy-dispersive x-ray spectroscopy (EDS) *Characterization of Nanoparticles* (Elsevier) pp 397–417
[72] Cheng Z, Wang C, Wu X and Chu J 2022 Review *in situ* transmission electron microscope with machine learning *J. Semicond.* **43** 081001
[73] Botifoll M, Pinto-Huguet I and Arbiol J 2022 Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook *Nanoscale Horiz.* **7** 1427–77

[74] Wei J, Blaiszik B, Scourtas A, Morgan D and Voyles P M 2023 Benchmark tests of atom segmentation deep learning models with a consistent dataset *Microsc. Microanal.* **29** 552–62
[75] Andronie M, Lăzăroiu G, Karabolevski O L, Ștefănescu R, Hurloiu I, Dijmărescu A and Dijmărescu I 2022 Remote big data management tools, sensing and computing technologies, and visual perception and environment mapping algorithms in the internet of robotic things *Electronics* **12** 22
[76] Grigorescu S, Trasnea B, Cocias T and Macesanu G 2020 A survey of deep learning techniques for autonomous driving *J. Field Robot.* **37** 362–86
[77] Frazier P I 2018 Bayesian optimization *Recent Advances in Optimization and Modeling of Contemporary Problems* (Informs) pp 255–78
[78] Wilson A G, Hu Z, Salakhutdinov R and Xing E P 2016 Deep kernel learning *Proc. 19th Int. Conf. on Artificial Intelligence and Statistics, Proc. Machine Learning Research*
[79] Williams C K I and Barber D 1998 Bayesian classification with Gaussian processes *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 1342–51
[80] Rana S, Li C, Gupta S, Nguyen V and Venkatesh S 2017 High dimensional Bayesian optimization with elastic Gaussian process *Proc. 34th Int. Conf. on Machine Learning, Proc. Machine Learning Research*
[81] Ziatdinov M, Dyck O, Maksov A, Hudak B M, Lupini A R, Song J, Snijders P C, Vasudevan R K, Jesse S and Kalinin S V 2018 Deep analytics of atomically-resolved images: manifest and latent features (arXiv:1801.05133)
[82] Sotres J, Boyd H and Gonzalez-Martinez J F 2021 Enabling autonomous scanning probe microscopy imaging of single molecules with deep learning *Nanoscale* **13** 9193–203
[83] Liu Y, Yang J, Lawrie B J, Kelley K P, Ziatdinov M, Kalinin S V and Ahmadi M 2023 Disentangling electronic transport and hysteresis at individual grain boundaries in hybrid perovskites via automated scanning probe microscopy *ACS Nano* **17** 9647–57
[84] Liu Y, Kelley K P, Funakubo H, Kalinin S V and Ziatdinov M 2022 Exploring physics of ferroelectric domain walls in real time: deep learning enabled scanning probe microscopy *Adv. Sci.* **9** 2203957
[85] Liu Y, Kelley K P, Vasudevan R K, Zhu W, Hayden J, Maria J-P, Funakubo H, Ziatdinov M A, Trolier-mckinstry S and Kalinin S V 2022 Automated experiments of local non-linear behavior in ferroelectric materials *Small* **18** 2204130
[86] Liu Y, Proksch R, Wong C Y, Ziatdinov M and Kalinin S V 2021 Disentangling ferroelectric wall dynamics and identification of pinning mechanisms via deep learning *Adv. Mater.* **33** 2103680
[87] Liu Y, Vasudevan R K, Kelley K P, Funakubo H, Ziatdinov M and Kalinin S V 2023 Learning the right channel in multimodal imaging: automated experiment in piezoresponse force microscopy *npj Comput. Mater.* **9** 34
[88] Liu Y, Yang J, Vasudevan R K, Kelley K P, Ziatdinov M, Kalinin S V and Ahmadi M 2023 Exploring the relationship of microstructure and conductivity in metal halide perovskites via active learning-driven automated scanning probe microscopy *J. Phys. Chem. Lett.* **14** 3352–9
[89] Xie W, Feng Q, Srinivasan R, Stevens A and Browning N D 2017 Acquisition of STEM images by adaptive compressive sensing *Microsc. Microanal.* **23** 96–97
[90] Liu Y, Kelley K P, Vasudevan R K, Funakubo H, Ziatdinov M A and Kalinin S V 2022 Experimental discovery of structure–property relationships in ferroelectric materials via active learning *Nat. Mach. Intell.* **4** 341–50
[91] Krull A, Hirsch P, Rother C, Schiffrin A and Krull C 2020 Artificial-intelligence-driven scanning probe microscopy *Commun. Phys.* **3** 54
[92] Zhu Z *et al* 2022 A deep-learning framework for the automated recognition of molecules in scanning-probe-microscopy images *Angew. Chem., Int. Ed.* **61** e202213503
[93] Kandel S, Zhou T, Babu A V, Di Z, Li X, Ma X, Holt M, Miceli A, Phatak C and Cherukara M J 2023 Demonstration of an AI-driven workflow for autonomous high-resolution scanning microscopy *Nat. Commun.* **14** 5501
[94] Szeremeta W K, Harniman R L, Bermingham C R and Antognozzi M 2021 Towards a fully automated scanning probe microscope for biomedical applications *Sensors* **21** 3027
[95] Rashidi M and Wolkow R A 2018 Autonomous scanning probe microscopy *in situ* tip conditioning through machine learning *ACS Nano* **12** 5185–9
[96] Kalinin S V, Mukherjee D, Roccapriore K, Blaiszik B J, Ghosh A, Ziatdinov M A, Al-Najjar A, Doty C, Akers S and Rao N S 2023 Machine learning for automated experimentation in scanning transmission electron microscopy *npj Comput. Mater.* **9** 227
[97] HyperSpy Open source Python framework for exploring, visualizing and analyzing multi-dimensional data (available at: https://hyperspy.org/)
[98] Beton J G An AFM image analysis program to batch process data and obtain statistics from images (available at: https://github.com/AFM-SPM/TopoStats)
[99] Ziatdinov M AtomAI is a Pytorch-based package for deep and machine learning analysis of microscopy data (available at: https://github.com/pycroscopy/atomai)
[100] ilastik The interactive learning and segmentation toolkit (available at: www.ilastik.org/)
[101] Roboflow *Roboflow* (available at: https://roboflow.com/)
[102] META Segment anything model (available at: https://ai.meta.com/sam2/)
[103] Company, N. Electron microscopy instrumentation (available at: www.nion.com/)
[104] Nanonis SPM applications (available at: www.specs-group.com/nanonis/products/)
[105] Nanonis S SpecsGroup Nanonis (available at: www.specs-group.com/nc/nanonis/products/detail/programming-interface/)
[106] LabVIEW *LabVIEW* (available at: www.ni.com/en/shop/labview.html)
[107] Nanosurf Nanosurf python scripting interface
[108] ORNL *AEcroscoPy* (available at: https://yongtaoliu.github.io/aecroscopy.pyae/welcome_intro.html)
[109] Instruments, O. Asylum research (available at: https://afm.oxinst.com/)
[110] Krull A DeepSPM
[111] Scientific T F Thermo fisher scientific (available at: www.thermofisher.com/us/en/home.html)
[112] JEOL *JEOL* (available at: www.jeol.com/)
[113] Israel U *et al* 2023 A foundation model for cell segmentation *BioRxiv* https://doi.org/10.1101/2023.11.17.567630
[114] Israel U *et al* 2025 CellSAM: a foundation model for cell segmentation *bioRxiv* https://doi.org/10.48550/arXiv.2311.11004
[115] Isola P, Zhu J-Y, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks *Proc. IEEE conf. on computer vision and pattern recognition* pp 1125–34

[116] Bulanadi R, Cordero-Edwards K, Tückmantel P, Saremi S, Morpurgo G, Zhang Q, Martin L W, Nagarajan V and Paruch P 2024 Interplay between point and extended defects and their effects on jerky domain-wall motion in ferroelectric thin films *Phys. Rev. Lett.* **133** 106801

[117] Liu Y, Checa M, Vasudevan R K, Liu Y, Checa M and Vasudevan R K 2024 Synergizing human expertise and AI efficiency with language model for microscopy operation and automated experiment design* *Mach. Learn.: Sci. Technol.* **5** 02LT01

[118] Yin X, Shi C, Fein-Ashley B, Shao Y-T, Han Y and Jiang Y 2025 Nodeology: creating graph-based agentic workflows for AI-assisted electron microscopy *Microsc. Microanal.* **31** ozaf048–1108