



Newsletter

Director's Welcome

It's been a productive and exciting year at the Global Computing Lab (GCLab), and I am delighted to share some of our recent accomplishments with you.

Our talented researchers collaborate with scientists around the world to solve some of the most pressing problems at the intersection between domain science and computer science research. Working across scientific domains from molecular dynamics to seismology, from chemical engineering to astrophysics, we develop innovative solutions that advance trustworthiness and efficiency in high performance, cloud, and edge computing.

From post-doctoral researchers to undergraduate students, our team works on democratizing data delivery (NSDF Pilot), using compute resources efficiently (A4NN), building a data-driven software ecosystem for managing growing data collections in environmental sciences (SOMOSPIE), and addressing reproducibility in HPC applications (ANACIN-X).

We're excited to be at the forefront of reimagining HPC, and we look forward to continuing our high-impact research in the coming year. We invite you to reach out for collaborations.

Michela Taufer, Ph.D.
Dongarra Professor in High Performance Computing
University of Tennessee, Knoxville, U.S.A.
Electrical Engineering and Computer Science



Dr. Taufer at ISC 2023



@TauferLab



<https://globalcomputing.group>



Analytics
for Neural
Networks

Project Spotlight



National
Science
Foundation

Methods, Workflows, and Data Commons for Reducing Training Costs in Neural Architecture Search on High-Performance Computing Platforms

Source of Support: NSF #2223704

Project Website: <https://analytics4neuralnetworks.ai/>

Neural networks are powerful artificial-intelligence models that capture embedded knowledge in scientific data automatically. Scientists can use the knowledge to solve problems in domains such as physics, materials science, neuroscience, and medical imaging, among others. Finding accurate neural networks for a specific scientific dataset or particular problem comes at a high training cost: it requires searching among thousands of neural networks on a large number of high-performance-computing resources.

In this project, we deliver methods, workflows, and a data commons for reducing the training cost of neural networks. Our methods are based on parametric modeling and enable rapid search termination early in the training process, making the search process faster and cheaper. Our workflows decouple the search from the accuracy prediction of neural networks for different datasets and problems. Our data commons shares the full provenance of the neural networks so other scientists can deploy the neural networks in their own research.



Dr. Taufer (left) is shown here with project members.

Advances in neural networks research have a far-reaching impact on many scientific applications. Our accurate neural networks can be used to extract structural information from raw microscopy data, predict performance of business processes, analyze cancer pathology data, map protein sequences to folds, and predict soil moisture or crop yield.

Our team is committed to build a broader community of high-performance-computing experts and to promote the design and use of efficiently-generated artificial-intelligence products. In our daily work, we promote increased participation of underrepresented students, particularly women, through mentoring of students in SysTers (the organization for women in Electrical Engineering and Computer Science at the University of Tennessee Knoxville).



Researcher's Corner

with

Jack Marquez

Research Assistant Professor

What got you interested in computer science?

Since my parents bought the first computer for our home, I was very interested in computer science. During high school, I had a lot of interaction with computers in general, and from there I met professors who did interesting and innovative things using these tools that gradually captured my attention. I decided to take this path, and I realized that it is a very interesting world that evolves at great speed. To keep up with it, I must follow something that I really like: to keep learning.

What is the main project you are focusing on currently?

Since I joined GCLab I've been working mostly on HPC reproducibility. ANACIN-X, which is the project I joined, advances the reproducibility study of HPC applications by proposing an open-source modular framework for automatic measurement, analysis, and visualization of non-determinism and root causes of non-determinism in MPI applications. Contributing to this project has been really meaningful to me because I've been interested in software reproducibility for a long time, and allowing reproducible science is something very important for our group. In addition to this project, I've contributed to the Analytics for Molecular Dynamics (A4MD) project. This project involves workflow management methods and high performance computing techniques to analyze molecular dynamics data as it is generated (in situ analytics).

What do you think are the most exciting areas in computer science today?

Computer science is a field that is continuously and rapidly evolving with the purpose of being able to provide solutions to a broader number of problems. New paradigms, new architectures, and new methods are always appearing. I think that right now, one of the topics that most attracts my attention is continuum computing, which seamlessly connects different types of computing resources, such as cloud computing, edge computing, fog computing, and HPC, into an ecosystem. This allows applications to run on the most appropriate resources for the task at hand, which can lead to improved performance, scalability, and security. There is a lot of work to be done because each of the individual systems has multiple challenges that need to be solved before integration. However, continuum computing can provide a large community with an environment in which to study different and new phenomena, advancing scientific discovery in exciting and innovative ways.

Tell us something about yourself that is not computer science related

I've been playing soccer for many years; according to my mom, I started playing when I was four years old. I've even been able to play as a doctoral student! Since I had two internships abroad, I was able to play in a Division 1 indoor soccer team in Canada and a Division 3 Futsal (indoor 5v5) soccer team in Spain. Playing soccer is one of the moments when I can find tranquility and joy.



Taufer to serve on CCC

Taufer was recently selected for the Computing Research Association's Computing Community Consortium (CCC) Council. The CCC's mission is "to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges." Taufer is eager to advocate for increased trustworthiness of AI in scientific discovery and for the inclusion of disadvantaged communities through greater democratization of data.

Taufer elected Program Chair of ISC High Performance

ISC High Performance is the world's oldest conference for the HPC, machine learning, and high performance data analytics communities. The event slogan for 2024 is "Reinventing HPC," and Taufer is excited to engage the ISC community to build a bold roadmap for reinventing HPC.

Taufer wins 2023 Provost's Award for Research and Creative Achievement

This award at the University of Tennessee recognizes faculty who have received national and international recognition in their field and have shown a significant impact on their discipline. The award notes Taufer's excellence in research, scholarship, and creative achievement, as well as her commitment to mentorship and broadening participation at UT and in the scientific community more generally.

Patel wins 2023 J. Wallace and Katie Dean Fellowship

Doctoral candidate Ria Patel was awarded the prestigious J. Wallace and Katie Dean Fellowship from the Graduate School at the University of Tennessee Knoxville. This award recognizes "extraordinary students" who "show promise for outstanding graduate work in excellent and demanding programs at UT."

Channing wins DeepMind Scholarship

Georgia Channing, GCLab research scientist, was awarded the prestigious DeepMind scholarship from the University of Oxford. Georgia is one of only two students from outside the UK to receive the award; it will enable her to pursue a fully-funded MSc in Advanced Computer Science at the University of Oxford.

Strong presence at SC22

Seven members of GCLab participated in a variety of activities at SC22 in Dallas. Ian Lumsden, Paula Olaya, and Nigel Tan all served as Lead Student Volunteers, and Vanessa Lama was a student volunteer. Michela Taufer, Silvina Caíno-Lores, Nigel Tan, Ian Lumsden, and Jakob Luetzgau gave booth talks, participated in BoF sessions, and presented workshop papers.

Taufer delivers three Keynote Addresses:

- Dec. 2022: In Situ Data Analytics For Next Generation Molecular Dynamics Workflows. 15 IEEE/ACM International Conference on Utility and Cloud Computing (UCC), Portland, OR, USA.
- Sept. 2022: Studying Degree And Sources Of Non-Determinism In MPI Applications Via Graph Kernels. Latin America High Performance Computing Conference (CARLA), Porto Alegre, RS, Brazil.
- Sept. 2022: In Situ Data Analytics For Next Generation Molecular Dynamics Workflows. Parallel Processing and Applied Mathematics (PPAM), Gdańsk, Poland.

Group Papers win Competitive Reproducibility Badges at ICPP 2023

A paper authored by GCLab members Nigel Tan, Jakob Leuttgau, Jack Marquez, and Michela Taufer was awarded three ACM reproducibility badges at the ICPP 2023 conference, and a second paper authored by GCLab members Georgia Channing, Ria Patel, Paula Olaya, Silvina Caíno-Lores, and Michela Taufer earned two ACM reproducibility badges at the ICPP 2023 conference. GCLab has long made reproducibility one of its priorities, and this recognition of two of the group's papers demonstrates that reproducibility can be a reality.

Nigel Tan, Bogdan Nicolae, Jakob Luettgau, Jack Marquez, Keita Teranishi, Nicolas Morales, Sanjukta Bhowmick, Michela Taufer, and Franck Cappello. Scalable Checkpointing of Applications with Sparsely Updated Data. In Proceedings of the 52nd International Conference on Parallel Processing (ICPP), pages 1–10, Salt Lake City, UT, USA, August 2023. ACM.



Georgia Channing, Ria Patel, Paula Olaya, Ariel Rorabaugh, Osamu Miyashita, Silvina Caino-Lores, Catherine Schuman, Florence Tama, and Michela Taufer. Composable Workflow for Accelerating Neural Architecture Search Using In Situ Analytics for Protein Characterization. In Proceedings of the 52nd International Conference on Parallel Processing (ICPP), pages 1–10, Salt Lake City, UT, USA, August 2023. ACM



Group contributions recognized with IEEE eScience awards 2022

Silvina Caíno-Lores won the Early Career Researcher Award at IEEE eScience in Salt Lake City in October 2022. A paper and a poster authored by members of GCLab were both finalists for awards at IEEE 2022:

Paula Olaya, Silvina Caino-Lores, Vanessa Lama, Ria Patel, Ariel Rorabaugh, Osamu Miyashita, Florence Tama, and Michela Taufer. Identifying Structural Properties of Proteins from X-ray Free Electron Laser Diffraction Patterns. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1–10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. Best paper candidate.

Ria Patel, Ariel Rorabaugh Keller, Paula Olaya, Silvina Caino-Lores, Georgia Channing, Catherine Schuman, Osamu Miyashita, Florence Tama, and Michela Taufer. A Methodology to Generate Efficient Neural Networks for Classification of Scientific Datasets. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1–2, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. (Short paper). Best poster candidate.

Dr. Michela Taufer,
Director

Jakob Luettgau, Ph.D.,
Research Assistant Professor

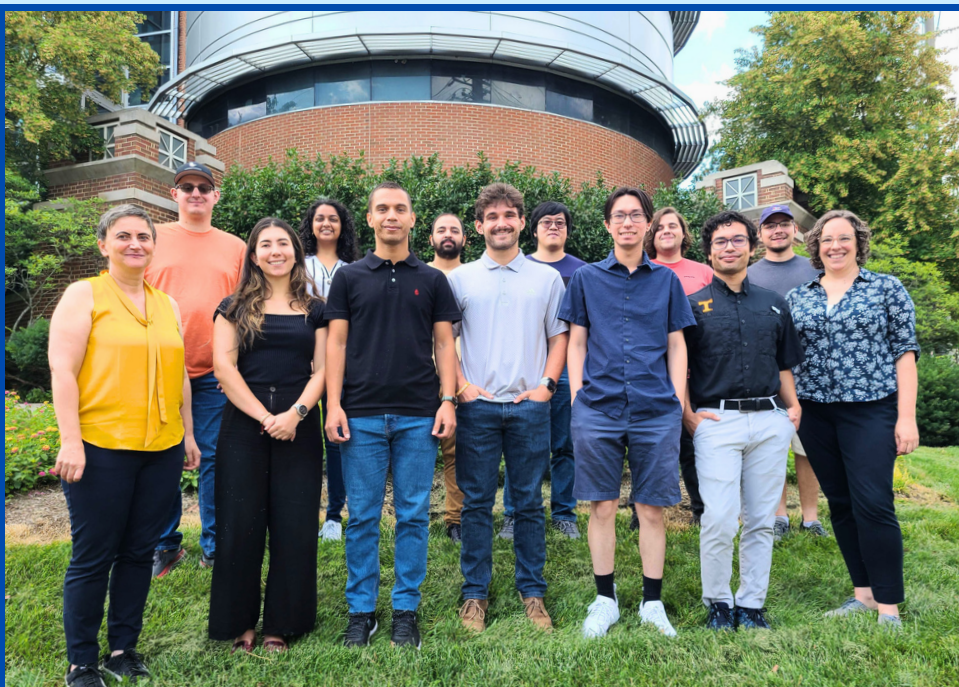
Jake Marquez, Ph.D.,
Research Assistant Professor

Dr. Lauren Whitnah, Associate
Director and Research Manager

Georgia Channing, Research Scientist
Heberth Martinez, Research Scientist

Ian Lumsden, Doctoral Student
Paula Olaya, Doctoral Student
Ria Patel, Doctoral Student
Nigel Tan, Doctoral Student

Barbara Fossum, Outreach Coordinator
Grace Wisser, Assistant Director



GCLab at the August 2023 retreat

Jay Ashworth, Master's Student
Gabriel Laboy, Master's Student
Andrew Mueller, Master's Student
Brandan Roachell, Master's Student

Seoyoung (Amy) An, Undergraduate Student
Befikir Bogale, Undergraduate Student
Blake Milstead, Undergraduate Student

GCLab Member Internships

RIKEN
Ria Patel

LOS ALAMOS NATIONAL LABORATORY
Nigel Tan & Befikir Bogale

LAWRENCE LIVERMORE NATIONAL
LABORATORY
Ian Lumsden & Julius Plehn

IBM
Paula Olaya

UNIVERSITY OF SOUTHERN CALIFORNIA
Maria Camila Roa



Ria Patel with the
Fugaku Supercomputer at RIKEN



2022-2023 Journal Publications

Naweiluo Zhou, Giorgio Scorzelli, Jakob Luettgau, Rahul Reddy Kancharla, Joshua Kane, Robert Wheeler, Brendan Croom, Pania Newell, Valerio Pascucci, and Michela Taufer. Orchestration of Materials Science Workflows for Heterogeneous Resources at Large Scale. *International Journal of High-Performance Computing Applications (IJHPCA)*, 3-4(37):260-271, 2023.
<https://doi.org/10.1177/10943420231167800>

Sanjukta Bhowmick, Parick Bell, and Michela Taufer. A Survey of Graph Comparison Methods with Applications to Nondeterminism in High-Performance Computing. *International Journal of High-Performance Computing Applications (IJHPCA)*, 3-4(37):306-327, 2023.
<https://doi.org/10.1177/10943420231166610>

Ria Patel, Brandan Roachell, Silvina Caíno-Lores, Charles Ketron, Jacob Leonard, Nigel Tan, Karan Vahi, Duncan Brown, Ewa Deelman, and Michela Taufer. Reproducibility of the First Image of a Black Hole in the Galaxy M87 from the Event Horizon Telescope (EHT) Collaboration. *IEEE Computing in Science and Engineering (CiSE)*, 5(24):42-52, 2022.
<https://doi.org/10.1109/MCSE.2023.3241105>

Tu Mai Anh Do, Loïc Pottier, Rafael Ferreira da Silva, Silvina Caíno-Lores, Michela Taufer, and Ewa Deelman. Performance Assessment of Ensembles of In Situ Workflows under Resource Constraints. *Journal of Concurrency and Computation: Practice and Experience (CCPE)*, 2022.
<https://doi.org/10.1002/cpe.7111>

Paula Olaya, Dominic Kennedy, Ricardo Llamas, Leobardo Valera, Rodrigo Vargas, Jay Lofstead, and Michela Taufer. Building Trust in Earth Science Findings through Data Traceability and Results Explainability. *IEEE Trans. Parallel Distributed Syst. (TPDS)*, 34(2):704-717, 2023.
<https://doi.org/10.1109/TPDS.2022.3220539>

Stephen Herbein, Tapasya Patki, Dong H. Ahn, Sebastian Mobo, Clark Hathaway, Silvina Caino-Lores, James Corbett, David Domyancic, Thomas R. W. Scogland, Bronis R. de Supinski, and Michela Taufer. An Analytical Performance Model of Generalized Multi-Level Scheduling. *International Journal of High-Performance Computing Applications (IJHPCA)*, 36(3):289-306, 2022.
<https://doi.org/10.1177/10943420211051039>

Ariel Keller Rorabaugh, Silvina Caíno-Lores, Travis Johnston, and Michela Taufer. Building High-throughput Neural Architecture Search Workflows via a Decoupled Fitness Prediction Engine. *IEEE Trans. Parallel Distributed Syst. (TPDS)*, 33(11):2913-2926, 2022.
<https://doi.org/10.1109/TPDS.2022.3140681>

2022-2023 Journal Publications

Michael R. Wyatt II, Stephen Herbein, Todd Gamblin, and Michela Taufer. AI4IO: A Suite of AI-Based Tools for IO-Aware Scheduling. *International Journal of High Performance Computing Applications (IJHPCA)*, 36(3):370–387, 2022.

<https://doi.org/10.1177/10943420221079765>

Nigel Tan, Robert Bird, Guangye Chen, Scott V. Luedtke, Brian Albright, and Michela Taufer. Analysis of Vector Particle-In-Cell (VPIC) Memory Usage Optimizations on Cutting-Edge Computer Architectures. *Journal of Computational Science*, 60:101566, 2022.

<https://doi.org/10.1016/j.jocs.2022.101566>

Ricardo M. Llamas, Leobardo Valera, Paula Olaya, Michela Taufer, and Rodrigo Vargas. Downscaling Satellite Soil Moisture Using a Modular Spatial Inference Framework. *Remote Sensing in Geology, Geomorphology and Hydrology*, 14(13):3137, 2022.

<https://doi.org/10.3390/rs14133137>

Robert F. Bird, Nigel Tan, Scott V. Luedtke, Stephen Lien Harrell, Michela Taufer, and Brian J. Albright. VPIC 2.0: Next Generation Particle-in-Cell Simulations. *IEEE Trans. Parallel Distributed Syst.*, 33(4):952–963, 2022.

<https://doi.org/10.1109/TPDS.2021.3084795>

2022-2023 Conference Publications

Harshita Sahni, Hector Carrillo-Cabada, Ekaterina Kots, Silvina Caino-Lores, Jack Marquez, Ewa Deelman, Michel Cuendet, Harel Weinstein, Michela Taufer, and Trilce Estrada. Online Boosted Gaussian Learners for in-situ Detection and Characterization of Protein Folding States in Molecular Dynamics Simulations. In *Proceedings of the 19th IEEE International Conference on e-Science (eScience)*, pages 1–10, Limassol, Cyprus, October 2023. IEEE Computer Society.

Nigel Tan, Bogdan Nicolae, Jakob Luetzgau, Jack Marquez, Keita Teranishi, Nicolas Morales, Sanjukta Bhowmick, Michela Taufer, and Franck Cappello. Scalable Checkpointing of Applications with Sparsely Updated Data. In *Proceedings of the 52nd International Conference on Parallel Processing (ICPP)*, pages 1–10, Salt Lake City, UT, USA, August 2023. ACM.

Georgia Channing, Ria Patel[#], Paula Olaya⁺, Ariel Rorabaugh, Osamu Miyashita, Silvina Caino-Lores, Catherine Schuman, Florence Tama, and Michela Taufer. Composable Workflow for Accelerating Neural Architecture Search Using In Situ Analytics for Protein Characterization. In *Proceedings of the 52nd International Conference on Parallel Processing (ICPP)*, pages 1–10, Salt Lake City, UT, USA, August 2023. ACM.

2022-2023 Conference Publications

Paula Olaya, Jakob Luettgau, Camila Roa, Ricardo Llamas, Rodrigo Vargas, Sophia Wen, I-Hsin Chung, Seetharami Seelam, Yoonho Park, Jay Lofstead, and Michela Taufer. Enabling Scalability in the Cloud for Scientific Workflows: An Earth Science Use Case. In Proceedings of IEEE CLOUD, pages 1-10, Chicago, IL, USA, June 2023. IEEE Computer Society.

Silvina Caino-Lores, Michel A. Cuendet, Jack Marquez, Trilce Estrada, Ewa Deelman, Harel Weinstein, and Michela Taufer. Runtime Steering of Molecular Dynamics Simulations Through In Situ Analysis and Annotation of Collective Variables. In Proceedings of Platform for Advanced Scientific Computing (PASC) Conference, pages 1-10, Davos, Switzerland, June 2023. ACM.

Camila Roa, Paula Olaya, Ricardo Llamas, Rodrigo Vargas, and Michela Taufer. GEOTiled: A Scalable Workflow for Generating Large Datasets of High-Resolution Terrain Parameters. In Proceedings of the 32nd International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC), pages 1-2, Orlando, Florida, USA, June 2023. ACM.

Jakob Luettgau, Heberth Martinez, Glenn Tarcea, Giorgio Scorzelli, Valerio Pascucci, and Michela Taufer. Studying Latency and Throughput Constraints for Geo-Distributed Data in the National Science Data Fabric. In Proceedings of the 32nd International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC), pages 1-2, Orlando, Florida, USA, June 2023. ACM.

Stephanie Brink, Michael McKinsey, David Boehme, W. Daryl Hawkins, Connor Scully-Allison, Ian Lumsden, Treece Burgess, Vanessa Lama, Katherine E. Isaacs, Jakob Luettgau, Michela Taufer, and Olga Pearce. Thicket: Seeing the Performance Experiment Forest for the Individual Run Trees. In Proceedings of the 32nd International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC), pages 1-10, Orlando, Florida, USA, June 2023. ACM.

Jakob Luettgau, Giorgio Scorzelli, Valerio Pascucci, and Michela Taufer. Development of Large-Scale Scientific Cyberinfrastructure and the Growing Opportunity to Democratize Access to Platforms and Data. In Proceedings of the 25TH International Conference On Human-Computer Interaction (HCI), Copenhagen, Denmark, July 2023. Springer.

Jakob Luettgau, Giorgio Scorzelli, Glenn Tarcea, Christine R. Kirkpatrick, Valerio Pascucci, and Michela Taufer. NSDF-Catalog: Lightweight Indexing Service for Democratizing Data Delivery. In Proceedings of the 15th IEEE/ACM International Conference on Utility and Cloud Computing (UCC), pages 1-10, Seattle, WA, USA, December 2022. IEEE Computer Society.



2022-23 Conference Publications

Tu Mai Anh Do, Loic Pottier, Rafael Ferreira da Silva, Frederic Suter, Silvina Caíno-Lores, Michela Taufer, and Ewa Deelman. Co-Scheduling Ensembles of In Situ Workflows. In Proceedings of the 17th Workshop on Workflows in Support of Large-Scale Science (WORKS), pages 1-10, Dallas, TX, USA, November 2022.

Ian Lumsden, Jakob Luettgau, Vanessa Lama, Connor Scully-Allison, Stephanie Brink, Katherine E. Isaacs, Olga Pearce, and Michela Taufer. Enabling Call Path Querying in Hatchet to Identify Performance Bottlenecks in Scientific Applications. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society.

Paula Olaya, Silvina Caino-Lores, Vanessa Lama, Ria Patel, Ariel Rorabaugh, Osamu Miyashita, Florence Tama, and Michela Taufer. Identifying Structural Properties of Proteins from X-ray Free Electron Laser Diffraction Patterns. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. Best paper candidate.

Jakob Luettgau, Silvina Caino-Lores, Kae Suarez, Dong H. Ahn, Stephen Herbein, and Michela Taufer. Reproducing and Extending Analytical Performance Models of Generalized Hierarchical Scheduling. In Proceedings of the Second Workshop on Reproducible Workflows, Data, and Security (ReWorDS), pages 1-6, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society.

Dominic Kennedy, Paula Olaya, Jay Lofstead, Rodrigo Vargas, and Michela Taufer. Augmenting Singularity to Generate Fine-grained Workflows, Record Trails, and Data Provenance. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-2, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society.

Yeom Jae-Seung, Dong H. Ahn, Ian Lumsden, Jakob Luettgau, Silvina Caino-Lores, and Michela Taufer. Ubique: A New Model for Untangling Inter-task Data Dependence in Complex HPC Workflows. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-2, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. (Short paper).

Glenn Tarcea, Brian Puchala, Tracy Berman, Giorgio Scorzelli, Valerio Pascucci, Michela Taufer, and John Allison. The Materials Commons Data Repository. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-2, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. (Short paper).

Ria Patel, Ariel Rorabaugh Keller, Paula Olaya, Silvina Caino-Lores, Georgia Channing, Catherine Schuman, Osamu Miyashita, Florence Tama, and Michela Taufer. A Methodology to Generate Efficient Neural Networks for Classification of Scientific Datasets. In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1-2, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society. (Short paper). Best poster candidate.

Collaborating Institutions



Sponsors

