May 16th, 2025

Masur Auditorium, Building 10

Join us for a day-long symposium exploring a broad range of AI approaches in biomedical science

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA
NIH National Institutes of Health
Turning Discovery Into Health
NIH National Heart, Lung, and Blood Institute
FAES
FOUNDATION FOR ADVANCED EDUCATION IN THE SCIENCES

NIH Artificial Intelligence Symposium

Friday, May 16th, 2025

Building 10, Masur Auditorium

Biomedical science is in the early phase of a technological revolution, driven in large part by innovations in deep learning neural network architecture and availability of computational power. These cutting-edge techniques are being applied to every sub-field of the biological sciences, and with novel ground-breaking advancements arriving every week it is challenging for researchers to stay current on what is available and possible. This one day symposium will bring together researchers from a broad range of disciplines to share their AI-related research, with the goal of disseminating the newest AI research, providing an opportunity to network, and to cross-pollinate ideas across disciplines in order to advance AI research in biomedicine.

2025 Planning Committee:

Ryan O'Neill (NHLBI)
Samar Samarjeet (NHLBI)
Vineeta Das (NEI)
Katerina Atallah-Yunes (NCI)
Tiarnan Keenan (NEI)
Chris Combs (NHLBI)
Kristen Morgan (NHLBI)
Chris Wanjek (OD)
Colby Lewallen (NEI)
Amy Stonelake (NCI)
Nick Asendorf (NHLBI)
Lana Yeganova (NLM/NCBI)

Sponsored by NHLBI and NIH Office of Intramural Research, in partnership with FAES

Artwork was created using generative AI, and no government funds were used for the artwork

NIH Artificial Intelligence Symposium – *May 16th, 2025 – Masur Auditorium, Building 10*

Agenda

8:30 – 9:00 Badge Pick-up

9:00 – 9:15 Opening Remarks – **Nina Schor, MD, PhD**, Deputy Director for Intramural Research, *NIH*

9:15 – 10:15 **Leo Anthony Celi, MD, MPH, MSc -** Senior Research Scientist, *Massachusetts Institute of Technology*; Associate Professor of Medicine, *Harvard Medical School* - "AI by Us, for All of Us"

10:15 – 10:45 **Zhiyong Lu, PhD, FACMI, FIAHSI -** Senior Investigator, *National Library of Medicine* - "Large Language Models in Medicine: From TrialGPT to GeneAgent"

10:45 – 11:00 *Break,* ***Coffee*** **(FAES)**

11:00 – 12:00 **Poster Session 1 – Odd numbers (FAES Terrace)**

12:00 – 12:30 Lunch

12:30 – 1:30 **Poster Session 2 – Even numbers (FAES Terrace)**

1:30 – 2:30 **Alexander Rives, PhD -** Chief Scientist and Co-founder, *EvolutionaryScale*; Core Institute Member, *Broad Institute*; Assistant Professor, *Massachusetts Institute of Technology* - "From Language to Life: AI's Role in Redesigning Biology"

2:30 – 3:15 **Short Talks**

2:30 – **Yoshitaka Inoue**, *NLM/NCI*, "Interpretable Drug Response and Drug-Target Interaction Prediction Using Artificial Intelligence"

2:45 – **Caroline Maclaren**, *NINDS*, "Deep Learning Approach to Video-based Behavioral Classification Through Human Pose Estimation"

3:00 – **Sungrim Moon, PhD**, *NCATS*, "Using Genomics Data for Basket Trial Design in Rare Diseases"

3:15 – 3:30 *Break*

3:30 – 4:00 **Ronald Summers, MD, PhD, FSAR, FAIMBE, FSPIE -** Senior Investigator, *NIH Clinical Center* - "The AI Revolution in Radiology Informatics"

4:00 – 4:45 **Short Talks**

4:00 – **Xiaoyu Duan, PhD**, *NIDDK*, "Deep learning cellular dynamics from single-cell RNA sequencing"

4:15 – **Gefei Lin**, *NHLBI*, "Multi-Task DeepHit: Simultaneous Prediction of Survival and Progressions of Risk Factors with An Application of Predicting Mortality in Sickle Cell Disease"

4:45 – 5:00 Final Remarks

Posters

P1 - **Lillich, A**; Mornini, J

"Empowering AI literacy: Library-led training and support at NIH"

P2 - **Mason, A**; Patel, K; Nachabe, F; Firrincieli, D; Saraiya, D; Meyer, A; Footer, K; Croghan, J; Rosenthal, A; Tartakovsky, M

"Transforming NIAID Operations with GenAI Tools"

P3 - **Bruno, FP**; Plevock Haase, KM,; Perez, L; Khan, S

"Harnessing AI in Implementation Science: Insights and Strategies from a Multidisciplinary Think Tank"

P4 - **Vineyard, N**; Gao, J; Gao, J; Mudd, L; Peng, G; Sen, S; Kano, C; Kinsinger, C; Resat, H

"NIH Common Fund Bridge2AI Program"

P5 - **Forsyth, AD**

"AI for Health Science in Low-Resource Settings: NIH Portfolio Landscape, Gaps, and Opportunities"

P6 - **Radujevic, A**; Wood, S; Muhoberac, M; Iyer, S; Parikh, A; Vakharia, N; Virani, S; Verma, M; Masquelin, T; Godfrey, A; Gardner, S; Rudnicki, D; Hall, MD; Klumpp-Thomas, C; Chopra, G

"SciBORGs: Scientific Bespoke Artificial Intelligence Agents Optimized for Research Goals"

P7 - **Wang, J**; Sra, A; Weiss, JC

"Active Learning for Forecasting Severity among Patients with Post Acute Sequelae of SARS-CoV-2"

P8 - **Kambara, MS**; Chukka, O; Choi, KJ; Tsenum, J; Gupta, S; English, NJ1,; Jordan, IK; Mariño-Ramírez, L

"Explainable machine learning for health disparities: type 2 diabetes in the All of Us research program"

Posters

P9 - **Egle, M**; Groechel, RC; Johansen, MC; Kucharska-Newton, AM; Gottesman, RF; Koton, S

"The role of morbidity clusters in midlife on stroke incidence and severity: The ARIC study"

P10 - **Olinger, B**; Anerillas, C; Herman, AB; Tsitsipatis, D; Banarjee, R; Tanaka, T; Candia, J; Walker, KA; Simonsick, EM; Gorospe, M; Basisty, N

"Machine Learning to Identify Tissue-Specific Clinical Associations of Senescence Signatures"

P11 - **Sun, S**; Do, AD; Zhu, Q

"AI-based biomarker discovery in CLN3"

P12 - **Campagnolo, EM**; Shulman, ED; Lodha, R; Stemmer, A; Jiang, P; Caldas, C; Knott, S; Hoang, DT; Aldape, K; Ruppin, E

"Path2Space: An AI approach for cancer biomarker discovery via histopathology inferred spatial transcriptomics"

P13 - **Salazar-Cavazos, E**; Jia, D; Missolo-Koussou, Y; Kenet, AL; Achar, S; Dada, H; Kondo, T; Krishnan, A; Taylor, N; Jiang, P; Waterfall, J; DeVoe DL; Altan-Bonnet, G

"Stochasticity in cancer immunotherapy maps with the rarity of critical Spark T cells"

P14 - **Huang, D**; Ovcharenko, I

"Deep Learning reveals the significant contribution of silencer variants to human diseases and traits"

P15 - **Ahrend, F**; Meister, G; Haase, AD

"Predicting piRNA cluster regions from genomic sequences using deep learning"

P16 - **Manzo, G**; Borkowski, K; Ovcharenko, I

"Comparative Analysis of Deep Learning Models for Predicting Causative Regulatory Variants"

Posters

P17 - **Srivastava, J**; Ovcharenko, I

"Regulatory plasticity of the human genome"

P18 - **Li, Q**; Hanchard, N

"DNA methylation differences in children of severe acute malnutrition suggest epigenetic networks at play via machine learning"

P19 - **Hudaiberdiev, S**; Ovcharenko, I

"Modeling cCREs using deep learning with applications to prioritizing candidate causal mutations from GWAS data"

P20 – **Um, S**; Mooney

"Variant Effect Predictions for *PTPN11* Missense Variants with MutPred2"

P21 - **Moon, S**; Maine J; Mathe, E; Zhu, Q

"Using Genomics Data and Literature for Basket Trial Design in Rare Diseases"

P22 - **Halder, S**; Periwal, V

"Donor-specific digital twin for living donor liver transplant recovery"

P23 - **Shi, G**; Nagarajan, V; Caspi, RR

"Identification of essential transcription factors by IAN: a new perspective on T cell licensing"

P24 - **Park, M**; Yan, C; Chen, Q; Khanna, R; Tanis, J; Meerzaman, D

"WSIomics: An Automated Pipeline for Training Multimodal AI Models to Classify therapy response of cancer patients using whole slide images and transcriptome data"

P25 - **Marini, N**; Liang, Z; Rajaraman, S; Xue, Z; Antani, S

"Combining Real and Synthetic Data to Overcome Limited Training Datasets in Multimodal Learning"

Posters

P26 - **Cordes, S**

"Towards a better CAR through *in vitro* and *in silico* Perturbations"

P27 - **Kelly, C**; Bahr, R; Zhu, W; Keyvanfar, K; Dagur, P; Cordes, S

"Optimizing CAR costimulatory domains using contrastive learning and optimal transport on high-throughput screening data"

P28 - **Duan, X**; Periwal, V

"Deep learning cellular dynamics from single-cell RNA sequencing"

P29 - Zeng, W; **Yadaw, AS**; Mehta, K; Sanjak, J; Nguyen, D-T; Huang, R; Mathé, EA

"Predicting Chemical Toxicity by Applying a Hierarchical Bayesian Approach with Priors to the Tox21 Assay Data"

P30 - **Inoue, YI**; Song, TS; Fu, TF; Luna, AL

"Interpretable Drug Response and Drug-Target Interaction Prediction Using Artificial Intelligence"

P31 - **Oyinloye, P**; Wu, F; Lee, KH; Shi, L

"Advancing antidepressant discovery through machine learning-based QSAR modelling and insights from SHAP features"

P32 - **Jain, SJ**; Yasgar, AY; Nilova, AN; Dalal, AD; Rai, GR; Zakharov, AZ

"AI-driven development of ALDH3A1 selective inhibitors"

P33 - Shah, P; Weber, C; Lim, G; Zhao, T; Sun, H; Jain, S; Zakharov, A; Siramshetty, V; Mathe, E; Xu, T; Huang, R; **Xu, X**

"*In silico* ADME models in drug discovery"

Posters

P34 - **Colelough, BC** ; Bartels, D; Demner-Fushman, D

"ClinIQLink: A Neuro-Symbolic Pipeline for QA generation with Crowd-Sourced Human-in-the-Loop Verification"

P35 - **Mollerus, P**; Seideman, J; Saraiya, D; Meyer, A; Footer, K; Chang, R; Nguyen, L; Croghan, J; Rosenthal, A; Klinkenberg, L; Meyers, J

"Scientific Review NLP Conflict of Interest Identification"

P36 - **Seideman, J**; Do, W; Tembo, M; Opsahl-Ong, L; Meyer, A; Saraiya, D; Footer, K; Desai, A; Lee, L; Nguyen, L; Croghan, J; Rosenthal, A; Tartakovsky, M

"Supervised Machine Learning for Scientific Coding Assistance"

P37 - **Piatkowski, GS**

"AI helped me write this: using AI to analyze NIH's AI and data science grant portfolio"

P38 - **Balci, H**; Luna, A

"Automating conversion of hand-drawn SBGN diagrams to SBGNML using large language models"

P39 - Rotenberg, NH; **Leaman, R**; Islamaj, R; Fluharty, B; Kuivaniemi, H; Richardson, S; Tromp, G; Lu, Z; Scheuermann, RH

"Cell phenotypes in the biomedical literature: First look at a new corpus"

P40 - **Kaiyrbekov, K**; Dobbins, NJ; Mooney, S

"Automated Survey Collection with LLM-based Conversational Agents"

P41 - Ornek, ME; Zahnen, CR; **Chen, M-C**

"Author and affiliated institution extraction from free-form letters using GenAI"

P42 - Heymann, D; **Mykins, M**; Zhou, N

"AI in action at the NICHD: Case studies and developmental pathways"

Posters

P43 - **Alodadi, M**; Lyons, E; Che, A; Watson, D; Tawa, GJ; Porter, F; Haugabook, SJ; Ottinger, E; Mudunuri, U

"RARe-SOURCE Literature AI: Rare Disease Genotype-Phenotype Associations from Biomedical Literature"

P44 - Jin, Q; Wang, Z; Floudas, CS; **Wan, N**; Chan, J; Chen, F; Gong, C; Bracken-Clarke, D; Xue, E; Fang, Y; Tian, S; Yang, Y; Sun, J; Lu, Z

"TrialGPT: matching patients to clinical trials with large language models"

P45 - **Soni, SS**; Demner-Fushman, DD

"A Dataset for Grounded Question Answering from Electronic Health Records to Relieve Clinician Burden"

P46 - **Kumar, SK**; Noroozizadeh, SN; Weiss, JCW

"Forecasting from Clinical Textual Time Series: Adaptations of the BERT and Decoder Families"

P47 - **Aston, SA**; Cheng, H

"Responsible Integration of Large Language Models in Biomedical Research"

P48 - **Liang, Z**; Rajaraman, S; Marini, N; Xue, Z; Antani, S

"Multi-Agent Cross-Modal Large Language Model Framework for Chest X-ray Analysis and Integrating COVID-19 Pneumonia Predictions"

P49 - **Nolte, S**; Saddler, TO; Reif, DM; Schmitt, CP; Auerbach, SS; Hsieh, J-H

"RAG2SQL"

P50 - **Erkan, CN**; Gu, G; Tandilashvili, E; Meigs, JM; Lee, K; Metcalf, O; Livinski, A; Pine, DS; Pereira, F; Brotman, MA; Henry, LA

"Leveraging Large Language Models for data extraction and quality assessment in psychiatry systematic reviews: A comparison of inter-rater reliability between Elicit and human coders"

Posters

P51 - **Lowekamp, BC**; Gabrielian, A; Hurt, DE; Rosenthal, A; Yaniv, Z

"Tuberculosis chest X-ray image retrieval system using deep learning based biomarker predictions"

P52 - **Arlova, A**; Weller, C; Nalls, M; Kelpsch, D; Faghri, F; Ryan, V

"Artificial Intelligence-based Segmentation of Neurites in High-Resolution Microscopy Images of iPSC-derived neurons"

P53 - **Lowekamp, B**; Yaniv, Z; Cobean, R; Hoppes, M; Rosenfeld, G; Grinev, A; Gabrielian, A; Hurt, D; Rosenthal, A; Tartakovsky, M

"Tuberculosis Portals AI in Image Processing and Abnormality Detection"

P54 - **May, CM**; Kasi, KK; Kobayashi, LK; Conway, BC; Pare, JP

"Machine Learning Classification of Clinical Edema"

P55 - Khanna, K; Chen, Q; Yan, C; **Meerzaman, D**

"Leveraging an MRI-Based Foundation Model to Enhance Predictions of Survival in Glioblastoma: A Multimodal Deep Learning Approach"

P56 - **MacLaren, CE**; Jackson, SN; Fruchet, OE; Volkman, RA; Inati, SK; Zaghloul, KA

"Deep learning approach to video-based behavioral classification through human pose estimation"

P57 - **Shive, HR**

"AI-based analysis of complex pigmentation phenotypes in zebrafish embryos"

P58 - **Lohmann, JJGL**; Witte, AW; Maier, AM; Saak, CCS; Sauter, GS; Zimmermann, MZ; Bonn, SB; Baumbach, JB

"Privacy-preserving and communication-efficient prediction of ISUP grades from prostate cancer histopathology images with foundation models"

Posters

P59 - **Cheng, J**; Flaharty, KA; Duong, D; Waikel, RL; Hu, P; Ledgister Hanchard SE; Solomon, BD

"The effects of syndromic facial feature editing on AI and clinician diagnosis of genetic conditions"

P60 - Kantipudi, K; Gabrielian, A; Hurt, DE; Rosenthal, A; **Yaniv, Z**

"Predicting tuberculosis from frontal chest X-rays: A Radiomics Analysis Portal research service"

P61 - **von Buchholtz, LJ**

"Deep learning assisted matrix factorization improves cell recognition in calcium imaging analysis"

P62 - **Patel, MH**; Stecko, H; Pramod, N; Esengur, O; Stevenson, E; Saini, J; Loebach, L; Blachman-Braun, R; Millan, B; Nethala, D; Gurram, S; Linehan, WM; Turkbey, B; Ball, MW

"Predicting Renal Tumor Pathology from Gross Appearance: An AI-based Pilot Study"

P63 - **Bhadra, S**; Liu, J; Summers, RM

"Weakly supervised learning for subcutaneous edema segmentation of abdominal CT using pseudo-labels and multi-stage nnU-Nets"

P64 - **Chan, S**; Mathai, TS; Balamuralikrishna, PTS; Batheja, V; Liu, J; Lubner, MG; Pickhardt, PJ; Summers, RM

"Staging Liver Fibrosis with Hepatic Perivascular Adipose Tissue as a CT Biomarker"

P65 - **Kantipudi, K**; Bui, V; Yu, H; Lure, YMF; Jaeger, S; Yaniv, Z

"Semantic segmentation of TB in chest X-rays: A new dataset and generalization evaluation"

P66 - **Joseph, TL**; Yu, ZX; Siddique, MAH; Chen, LY; Elinoff, JM

"Developing a deep learning algorithm to quantify pulmonary vascular remodeling in a pre-clinical model of pulmonary arterial hypertension and comparing performance to formal histopathological assessment"

Posters

P67 - **Harouni, M**; Voss TC

"Scalable deep learning-based vessel segmentation and morphological quantification"

P68 - **Bhardwaj, A**; Narayan, K

"Empanada - a napari plugin with pre-packaged segmentation models for nuclei, lipid droplets and mitochondria"

P69 - **Elsawy, A**; Keenan, T; Chew, EY; Lu, Z

"Optical Coherence Tomography: A Reliable Imaging Modality for Detecting Age-Related Macular Degeneration Features"

P70 - **Mathai, TS**; Balamuralikrishna, PTS; Batheja, V; Kassin, M; Hannah, C; Ukeh, I; Hernandez, J; Summers, RM

"Deep Learning-based Contouring of Couinaud Segments on CT: Utility for Volumetric Analysis of Future Liver Remnant"

P71 - **Aggarwal, M**; Cogan, N; Periwal, V1

"Sensitivity based model agnostic scalable explanations of deep learning"

P72 - **Lita, A**; Sjöberg, J; Păcioianu, D; Siminea, N; Celiku, O; Dowdy, T; Păun, A; Gilbert, MR; Noushmehr, H; Petre, I; Larion, M1

"Raman-based machine-learning platform reveals unique metabolic differences between IDHmut and IDHwt glioma"

P73 - **Bodosa, J**; Pastor, R

"Understanding and simulating membrane pore formation by piscidin1 using AI informed enhanced sampling"

P74 - **Weaver, A**; Tuvikene, J; Koivomagi, M

"AlphaFold2 screen reveals novel G1 cyclin docking modalities"

Posters

P75 - **Sahakyan, HK**; Babajanyan, SG; Wolf, YI; Koonin, EV

"*In silico* evolution of globular protein folds from random sequences"

P76 - **Tuvikene, J**; Esvald, EE; Heidebrink G; Koivomagi, M

"Computational modeling of Cyclin D1 protein-protein interactions"

P77 - **Kanno, T**; Kalchschmidt JS; Brooks, SR; Sun, H

"Integrating Network Analysis and Localization Prediction Using B-LEARN and ProtGPS"

P78 - **Nguyen, TH**; Ghedin, E; Sormanni, P

"Efficient Computational Prioritization of Local Host Structures Mimicking Pathogen Antibody Epitopes"

Keynote Speakers

Leo Anthony Celi, MD, MPH, MSc

Principal Research Scientist, Massachusetts Institute of Technology; Associate Professor of Medicine (Part Time), Harvard Medical School; Instructor, Harvard T.H. Chan School of Public Health; Associate Program Director, Department of Medicine, Beth Israel Deaconess Medical Center; Co-Director, Sana (MIT)

Leo Anthony Celi is a global leader in AI-driven healthcare, renowned for his groundbreaking work in leveraging data science to improve clinical outcomes and promote health equity. With a medical career spanning three continents, he brings a unique, inclusive perspective to his mission of transforming healthcare through technology. He is the creator of the Medical Information Mart for Intensive Care (MIMIC) database, a publicly accessible resource that has become a cornerstone for AI research in critical care, enabling thousands of researchers in over 30 countries to develop innovative solutions. His partnership with Philips produced the eICU Collaborative Research Database, which provides comprehensive data on over 2 million ICU patients, further expanding the foundation for AI-driven advancements in medicine.

Through his leadership in Sana, an MIT-based initiative, Dr. Celi develops open-source mobile health technologies that empower healthcare providers in underserved regions, delivering critical care to communities with limited resources. This work has earned international acclaim, including first place in the 2012 Mobile Health University Challenge and a finalist position for the 2011 INDEX: Award for Design to Improve Life. His research, cited over 36,000 times, reflects his profound impact on the field, with contributions that bridge clinical practice, data science, and global health policy.

Dr. Celi is also a passionate advocate for ethical AI, pushing for sustainable practices that mitigate the environmental impact of AI infrastructure, such as the energy demands of data centers, and ensure equitable access to technological benefits. He emphasizes the importance of understanding clinical data's context to build trustworthy AI models, challenging the field to prioritize transparency and inclusivity. His vision for a healthcare system where AI serves all communities makes him a vital voice in shaping the future of biomedicine.

https://imes.mit.edu/people/celi-leo

https://www.youtube.com/watch?v=3StkjrQ8n5Y

Alexander Rives, PhD

Chief Scientist and Co-founder, EvolutionaryScale; Core Institute Member, Broad Institute of MIT and Harvard; Assistant Professor, Department of Electrical Engineering and Computer Science, MIT

Alexander Rives, PhD, is a trailblazer in the integration of artificial intelligence and biology, driving transformative advancements in biomedicine through computational innovation. At EvolutionaryScale, the public-benefit startup he co-founded, he spearheads the development of advanced AI models like ESM3, which has demonstrated remarkable capabilities in generating novel proteins such as esmGFP. These innovations hold immense potential for accelerating scientific discovery and developing new therapeutics, reshaping how we approach biological research and healthcare solutions.

Prior to EvolutionaryScale, Rives led the Evolutionary Scale Modeling (ESM) project at Meta's AI research lab, where he pioneered the creation of the first large-scale transformer language models for proteins. These models have been widely adopted by the global scientific community, enabling breakthroughs in drug design, predicting the clinical effects of genetic mutations, and modeling cellular processes. His research, cited over 9,800 times, underscores his profound influence on the field, with publications like the 2019 paper "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences" setting new benchmarks in computational biology.

Rives's entrepreneurial vision extends to co-founding biotech companies such as Fate Therapeutics and Syros Pharmaceuticals, both publicly traded on NASDAQ, and Kallyope, demonstrating his ability to translate cutting-edge research into impactful commercial ventures. In his academic roles at the Broad Institute and MIT, he fosters interdisciplinary collaboration, developing AI systems that support global scientific efforts and mentoring the next generation of researchers at the intersection of AI and biology. With a PhD in Computer Science from New York University and a B.S. in Philosophy and Biology from Yale University, Rives brings a multidisciplinary perspective to his work, ensuring that AI advancements in biomedicine are not only technically groundbreaking but also ethically responsible. His leadership continues to shape the future of AI-driven healthcare, making him a pivotal voice in the field.

https://www.evolutionaryscale.ai/

https://www.broadinstitute.org/bios/alex-rives

https://www.youtube.com/watch?v=TiDo7xXMbUI

Zhiyong Lu, PhD, FACMI, FIAHSI

Senior Investigator, National Library of Medicine; Deputy Director for Literature Search, National Center for Biotechnology Information; Adjunct Professor, University of Illinois Urbana-Champaign

Zhiyong Lu, PhD, FACMI, FIAHSI, is a leading innovator in biomedical informatics, renowned for his pioneering work in applying artificial intelligence and machine learning to enhance biomedical research and healthcare. At the National Library of Medicine (NLM), he drives advancements in text mining and information retrieval, significantly improving access to scientific literature through widely used platforms like PubMed and LitCovid. These resources empower millions of researchers and clinicians worldwide by providing efficient, AI-driven tools to navigate vast biomedical datasets.

Dr. Lu's research has led to the development of transformative tools such as LitVar 2.0, which tracks genetic variants in biomedical literature, and TrialGPT, a large language model that streamlines patient matching for clinical trials. His work on advanced AI models, including GeneAgent, showcases his expertise in harnessing large language models to address complex challenges in precision medicine and genetic research. As an Adjunct Professor at the University of Illinois Urbana-Champaign, he mentors students in computational biology, fostering the next generation of AI researchers.

Recognized as a Fellow of the American College of Medical Informatics and the International Academy of Health Sciences Informatics, Dr. Lu's contributions have earned him international acclaim. He plays a pivotal role in shaping AI policy, serving on the US federal AI R&D inter-agency committee and the NIH Intramural Research Program AI task force. His leadership in organizing workshops and editing special issues, such as the 2024 JAMIA special issue on large language models in biomedicine, underscores his commitment to advancing the field. Dr. Lu's work continues to bridge technology and medicine, driving innovation in AI applications for healthcare.

https://irp.nih.gov/pi/zhiyong-lu

https://www.youtube.com/watch?v=AHPPGECs7KQ

Ronald M. Summers, MD, PhD, FSAR, FAIMBE, FSPIE

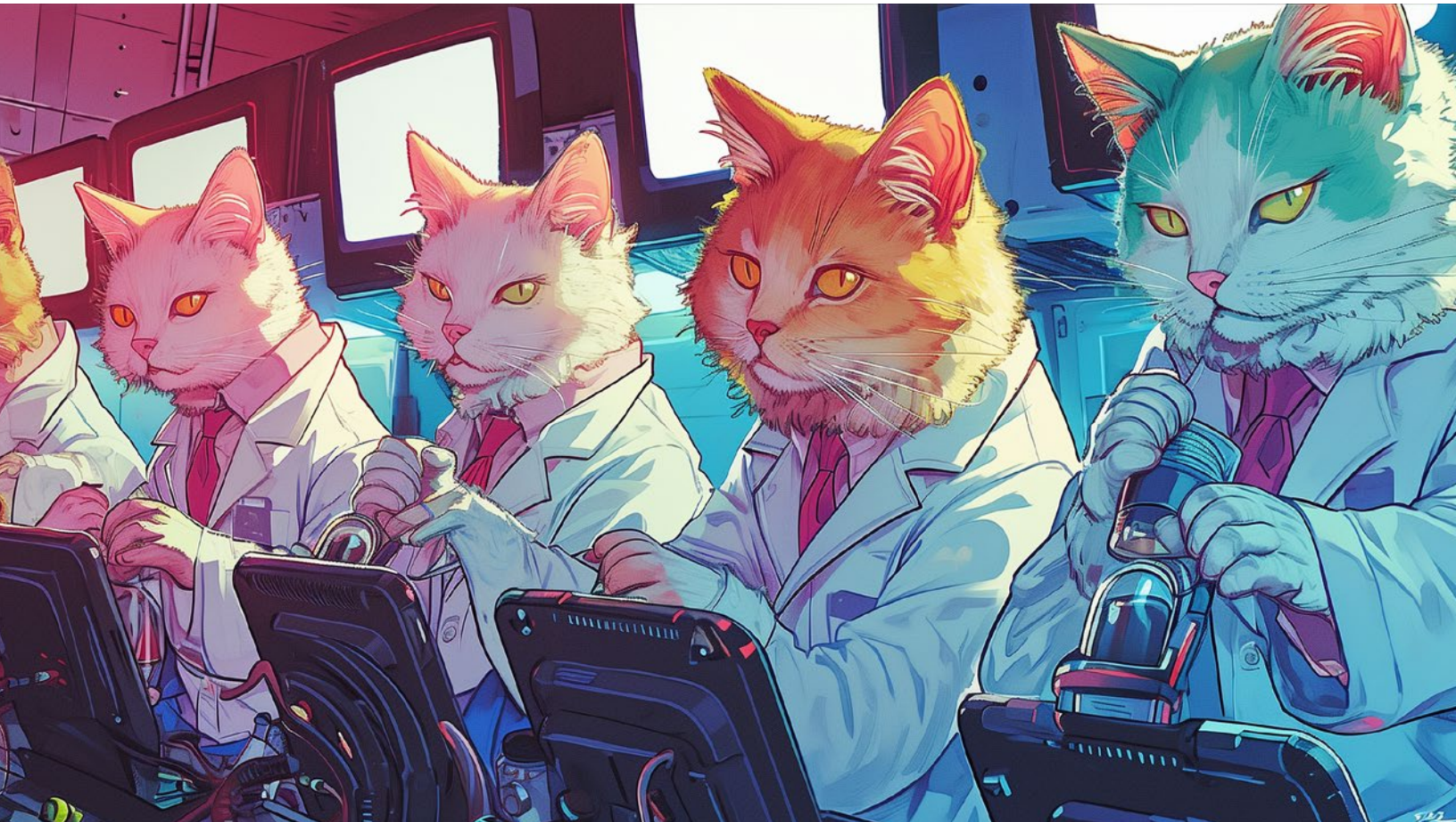
Senior Investigator, NIH Clinical Center; Chief, Clinical Image Processing Service; Director, Imaging Biomarkers and Computer-Aided Diagnosis Laboratory

Ronald M. Summers, MD, PhD, is a visionary radiologist whose pioneering work in artificial intelligence has revolutionized medical imaging, particularly in cancer diagnosis and treatment. His research harnesses deep learning to develop advanced computer-aided diagnosis systems, significantly improving the detection of lung, colon, and other cancers through techniques like virtual colonoscopy. By creating large-scale radiologic image databases, he has provided the global research community with critical resources to train AI models, fostering innovation in radiology informatics. His prolific output, with over 500 publications cited more than 57,700 times and 12 patents, underscores his transformative impact on the field.

Summers's expertise in thoracic and abdominal radiology and body cross-sectional imaging informs his AI-driven innovations, which have set new standards for precision and efficiency in clinical practice. His leadership extends to mentoring dozens of radiology fellows, many of whom have become leaders in the field, and serving on editorial boards for prestigious journals like *Radiology: Artificial Intelligence* and *Journal of Medical Imaging*. Recognized with the Presidential Early Career Award for Scientists and Engineers in 2000 and the NIH Director's Award in 2012, he holds fellowships from the Society of Abdominal Radiologists, the American Institute for Medical and Biological Engineering, and the Society of Photo-Optical Instrumentation Engineers. His global influence is further evidenced by his roles in shaping AI standards through conferences like SPIE Medical Imaging, where he has served as program co-chair. With a career bridging clinical expertise and technological innovation, Summers continues to drive the AI revolution in radiology, advancing the future of precision medicine.

https://www.cc.nih.gov/meet-our-doctors/rsummers

Abstracts - Short Talks



Short Talk Abstracts

Interpretable Drug Response and Drug-Target Interaction Prediction Using Artificial Intelligence

Inoue, YI[1,2,3]; Song, TS[1]; Fu, TF[4]; Luna, AL[2,3]

1. Computer Science, University of Minnesota, Minneapolis, MN
2. Computational Biology Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD
3. Developmental Therapeutics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD
4. Department of Computer Science, Nanjing University Nanjing, Jiangsu, China

A challenge of using machine learning (ML) in biomedical research is a lack of interpretability, which limits its support of data-driven decisions with explanations. We explore this topic here, focusing on cancer drug response and mechanism prediction. We introduce two components: GraphPINE (Graph Propagating Importance Network for Explanation) and DrugAgent. GraphPINE is a graph neural network (GNN) model for drug response prediction using multi-omics data (e.g., gene expression) and interaction networks (e.g., protein-protein). The novelty of GraphPINE lies in its initialization of importance scores using biological prior knowledge (drug-target interactions, DTIs) from literature and a dynamic updating mechanism. We build on concepts from LSTM (Long Short-Term Memory), relying on previous predictions as hidden states to advance GNNs such that GraphPINE initializes importance scores using prior knowledge and updates these scores during model training. We apply GraphPINE to NCI60 data; GraphPINE achieves AUROC of 0.796 and AUPRC of 0.894 for 952 drugs. Separately, we developed DrugAgent, a multi-agent system integrating knowledge graphs, internet searches, ML methods, and large language models (LLMs) to improve DTI prediction. DrugAgent was evaluated using 178 kinase inhibitors against 300 kinases; DrugAgent achieves superior performance (AUROC: 0.905 and AUPRC: 0.529). Interpretable subgraphs accompany GraphPINE results, while DrugAgent results are enriched with prior knowledge. Multiple lines of evidence must support conclusions in biomedical research. The bioinformatics efforts here build on this fundamental notion to draw in additional data from heterogeneous sources uniformly and transparently as part of ensemble results presented to users.

See also Poster P31

Short Talk Abstracts

Deep learning approach to video-based behavioral classification through human pose estimation

MacLaren, CE[1]; Jackson, SN[1]; Fruchet, OE[1]; Volkman, RA[1]; Inati, SK[2]; Zaghloul, KA[1]

1. Functional Neurosurgery Section, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD
2. Neurophysiology of Epilepsy Unit, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

There exist current methods for identifying human action in videos, but little advancements in behavior in a hospital environment, where patients are monitored 24/7 via a live-stream camera. With new advances in machine learning and computer vision, different deep learning models can now identify objects and track their movement throughout a video. Through such, human movement – described as human pose – can be extracted in videos, creating opportunity for tracking and classifying different actions. We are interested in applying these methods for our own video data, where we record 24/7 clinical footage for epilepsy patients admitted at the NIH for seizure monitoring. Pre-trained human pose models achieve very high mean average precision (mAP) and are useful for transferring to different datasets. Utilizing a pre-trained network and fine-tuning for refined features, we can identify more positional information to the standard pre-trained network for our non-uniform environments, where patients may be in different settings with various obstructions, such as staff and family interruptions, blankets, tables, etc. We are able to achieve a mAP of 0.78147, identifying 18 different points of interest on the human body, and a precision of 0.99668 for identifying the proper boundaries of our patient. By correctly identifying human movement in videos, we can cluster different behaviors to classify a patient's unique behaviors. With this annotated data, we can extract neural correlates for precise behaviors throughout a patient's entire stay.

See also Poster P57

Short Talk Abstracts

Using Genomics Data for Basket Trial Design in Rare Diseases

Moon, S[1]; Maine, J[1]; Mathe, E[1]; Zhu, Q[1]

1. Division of Pre-Clinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD

Gaining insight into the underlying molecular etiologies of rare diseases can aid cross-disease research, inform the design of basket trials, and identify drug repurposing opportunities. In our preliminary study, we identified 36 rare disease clusters based on common genetic causes and biological mechanisms. However, there clusters were too broad for basket trial applications. In this study, we refined these clusters by collecting allelic variant data from the Online Mendelian Inheritance in Man (OMIM), along with corresponding Sorting Intolerant From Tolerant (SIFT) scores for single nucleotide polymorphisms (SNPs) and transcripts from Ensemble validated from the Medical Genomics Japan Variant Database (MGeND). We assessed the functional impact of gene mutations using SIFT scores, calculating the ratio of deleterious to tolerated cases (deleterious cases / (deleterious cases + tolerated cases)). We generated an matrix with imputed data by extracting the deleterious level of genetic and mutation data for each rare disease, and identified shared mutations across diseases. Then, we applied Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to the imputed matrix, creating sub-clusters on the top of the 36 clusters. Our results illustrate consistent findings with the published studies of basket trial design for instance, a subcluster of NLRP3 mutation-related diseases including Neonatal Onset Multisystem Inflammatory Disease, Familial Cold Autoinflammatory Syndrome, and Muckle-Wells Syndrome.

See also Poster P22

Short Talk Abstracts

Deep learning cellular dynamics from single-cell RNA sequencing

Duan, X[1]; Periwal, V[1]

1. Lab of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

Single-cell RNA sequencing (scRNA-seq) provides a powerful framework for studying cellular heterogeneity, transitions, and regulatory networks. However, reconstructing the underlying dynamical processes governing these transitions remains a major challenge due to the high-dimensional nature of gene expression data. To address this, we develop a variational autoencoder (VAE)-based approach that learns a low-dimensional latent representation of cellular states and models their temporal evolution. We apply our framework to gene expression data from Drosophila melanogaster blastoderm embryos, compiled by Fowlkes et al., which includes measurements across multiple time points using a registration technique. In this approach, gene expression profiles are encoded into a low-dimensional latent space, where we train a neural stochastic differential equation (SDE) network to capture the continuous dynamics of latent states over developmental time. The learned neural SDE models the progression of cellular states, and a decoder subsequently maps these evolving latent representations back to the original high-dimensional gene expression space, allowing for both accurate reconstruction of observed transcriptional patterns and insight into the underlying dynamical processes. Future directions include the use of symbolic regression to extract dynamical models from the inferred trajectories.

See also Poster P29

Short Talk Abstracts

Multi-Task DeepHit: Simultaneous Prediction of Survival and Progressions of Risk Factors with An Application of Predicting Mortality in Sickle Cell Disease

Lin, GL[1,4]; Tian, XT[2]; Wu, COW[2]; Thein, SLT[3]; Miao, RM[2]

1. Office of Clinical Director, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD
2. Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD
3. Sickle Cell Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD
4. Department of Statistics, The George Washington University, Washington, DC

Survival analysis aims to estimate time-to-event outcomes but is frequently challenged by censoring due to limited study duration, loss to follow-up, and competing risks, reducing the number of observed events. Traditional approaches usually rely on surrogate or composite endpoints can diminish interpretability, and the proportional hazards assumption in commonly used Cox models is often violated in real-world settings. Progressions of risk factors, such as laboratory measurements, offer meaningful signals for survival prediction but are typically modeled separately, which may lead to inconsistencies in variable selection or effect directions.

To overcome these challenges, we develop a deep learning-based multi-task model (Multi-task DeepHit) that simultaneously predicts long-term mortality and short-term trajectories of risk factors. The model includes a shared representation network with a masked attention-based encoder pre-trained on baseline variables, and task-specific networks for predicting yearly mortality and trajectories of risk factors. This architecture fully utilizes available variables and allows partial missingness.

We applied our approach to a Sickle Cell Disease dataset of 598 patients with 68 baseline covariates and 13 key intermediate risk factors. Compared with Cox models using top variables selected by random survival forest and original DeepHit with baseline data, our model achieved superior discrimination and calibration, with a higher bootstrap-corrected C-statistic of 0.8648 (compared to 0.7324) and a lower 4-year integrated Brier score of 0.0364 (compared to 0.05). The model remains robust even when partial missingness exists during evaluation. At both the population and individual levels, we further explain our model using SHAP values to identify important variable contributions to mortality risk, which are clinically interpretable.

Abstracts - Posters

Poster Abstracts

P1

Empowering AI literacy: Library-led training and support at NIH

Lillich, A[1]; Mornini, J[1]

1. ORS Division of Library Services, National Institutes of Health, Bethesda, MD

As artificial intelligence (AI) continues to shape biomedical research and administrative processes, AI literacy has become essential for professionals at NIH. The NIH Library plays a critical role in supporting AI adoption by providing training, consultations, and curated resources to help researchers, administrators, and technical staff integrate AI tools responsibly and effectively.

This poster presents the NIH Library's approach to AI literacy, outlining key training programs, personalized support services, and collaborative efforts. Through workshops, events, one-on-one consultations, and learning materials, the NIH Library equips NIH staff with practical AI skills while addressing ethical considerations and best practices.

We highlight the impact of these NIH Library initiatives, showcasing engagement metrics and user feedback that demonstrate AI's transformative potential in research and operations. Additionally, we discuss common challenges in AI adoption and the strategies we use to overcome them. By sharing insights from the NIH Library's AI literacy program, this poster aims to inform best practices for integrating AI support in a federal research environment and inspire further collaboration on AI education at NIH.

Poster Abstracts

P2

Transforming NIAID Operations with GenAI Tools

Mason, A[1]; Patel, K[1]; Nachabe, F[1]; Firrincieli, D[1]; Saraiya, D[1]; Meyer, A[1]; Footer, K[1]; Croghan, J[1]; Rosenthal, A[1]; Tartakovsky, M[1]

1. Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

The introduction of GenAI tools at the National Institute of Allergy and Infectious Diseases (NIAID) marks a significant advancement in artificial intelligence application. Incorporating the open-source AI middleware GovConnect.ai, NIAID swiftly deployed applications including GenAI Chat, GenAI Doc Bot, and GenAI SummarizeIt, empowering over 5,000 staff to enhance productivity through large language models (LLMs). These tools are used for content generation and summarization, translation, document comparison, and information synthesis. The GovConnect.ai (platform) is specifically designed to enable enterprise AI adoption by providing connectors to essential AI components integrating diverse databases, cloud storage, and transformation code blocks for unified data orchestration. It ensures secure access and offers observability tools for cost monitoring. As a cloud-agnostic platform, GovConnect.ai boosts software development efficiency, promotes experimentation, and accelerates AI innovation at NIAID.

GenAI Chat serves as an advanced chatbot, enabling users to interact with LLMs by posing questions through prompts. It delivers detailed and precise answers, enhancing decision-making and expediting information retrieval.

GenAI Doc Bot enables users to upload multiple documents like PDFs, PowerPoints, and Word files for comprehensive insights, improving document analysis and understanding with contextually accurate responses.

GenAI SummarizeIt optimizes productivity by summarizing lengthy documents, such as policies or publications, reducing review time and enabling NIAID personnel to focus on more critical tasks.

Collectively, these GenAI tools represent a transformative leap in applying artificial intelligence to everyday business processes, significantly enhancing operational efficiency and productivity at NIAID, with broader implications for the National Institutes of Health.

Poster Abstracts

P3

Harnessing AI in Implementation Science: Insights and Strategies from a Multidisciplinary Think Tank

Bruno, FP[1]; Plevock Haase, KM,[1]; Perez, L[1]; Khan, S[1]

1. Center for Translation Research and Implementation Science, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

Background: Integrating artificial intelligence (AI) and implementation science (IS) offers critical opportunities to advance evidence-based practices and improve health outcomes. A recent think tank convened by the Center for Translation Research and Implementation Science (CTRIS) brought together clinicians, researchers, and AI experts to examine challenges and explore future directions. Objectives: This abstract summarizes key discussed themes, including ethical considerations, innovative applications, data management, and strategies for further growth.

Discussion:

• Applications: AI facilitates personalized healthcare, supports guideline adherence, and streamlines EHR management using large language models. Risk-prediction tools can also inform targeted interventions for hospital readmission.

• Data and privacy: Transparency and privacy remain paramount. Explainable AI fosters clarity and trust, while federated learning safeguards data. Evaluation frameworks, such as CONSORT-AI and RE-AIM, support responsible innovation. Community-led governance and Community-Based Participatory Research are essential for building trust and safeguarding data sovereignty.

• Challenges, Gaps, and Opportunities: Conflicting priorities between market-driven and academic approaches underscore the need for adaptable, transparent practices. Obstacles include the “black box” challenge, limited AI literacy among stakeholders, and privacy concerns. However, locally-tailored frameworks, collaborative knowledge-sharing, and balanced oversight can enhance capacity-building and accelerate health translation across diverse populations.

• Future steps: Addressing population-specific biases, integrating context-specific factors influencing health outcomes, and developing adaptive AI platforms can advance comprehensive care in under-resourced settings.

Conclusions: Realizing AI’s potential in IS requires robust interdisciplinary collaboration, transparent governance, and sustained community engagement. Researchers emphasize co-created frameworks that leverage diverse data sources, incorporate real-world contexts, and promote equitable adoption for improved healthcare outcomes.

Poster Abstracts

P4

NIH Common Fund Bridge2AI Program

Vineyard, N[1]; Gao, J[2]; Gao, J[3]; Mudd, L[4]; Peng, G[5]; Sen, S[2]; Kano, C[1]; Kinsinger, C[1]; Resat, H[1]

1. Office of the Director/Office of Strategic Coordination, National Institutes of Health, Bethesda, MD
2. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
3. National Eye Institute, National Institutes of Health, Bethesda, MD
4. National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, MD
5. National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD

The NIH Common Fund's Bridge to Artificial Intelligence (Bridge2AI) program aims to propel biomedical research forward by setting the stage for widespread adoption of artificial intelligence and machine learning (AI/ML) to tackle complex biomedical and behavioral research challenges. The Bridge2AI program bridges the gap between the biomedical and behavioral research communities through a consortium of experts to set the stage for widespread adoption of AI/ML in medicine. The Bridge2AI program is generating flagship data, developing best practices for making data ready for use in AI/ML models, and developing the workforce for the next generation of AI/ML specialists. Bridge2AI established four grand challenges to motivate these goals. The data from these grand challenges include: voice and speech for predictive diagnostics; functional genomic mapping of human cells; clinical data for diagnosis and risk prediction in acute care settings; and salutogenesis of Type II diabetes as a model for health restoration. To-date, the program has achieved several milestones such as: assembling large multidisciplinary teams of experts; hosting jamborees and hackathons of data resources; providing hands-on training to teach the next generation of researchers; hosting symposiums; creating educational modules for skills and workforce development; and publishing best practice guidelines on the program's lessons learned.

Bridge2AI released preliminary pilot data in 2024 with ongoing additional data releases occurring. Additional data releases are scheduled until the end of program in 2026. The program will also release best practices guidelines for the collection and preparation of AI/ML-ready data, considerations for data collection and (re)use, and how to make the future data collection projects AI/ML ready.

Poster Abstracts

P5

AI for Health Science in Low-Resource Settings: NIH Portfolio Landscape, Gaps, and Opportunities

Forsyth, AD[1]

1. Fogarty International Center, National Institutes of Health, Bethesda, MD

Artificial Intelligence (AI) demonstrates potential to deliver innovative, cost-efficient solutions in resource-constrained settings by addressing local health priorities while strengthening health systems, disease surveillance, and responses to critical challenges, regardless of geographic settings. AI investments can also bolster U.S. health security by detecting threats before they become domestic crises and sparking health advances that benefit all.

The Fogarty International Center (FIC) has launched a strategic initiative to catalyze ethical and responsible AI use in low-resource settings, ranging from rural America to low- and middle-income countries (LMICs), through a scoping literature review, NIH portfolio analysis, and partner mapping. A review of nearly 80 peer-reviewed articles enabled AI-assisted categorization of ~2,000 NIH-funded grants by eight primary use-cases across country income status. Although only 5.2% of NIH's AI portfolio supports LMIC-based research, almost 70% focuses on diagnostics and treatment, mirroring early priorities for the technology's use.

The current distribution of grants also shows notable gaps in disease surveillance, health systems optimization, and remote care, particularly in LMICs; further, it does not align with the health conditions that account for the highest probability of premature mortality. This suggests a mismatch between research investment and global health burden. FIC's initiative aims to inform the strategic use of accessible, fair, and trustworthy AI to advance health that maximizes NIH's impact across all resource-limited settings. By identifying gaps and opportunities, the project seeks to foster coordination, knowledge-sharing, and collaborative research to ensure AI delivers sustainable benefits for domestic and international health.

Poster Abstracts

P6

SciBORGs: Scientific Bespoke Artificial Intelligence Agents Optimized for Research Goals

Radujevic, A[1]; Wood, S[1]; Muhoberac, M[2]; Iyer, S[2]; Parikh, A[3]; Vakharia, N[4]; Virani, S[2]; Verma, M[1]; Masquelin, T[1]; Godfrey, A[1]; Gardner, S[1]; Rudnicki, D[1]; Hall, MD[1]; Klumpp-Thomas, C[1]; Chopra, G[2,4,5]

1. National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD
2. Department of Chemistry, Purdue University, West Lafayette, IN
3. Department of Statistics, Purdue University, West Lafayette, IN
4. Department of Computer Science, Purdue University, West Lafayette, IN
5. Purdue Institute for Drug Discovery, Integrative Data Science Institute, Purdue Center for Cancer Research, Purdue Institute for Inflammation, Immunology, and Infectious Disease, Purdue Institute for Integrative Neuroscience

The integration of artificial intelligence (AI) agents to assist scientists in planning, executing, and analyzing experiments is one of the core objectives of the ASPIRE initiative. A key milestone in this endeavor has been connecting AI agents with modern automation and laboratory equipment, demonstrating that rather than replacing human input, AI serves as a powerful tool to tackle some of the most complex challenges in drug discovery programs.

One of our significant achievements has been the development of cross-communication between an AI agent and a microwave reactor. By using text prompts, we successfully controlled the microwave reactor via the AI agent, turning the AI into a kind of co-pilot for microwave synthesis. In this context, the AI operates much like ChatGPT but for guiding microwave reactor experiments.

This integration allowed us to achieve two major objectives:

1. We successfully executed a benchmark reaction, N-alkylation, on the Biotage Initiator by inputting commands in the form of text prompts through the AI co-pilot.

2. We conducted a reaction optimization on the Biotage Initiator+, where the AI agent employed an optimization algorithm to improve percent conversion. The system intelligently searched for optimal conditions, adjusting key parameters like reaction temperature, duration, reagent selection (such as bases and catalysts), and solvent choice.

These advancements exemplify how AI can enhance experimental workflows, empowering scientists to unlock new levels of efficiency and precision in chemical research.

Poster Abstracts

P7

Active Learning for Forecasting Severity among Patients with Post Acute Sequelae of SARS-CoV-2

Wang, J[1]; Sra, A[2]; Weiss, JC[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD
2. George Washington University, Washington, DC

The long-term effects of Postacute Sequelae of SARS-CoV-2, known as PASC, pose a significant challenge to healthcare systems worldwide. Accurate identification of progression events—such as hospitalization and reinfection—is essential for effective patient management and resource allocation. However, traditional models trained on structured data struggle to capture the nuanced progression of PASC. In this study, we introduce the first publicly available cohort of 18 PASC patients, with text time series features based on Large Language Model Llama-3.1-70B-Instruct and clinical risk annotated by clinical expert. We propose an Active Attention Network to predict the clinical risk and identify progression events related to the risk. By integrating human expertise with active learning, we aim to enhance clinical risk prediction accuracy and enable progression events identification with fewer number of annotation. The ultimate goal is to improves patient care and decision-making for SARS-CoV-2 patient.

Poster Abstracts

P8

Explainable machine learning for health disparities: type 2 diabetes in the All of Us research program

Kambara, MS[1]; Chukka, O[2]; Choi, KJ[1]; Tsenum, J[2]; Gupta, S[1]; English, NJ[1,3]; Jordan, IK[2,3]; Mariño-Ramírez, L[1]

1. National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD
2. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA
3. IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA

Type 2 diabetes (T2D) is a disease with high morbidity and mortality and a disproportionate impact on minority groups. Machine learning (ML) is increasingly used to characterize T2D risk factors; however, it has not been used to study T2D health disparities. Our objective was to use explainable ML methods to discover and characterize T2D health disparity risk factors. We applied SHapley Additive exPlanations (SHAP), a new class of explainable ML methods that provide interpretability to ML classifiers, to this end. ML classifiers were used to model T2D risk within and between self-identified race and ethnicity (SIRE) groups, and SHAP values were calculated to quantify the effect of T2D risk factors. We then stratified SHAP values by SIRE to quantify the effect of T2D risk factors on prevalence differences between groups. We found that ML classifiers (random forest, lightGBM, and XGBoost) accurately modeled T2D risk and recaptured the observed prevalence differences between SIRE groups. SHAP analysis showed the top seven most important T2D risk factors for all SIRE groups were the same, with the order of importance for features differing between groups. SHAP values stratified by SIRE showed that income, waist circumference, and education best explain the higher prevalence of T2D in the Black or African American group, compared to the White group, whereas income, education and triglycerides best explain the higher prevalence of T2D in the Hispanic or Latino group. This study demonstrates that explainable ML can be used to elucidate health disparity risk factors and quantify their group-specific effects.

Poster Abstracts

P9

The role of morbidity clusters in midlife on stroke incidence and severity: The ARIC study

Egle, M[1]; Groechel, RC[1]; Johansen, MC[2]; Kucharska-Newton, AM[34]; Gottesman, RF[1]; Koton, S[56]

1. National Institute of Neurological Disorders and Stroke, Intramural Research Program, National Institutes of Health, Bethesda, MD
2. Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD
3. Department of Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC
4. Department of Epidemiology, University of Kentucky, Lexington, KY
5. Department of Nursing, The Stanley Steyer School of Health Professions, Tel Aviv University, Tel Aviv, Israel
6. Department of Epidemiology, Johns Hopkins University School of Public Health, Baltimore, MD

OBJECTIVE: Standardized scores summarizing clinical information have found application in risk stratification but their ability to capture stroke-related risk factors in midlife and predict risk in biracial populations remains less explored. This study employed a cluster analysis approach to group individuals into clusters based on similar clinical profiles in midlife and assessed the clusters' association with stroke risk and severity in a community-based prospective cohort.

METHODS: Participants (N=15,404) without prevalent stroke from the Atherosclerosis Risk in Communities (ARIC) study were included. An hierarchical clustering approach was used to allocate participants into clusters based on clinical information. In Cox proportional hazard models, the association of the clusters with overall ischemic stroke incidence was tested while accounting for age, sex, education, and race-center.

RESULTS: Of 1424 incident ischemic strokes diagnosed from baseline to 2020, 1104 included NIHSS grading. The cluster analysis identified 9 distinct midlife clusters in the population with the following defining features: cluster 1 (relatively healthy); cluster 2 (smoking); cluster 3 (cancer); cluster 4 (peripheral artery disease); cluster 5 (obesity, diabetes, hypertension, and hypertriglyceridemia); cluster 6 (coronary heart disease); cluster 7 (atrial fibrillation); cluster 8 (heart failure); cluster 9 (renal dysfunction). Clusters 2-9, compared to cluster 1 were each associated with a greater stroke risk, with the greatest hazard ratio (HR) for cluster 9 (HR(95%CI)= 3.00 (2.00,4.50)).

INTERPRETATION: The findings emphasizes the importance of morbidity clusters in midlife for stroke. Cluster analysis may be a powerful tool when stratifying large diverse populations based on morbidity burden.

Poster Abstracts

P10

Machine Learning to Identify Tissue-Specific Clinical Associations of Senescence Signatures

Olinger, B[1,2]; Anerillas, C[5]; Herman, AB[4]; Tsitsipatis, D[4]; Banarjee, R[1]; Tanaka, T[1]; Candia, J[1]; Walker, KA[3]; Simonsick, EM[1]; Gorospe, M[4]; Basisty, N[1]

1. Translational Gerontology Branch, National Institute on Aging, NIH, Baltimore, MD
2. Department of Biology, Johns Hopkins University, Baltimore, MD
3. Laboratory of Behavioral Neuroscience, National Institute on Aging, NIH, Baltimore, MD
4. Laboratory of Genetics and Genomics, National Institute on Aging, NIH, Baltimore, MD
5. Tissue and Organ Homeostasis Program, Centro de Biología Molecular Severo Ochoa (CBM), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

Purpose: Cellular senescence is a hallmark of aging, and a key contributor to age-related disease. Senescence-Associated Proteins (SAPs), when measured in plasma, are promising biomarkers for assessing senescence burden; however, senescent cells are numerous and heterogeneous by cell type and the relative importance of SAPs originating from different tissues is unknown. This study leverages the SenCat, a novel database of cell-type specific senescence signatures, to evaluate the clinical relevance of tissue-specific senescence burden.

Methods: This study uses machine learning to investigate clinical associations of tissue-specific SAPs in circulating plasma in two longitudinal studies, including 1275 individuals from the Baltimore Longitudinal Study of Aging (BLSA) and 997 from the Italian InCHIANTI study. Tissue-specific SAPs were identified using mass spectrometry from 15 cell types including preadipocytes, astrocytes, and PBMCs, among others. Plasma levels of these tissue-specific senescence signatures were assessed for associations with a broad range of clinical parameters including mobility and disease status.

Results: Tissue-specific senescence burden showed unique clinical associations that map to their corresponding health domain. For example, renal senescence best associated with kidney disease and lung senescence best predicted pulmonary disease. Additionally, a panel of SAP were identified as a high-impact panel that predicted many clinical traits across several health domains.

Conclusion: These findings demonstrate that health status can be modeled non-invasively and with higher resolution than previously determined, and that senescence signatures can serve as biomarkers to inform clinical studies.

Poster Abstracts

P11

AI-based biomarker discovery in CLN3

Sun, S[1]; Do, AD[2]; Zhu, Q[1]

1. Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD
2. Unit on Cellular Stress in Development and Diseases, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD

CLN3, also known as juvenile neuronal ceroid lipofuscinosis, is a rare and progressive neurodegenerative disorder characterized by the accumulation of lipopigments in the brain, leading to cognitive decline, seizures, and vision loss. This devastating condition primarily affects children and young adults, with symptoms typically appearing between the ages of 4 and 10. The pathogenesis of CLN3 disease involves variants in the CLN3 gene, which encodes a protein of unclear function but is crucial for normal cellular processes. Current treatments are largely symptomatic and supportive, including seizure management and physical therapy, but there is no cure or disease-modifying therapy available. Biomarkers are critical for understanding disease mechanisms, monitoring disease progression, and evaluating therapeutic responses. In this study, we developed various machine learning models to systematically predict potential novel proteins biologically relevant to CLN3, by analyzing proteomics and laboratory data collected via a prospective CLN3 natural history study protocol (NCT03307304). To further examine the biological mechanism of those predicted proteins to CLN3 disease, we performed two different approaches, 1) conducting KEGG pathway enrichment analysis to obtain any enriched pathways associated with CLN3; and 2) building Protein-Protein Interaction (PPI) network with the proteins from those enriched pathways to identify hub proteins, such as EGFR, HIF1A, ACAN, and BSG, as biomarker candidates based on calculated network measurements. The biological associations of the identified biomarkers with CLN3 will be further evaluated via biological experiments.

Poster Abstracts

P12

Path2Space: An AI approach for cancer biomarker discovery via histopathology inferred spatial transcriptomics

Campagnolo, EM[1]; Shulman, ED[1]; Lodha, R[1]; Stemmer, A[1]; Jiang, P[1]; Caldas, C[2]; Knott, S[3]; Hoang, DT[1]; Aldape, K[4]; Ruppin, E[1]

1. Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, Bethesda, MD

2. Department of Clinical Biochemistry and Institute of Metabolic Science, University of Cambridge, Cambridge, UK

3. Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA

4. Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD

Spatial transcriptomics (ST) assays are transforming our understanding of tumor heterogeneity by enabling high-resolution, location-specific mapping of gene expression across the tumor microenvironment. However, the high cost of these assays has limited their application to modest cohort sizes, limiting their application in large-scale spatial biomarker discovery. Here we present Path2Space, a deep learning approach that predicts spatial gene expression directly from histopathology slides. Trained on substantial breast cancer ST data, it robustly predicts the spatial expression of over 4,300 genes in independent validation cohorts, significantly outperforming 12 state-of-the-art ST prediction methods. Path2Space accurately infers cell-type abundances in the tumor microenvironment (TME) based on the inferred ST data. It characterizes the TME of ~1,100 TCGA breast tumors, identifying three new spatially-defined breast cancer subgroups with distinct survival rates. Notably, Path2Space H&E-inferred TME landscapes enable more accurate predictions of patient response to both chemotherapy and trastuzumab than those obtained by established sequencing-based biomarkers. By enabling spatial transcriptomic profiling directly from widely available histopathology slides, Path2Space provides a scalable and cost-effective alternative to sequencing-based assays. It opens new avenues for large-cohort spatial biomarker discovery, and facilitates clinically actionable insights into tumor biology, prognosis, and therapy response.

Poster Abstracts

P13

Stochasticity in cancer immunotherapy maps with the rarity of critical Spark T cells

Salazar-Cavazos, E[1]; Jia, D[1]; Missolo-Koussou, Y[2,3]; Kenet, AL[1]; Achar, S[1,4]; Dada, H[1,4]; Kondo, T[5]; Krishnan, A[1,4]; Taylor, N[5]; Jiang, P[6]; Waterfall, J[2]; DeVoe DL[7]; Altan-Bonnet, G[1]

1. Immunodynamics Group, Laboratory of Integrative Cancer Immunology, Center for Cancer Research, National Cancer Institute, Bethesda, MD

2. Inserm U830 and Translational Research Department, Institut Curie, PSL Research University, Paris, France

3. Inserm U932, Institut Curie, PSL Research University, Paris, France

4. Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

5. Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD

6. Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, Bethesda, MD

7. Department of Mechanical Engineering, University of Maryland, College Park, MD

Cancer immunotherapies result in highly variable responses in different patients even when the treatment is the same. Strikingly, some murine tumor models show large variability in the outcome of cancer immunotherapies, even when the mice, tumor cells and anti-tumor immune cells injected into mice are all genetically identical. Here, we sought to analyze this variability in adoptive cell therapies to identify the immune cell subset driving this variability.

To quantify variability ex vivo, we extracted mouse TCR-transgenic CD8+ T cells, and co-cultured them with antigen-expressing tumor cells in a high-throughput robotic system. We used multiplexed in vitro assays and single-cell analysis of thousands of samples to analyze the inter-replicate variability of tumor cell killing and immune activation. We identified conditions (e.g. cell numbers, antigen quality, tumor cell types) where macroscopic large variations in the immune response against cancer cells are observed between highly-controlled technical replicates. Stochastic activation of a rare subset of CD8+ T cells, so-called Spark T cells, coupled to a paracrine IFN-γ-driven positive feedback explains this measured “noise” in immunotherapeutic reactions. We then developed a custom-designed machine-learning pipeline (Stochasticity-based Identification of Cell Subsets a.k.a. StoICS) to identify the subset of immune cells responsible for the immunotherapeutic variability. We applied StoICS to identify the Spark T cells in murine naïve T cells, and in human TCR-engineered T cell blasts prepared as for adoptive T cell therapy. We then show that diverse levels of Spark T cells in tumor samples explain variable outcomes in cancer immunotherapies.

Poster Abstracts

P14

Deep Learning reveals the significant contribution of silencer variants to human diseases and traits

Huang, D[1]; Ovcharenko, I[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD

Although disease-causal genetic variants have been found within silencer sequences, we still lack a comprehensive analysis of the association of silencers with diseases. Here, we profiled 2.8 million silencers across 97 human samples derived from a diverse panel of tissues and developmental time points, using deep learning models.

These enhancers exhibit strong enrichment in disease-associated variants, which are comparable in overrepresentation but differ in functional characteristics from their target genes compared to enhancers, highlighting the distinguishing role of silencers in human health. For example, in neuronal biosamples, Parkinson's disease variants exhibit an average of over 2 times enrichment within silencers compared to enhancers. The disruption of apoptosis in neuronal cells is associated with both schizophrenia and bipolar disorder and can largely be attributed to variants within silencers, similar to the disruption of GABAergic interneurons, a central factor in the onset of schizophrenia. Our model permits a mechanistic explanation of causative SNP effects by identifying altered binding of tissue-specific repressors and activators, validated with 70% of directional concordance between our predictions and SNP-SELEX experimental quantification of transcription factor binding affinity.

In summary, our results indicate that advances in deep learning models for silencers enable the discovery of disease-causal silencer mutations on a whole-genome scale, effectively 'doubling' the number of functionally characterized GWAS variants. This provides a basis for explaining mechanisms of action and designing novel diagnostics and therapeutic methods addressing dysregulated pathways with disrupted silencers.

Poster Abstracts

P15

Predicting piRNA cluster regions from genomic sequences using deep learning

Ahrend, F[1,2]; Meister, G[2]; Haase, AD[1]

1. National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD
2. Regensburg Center for Biochemistry (RCB), Laboratory for RNA Biology, University of Regensburg, Regensburg, Germany

piRNA clusters are discrete genomic regions that give rise to PIWI-interacting RNAs (piRNAs), a specialized class of small non-coding RNAs essential for safeguarding germ cells against the mutagenic activity of transposable elements. piRNA clusters serve as the information source for the small RNA defense system, producing thousands of piRNAs that collectively recognize and silence genomic threats. While their biological role is well established, the sequence features and targeting rules that govern piRNA cluster activity remain poorly understood. This early-stage project aims to systematically decode these patterns using machine learning. Leveraging annotated piRNA clusters across multiple mammalian species, we plan to train convolutional neural networks (CNNs) to learn how factors like sequence composition, piRNA read coverage, mismatch tolerance, and positional bias contribute to effective silencing.

Our approach will use sigmoid-based output layers to predict not just the presence of clusters but their relative strength – potentially revealing design principles of this selective RNA-guided system. We are currently shaping the modeling pipeline and welcome feedback on CNN architectures, input feature selection, and cross-species generalization strategies. We're especially eager to connect with colleagues interested in transfer learning, genome annotation, or piRNA biology.

Poster Abstracts

P16

Comparative Analysis of Deep Learning Models for Predicting Causative Regulatory Variants

Manzo, G[1]; Borkowski, K[1]; Ovcharenko, I[1]

1. Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD

Motivation: Genome-wide association studies (GWAS) have identified many noncoding variants linked to complex traits and diseases. However, distinguishing causal variants from those merely associated remains a major challenge. A small subset of noncoding variants has true regulatory effects, which can only be detected through accurate models assessing DNA sequence variation. Deep learning approaches—especially Convolutional Neural Networks (CNNs) and transformer-based models—have shown promise in predicting these effects, particularly in enhancers, by leveraging genomic and epigenomic patterns. Yet, the absence of standardized benchmarks and consistent evaluation criteria across studies makes model selection difficult.

Results: This study benchmarks cutting-edge deep learning models for predicting the regulatory effects of genetic variants on enhancers. Using nine datasets derived from MPRA, raQTL, and eQTL experiments, we assessed 54,859 SNPs across four human cell lines. Our findings show that CNN-based models, such as TREDNet and SEI, consistently achieve high accuracy in predicting SNP effects. Meanwhile, hybrid CNN-transformer models like Borzoi excel at identifying causal variants within linkage disequilibrium blocks. Although fine-tuning improves the performance of transformer models, they generally remain less effective than CNN and hybrid models under optimized training conditions.

Poster Abstracts

P17

Regulatory plasticity of the human genome

Srivastava, J[1]; Ovcharenko, I[1]

1. Computational Biology Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD

Evolutionary turnover in unconstrained enhancers has driven phenotypic divergence during past speciation events and continues to facilitate environmental adaptation through variants. Even closely related species like humans and chimpanzees with ~98% genomic identity show significant differences in morphology and cognition which are largely attributed to enhancer divergence, raising key questions: What fraction of the genome undergoes regulatory turnover due to accumulating variants? Are certain genes or transcription factor binding sites more prone to changes? We addressed these questions using a deep learning model to identify substrates of regulatory turnover using genome wide mutations mimicking three evolutionary pathways: recent history (human-chimp substitutions), modern population (human population variation), and mutational susceptibility (random mutations). We observed that >80% of the novel activity arises from repurposing of enhancers between cell-types. Turnover in a cell-type is remarkably similar across the three pathways, despite only ∼19% overlap in the source regions. The highest turnover occurring near neurodevelopmental genes including CNTNAP2, NPAS3, and AUTS2. Flanking enhancers of these genes undergo high turnover irrespective of the mutational model pathway, suggesting their intrinsic plasticity in cognitive evolution and recurrent roles in neuropsychiatric diseases. Based on susceptibility to random mutations, these enhancers were identified as vulnerable by nature and feature a higher abundance of cell type-specific TFBSs, whereas mutationally robust enhancers are enriched for globally expressed general transcription regulators. Our findings suggest that while robust enhancers contribute to the stability of core regulatory networks, enhancer repurposing within vulnerable loci serve as hotspots of cell type- and species-specific regulatory innovation.

Poster Abstracts

P18

DNA methylation differences in children of severe acute malnutrition suggest epigenetic networks at play via machine learning

Li, Q[1]; Hanchard, N[1]

1. Childhood Complex Disease Genomics Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Childhood Severe Acute Malnutrition (SAM) poses serious health risk in children worldwide. The two primary forms of SAM, the Kwashiorkor (ESAM) and the milder form Marasmus (NESAM) have distinct differences in morbidity and mortality risk. However, those difference cannot be attributed to dietary nor environmental factors alone. In our previous study of children with SAM from Jamaican (N=110) and Malawi (N=191), we found strong evidence of lower DNA methylation at >800 CpG sites across the genome in individuals with ESAM relative to those without. Here, we employed a random forest (RF) approach to identify top CpG sites with the highest prediction scores, and correlated them with biological evidence to hypothesize a network of genes that collectively, indicates an elevated risk of ESAM.

Given ~486,000 CpG sites, we tested various tuning parameters to allow for a deep search, and used permutation test to rank the predictors. To account for population stratification and confounding variables, we included covarying factors as fixed features besides randomly assigned features. Those factors included sex, the principal components (PCs) estimated from genotypes and from DNA methylation data. At average prediction error of ~.37 across various models, we found that XYLT1 on chr16, MFAP2 and ZC3H12A on chr1, and SOX7 on chr8 as the top CpG predictors. We then searched the literature using PubTator3 to find interactions between the top CpG sites and related diseases, in order to identify potential biological networks. Our results are a promising foundation for employing a multi-omics machine learning model.

Poster Abstracts

P19

Modeling cCREs using deep learning with applications to prioritizing candidate causal mutations from GWAS data.

Hudaiberdiev, S[1]; Ovcharenko, I[1]

1. National Center for Biotechnology and Information, National Library of Medicine, National Institutes of Health, Bethesda MD

Developmental stuttering is a speech disorder affecting 7-15% of children aged 2-4 and around 1% of adult population. Varying heritability estimates of stuttering have been reported to range from 0.42 to 0.84 in twin studies. The biggest evidence on the genetic components of stuttering came from familial genetic studies. However, two different GWAS studies on stuttering did not produce hits within the significance threshold level of p<1-E8 accepted in the field. In addition, these two GWAS studies resulted in completely disjoint set of suggestive SNP sets associated with the trait, highlighting the heterogeneity and complex nature of the phenotype. To aid in search for candidate causal mutations, we employed a method that we have developed and successfully applied for analyzing Type 2 Diabetes GWAS data. First, we conducted wide range of analyses for detecting implicated tissues and cell types, using the data from GTEx, PsychENCODE and ENCODE Projects. We observed that stuttering-related mutations are active in wide range of tissues, resulting in only 4 brain and neural cell types out of 23 significantly enriched cell types. Crucially, the astrocyte of cerebellum were among the top enriched tissues, dysfunctions in which have been shown to be correlated with stuttering. Using our model trained to recognize enhancers, we prioritized SNPs from fine-mapped GWAS sets for each locus, and short-listed 11 GWAS SNPs in the enhancer regions of astrocytes with increased likelihood affecting expression of genes previously linked to stuttering.

Poster Abstracts

P20

Variant Effect Predictions for *PTPN11* Missense Variants with MutPred2

Um[1]; Mooney[1]

1. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

The PTPN11 gene encodes the Src homology 2 domain-containing protein tyrosine phosphatase (SHP2), a key regulator of cell growth, differentiation, and apoptosis through its modulation of various signaling pathways, including the RAS/ERK signaling pathway. Missense variants in PTPN11 disrupt SHP2's proper catalytic activity and the regulation of signaling pathways, leading to conditions such as Noonan syndrome, LEOPARD syndrome, or juvenile myelomonocytic leukemia. SHP2 variants have a wide spectrum of molecular disruptions leading to both gains and losses of function at the molecular level as well as gains and losses of function at the phenotypic level. While NS and JMML are associated with gain-of-function variants of SHP2, loss-of-function variants are thought to underlie LEOPARD syndrome. In this study, we model the underlying causes of pathogenicity of missense variants in PTPN11 and compare gain and loss of function variants that cause disease.

Poster Abstracts

P21

Using Genomics Data and Literature for Basket Trial Design in Rare Diseases

Moon, S[1]; Maine J[1]; Mathe, E[1]; Zhu, Q[1]

1. Division of Pre-Clinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD

Gaining insight into the underlying molecular etiologies of rare diseases can aid cross-disease research, inform the design of basket trials, and identify drug repurposing opportunities. In our preliminary study, we identified 36 rare disease clusters based on common genetic causes and biological mechanisms. However, there clusters were too broad for basket trial applications. In this study, we refined these clusters by collecting allelic variant data from the Online Mendelian Inheritance in Man (OMIM), along with corresponding Sorting Intolerant From Tolerant (SIFT) scores for single nucleotide polymorphisms (SNPs) and transcripts from Ensemble validated from the Medical Genomics Japan Variant Database (MGeND). We assessed the functional impact of gene mutations using SIFT scores, calculating the ratio of deleterious to tolerated cases (deleterious cases / (deleterious cases + tolerated cases)). We generated an matrix with imputed data by extracting the deleterious level of genetic and mutation data for each rare disease, and identified shared mutations across diseases. Then, we applied Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to the imputed matrix, creating sub-clusters on the top of the 36 clusters. Our results illustrate consistent findings with the published studies of basket trial design for instance, a subcluster of NLRP3 mutation-related diseases including Neonatal Onset Multisystem Inflammatory Disease, Familial Cold Autoinflammatory Syndrome, and Muckle-Wells Syndrome.

Poster Abstracts

P22

Donor-specific digital twin for living donor liver transplant recovery

Halder, S[1]; Periwal, V[1]

1. Laboratory of Biological Modeling, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

The remarkable capacity of the liver to regenerate its lost mass after resection makes living donor liver transplantation a successful treatment option. However, donor heterogeneity significantly influences recovery trajectories, highlighting the need for individualized monitoring. With the rising incidence of liver diseases, safer transplant procedures and improved donor care are urgently needed. Current clinical markers provide only limited snapshots of recovery, making it challenging to predict long-term outcomes. Following partial hepatectomy, precise liver mass recovery requires tightly regulated hepatocyte proliferation. We identified distinct gene expression patterns associated with liver regeneration by analyzing blood-derived gene expression measurements from twelve donors followed over a year using weighted gene co-expression network analysis. Using a deep learning-based framework, we integrated these patterns with a mathematical model of hepatocyte transitions to develop a Personalized Progressive Mechanistic Digital Twin - a virtual liver model predicting patient- specific recovery trajectories. This approach integrates clinical genomics and computational modeling to enhance post-surgical care, ensuring safer transplants and improved donor recovery.

Poster Abstracts

P23

Identification of essential transcription factors by IAN: a new perspective on T cell licensing

Shi, G[1]; Nagarajan, V[1]; Caspi, RR[1]

1. Laboratory of Immunology, National Eye Institute, National Institutes of Health, Bethesda, MD

The rapid accumulation of "Omics" data has revolutionized biological research, enabling researchers to explore complex biological systems with unprecedented depth. In transcriptomic studies, analysis typically involves generating a list of differentially expressed genes (DEGs) and then performing enrichment analysis to gain an overall understanding of the system. Despite numerous enrichment analysis tools and methods available, along with thousands of gene sets to select from, current methods often produce biased results. IAN (Intelligent System for Omics Data Analysis) is an R package addressing the challenge of integrating, analyzing, and interpreting high-throughput Omics data using a multi-agent artificial intelligence (AI) system. We performed enrichment analysis on a publicly available dataset (GSE38645) regarding T cell licensing, using IAN. T cell "licensing" is the process by which circulating T cells specific to brain or eye antigens transiently lodge in the spleen or lung, where they undergo gene-expression and functional reprogramming that enables them to migrate and cross the blood-brain or blood-retinal barrier, and initiate autoimmune inflammation. Our findings indicate that cytokine/chemokine signaling and cell adhesion molecules are positively associated with licensed T cells, whereas cell cycling, metabolism, and protein processing are negatively correlated. Notably, three transcription factors – FOXO1, MECOM, and JUN – act as key regulators, bridging those positively and negatively related pathways. The roles of these transcription factors were further validated using single sample gene set enrichment analysis (ssGSEA) in an independent publicly available dataset (GSE57098). Collectively, these findings provide new insights into the molecular mechanisms underlying T cell licensing.

Poster Abstracts

P24

WSIomics: An Automated Pipeline for Training Multimodal AI Models to Classify therapy response of cancer patients using whole slide images and transcriptome data

Park, M[1]; Yan, C[1]; Chen, Q[1]; Khanna, R[1]; Tanis, J[1]; Meerzaman, D[1]

1. Center for Biomedical Informatics and Information Technology, National Cancer Institute, National Institutes of Health, Bethesda, MD

Prediction of therapy responses in cancer patients is essential for personalized treatment strategies to improve treatment outcomes. Multimodal AI models that integrate various data types, such as whole slide images (WSIs), transcriptome data, genetic data, and clinical data, are reported to perform better than single modalities. In this study, we have developed a fully automated pipeline to train multimodal AI models to predict therapy responses of cancer patients. WSIs and transcriptome datasets were used to train multimodal AI models. The WSI training pipeline was developed based on the CLAM pipeline, which uses attention algorithms. Transcriptome and multimodal models were developed based on simple multi-layer perceptron models. For high-throughput AI model development, the marker genes for transcriptome modality were automatically identified in the pipeline by analyzing the trendline between expression values and progression-free interval values. The multimodal approach was applied to cohorts of eleven cancer types in the TCGA database, which reported neo-adjuvant treatment. The models trained with multimodal data revealed superior performance in seven of the 11 cancer types compared to single-modality models. This study underscores the importance of multimodal approaches in advancing precision oncology and provides a foundational framework for further exploration of multimodal data integration in cancer research.

Poster Abstracts

P25

Combining Real and Synthetic Data to Overcome Limited Training Datasets in Multimodal Learning

Marini, N[1]; Liang, Z[1]; Rajaraman, S[1]; Xue, Z[1]; Antani, S[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD

Biomedical data are inherently multimodal capturing complementary aspects of patient health, where, for example, images provide "low-level" visual features, while associated textual reports summarize "high-level" diagnostic findings. Artificial intelligence (AI) algorithms that integrate multiple modalities into a single data representation can significantly improve clinical decision-making. For example, multimodal foundation models, that are generally trained unsupervised, can integrate information from multiple data types and effectively perform a wide variety of tasks.

However, the development of reliable multimodal AI requires use of large training datasets with samples from modalities of interest. Biomedical datasets tend to be unimodal, skewed, often including just basic class labels. Case in point, consider various publicly available skin lesion data sets that lack of annotated reports paired with the images.

The goal of this work is to present a multimodal architecture that encodes fine-grained text representations within image embeddings to build a robust representation of skin lesion data, exploiting real and synthetized data. Large language models (LLMs) are used to synthesize textual descriptions that are paired with the original skin lesion images and used for model development. The architecture is evaluated on three tasks: skin lesion image classification, multi-modal data retrieval, and the linkages between visual and textual concepts. The latter two tasks are a consequence of architectural design and do not need supervised training.

The proposed multimodal representation outperforms the unimodal one on the classification of skin lesion images and allows the extraction of knowledge from datasets without the need for additional annotations.

Poster Abstracts

P26

Towards a better CAR through *in vitro* and *in silico* Perturbations

Cordes, S[1]

1. Laboratory of Immune Regulatory Genomics and Epigenomics, Translational Stem Cell Biology Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD

The durability of gene and cellular therapies depends on the persistence of adoptively transferred cells, while their efficacy relies on mature effector function. Physiologically, cells rarely exhibit indefinite self-renewal concurrently with mature functionality. Thus, optimizing cellular therapies is inherently a dynamical challenge.

Genetic knockouts of transcription factors and epigenetic modifiers broadly alter cellular function. In contrast, perturbations of cis-regulatory elements (CREs) generally exert localized effects, providing more precise control of cellular states. However, systematically screening numerous candidate CREs remains experimentally impractical.

We developed a machine learning model, trained on single-cell multi-omics data, to predict cellular state dynamics following CRE perturbations. The model prioritizes therapeutic interventions by computing metrics directly relevant to the persistence and efficacy of adoptive cells.

Model predictions indicate that knocking out the transcription factor TCF7 enhances transitions from naïve and memory T cells to effector states. Conversely, disruption of a candidate CRE near KLF10 (chr8:101,790,788-101,791,270) significantly promotes transitions into naïve and memory subsets, potentially enhancing persistence.

Our computational approach efficiently prioritizes an experimentally intractable number of candidate perturbations using therapeutically relevant metrics. This targeted method enhances experimental feasibility and accelerates discovery.

Future work includes integrating additional single-cell multi-omic modalities, such as single-cell Hi-C, to further elucidate regulatory dynamics. This integration promises to refine predictions and facilitate the design of more effective, persistent cellular therapies.

Poster Abstracts

P27

Optimizing CAR costimulatory domains using contrastive learning and optimal transport on high-throughput screening data

Kelly, C[1]; Bahr, R[2]; Zhu, W[2]; Keyvanfar, K[3]; Dagur, P[3]; Cordes, S[2]

1. School of Medicine, Georgetown University, Washington, DC
2. Laboratory of Immune Regulatory Genomics and Epigenomics, Translational Stem Cell Biology Branch, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD
3. Flow Cytometry Core, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD

Chimeric antigen receptor (CAR) T cells have established themselves as therapies for B cell malignancies, where they induce prolonged remission in most patients, though fewer than half achieve durable control of their disease. CARs follow a modular design, with an extracellular portion conferring target-specificity and two or more intracellular signaling domains. One of the signaling components, called the costimulatory domain, is crucial for ensuring proper and controlled immune responses. Absence of this domain results in T cell anergy, but the optimal costimulatory domains remain unknown.

To gain a predictive understanding of immunophenotypes resulting from different costimulatory domains, we trained a contrastive learning model to learn a shared embedding between ESM-2 representations and our experimentally determined immunophenotypes under different co-culture conditions. We trained two dense neural networks to project to a shared embedding using 90% of our data, we achieved a moderate Fraction of Samples Closer than the True Match (FOSCTTM) of 0.45 in our validation sample. We further refined this shared embedding with fused Gromov-Wasserstein optimal transport to achieve a respectable FOSCTTM of 0.13.

Integration of cellular therapy experimental results into a pre-trained PLM via transfer learning permits identification of latent dimensions representing the complex patterns governing cellular behavior that depend on costimulatory domain sequence. This insight allows interpolation within that latent space to locations of desired immunophenotypes and then use any decoder to generate optimal synthetic costimulatory domain sequences which induce such phenotypes. We plan arrayed testing of synthetic domains predicted to optimize CAR T cell immunophenotype proportions.

Poster Abstracts

P28

Deep learning cellular dynamics from single-cell RNA sequencing

Duan, X[1]; Periwal, V[1]

1. Lab of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Bethesda, MD

Single-cell RNA sequencing (scRNA-seq) provides a powerful framework for studying cellular heterogeneity, transitions, and regulatory networks. However, reconstructing the underlying dynamical processes governing these transitions remains a major challenge due to the high-dimensional nature of gene expression data. To address this, we develop a variational autoencoder (VAE)-based approach that learns a low-dimensional latent representation of cellular states and models their temporal evolution. We apply our framework to gene expression data from Drosophila melanogaster blastoderm embryos, compiled by Fowlkes et al., which includes measurements across multiple time points using a registration technique. In this approach, gene expression profiles are encoded into a low-dimensional latent space, where we train a neural stochastic differential equation (SDE) network to capture the continuous dynamics of latent states over developmental time. The learned neural SDE models the progression of cellular states, and a decoder subsequently maps these evolving latent representations back to the original high-dimensional gene expression space, allowing for both accurate reconstruction of observed transcriptional patterns and insight into the underlying dynamical processes. Future directions include the use of symbolic regression to extract dynamical models from the inferred trajectories.

Poster Abstracts

P29

Predicting Chemical Toxicity by Applying a Hierarchical Bayesian Approach with Priors to the Tox21 Assay Data

Zeng, W[1]; Yadaw, AS[1]; Mehta, K[1]; Sanjak, J[2]; Nguyen, D-T[3]; Huang, R[1]; Mathé, EA[1]

1. Division of Pre-Clinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD
2. Booz Allen Hamilton, Bethesda, MD
3. Digital R&D Solutions, Pfizer, Inc., Maryland

In vitro testing for drug activity and toxicity experimentally is an expensive and time-consuming process yet critical for identifying candidate treatments that will translate into the clinic. In silico activity and toxicity predictions using machine learning can speed up the drug development process by generating accurate predictions. Resulting models can also highlight the most important factors that contribute to predicting outcomes. The cell viability assay has been commonly used as a counter screen for compound cytotoxicity in high throughput screening assays. In this paper, the cell viability counter-screen data generated from screening 50 in vitro assays against the "Tox21 10K library" which consists of 8,947 unique compounds, each of which can be represented using 208 chemical descriptors, was used for the first time to develop predictive models for in vitro cytotoxicity of small molecules. This library was tested on 13 different cell lines, comprising tumor type (primary, metastatic, normal), 11 different tissue types, cell types (epithelial, lymphoblast, fibroblast, or epithelial-like), sex and organism (animal cell and human cell). Predictions were assessed using state-of-the-art machine learning approaches, including logistic regression, a naïve Bayes and a hierarchical Bayesian model. In our approach, we evaluated a) improvements in prediction capacity between models within individual assays and across assays with and without the use of biological meta information (gender & organism); b) the chemical descriptors that contributed most to predictions. The resulting balanced accuracies range from 0.66 to 0.81, depending on the assay and model used. Logistic regression and hierarchical Bayesian models resulted in similar balanced accuracies, and considering sex and organism in the models did not substantially improve model predictions. Overall, our results demonstrate the utility of the Tox21 in vitro assay data in predicting in vitro cytotoxicity. These predictive models are critical for prioritizing chemicals as drug candidates for further clinical evaluation.

Poster Abstracts

P30

Interpretable Drug Response and Drug-Target Interaction Prediction Using Artificial Intelligence

Inoue, YI[1,2,3]; Song, TS[1]; Fu, TF[4]; Luna, AL[2,3]

1. Computer Science, University of Minnesota, Minneapolis, MN
2. Computational Biology Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD
3. Developmental Therapeutics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD
4. Department of Computer Science, Nanjing University Nanjing, Jiangsu, China

A challenge of using machine learning (ML) in biomedical research is a lack of interpretability, which limits its support of data-driven decisions with explanations. We explore this topic here, focusing on cancer drug response and mechanism prediction. We introduce two components: GraphPINE (Graph Propagating Importance Network for Explanation) and DrugAgent. GraphPINE is a graph neural network (GNN) model for drug response prediction using multi-omics data (e.g., gene expression) and interaction networks (e.g., protein-protein). The novelty of GraphPINE lies in its initialization of importance scores using biological prior knowledge (drug-target interactions, DTIs) from literature and a dynamic updating mechanism. We build on concepts from LSTM (Long Short-Term Memory), relying on previous predictions as hidden states to advance GNNs such that GraphPINE initializes importance scores using prior knowledge and updates these scores during model training. We apply GraphPINE to NCI60 data; GraphPINE achieves AUROC of 0.796 and AUPRC of 0.894 for 952 drugs. Separately, we developed DrugAgent, a multi-agent system integrating knowledge graphs, internet searches, ML methods, and large language models (LLMs) to improve DTI prediction. DrugAgent was evaluated using 178 kinase inhibitors against 300 kinases; DrugAgent achieves superior performance (AUROC: 0.905 and AUPRC: 0.529). Interpretable subgraphs accompany GraphPINE results, while DrugAgent results are enriched with prior knowledge. Multiple lines of evidence must support conclusions in biomedical research. The bioinformatics efforts here build on this fundamental notion to draw in additional data from heterogeneous sources uniformly and transparently as part of ensemble results presented to users.

Poster Abstracts

P31

Advancing antidepressant discovery through machine learning-based QSAR modelling and insights from SHAP features

Oyinloye, P[1]; Wu, F[1]; Lee, KH[1]; Shi, L[1]

1. Computational Chemistry and Molecular Biophysics Section, Molecular Targets and Medications Discovery Branch, National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD

The serotonin transporter (SERT), a member of the neurotransmitter sodium symporter family, is essential for regulating extracellular serotonin levels in the brain, influencing mood, emotion, motivation, and memory. By regulating serotonin availability, SERT not only contributes to the therapeutic effects of certain antidepressants but also holds potential for mitigating opioid abuse. In contrast, several inhibitors of the dopamine transporter (DAT), which shares significant homology with SERT, are widely abused psychostimulants. In this study, we conducted a comprehensive data extraction from ChEMBL and DrugBank to retrieve and filter compounds with significant affinities for both SERT and DAT. We found that antidepressant drugs are typically associated with higher affinities in SERT and lower affinities in DAT. We further assembled datasets of both SERT and DAT ligands from ChEMBL to build machine learning-based Quantitative Structure-Activity Relationship (QSAR) models. Additionally, selectivity models were developed for SERT in comparison to DAT. These models showed robust predictive performance in predicting the affinity at SERT, DAT, and the selectivity for SERT. To further interpret the outputs of these models, we applied SHapley Additive exPlanations (SHAP) to identify key chemical features that distinguish SERT-selective from DAT-selective compounds. This research enhances our understanding of the structure-activity relationships of SERT ligands and lays the foundation for the rational design of the SERT selective ligands with reduced abuse liability.

Poster Abstracts

P32

AI-driven development of ALDH3A1 selective inhibitors

Jain, SJ[1]; Yasgar, AY[1]; Nilova, AN[1]; Dalal, AD[1]; Rai, GR[1]; Zakharov, AZ[1]

1. National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD

Human Aldehyde Dehydrogenase 3A1 (ALDH3A1), a member of the ALDH enzyme family, plays a critical role in metabolizing aliphatic and aromatic aldehydes, protecting cells from oxidative stress, and maintaining homeostasis. Emerging evidence highlights its elevated expression in cancer stem cells, where it may promote survival through oxidation of lipid-derived aldehydes and support the tumor microenvironment. Therefore, pharmacological inhibition of ALDH3A1 represents a promising therapeutic strategy in oncology. In this study, we present a computationally driven hit-to-lead workflow combining reaction-based enumeration, molecular docking, and AI/ML models to identify and optimize potent ALDH3A1 inhibitors. A structure-based virtual screen of our internal compound library yielded 47 active compounds from 255 virtual hits (hit rate ~18%). Several chemotypes were further validated using the AldeFluor cell assay, leading to the selection of two series for medicinal chemistry optimization. To expand the chemical space around these hits, we performed reaction-based enumeration using the Enamine building block collection (>1 million BBs), generating ~40,000 virtual analogs. These were prioritized using a deep learning model trained on in-house ALDH3A1 data. Over 100 compounds were synthesized and experimentally tested, resulting in several inhibitors with sub-100 nM potency. This study demonstrates the utility of integrating in-silico reaction-based enumeration with AI-driven prioritization to accelerate the discovery of potent ALDH3A1 inhibitors. The identified compounds serve as promising chemical probes and potential leads for therapeutic development targeting cancer-related ALDH3A1 activity.

Poster Abstracts

P33

In silico ADME models in drug discovery

Shah, P[1]; Weber, C[1]; Lim, G[1]; Zhao, T[1]; Sun, H[1]; Jain, S[1]; Zakharov, A[1]; Siramshetty, V[1]; Mathe, E[1]; Xu, T[1]; Huang, R[1]; Xu, X[1]

1. Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD

Drug discovery and development is a long, expensive venture and the cost of bringing a new drug candidate into the market has increased steadily in the past few decades. It costs between 1.6 and 2.8 billion US$ and takes between 10 and 15 years. While thousands of compounds are screened in preclinical discovery, it is estimated that only 10 out of 1000 screened compounds ever become optimized leads that progress into preclinical in vivo testing. Additionally, the clinical attrition rate is also extremely high with <10% candidates reaching the market after entering clinical development. A central piece of the drug discovery and development puzzle includes absorption, distribution, metabolism, and elimination (ADME) studies. This is highlighted by the fact that in the past, >40% of drug candidates failed due to poor pharmacokinetics. While Pharma companies have large databases and computational models on the in vitro ADME properties, these models are not readily available to all drug discovery research groups. To support drug discovery efforts at NCATS, we have collected in vitro ADME data on “drug-like” properties for over 30K compounds synthesized by MedChem scientists. These datasets cover a broad chemical space from over 200 NCATS projects and are used to build the in silico ADME models which can be used to guide the design of new molecules in drug discovery. This work introduces ADME@NCATS, an in silico ADME prediction platform developed at NCATS, along with a curated list of open-access ADME tools.

Poster Abstracts

P34

ClinIQLink: A Neuro-Symbolic Pipeline for QA generation with Crowd-Sourced Human-in-the-Loop Verification

Colelough, BC[1,2]; Bartels, D[1]; Demner-Fushman, D[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD
2. University of Maryland, College Park, MD

The rapid advancements in the area of generative large language models (LLMs) have introduced new opportunities for the automation of question-answer (QA) dataset generation, particularly in high-risk application domains such as medicine. However, the issue of maintaining factual accuracy and making the process of knowledge recall transparent remains open. In this work, we present ClinIQLink: A Neuro-Symbolic Pipeline for QA Generation with Crowd-Sourced Human-in-the-Loop Verification, an automated framework to generate a novel medical QA dataset from subject-matter expert source literature with explicit linkages to the underlying references. We integrate an open-source LLM (LLaMA 3.3 - 70B) into a neuro-symbolic pipeline to structure, extract, and generate atomic QA pairs from medical texts. To enhance the validity of these generated QA pairs, we employed a crowd-sourced human validation process using volunteer physicians from the NIH and US medical schools. The human annotators assessed the factuality and relevance of the generated QA pairs through an interactive web-based interface. To ensure maximum participation and maintain a high level of focus while evaluating, we incorporated gamification elements in the interface design. Our results demonstrate that this hybrid neuro-symbolic and human-in-the-loop approach effectively realizes automation effectiveness and expert validation with the result of high-quality, transparent, and verifiable medical QA data. Our work advances the research toward improved factual grounding of LLM-produced medical content to ensure that AI-based knowledge retrieval complies with the standards of medical accuracy and credibility.

Poster Abstracts

P35

Scientific Review NLP Conflict of Interest Identification

Mollerus, P[1]; Seideman, J[1]; Saraiya, D[1]; Meyer, A[1]; Footer, K[1]; Chang, R[1]; Nguyen, L[1]; Croghan, J[1]; Rosenthal, A[1]; Klinkenberg, L[2]; Meyers, J[2]

1. Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD
2. Scientific Review Program, Division of Extramural Activities, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

In the scientific grant review lifecycle, ensuring a transparent and unbiased evaluation is paramount. Scientific Review Officers (SROs) must meticulously identify potential conflicts of interest (COI) between grant applicants and reviewers, a task that typically necessitates manually sifting through all grant applications to find relevant names. This process is both time-consuming and susceptible to human error. To address this challenge, we developed a machine learning-based solution that leverages natural language processing (NLP) to facilitate identification of potential COI in grant applications. The COI NLP module employs named entity recognition (NER), entity resolution, and optical character recognition (OCR) algorithms to process, identify, and extract names from machine-readable documents and scanned images, with the capacity to process scanned supporting documents. The solution was carefully tuned to err on the side of more false positives rather than miss any actual entities. This COI NLP solution is integrated into a custom, enterprise-wide web application – the Scientific Review Data Management System (SRDMS) – with over 400 users across 23 different NIH ICs. By using NLP to identify potential COI in each grant application, this solution saves SROs an average of 45 minutes of manual review time per application and significantly enhances the efficiency, reliability, and fairness of the grant review process.

Poster Abstracts

P36

Supervised Machine Learning for Scientific Coding Assistance

Seideman, J[1]; Do, W[1]; Tembo, M[1]; Opsahl-Ong, L[1]; Meyer, A[1]; Saraiya, D[1]; Footer, K[1]; Desai, A[2]; Lee, L[2]; Nguyen, L[1]; Croghan, J[1]; Rosenthal, A[1]; Tartakovsky, M[1]

1. Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD
2. Office of Strategic Planning, Initiative Development, and Analysis, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

The NIAID Office of Strategic Planning, Initiative Development, and Analysis (OSPIDA) assigns scientific codes to extramural grants to enable detailed financial reporting of funding by pathogen, disease, immunology, and other categories. This reporting is needed for congressional inquiries, data requests, and interagency decision making. Assigning codes accurately is imperative to support these activities. Currently, this process involves manual curation and classification of scientific content for thousands of grants each year. To assist with this process, we developed the Coding Assistant Tool (CAT) – a supervised machine learning solution that ingests grant application text and returns recommendations for scientific codes. CAT leverages natural language processing (NLP) through a multi-layer perceptron neural network trained on years of prior grant application text and OSPIDA-assigned scientific codes. Testing demonstrated that, after model training, CAT predicts infectious disease codes with accuracy, recall, and precision above 85%. CAT scientific code recommendations are currently being integrated into an existing custom web application, for OSPIDA staff to leverage more seamlessly. The implementation of CAT is saving OSPIDA staff time and effort and enhancing efficiency and consistency in the scientific coding process.

Poster Abstracts

P37

AI helped me write this: using AI to analyze NIH's AI and data science grant portfolio

Piatkowski, GS[1]

1. Obstetric and Pediatric Pharmacology and Therapeutics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

Artificial intelligence (AI) and data science are rapidly transforming biomedical research. NIH's investment in these areas has grown significantly, but identifying relevant grants across all Institutes and Centers remains challenging due to inconsistent terminology and limitations in existing classification methods. In this study, we evaluate and compare multiple approaches to identify data science and AI-related grants funded by NIH from FY2007 to FY2024.

We first apply standard filters using RCDC categories (e.g., "Artificial Intelligence") and keyword heuristics (e.g., "machine learning," "algorithm") to extract baseline sets. To improve accuracy and capture projects that may be missed or misclassified, we test large language models (LLMs) including GPT-4 and open-source LLaMA models. Each model classifies whether a project is AI/data science–related based on Title, Abstract, and Project Terms, and optionally provides subcategories (e.g., NLP, imaging, predictive modeling) and rationale.

Preliminary findings suggest LLMs improve classification precision over keyword-only methods, especially in edge cases or when terminology is vague. The approach also enables flexible classification—for example, distinguishing when AI is central to a project vs. peripherally mentioned. Aggregated results will be used to analyze funding trends over time, by Institute, and by activity code and show how NIH's data science or AI portfolio has changed over time.

This work demonstrates how combining rule-based filters with generative AI can enhance portfolio analysis and support more nuanced tracking of NIH investments in emerging technologies.

Poster Abstracts

P38

Automating conversion of hand-drawn SBGN diagrams to SBGNML using large language models

Balci, H[1]; Luna, A[1,2]

1. Computational Biology Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD
2. Developmental Therapeutics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD

In systems biology, researchers often use diagrams to map how biological processes interact. These diagrams can be created using visual languages such as Systems Biology Graphical Notation (SBGN), which standardizes the representation of biological pathways, with Systems Biology Graphical Notation Markup Language (SBGNML) as the file format for sharing them. Using machine-readable formats is key for enabling data reuse and computational analysis. However, existing software tools for creating SBGN diagrams are often difficult to use for first-time users, due to complex toolbars. Additionally, converting hand-drawn SBGN diagrams into SBGNML after brainstorms is time-consuming. This research investigates large language models (LLMs) to automate this conversion process, focusing on GPT-4o and Gemini 1.5 Pro models. We curated a crowdsourced dataset of 1,000 hand-drawn SBGN images, evenly split between two common representations: Process Description (PD), detailing step-by-step biological events, and Activity Flow (AF), showing the influence of one activity on another. We evaluate model performances in extracting node types, edge connections, and labels. Preliminary results show strong performance in node and label conversion, but edge conversion remains challenging. Gemini 1.5 Pro achieved up to 85.9% accuracy in node conversion and 92% in label extraction for PD, with GPT-4o slightly behind. Edge conversion remains less accurate, with 53% as the highest. We observed similar trends in AF diagrams. This work can enhance SBGNML accessibility by automating diagram digitization and could extend to other fields like engineering and software design, highlighting LLMs' potential for converting conceptual sketches into machine-readable formats.

Poster Abstracts

P39

Cell phenotypes in the biomedical literature: First look at a new corpus

Rotenberg, NH[1]; Leaman, R[1]; Islamaj, R[1]; Fluharty, B[1]; Kuivaniemi, H[1]; Richardson, S[1]; Tromp, G[1,2]; Lu, Z[1]; Scheuermann, RH[1]

1. Division of Intramural Research, National Library of Medicine (NLM), National Institutes of Health, Bethesda, MD
2. Division of Immunology, Department of Biomedical Sciences, Biomedical Research Institute, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Single-cell technologies are enabling the discovery of many novel cell phenotypes, but this growing body of knowledge remains fragmented across the scientific literature. While natural language processing (NLP) offers a promising approach to extract this information at scale, the current annotated datasets required for NLP system development and evaluation do not reflect the complex assortment of cell phenotypes described in recent studies.

We present a new corpus of excerpts from recent articles, manually annotated with mentions of human and mouse cell phenotypes. The corpus distinguishes three types: (1) specific cell phenotypes (cell types and states), (2) heterogenous cell populations, and (3) vague cell population descriptions. Mentions of the first two types were linked to Cell Ontology identifiers where possible, using their meaning in context, with matches labeled as exact or related. Annotation was performed by four cell biologists using a multi-round process, with automated pre-annotation and extensive quality control.

The corpus contains over 22,000 annotations across more than 3,000 passages selected from 2,700 articles, covering nearly half the concepts in the current Cell Ontology. Fine-tuning BiomedBERT in a simplified named entity recognition task on this corpus resulted in substantially higher performance than the same configuration fine-tuned on previously annotated datasets.

This corpus is a valuable resource for developing automated systems to identify cell phenotype mentions in the biomedical literature and a foundation for the future extraction of relationships between cell types and key biomedical entities, including genes, anatomical structures, and diseases.

Poster Abstracts

P40

Automated Survey Collection with LLM-based Conversational Agents

Kaiyrbekov, K[1]; Dobbins, NJ[2]; Mooney, S[1]

1. Cyberinfrastructure and Artificial Intelligence Platforms Section, Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
2. Biomedical Informatics & Data Science, Department of Medicine, Johns Hopkins University, Baltimore, MD

Objective: Phone surveys are crucial for collecting health data but are expensive, time-consuming, and difficult to scale. To overcome these limitations, we propose a survey collection approach powered by conversational Large Language Models (LLMs).

Materials and Methods: Our framework leverages an LLM-powered conversational agent to conduct surveys and transcribe conversations, along with an LLM (GPT-4o) to extract responses from the transcripts. We evaluated the framework's performance by analyzing transcription errors, the accuracy of inferred survey responses, and participant experiences across 40 surveys.

Results: GPT-4o extracted responses to survey questions with an average accuracy of 98%, despite an average transcription word error rate of 7.7%. Participants reported occasional errors by the conversational agent but praised its ability to demonstrate comprehension and maintain engaging conversations.

Conclusions: Our study showcases the potential of LLM agents to enable scalable, AI-powered phone surveys, reducing human effort and advancing healthcare data collection.

Poster Abstracts

P41

Author and affiliated institution extraction from free-form letters using GenAI

Ornek, ME[1]; Zahnen, CR[1]; Chen, M-C[1]

1. Division of Planning, Analysis, and Information Management, Center for Scientific Review, National Institutes of Health, Bethesda, MD

Efficient and effective processing of Letters of Support (LoS) in NIH grant applications is critical for ensuring compliance and mitigating potential conflicts of interest during the peer review process. This project introduces an AI-assisted approach that leverages GPT-4o to automate the identification and extraction of authors and their institutional affiliations from LoS documents.

The workflow begins by preprocessing each combined PDF submission into image and text pages to enable targeted input for GPT-4o. The automation then proceeds in two key stages. First, GPT-4o is prompted to analyze the image pages to determine the start and end of each individual LoS, isolating them from multi-letter files. Next, it is prompted to interpret the corresponding letter text to extract author names and institutional affiliations. Traditional Natural Language Processing (NLP) methods showed limited effectiveness due to the unstructured nature and formatting variability of LoS documents. In contrast, utilizing GPT-4o demonstrated significantly improved performance in handling these complexities.

By automating detection and extraction, this approach reduces manual effort required to screen LoS submissions, minimizes human error, and improves consistency in data handling. The resulting accuracy and repeatability enhance administrative efficiency and support a more transparent and reliable grant review process. This project highlights the transformative potential of generative AI tools in biomedical research administration, enabling streamlined workflows and reinforcing the integrity of peer review operations.

Poster Abstracts

P42

AI in action at the NICHD: Case studies and developmental pathways

Heymann, D[1]; Mykins, M[1]; Zhou, N[1]

1. Referral and Program Analysis Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

The Referral and Program Analysis Branch (RPAB) at the NICHD is leveraging advanced data automation solutions to solve complex problems, streamline processes, and enhance operational efficiency to benefit the broader NICHD community. One of RPAB's responsibilities is to triage incoming referrals and assign them to the appropriate scientific branch. This process is challenging due to the time, and expertise needed to navigate the complexity of overlapping scientific interests across branches.

To overcome these challenges, we developed and implemented an AI/ML Application Referral System that augments and increases efficiency of the grant referral process. The versatility and adaptability of our model also permitted us to repurpose our model framework for other use cases to 1) increase the efficiency of NICHD's study branch section assignment, and 2) handle mission critical items with short turnarounds. To complement our AI/ML models, we are utilizing NIH LLMs, such as ChatBot for Intramural Research Program (ChIRP), for additional refinement of our model predictions.

Throughout our development process we continuously integrate human-in-the-loop feedback from our SMEs to refine and validate model outputs. Our approach of continuously integrating human feedback in our DevOps allows us to deliver tangible benefits such as expedited turnaround, enhanced accuracy, reduction in manual errors and administrative burden, and real-time insights for improved decision-making for our stakeholders. The system is designed to adapt to evolving scientific landscapes and changes in NICHD research priorities. Our RPAB cloud environment structure ensures flexibility and scalability to meet future requirements and expand as needed.

Poster Abstracts

P43

RARe-SOURCE Literature AI: Rare Disease Genotype-Phenotype Associations from Biomedical Literature.

Alodadi, M[1]; Lyons, E[1]; Che, A[1]; Watson, D[1]; Tawa, GJ[2]; Porter, F[3]; Haugabook, SJ[2]; Ottinger, E[2]; Mudunuri, U[1]

1. Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, MD
2. Therapeutic Development Branch, Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD
3. Division of Translational Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

Rare diseases affect millions globally, but information and resources are often scarce [1]. To address this need, the Therapeutic Development Branch in the intramural Division of Preclinical Innovation at NCATS conceptualized Rare-SOURCE™ and in collaboration with the Advanced Biomedical and Computational Science developed and launched this user-centric bioinformatic resource platform for rare disease information. The main objective is to facilitate data mining through a searchable interface that integrates bioinformatics databases and enables users to navigate disease-gene-variant information quickly and efficiently.

Determining genotype-phenotype connections is essential in preclinical and clinical research environments, as it helps in fully understanding the relationship between genetic variations and disease susceptibility. Biomedical literature is a knowledge source with valuable information. However, as the data is largely unstructured, extracting relevant details is challenging, and mining for contextual information such as variant pathogenicity [3], clinical and phenotypic details, and their relations to the disease are not yet resolved.

Advancements in BioNLP and text-mining, powered by machine learning and transformer models like BERT, have significantly improved information extraction from biomedical texts, advancing named-entity recognition, relation extraction, and document classification [4-12]. RARe-SOURCE™ aims to make rare disease literature scalable, disease-agnostic and accessible through its LiteratureAI feature, which leverages NLP to scan titles and abstracts for disease and gene mentions while also integrating other resources for synonyms and aliases. Future enhancements will incorporate GenAI to extract genetic variants along with their functional impact and clinical significance by accurately capturing clinical context, potentially transforming variant pathogenicity predictions for the rare disease community.

Poster Abstracts

P44

TrialGPT: matching patients to clinical trials with large language models

Jin, Q[1]; Wang, Z[2]; Floudas, CS[3]; Wan, N[1]; Chan, J[1]; Chen, F[4]; Gong, C[5]; Bracken-Clarke, D[3]; Xue, E[3]; Fang, Y[1]; Tian, S[1]; Yang, Y[1]; Sun, J[2]; Lu, Z[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD
2. University of Illinois Urbana-Champaign, Champaign, IL
3. National Cancer Institute, National Institutes of Health, Bethesda, MD
4. University of Pittsburgh, Pittsburgh, PA
5. Albert Einstein College of Medicine, Bronx, NY

Introduction

Trial recruitment is a persistent challenge in clinical research, hindered by increasingly complex patient data and eligibility criteria. Traditionally, automatic trial matching relies either on embedding-based techniques—demanding training data and often lacking interpretability—or structuring-based methods––transforming criteria into structured queries. Such approaches can be resource-intensive, necessitating a flexible, explainable solution that increases recruitment efficiency.

Methods

We present TrialGPT,1 a framework leveraging large language models (LLMs) for streamlined patient-to-trial matching. TrialGPT contains three modules: (1) TrialGPT-Retrieval, which uses LLM-generated keywords and hybrid lexical-semantic search to filter trials from databases like ClinicalTrials.gov; (2) TrialGPT-Matching, which analyzes shortlisted patient–trial pairs, providing explanations and citing patient-level sentences on a criterion-by-criterion basis; and (3) TrialGPT-Ranking, which integrates criterion-level predictions, ranking trials by the extent to which criteria are met. Notably, TrialGPT-Retrieval can be bypassed when trial search spaces are relatively small, such as hundreds of institution-specific trials.

Results and Conclusion

On three publicly available cohorts of 183 synthetic patients with over 75,000 trial eligibility annotations from TREC challenges,2 TrialGPT-Retrieval recalls over 90% of relevant trials using less than 6% of the initial collection. Manual evaluations on 1,015 patient-criterion pairs show that TrialGPT-Matching achieves an accuracy of 87.3% with faithful explanations, near expert performance of 88.7%-90.0%. The TrialGPT-Ranking scores are highly correlated with human judgments and outperform the best-competing models by 29.8% in ranking trials. Furthermore, our pilot user study reveals that TrialGPT can reduce screening time by 42.6% in recruitment. Overall, these results demonstrate promising opportunities for patient-to-trial matching with TrialGPT.

Poster Abstracts

P45

A Dataset for Grounded Question Answering from Electronic Health Records to Relieve Clinician Burden

Soni, SS[1]; Demner-Fushman, DD[1]

1. Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD

Patient requests for medical information via patient portals are rising, contributing significantly to desktop medicine and clinician burden. An approach to handling the increasing messaging burden is to assist clinicians in formulating the responses. Thus, automatically generating answers to patients' questions considering their electronic health records (EHRs) is important.

We introduce a novel dataset to develop and evaluate systems that answer patients' questions using clinical evidence from EHRs. It comprises hand-curated patient questions (reflective of portal messages), clinician-identified focus areas in questions, clinician-rewritten questions (to aid in formulating responses), and clinical note excerpts providing context (from MIMIC-III and IV databases). Each sentence in the note excerpt is manually annotated to mark its importance in answering the question as "essential" (must use), "supplementary" (may provide support), or "not-relevant". We evaluate system-generated responses on "Factuality"—measured by F1 scores of cited evidence under strict (considers only "essential" sentences as answers) and lenient (considers both essential and supplementary) criteria—and "Relevance"—assessed by comparing answers to "essential" evidence and the original question.

The dataset contains 130 questions, with a mean note excerpt length of 18.2 (sd-12.1) sentences—including 5.7 (sd-4.0) essential, 1.9 (sd-2.9) supplementary, and 10.6 (sd-8.2) not-relevant sentences. The baseline model, LLaMa 3.3 70B, achieved Factuality F1 of 55.6 (strict) and 57.3 (lenient), and Relevance scores: ROUGE (19.0), SARI (53.5), BERTScore (83.8), AlignScore (52.2), and MEDCON (26.4). This underscores the challenging nature of the dataset, offering a robust benchmark for automated patient messaging systems leveraging EHR data.

Poster Abstracts

P46

Forecasting from Clinical Textual Time Series: Adaptations of the BERT and Decoder Families

Kumar, SK[1]; Noroozizadeh, SN[1]; Weiss, JCW[1]

1. Computational Health Research Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD

Clinical case reports document patient trajectories for communication of findings and best practices across disciplines. While they contain a richer set of information than that of tabular data streams in Electronic Health Records (EHRs), they lack the regularity of inputs and outputs that classical machine learning algorithms typically use. To use these patient trajectories in their textual form, we propose the forecasting task from textual time series, where the inputs are timestamped clinical findings. We define evaluation measures appropriate for time-ordered text setup and test a large suite of large language models from both the BERT and decoder model families. We find that finetuned decoder based models perform the best at forecasting in the near-horizon. We further demonstrate the importance of time ordering, which requires clinical time series construction, as compared to text ordering, the format of the text inputs that LLMs are classically trained on. This highlights the additional benefit that can be ascertained from time-ordered corpora, with implications for temporal tasks in the era of widespread LLM use.

Poster Abstracts

P47

Responsible Integration of Large Language Models in Biomedical Research

Aston, SA[1]; Cheng, H[1]

1. Data Science Core, National Institute on Deafness and Communication Disorders, National Institutes of Health, Bethesda, MD

Large language models (LLMs), such as OpenAI's ChatGPT, are transforming professional workflows, including biomedical research. These AI models, trained on vast textual datasets, can assist in literature review, hypothesis generation, and coding support. However, their integration into scientific research requires careful consideration of key limitations, including limited access to recent findings due to fixed training datasets, no direct access to databases or proprietary content, a lack of real understanding and gaps in reasoning, biases inherited from training data, hallucinations (plausible but false information), and potential plagiarism.

To address these concerns, the NIH and scientific journals have established guidelines to ensure transparency and mitigate risks. Rapid advancements in LLM development also may help overcome some of these challenges.

In this presentation, we explore the responsible use of LLMs in biomedical research by:

1. Presenting results from a survey of NIDCD researchers on their experiences, concerns, and questions regarding LLMs.

2. Reviewing NIH and journal policies on AI-assisted research and publication.

3. Demonstrating examples of LLM applications in biomedical research.

By fostering discussion on best practices, we aim to equip researchers with strategies for leveraging LLMs responsibly, maximizing their potential while minimizing risks in life-changing research.

Poster Abstracts

P48

Multi-Agent Cross-Modal Large Language Model Framework for Chest X-ray Analysis and Integrating COVID-19 Pneumonia Predictions

Liang, Z[1]; Rajaraman, S[1]; Marini, N[1]; Xue, Z[1]; Antani, S[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD

We present a novel multi-agent large language model (LLM) framework to enhance the accuracy and robustness of diagnosing COVID-19 pneumonia on chest X-ray (CXR) images. The framework leverages the Generative Pre-trained Transformer (GPT)-o1 and integrates several specialized artificial intelligence (AI) agents for classification and regression tasks. Using a chain-of-thought reasoning process, GPT-o1 analyzes both CXRs and expert annotations, synthesizes outputs from the agents and assigns confidence weights based on performance metrics to ensure reliable predictions. Our evaluation of the model confirms its capability to grade pneumonia severity while mitigating noise in regression tasks. Our approach addresses challenges introduced due to the variety in the data and any modality-specific limitations. The proposed LLM-based architecture is shown to outperform conventional AI models in COVID-19 pneumonia detection and severity assessment. Our work highlights the potential benefits of using multi-agent LLMs to enhance AI support for clinical decision-making through robust, efficient, and comprehensive AI-assisted diagnostic tools.

Poster Abstracts

P49

RAG2SQL

Nolte, S[1]; Saddler, TO[2]; Reif, DM[2]; Schmitt, CP[1]; Auerbach, SS[2]; Hsieh, J-H[2]

1. Office of Data Science, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, NC
2. Division of Translational Toxicology, National Institute of Environmental Health Sciences, National Institutes of Health, Durham, NC

Toxicity-related databases can be challenging to navigate, requiring users to possess both domain-specific knowledge and proficiency in query languages such as Structured Query Language (SQL). Writing effective queries can therefore demand substantial time and effort. Large Language Models (LLMs) offer a promising solution by enabling natural language interfaces to databases through Text-to-SQL, which translates user questions into SQL queries. This capability can be further enhanced using a Retrieval-Augmented Generation (RAG) framework.

We evaluated a RAG-to-SQL approach using a SQLite database containing zebrafish larval behavior developmental neurotoxicity data following chemical exposure. The LLM used was GPT-4o via Azure OpenAI. Using 10 crafted prompts randomly selected from a prompt pool, repeated over three iterations, we assessed the accuracy of generated queries under four conditions: (1) providing an image of the database's entity-relationship (ER) diagram, (2) supplying schema information derived solely from the database's Data Definition Language (DDL), (3) DDL schema plus supplementary database documentation, and (4) DDL schema plus contextually relevant SQL examples.

Condition (1) substantially underperformed (mean accuracy = 0.46), while the other three conditions produced comparable results, each achieving an average accuracy above 0.90. Errors in the top-performing conditions typically reflected correct logical reasoning but misinterpretation of field names in the prompts—suggesting that additional prompt tuning could further enhance performance. Ongoing work includes testing with more complex database schemas, applying alternative LLMs (e.g., Gemini), and developing web-based applications for real-time data visualization.

Poster Abstracts

P50

Leveraging Large Language Models (LLMs) for data extraction and quality assessment in psychiatry systematic reviews: A comparison of inter-rater reliability between Elicit and human coders

Erkan, CN[1]; Gu, G[1]; Tandilashvili, E[1]; Meigs, JM[3]; Lee, K[1]; Metcalf, O[4]; Livinski, A[2]; Pine, DS[1]; Pereira, F[1]; Brotman, MA[1]; Henry, LA[1]

1. National Institute of Mental Health, National Institutes of Health, Bethesda, MD
2. National Institute of Health Library, Office of the Director, National Institutes of Health, Bethesda, MD
3. The Catholic University of America, Washington, DC
4. The University of Melbourne, Parkville, Melbourne, Australia

Background

Data extraction is a laborious and error-prone part of systematic reviews (Gartlehner et al., 2024). Large language models (LLMs) may improve efficiency in reviews (Amirian et al., 2024), but their accuracy varies. In our systematic review, we used an LLM, Elicit, as a secondary coder. We compared its accuracy in data extraction and quality assessment to humans, hypothesizing that inter-rater reliability would be higher for human-human than for human-Elicit coders.

Method

We reviewed 229 studies on ecological momentary assessment (EMA). Research assistants extracted 176 data points (e.g., demographics) and assessed 9 quality items (e.g., validity of measures). Articles were double-coded: 99 by two human coders, 130 by one human coder and Elicit. Inter-rater reliability was calculated for data extraction as [100 - ((# of discrepant data points/total # of extracted data points) * 100)] and for quality assessment as ((# of items agreed upon/total # of items)*100). Independent samples t-tests compared reliabilities between groups.

Results

For data extraction, human-human coders showed higher inter-rater reliability (M=87.35, SD=5.97, range = 72.73 – 97.16) than human-Elicit coders (M=82.29, SD=7.83, range = 55.68 - 94.89), t(226)=5.33, p<.001. Quality assessment reliability was similar between groups (human-human: M=72.17, SD=14.97, range = 33.33 – 100.00; human-Elicit: M=68.63, SD=16.22, range = 22.22 – 100.00, t(225)=1.68, p=0.094.

Conclusions

LLMs may reduce labor and errors in systematic reviews. Elicit's quality assessment performance was similar to human coders. While its data extraction performance is not yet at the level of human coders, it shows promise for improving efficiency in evidence synthesis.

Poster Abstracts

P51

Tuberculosis chest X-ray image retrieval system using deep learning based biomarker predictions

Lowekamp, BC[1]; Gabrielian, A[1]; Hurt, DE[1]; Rosenthal, A[1]; Yaniv, Z[1]

1. National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

It is estimated that in 2023, 10.8 million people fell ill with Tuberculosis (TB) and 1.25 million died from it. Additionally, there were about 400,000 new drug-resistant cases reported. These are especially challenging, as treatment is complex, and outcomes are often poor. The NIAID TB Portals program is an international consortium with a primary focus on patient centric data collection and analysis for drug resistant TB. The data includes images, their associated radiological findings, clinical records, and socioeconomic information. This work describes a TB Portals' Chest X-Ray (CXR) based image retrieval system which enables precision medicine. An input image is used to retrieve similar images and the associated patient specific information, facilitating inspection of outcomes and treatment regimens from comparable patients. Image similarity is defined using clinically relevant biomarkers: sex, age, body mass index (BMI), and the percentage of lung affected per sextant. The biomarkers are predicted using variations of the DenseNet169 convolutional neural network. A multi-task approach is used to predict sex, age and BMI incorporating transfer learning from an initial training on the NIH Clinical Center CXR dataset to the TB portals dataset. The resulting sex AUC, age and BMI mean absolute errors were 0.9854, 4.03years and 1.67kg/m2. For the percentage of sextant affected by lesions the mean absolute errors ranged between 7% to 12% with higher error values in the middle and upper sextants which exhibit more variability than the lower sextants. The retrieval system is currently deployed as part of the TB Portals Radiomics Analysis Portal.

Poster Abstracts

P52

Artificial Intelligence-based Segmentation of Neurites in High-Resolution Microscopy Images of iPSC-derived neurons

Arlova, A[1]; Weller, C[1]; Nalls, M[1]; Kelpsch, D[1]; Faghri, F[1]; Ryan, V[1]

1. Center for Alzheimer's and Related Dementias, National Institute on Aging, National Institutes of Health, Bethesda, MD

Dysfunction of mRNA transport and local translation in neurons, a potential factor involved in frontotemporal dementia/amyotrophic lateral sclerosis (FTD/ALS), is well-suited to be assayed by high-resolution microscopy in cultured cells derived from induced pluripotent stem cells (iPSCs). Analysis of sub-organellar particles can provide insight into effects of various mutations on these transcription and translation processes. Therefore, accurate and fast segmentation of neurites in confocal fluorescent microscopy images is a necessary step to determine boundaries of neurites and facilitate downstream analysis of sub-organellar structures. Manual segmentation is time-consuming, subject to variability among annotators, and does not allow for high-throughput image processing. Other methods, such as automated thresholding with imaging software, do not always produce satisfactory results as image quality can vary among images in a batch, and fluorescent signal produced by imaging markers is often confounded by noise. Artificial Intelligence-based solutions can help speed up neurite segmentation and alleviate biases introduced by other segmentation methods. In this work, we investigate application of convolutional neural networks (CNNs) to segment neurites in confocal fluorescent microscopy images. In our preliminary findings, a UNet-based model achieved a Dice-Sorensen similarity coefficient score (Dice) of 0.55, while NIS-Elements software thresholding workflow achieved a Dice of 0.43, and NIS-Elements proprietary AI model achieved a Dice of 0.46. This experiment is a proof of concept that a well-performing CNN can be trained on a relatively small manually annotated dataset. It is a viable neurite segmentation method that can be implemented as the first step of an image analysis pipeline.

Poster Abstracts

P53

Tuberculosis Portals AI in Image Processing and Abnormality Detection

Lowekamp, B[1]; Yaniv, Z[1]; Cobean, R[1]; Hoppes, M[1]; Rosenfeld, G[1]; Grinev, A[1]; Gabrielian, A[1]; Hurt, D[1]; Rosenthal, A[1]; Tartakovsky, M[1]

1. Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

The NIAID Tuberculosis (TB) Portals Program is a global initiative that enhances TB research through open-access, multi-domain data and tools. The TB Case Browser collects, stores, and provides deidentified TB clinical, pathogen genomics, and imaging data for research. Researchers from over 20 countries upload data including chest radiographs (CXR) and computed tomography (CT) studies.

Medical image analysis offers critical insights into disease diagnosis, progression, and treatment efficacy, which led the TB Portals program to include radiologist/clinical reads in its image collection. To maximize available information on disease pathology research, TB Portals partnered with Qure AI, CADDIETECH, and UIIP, all providing researchers with AI-predicted abnormality labels and segmentations. These efforts are essential in optimizing AI assistance for practicing radiologists in clinical settings.

The TB Portals program utilizes AI and advanced machine learning in its data quality pipelines to provide high quality public datasets. To prevent ingestion of unwanted protected health information (PHI), TB Portals uses AWS Rekognition with Optical Character Recognition (OCR) to screen images for text, significantly reducing manual effort. Given the diverse quality of images from other countries, the TB Portals program also implemented a “Chest-X-Ray Outlier” algorithm. This algorithm evaluates pixel data and classifies images at the distribution tails as outliers, helping to identify low quality images and in general any image that does not look like a frontal chest X-Ray.

Poster Abstracts

P54

Machine Learning Classification of Clinical Edema

May, CM[1]; Kasi, KK[1]; Kobayashi, LK[2]; Conway, BC[1]; Pare, JP[2]

1. National Eye Institute, National Institutes of Health, Bethesda, MD
2. Brown University Emergency Department

Hyperspectral imaging (HSI) is a powerful technique that captures and analyzes data from across the electromagnetic spectrum, providing significantly more detailed information than traditional RGB imaging, which is limited to just three-color channels. By collecting hundreds of narrow, contiguous spectral bands, HSI allows for the identification of unique spectral signatures based on how materials interact with different wavelengths of light. In medicine, HSI has been applied to distinguish between conditions such as kidney stones and cancers, though most studies focus on pixel-level classification (e.g., detecting cancer in specific pixels), rather than whole-image classification methods commonly used in convolutional machine learning (e.g., determining if an entire image shows signs of disease). In this study, we used a hyperspectral camera (SOC-710, Surface Optics Corporation) to capture images of patients presenting with edema to the Brown University emergency department. Using principal components analysis for dimensionality reduction and continuum removal for denoising, we achieved strong classification accuracy. We applied support vector machine using a linear kernel, achieving a 100% accuracy for cellulitis. We also performed 1D classification of wavelengths using four different algorithms, again achieving perfect classification of cellulitis. We also determined relative melanin and hemoglobin saturation maps within each patient, which has potential applications for classification of disease states using hyperspectral imaging.

Poster Abstracts

P55

Leveraging an MRI-Based Foundation Model to Enhance Predictions of Survival in Glioblastoma: A Multimodal Deep Learning Approach

Khanna, K[1]; Chen, Q[1]; Yan, C[1]; Meerzaman, D[1]

1.Computational Genomics and Bioinformatics, Center for Biomedical Informatics and Information Technology, National Cancer Institute, National Institutes of Health, Bethesda, MD

Glioblastoma multiforme (GBM) is the most common primary brain tumor in adults and has very poor survival outcomes. Accurate prediction of overall survival could significantly improve patient stratification in clinical trials and treatment planning. In this study, we demonstrate that a pretrained 3D vision transformer (SwinViT) can effectively predict overall survival in patients with GBM using preoperative MRIs. Using the UPENN-GBM dataset (n=520), we adapted and finetuned the BrainSegFounder, a 3D foundation model trained using MRI volumes from 41,400 healthy participants and 1,251 patients with brain tumors, in order to predict overall survival. Our finetuned model significantly outperformed models trained from scratch (C-Index: 0.672 ± 0.036 vs. 0.643 ± 0.431, Wilcoxon p=0.0488), with particularly strong performance in stratifying the highest-risk patients. We further explored integration of diffusion tensor imaging (DTI) and clinical variables with proven prognostic value, finding that multimodal approaches combining imaging derived risk scores, age, MGMT methylation status, and extent of surgical resection achieved the highest performance (C-Index: 0.714 ± 0.066). While imaging features performed well in identifying high-risk patients, the addition of clinical markers provided the largest improvements when distinguishing long-term (low risk) survivors. Our results demonstrate the immense potential of foundation models to improve medical imaging analysis, particularly when labeled data is scarce, and emphasizes the synergy between imaging and clinical data in GBM survival prediction.

Poster Abstracts

P56

Deep learning approach to video-based behavioral classification through human pose estimation.

MacLaren, CE[1]; Jackson, SN[1]; Fruchet, OE[1]; Volkman, RA[1]; Inati, SK[2]; Zaghloul, KA[1]

1. Functional Neurosurgery Section, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD
2. Neurophysiology of Epilepsy Unit, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

There exist current methods for identifying human action in videos, but little advancements in behavior in a hospital environment, where patients are monitored 24/7 via a live-stream camera. With new advances in machine learning and computer vision, different deep learning models can now identify objects and track their movement throughout a video. Through such, human movement – described as human pose – can be extracted in videos, creating opportunity for tracking and classifying different actions. We are interested in applying these methods for our own video data, where we record 24/7 clinical footage for epilepsy patients admitted at the NIH for seizure monitoring. Pre-trained human pose models achieve very high mean average precision (mAP) and are useful for transferring to different datasets. Utilizing a pre-trained network and fine-tuning for refined features, we can identify more positional information to the standard pre-trained network for our non-uniform environments, where patients may be in different settings with various obstructions, such as staff and family interruptions, blankets, tables, etc. We are able to achieve a mAP of 0.78147, identifying 18 different points of interest on the human body, and a precision of 0.99668 for identifying the proper boundaries of our patient. By correctly identifying human movement in videos, we can cluster different behaviors to classify a patient's unique behaviors. With this annotated data, we can extract neural correlates for precise behaviors throughout a patient's entire stay.

Poster Abstracts

P57

AI-based analysis of complex pigmentation phenotypes in zebrafish embryos

Shive, HR[1]

1. Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD

Pigmentation is a complex process that can be dysregulated by diverse factors, including both external (UV exposure, medications) and internal (genetic disorders, inflammation) variables. The zebrafish (Danio rerio) is well-established as a facile comparative model for analyzing pigmentation disorders, and melanophore development and melanin synthesis occur through conserved signaling pathways that are comparable to humans. Phenotypic characterizations of pigmentation in zebrafish are frequently conducted in embryonic or larval stages, which facilitates rapid, high-throughput screening of large animal numbers and whole-animal imaging. However, downstream image-based analyses currently used to assess pigmentation are cumbersome, labor-intensive, and difficult to use quantitatively. Here I describe a method to analyze pigmentation patterns in zebrafish embryos using HALO (Indica Labs). Although HALO modules are designed as an image analysis platform for digital histologic images, I have developed an AI-based approach to analyze digital stereomicroscopic images of zebrafish embryos. This approach applies a convolutional neural network (CNN) algorithm to classify pigmented regions in embryos, which subsequently enables accurate quantitative analyses of traits such as size of pigmented regions, pigmentation pattern, and distance between zones of pigmentation. I applied this CNN to quantify aberrant pigment patterns in zebrafish embryos with mutations in neurofibromin 1 (NF1). NF1 mutations in humans cause neurofibromatosis type 1, a genetic disorder for which aberrant pigmentation is a key diagnostic criterion. This work demonstrates the flexibility in creative use of CNN for analyzing digital images and provides a novel approach that will be highly useful for assessing pigmentation phenotypes in various zebrafish models.

Poster Abstracts

P58

Privacy-preserving and communication-efficient prediction of ISUP grades from prostate cancer histopathology images with foundation models

Lohmann, JJGL[1,2]; Witte, AW[3]; Maier, AM[1]; Saak, CCS[1]; Sauter, GS[4]; Zimmermann, MZ[3]; Bonn, SB[3,5]; Baumbach, JB[1,6]

1. Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany
2. Cyberinfrastructure and Artificial Intelligence Platforms Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
3. Institute of Medical Systems Bioinformatics, Center for Biomedical AI (bAIome), Center for Molecular Neurobiology Hamburg (ZMNH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
4. Institute of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
5. Spearpoint Analytics AB, Stockholm, Sweden
6. Computational Biomedicine Lab, Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

Background: To provide the best possible treatment for prostate cancer patients, several machine learning (ML) models have been developed to automate the classification of cancer severity using the International Society of Urological Pathology (ISUP) grades based on histopathology images of prostate biopsies. While ML models generally benefit from including more training data, aggregation of distributed data, e.g. in a cloud, can be difficult due to privacy regulations.

Methods: To overcome data privacy issues, we utilize federated learning (FL), a privacy-preserving technique to train ML models. Specifically, we have built a resource-efficient federated model for ISUP grade prediction, utilizing foundation models. The performance of the FL-based model was evaluated by comparing it to training on centralized cohorts and locally isolated cohorts. To test model robustness, the change in performance was assessed on heterogeneous data distributions and additionally for an increasing number of participating data providers.

Results: Compared to the centralized model the FL-based models perform equally well on the different data distributions. Also, the model performance is robust against increasing numbers of participating sites. By incorporating preprocessing by foundation models the model size used in training was reduced by up to 98%, possibly leading to decreased communication overhead in real-world applications.

Conclusions: The presented FL approach competes with a centralized model when performing ISUP grade prediction of prostate cancer without the need to aggregate the underlying data. Additionally, image embedding using foundation models during preprocessing reduces the model size during federated training.

Poster Abstracts

P59

The effects of syndromic facial feature editing on AI and clinician diagnosis of genetic conditions

Cheng, J[1]; Flaharty, KA[1]; Duong, D[1]; Waikel, RL[1]; Hu, P[1]; Ledgister Hanchard SE[1]; Solomon, BD[1]

1. Medical Genomics Unit, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Clinicians recognize genetic conditions based on the presence of distinct phenotypic features. In clinical practice, they may use artificial intelligence tools to help them with this task. These phenotypic features are catalogued in the Human Phenotype Ontology (HPO), but how each HPO feature contributes to the recognition of a genetic condition remains unclear.

Advances in deep learning offer new possibilities for investigating clinical genetic datasets to better understand how HPO features may contribute to condition recognition, for example, through super-resolution and inpainting techniques. By leveraging BrushNet, a deep learning-based inpainting diffusion model, we can create high-resolution images of patients affected by genetic disorders with and without selected HPO features.

After systematically editing HPO features from syndromic face images across a variety of genetic conditions, we evaluated the diagnostic accuracy of clinicians and of state-of-the-art deep learning classifiers such as Face2Gene and GestaltMatcher.

This study provides insight into the relative importance of specific syndromic facial features, highlights the potential and implications of AI-driven image editing in clinical settings, and provides new ways of creating high-quality counterfactual facial image datasets for clinical genetics studies.

Poster Abstracts

P60

Predicting tuberculosis from frontal chest X-rays: A Radiomics Analysis Portal research service

Kantipudi, K[1]; Gabrielian, A[1]; Hurt, DE[1]; Rosenthal, A[1]; Yaniv, Z[1]

1. National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

According to the 2024 World Health Organization report, in 2023 10.8 million people were diagnosed with Tuberculosis (TB). Chest X-rays have shown their utility both as a screening tool to identify individuals exhibiting TB related abnormalities and as a triaging decision tool, guiding referral for additional testing. The NIAID TB-Portals program is an international consortium focused on patient centric data collection and analysis. The program's Radiomics Analysis Portal enables interaction with the imaging data and provides a web-based TB/not-TB classification service. The service offers three AI classification options. All models were trained using publicly available datasets: Shenzhen and Montgomery from the NLM, the TBX11K challenge and the TB-Portals program and derived PG-GAN images. Two lung segmentation models were trained, one including the heart and one excluding it. Cropped images derived from lung segmentation predictions were used to train DenseNet121 classification models. Classification training was conducted using a five-fold cross validation framework with ImageNet weights initialization and the cross-entropy loss function. The models with the highest area under the receiver operating characteristic curve (AUC) were incorporated into the service. Additionally, all five DenseNet121 models were incorporated into the service as an ensemble. The generalization performance of all three approaches was evaluated using a private segregated TB screening dataset (279 TB/ 9287 not-TB) and a public dataset (125 TB/ 153 not-TB). The AUC for the models on the private dataset ranged between 0.76 and 0.79. The AUC for the models on the public dataset ranged between 0.74 and 0.84.

Poster Abstracts

P61

Deep learning assisted matrix factorization improves cell recognition in calcium imaging analysis

von Buchholtz, LJ[1]

1. National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD

Most current methods in the analysis of calcium imaging data employ a form of matrix factorization that models a calcium imaging movie as the sum of cellular components that are each defined by the multiplication of a spatial footprint with an activity trace over time. While some algorithms have put constraints on the temporal profile of cell activity, all current methods treat pixels as independent data points and don't incorporate any prior knowledge about cell shape. As a result, these methods have problems detecting cells with low temporal complexity and tend to artificially conflate cells that are spatially connected and highly temporally correlated with each other.

I propose a model in which a Variational Autoencoder (VAE) is trained on cell image data to generate realistic cell shapes from a low-dimensional latent vector. After training the VAE, the optimal corresponding latent vector can then be calculated to match any given novel cell image by gradient descent and backpropagation.

This VAE can also be used to generate spatial cell footprints that are then multiplied with randomly initialized temporal activity traces. Simultaneous optimization for the latent vectors underlying the spatial footprints and the temporal activity patterns by gradient descent backpropagation allows reliable analysis of real-life calcium imaging data. Preliminary comparison with existing methods shows that the VAE-assisted matrix factorization is computationally more expensive, therefore slower but detects and separates more true cells. It therefore also removes neighboring undetected cells as a major source of signal contamination.

Poster Abstracts

P62

Predicting Renal Tumor Pathology from Gross Appearance: An AI-based Pilot Study

Patel, MH[1]; Stecko, H[2]; Pramod, N[3]; Esengur, O[2]; Stevenson, E[2]; Saini, J[1]; Loebach, L[1]; Blachman-Braun, R[1]; Millan, B[1]; Nethala, D[1]; Gurram, S[1]; Linehan, WM[1]; Turkbey, B[2]; Ball, MW[1]

1. Urologic Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD
2. Molecular Imaging Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD
3. Genitourinary Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD

Introduction: Renal tumor pathology can guide surgical management and treatment decisions, however, conventional imaging is imperfect to predict histology subtype and renal biopsy is under-utilized. This project aims to develop a convolutional neural network (CNN) model to predict renal tumor histologic subtype from intra-operative gross imaging data, to create an objective tool to assist in real-time clinical decision-making.

Methods: Images selected for AI model development contained encapsulated renal tumors captured from live surgical recordings for patients undergoing partial nephrectomy between the years 2008-2024. Selected images were assigned labels based upon tumor histology and randomized patient-wise into training and testing groups. A CNN, based on DenseNet121 architecture from the MONAI framework, was trained using the image sets.

Results: A total of 287 intraoperative images were utilized from 95 unique patients. A total of 48 patients were randomized to the training dataset while 47 were randomized to the testing dataset. Patient histology included clear cell (n = 123), chromophobe (n = 33), papillary (n = 48), angiomyolipoma (n = 23), hybrid (n = 21), and oncocytoma (n = 39). AUC for each class was: 0.72 for clear cell carcinoma, 0.69 for papillary type 1, 0.59 for oncocytoma, 0.61 for hybrid, 0.66 for chromophobe, and 0.72 for angiomyolipoma.

Conclusion: An AI model based on DenseNet121 architecture can reasonably predict the histology of the most common kidney tumor types based on intraoperative images. Further development of the AI model with a larger training set may aid in expediting differential management based on tumor type.

Poster Abstracts

P63

Weakly supervised learning for subcutaneous edema segmentation of abdominal CT using pseudo-labels and multi-stage nnU-Nets

Bhadra, S[1]; Liu, J[1]; Summers, RM[1]

1. Clinical Center, National Institutes of Health, Bethesda, MD

Anasarca refers to the excessive accumulation of interstitial fluids within subcutaneous adipose tissue, causing generalized edema commonly due to heart, kidney, or liver dysfunction. Automated volumetric assessment of edema from abdominal CT scans offers valuable insight for monitoring disease progression, particularly in ICU settings; however, manual annotation for supervised segmentation is impractical due to edema's diffuse nature. While a recent unsupervised deep learning method leveraging intensity priors was introduced, it resulted in frequent false positives or under-segmentation errors. To address these limitations, we propose a weakly supervised segmentation framework utilizing multi-class pseudo-labels, which combine edema intensity prior-based pseudo-labels with adipose tissue and muscle pseudo-labels for additional anatomical context. Our two-stage approach employs nnU-Net as the segmentation backbone. In Stage 1, muscle and fat pseudo-labels were generated from 101 contrast-enhanced CT scans of patients without edema (52F, 49M, age: 66.6 ± 5.1 years) using existing body composition software annotations. Stage 2 training involved 99 CT scans without edema (45F, 54M, age: 48.1 ± 17.7 years), incorporating combined multi-class pseudo-labels from Stage 1 and the intensity prior method. Evaluation on 16 edema-positive patient scans (10F, 6M, age: 52.4 ± 8.7 years), using five manually annotated slices per scan supervised by an experienced radiologist, demonstrated significant improvements. Our method enhanced the average Dice Similarity Coefficient and relative volume difference by 4–5% ($p<0.05$) compared to intensity prior segmentation alone. Qualitatively, this weak supervision significantly reduced false positives and under-segmentation errors, confirming its efficacy for accurate edema quantification in monitoring anasarca.

Poster Abstracts

P64

Staging Liver Fibrosis with Hepatic Perivascular Adipose Tissue as a CT Biomarker

Chan, S[1]; Mathai, TS[1]; Balamuralikrishna, PTS[1]; Batheja, V[1]; Liu, J[1]; Lubner, MG[2]; Pickhardt, PJ[2]; Summers, RM[1]

1. Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD
2. Department of Radiology, University of Wisconsin School of Medicine & Public Health, Madison, WI

Cirrhosis is the 12th leading cause of death in the US. There are several CT imaging signs of late fibrosis, such as redistribution of liver segment volume, increased liver nodularity, and periportal space widening. Timely intervention can reverse the progression of early hepatic fibrosis, but later stages are irreversible. We hypothesize that the perivascular adipose tissue (PVAT) around the portal vein arising from periportal space widening may also be predictive of liver fibrosis. In this work, a fully automated pipeline was developed to segment the liver, spleen, portal vein and its branches. The PVAT in the vicinity of the portal vein was identified. From these structures, CT imaging biomarkers (volume, attenuation, fat fraction) were computed. They were used to build uni- and multivariate logistic regression models for diagnosing advanced fibrosis and cirrhosis. The best multivariate model for cirrhosis achieved 93.3% AUC, 78.9% sensitivity, and 93.4% specificity. For advanced fibrosis, the multivariate model obtained 88.7% AUC, 84.2% sensitivity, and 73.7% specificity. The automated approach may be useful for population-based studies of metabolic disease and opportunistic screening.

Poster Abstracts

P65

Semantic segmentation of TB in chest X-rays: A new dataset and generalization evaluation

Kantipudi, K[1]; Bui, V[2]; Yu, H[2]; Lure, YMF[3]; Jaeger, S[2]; Yaniv, Z[1]

1. National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD
2. National Library of Medicine, National Institutes of Health, Bethesda, MD
3. MS Technologies Corp, Rockville, MD

According to the 2023 World Health Organization report, an estimated 7.5 million people were diagnosed with tuberculosis (TB) in 2022. TB triaging is often performed using chest X-rays (CXRs), with significant efforts invested in automating this task using deep learning. A key concern with algorithms that output image-level labels, in our context TB/not-TB, is that they do not provide an explicit explanation with respect to how the output was obtained, limiting the ability of user oversight. Semantic segmentation of TB lesions can enable human supervision as part of the diagnosis process. This work presents a new dataset, TB-Portals SIFT, which enables semantic segmentation of TB lesions in CXRs (6,328 images with 10,435 pseudo-label lesion instances). Using this data, ten semantic segmentation models from the UNet and YOLOv8-seg architectures were evaluated in a five-fold cross validation study. The best performing segmentation models from each architecture, nnUNet(ResEnc XL) and YOLOv8m-seg and their ensemble were then evaluated for generalization on related classification and object detection tasks. Additionally, several binary DenseNet121 classifiers were trained, and their classification generalization performance was compared to that of the semantic segmentation-based classifier. Results show that the segmentation-based approach achieved better generalizability than the DenseNet121 classifiers and that the ensemble of the models from the two architectures was the most stable, closely matching or exceeding the performance of all other models across the tasks of segmentation, classification, and object detection. The dataset is publicly available from the NIAID TB Portals program after signing a data usage agreement.

Poster Abstracts

P66

Developing a deep learning algorithm to quantify pulmonary vascular remodeling in a pre-clinical model of pulmonary arterial hypertension and comparing performance to formal histopathological assessment

Joseph, TL[1]; Yu, ZX[1]; Siddique, MAH[1]; Chen, LY[1]; Elinoff, JM[1]

1. Critical Care Medicine and Pulmonary Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

Rationale: Pulmonary arterial hypertension (PAH) is a rare, female-predominant disease characterized by an inflammatory, proliferative arteriopathy that results in progressive narrowing of pre-capillary pulmonary arterioles and eventually death from right heart failure. Pre-clinical PAH animal models are critical to the development of novel therapeutics and the most common primary endpoint of these studies is improvement in histopathological lung vessel remodeling. However, current methods for evaluating pulmonary vascular histopathology are time-consuming and inconsistently applied across studies. To address these limitations, our group sought to develop a deep learning algorithm with the ability to perform rigorous and non-biased histopathological assessments in the rat SU-5416/Hypoxia (SuHx) model of PAH.

Methods: A multi-layered feature-detecting neural network (Visiopharm tissue image analysis software) was trained using manually annotated Masson Trichrome stained lung sections from SuHx and control rats to develop a deep learning model that recognizes pulmonary arterioles and determines the extent of vessel occlusion.

Results: When evaluated on lung sections which were not used for training, the model achieved an F1 score of 0.54 (F1 score >0.5 is average, >0.7 is good). Model performance was better on lungs from SuHx animals (F1=0.63) than from controls (F1=0.45). Notably, many of the “false positives” identified by the model were indeed arterioles that were inadvertently or intentionally excluded from the validation set due to pre-specified criteria.

Conclusion: Lung vessel segmentation using deep learning is feasible, however, further refinements are required to improve model validity. A more exhaustive validation set may be more appropriate for assessing model performance.

Poster Abstracts

P67

Scalable deep learning-based vessel segmentation and morphological quantification

Harouni, M[1]; Voss TC1[1]

1. Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD

Accurate segmentation of vascular structures is critical for understanding biological processes and disease progression. Furthermore, automated vessel segmentation enables large-scale, quantitative analysis of vascular features in biological imaging. However, challenges such as image heterogeneity, dense vessel networks, and large dataset sizes complicate traditional approaches. In this work, we present a scalable, deep learning-based vessel analysis pipeline designed for high-throughput biological imaging studies. Our approach begins with automated region-of-interest (ROI) selection, leveraging local contrast normalization and statistical density estimation, followed by morphological validation through cluster-based analysis of vessel width and perimeter distributions.

The developed deep learning segmentation model, trained on a limited set of curated annotations, achieved a Dice similarity coefficient of 0.964 across heterogeneous datasets comprising more than 54,000 images. Post-processing steps including skeletonization, hole-filling, and graph-based modeling enabled detailed extraction of vascular topology, such as junction density, branch type distribution, and vessel orientation. Spatial features were visualized through heatmaps and validated against original imaging data to ensure interpretability.

This pipeline not only make more efficient vessel segmentation at scale but also provides rich quantitative metrics supporting phenotype discovery, disease modeling, and therapeutic evaluation. Our integrated framework highlights the power of combining advanced image analysis with deep learning to deliver reproducible, biologically meaningful understandings in complex vascular growing structures.

Poster Abstracts

P68

Empanada - a napari plugin with pre-packaged segmentation models for nuclei, lipid droplets and mitochondria

Bhardwaj, A[1]; Narayan, K[1]

1. CCR Volume Electron Microscopy, Frederick National Laboratory, Frederick, MD

Understanding cellular structures hinges on the effective segmentation of organelles in volume electron microscopy (vEM) images—a task traditionally marked by labor-intensive and time-consuming manual efforts. The advent of deep learning has revolutionized this process, automating segmentation and markedly enhancing efficiency. Convolutional neural networks (CNNs) have demonstrated particular success in delineating organelles such as mitochondria and nuclei within vEM datasets. However, the widespread adoption of these methods is often hindered by the substantial computational resources and extensive annotated datasets they require.

Addressing these challenges, we have developed empanada as a user-friendly plugin for the napari image viewer, designed to facilitate deep learning-based organelle segmentation. empanada now integrates pre-trained models NucleoNet, DropNet and MitoNet, and also offers users the flexibility to fine-tune existing models or train new ones with their own data, thereby accommodating a diverse array of electron microscopy images. The models within empanada are trained on large, heterogeneous, and meticulously curated datasets, which are freely accessible to the scientific community.

Empanada streamlines the segmentation workflow by providing dedicated modules for training, inference, and post-processing, opening deep learning applications to non-experts. By democratizing access to advanced segmentation tools, empanada empowers researchers to efficiently analyze large EM and vEM datasets, accelerating discoveries in cell biology through precise and high-throughput analysis of organelle structures.

Poster Abstracts

P69

Optical Coherence Tomography: A Reliable Imaging Modality for Detecting Age-Related Macular Degeneration Features

Elsawy, A[1]; Keenan, T[2]; Chew, EY[2]; Lu, Z[1]

1. National Library of Medicine, National Institutes of Health, Bethesda, MD
2. National Eye Institute, National Institutes of Health, Bethesda, MD

Age-related macular degeneration (AMD) is an irreversible and progressive retinal disease that affects the macula. AMD is a leading cause of central vision loss for the elderly in developed countries. AMD is predicted to affect 288 million people worldwide. Thus, early detection of AMD is important for slowing its progression and preserving vision. AMD is categorized into early, intermediate, and late stages. The early features of AMD include drusen, i.e., retinal deposits below the retinal pigment epithelium (RPE) layer of the retina, the intermediate features include reticular pseudodrusen, i.e., subretinal deposits above the RPE layer, and the late features include geographic atrophy (GA), i.e., the defining lesion of atrophic AMD. Optical coherence tomography (OCT) is a non-contact non-invasive imaging technology that provides high-resolution cross-sectional images in vivo. Hence, it is suitable for imaging the eye. To study AMD on OCT, we used OCT datasets from the age-related eye diseases study 2 (AREDS2) and the dark adaptation in AMD study (DAAMD). We developed Deep-GA-Net and Deep-RPD-Net for detecting GA and RPD on OCT scans. We compared the developed models to retina specialists and visualized their predictions. Results showed an area under the receiver operating characteristic curve of 0.94 for detecting GA on AREDS2, 0.91 and 0.84 for detecting RPD on AREDS2 and DAAMD, respectively. The developed models outperformed the retina specialists and could highlight interpretable features. These results suggest that AMD features can be reliably identified on OCT. Thus, OCT can provide a valuable imaging modality in managing the AMD progression.

Poster Abstracts

P70

Deep Learning-based Contouring of Couinaud Segments on CT: Utility for Volumetric Analysis of Future Liver Remnant

Mathai, TS[1]; Balamuralikrishna, PTS[1]; Batheja, V[1]; Kassin, M[1]; Hannah, C[1]; Ukeh, I[1]; Hernandez, J[1]; Summers, RM [1]

1. Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD
2. Center for Cancer Research , National Cancer Institute , National Institutes of Health, Bethesda, MD

BACKGROUND:

Hepatocellular carcinoma (HCC) is a common primary liver cancer. High tumor burden, proximity to hepatic vessels, and other comorbidities mean only 30% of patients are candidates for curative surgical resection, which requires a 20% future liver remanent (FLR) to avoid post-operative complications. An automated liver Couinaud segmentation tool was developed for FLR volumetric analysis and targeted localization of tumors.

METHODS:

Three CT datasets were used: 1) public MSD Hepatic Vessels (161 patients), 2) NIH (43 patients with cirrhosis, ascites and splenomegaly), and 3) public TCIA Colorectal Liver Metastasis (CRLM, 197 patients). FLR was annotated by an expert radiologist and the Couinaud segments were manually annotated by two physicians. MSD and NIH datasets were used for model training, while CRLM was reserved for testing. A 3D nnU-Net model outlined the Couinaud segments. Performance was compared to a prior 3D U-Net model and evaluated with Dice Similarity Coefficient (DSC), Hausdorff Distance (HD) error (in mm), and volume error (in cc).

RESULTS:

3D nnU-Net obtained a DSC of 0.99 ± 0.01 (IQR: 0.991, 0.998), HD error of 0.87 ± 1.83 mm (IQR: 0, 1.02), and volume error of 13.7 ± 28.1 cc (IQR: 3.4, 15.3). Compared to U-Net, it was significantly different for DSC ($p < 0.001$, effect size 0.86), HD error ($p < 0.001$, effect size 0.87), and volume error ($p < 0.001$, effect size 0.91).

CONCLUSION:

The model generalized well to an external dataset and may be used for volumetric analysis on patients undergoing portal vein embolization.

Poster Abstracts

P71

Sensitivity based model agnostic scalable explanations of deep learning

Aggarwal, M[1]; Cogan, N[2]; Periwal, V[1]

1. Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD
2. Department of Mathematics, Florida State University, FL

Deep neural networks (DNNs) are powerful tools for data-driven predictive machine learning, but their complex architecture obscures mechanistic relations that they have learned from data. This information is critical to the scientific method of hypotheses development, experiment design, and model validation, especially when DNNs are used for biological and clinical predictions that affect human health. We design SensX, a model agnostic explainable AI (XAI) framework that outperformed current state-of-the-art XAI in accuracy (up to 52% higher) and computation time (up to 158 times faster), with higher consistency in all cases. It also determines an optimal subset of important input features, reducing dimensionality of further analyses. SensX scaled to explain vision transformer (ViT) models with more than 150,000 features, which is computationally infeasible for current state-of-the-art XAI. SensX validated that ViT models learned justifiable features as important for different facial attributes of different human faces. SensX revealed biases inherent to the ViT architecture, an observation possible only when importance of each feature is explained. We trained DNNs to annotate biological cell types using single-cell RNA-seq data and SensX determined the sets of genes that the DNNs learned to be important to different cell types.

Poster Abstracts

P72

Raman-based machine-learning platform reveals unique metabolic differences between IDHmut and IDHwt glioma

Lita, A[1]; Sjöberg, J[2]; Păcioianu, D[3]; Siminea, N[3]; Celiku, O[1]; Dowdy, T[1]; Păun, A[4]; Gilbert, MR[1]; Noushmehr, H[5]; Petre, I[2]; Larion, M[1]

1. National Cancer Institute, Neuro-Oncology Branch, National Institutes of Health, Bethesda, MD
2. Department of Mathematics and Statistics, University of Turku, Turku, Finland
3. Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest, Romania
4. Department of Bioinformatics, National Institute for Research and Development in Biological Sciences, Bucharest, Romania
5. Department of Neurosurgery, Henry Ford Health System, Detroit, MI

BACKGROUND: Formalin-fixed, paraffin-embedded (FFPE) tissue slides are routinely used in cancer diagnosis, clinical decision-making, and stored in biobanks, but their utilization in Raman spectroscopy-based studies has been limited due to the background coming from embedding media.

METHODS: Spontaneous Raman spectroscopy was used for molecular fingerprinting of FFPE tissue from 46 patient samples with known methylation subtypes. Spectra were used to construct tumor/non-tumor, IDH1WT/IDH1mut, and methylation-subtype classifiers. Support vector machine and random forest were used to identify the most discriminatory Raman frequencies. Stimulated Raman spectroscopy was used to validate the frequencies identified. Mass spectrometry of glioma cell lines and TCGA were used to validate the biological findings.

RESULTS: Here we develop APOLLO (rAman-based PathOLogy of maLignant gliOma)- a computational workflow that predicts different subtypes of glioma from spontaneous Raman spectra of FFPE tissue slides. Our novel APOLLO platform distinguishes tumors from nontumor tissue and identifies novel Raman peaks corresponding to DNA and proteins that are more intense in the tumor. APOLLO differentiates isocitrate dehydrogenase 1 mutant (IDH1mut) from wildtype (IDH1WT) tumors and identifies cholesterol ester levels to be highly abundant in IDH1mut glioma. Moreover, APOLLO achieves high discriminative power between finer, clinically relevant glioma methylation subtypes, distinguishing between the CpG island hypermethylated phenotype (G-CIMP)-high and G-CIMP-low molecular phenotypes within the IDH1mut types.

CONCLUSIONS: Our results demonstrate the potential of label-free Raman spectroscopy to classify glioma subtypes from FFPE slides and to extract meaningful biological information thus opening the door for future applications on these archived tissues in other cancers.

Poster Abstracts

P73

Understanding and simulating membrane pore formation by piscidin1 using AI informed enhanced sampling

Bodosa, J[1,2]; Pastor, R[1]

1. Membrane Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD
2. Biophysics, University of Maryland, College Park, MD

Antimicrobial peptides (AMPs) are short peptides found in organisms often as part of the innate immune host defense. They are capable of disrupting the bacterial membrane thus leading to cell death. These peptides aggregate on the membrane surface and form pores which allow leakage of the cell content. Pisidin 1 is a fish AMP which has antimicrobial, antifungal properties. Rice et al. have used molecular dynamics (MD) simulations to study defect formation by piscidin 1 in different membrane compositions. Some challenges to the study of membrane pore formation by peptides are - time scale of peptide aggregation, pore formation and pore persistence. MD is limited to studies which can be simulated within reasonable time in all-atom resolution. If the event occurs over a long timescale (> 10□s) then it is unlikely for one to observe statistically significant results using normal MD. We want to use an AI informed method developed by Tiwari et al. to simulate peptide-mediated pore formation and obtain the free energy associated with it.

Poster Abstracts

P74

AlphaFold2 screen reveals novel G1 cyclin docking modalities

Weaver, A[1]; Tuvikene, J[1]; Koivomagi, M[1]

1. Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD

The cell cycle is a series of tightly regulated events that drive cellular growth, DNA replication, and division. Regulation of this cycle is tightly controlled by cyclins and CDKs, which function together to phosphorylate a specific set of substrates. Understanding these interactions involved in cell cycle transitions is crucial for discovering the molecular mechanisms of cell cycle regulation. However, identifying specific docking interfaces has been challenging due to the dynamic nature of these interactions. Traditional methods, such as alanine scanning, are time-consuming and rely on low-throughput, one-by-one validation approaches, making it difficult to systematically identify these interactions.

We present a computational pipeline leveraging AlphaFold2 to predict docking interfaces between yeast cyclin, Cln3, and the full nuclear proteome of Saccharomyces cerevisiae. This allows for identification of candidate docking interactions at a proteome-wide scale. Predicted AlphaFold2 complexes were analyzed based on spatial and confidence features to define potential docking regions. To validate, we generated yeast strains expressing mutant versions of Cln3 with targeted disruptions in the predicted interfaces. Functional impact was assessed using a cell size assay as it's a proxy for Cln3-activity during G1 phase progression. Mutations in select docking regions resulted in altered cell size distributions, supporting their role in mediating Cln3's function in driving cell cycle progression.

Our results provide proof of principle for using AlphaFold2 to identify novel docking interfaces between cyclins and their target proteins across model systems. This allows for the development of targeted therapeutics that can disrupt specific docking interactions to halt cell division.

Poster Abstracts

P75

In silico evolution of globular protein folds from random sequences

Sahakyan, HK[1]; Babajanyan, SG[1]; Wolf, YI[1]; Koonin, EV[1]

1. Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD

The origin and evolution of protein folds are among the most challenging, long-standing problems in biology. Although many plausible scenarios of early protein evolution leading to fold nucleation have been proposed, realistic simulation of this process was not feasible because of the lack of efficient approaches for protein structure prediction, a situation that changed with the advent of powerful AI-based tools for fast and robust protein structure prediction. We developed a computational approach for protein fold evolution simulations (PFES) with atomistic details that provide insights into the mechanisms of evolution of globular folds. PFES introduces random mutations in a population of proteins, evaluates the effect of mutations, and selects a new set of proteins for further evolution. Repeating this process iteratively allows tracking the evolutionary trajectory of a changing protein fold that evolves under a selective pressure. We employed PFES to show how stable, globular protein folds could evolve from random amino acid sequences in various scenarios. The simulations reproduce the evolution of many simple folds of natural proteins as well as the evolution of distinct folds not known to exist in nature. These findings could shed light on the enigma of the rapid evolution of protein fold diversity at the earliest stages of life evolution. PFES tracks the complete evolutionary history from simulations that describes intermediate states and can be used to test versatile hypotheses on protein fold evolution.

Poster Abstracts

P76

Computational modeling of Cyclin D1 protein-protein interactions

Tuvikene, J[1]; Esvald, EE[1]; Heidebrink G[1]; Koivomagi, M[1]

1. Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD

Cyclin D1, forming complexes with cyclin-dependent kinases CDK4/6, is a key regulator of cell cycle progression, with its dysregulation implicated in numerous cancers. However, only a few substrates of the Cyclin D1-CDK4/6 complex are currently known, with the pocket proteins Rb, p107, and p130 being the best characterized. To further explore Cyclin D1 interactions, we employed AlphaFold-Multimer to predict interactions between Cyclin D1 and the entire human proteome. Next, we subjected the high confidence models to a computational pipeline developed in our lab to identify interacting regions and hotspots on Cyclin D1. To experimentally validate our predictions, we used a chemical-genetic approach combined with mass spectrometry, labeling Cyclin D1-CDK4/6 substrates using different Cyclin D1 docking region mutants. We estimate that approximately 30%-40% of the interactions predicted by AlphaFold-Multimer could be validated experimentally.

Our research uncovers a wide range of potential novel substrates and docking sites, providing a detailed map of Cyclin D1 interactions and enhancing our understanding of the molecular mechanisms controlling cell cycle progression. Our study demonstrates the power of integrating computational predictions with experimental validation to identify critical protein-protein interactions, offering new opportunities for cancer research and therapeutic development.

Poster Abstracts

P77

Integrating Network Analysis and Localization Prediction Using B-LEARN and ProtGPS

Kanno, T[1]; Kalchschmidt JS[2]; Brooks, SR[1]; Sun, H[1]

1. Biodata Mining and Discovery Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD
2. Genomics and Immunity section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

Investigating regulatory interactions and subcellular localization provides crucial insights into protein function. While it would be ideal to have bespoke experimental data for each protein of interest, data repositories from previous experiments, combined with state-of-the art AI prediction tools, present a fast and computationally efficient way to predict outcomes and refine hypotheses. Here, we present two recent tools, B-LEARN and ProtGPS that serve this purpose.

Our group has recently developed B-LEARN, an interactive online data portal designed for the intuitive search and visualization of 4,400 B cell regulators and 17,638 regulatory connections identified in B cells. The data consists of 47 genome-wide loss of function sgRNA screens, providing a high order view of transcriptional networks across a wide range of genes.

ProtGPS, a neural network classifier recently published by the Whitehead Institute and MIT, can accurately predict the subcellular compartmentalization of proteins from amino acid sequences. At BMDS, we have converted ProtGPS from the original proof of concept Jupyter Notebook into an easy-to-use tool on our internal Shiny Server, making data processing and sharing intuitive and accessible while maintaining data security.

We explore a potential workflow integrating both tools on the well-studied coregulatory network between RUNX1 and CBFB, alongside fusion candidates presented in fusionPDB. The predictions agree well with published imaging and provide a promising pathway to narrow down other fusion proteins to find candidates for further experimental study.

Poster Abstracts

P78

Efficient Computational Prioritization of Local Host Structures Mimicking Pathogen Antibody Epitopes.

Nguyen, TH[1,2]; Ghedin, E[1]; Sormanni, P[2]

1. Laboratory of Parasitic Diseases, National Institute of Allergies and Infectious Diseases, National Institutes of Health, Bethesda, MD
2. Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge, UK

Molecular mimicry, where pathogen proteins resemble host structures, is a key hypothesis for hijacking cellular pathways and triggering autoimmune responses. Proposed mechanisms involve promiscuous T cells and antibodies cross-reacting due to sequence or structural similarity. However, detecting subtle, local patch-level mimicry often missed by domain-level searches remains challenging. We present a systematic computational method wherein proteins from known antibody-antigen complexes are segmented and rapidly searched against host databases for local structural and chemical similarity. Putative mimics are further evaluated using antibody docking and prioritized via AlphaFold3 co-folding to assess complex formation feasibility. Promising candidates proceed to experimental validation from high-throughput enzyme-linked immunosorbent assays (ELISA) to biolayer interferometry (BLI). Preliminary computational hits to the SARS-CoV-2 proteome with solved antibody structures yielded fewer than 30 candidate mimic complexes. This stringent approach aims to reduce the high false-positive rates associated with simpler sequence or structure comparisons, potentially uncovering novel antibody-mediated mechanisms underlying conditions like Post-Acute Sequelae of COVID-19 (PASC).

NIH Campus Directions

NIH Artificial Intelligence Symposium – *May 16th, 2025 – Masur Auditorium, Building 10*

Directions

Public Transportation

Visitor parking is extremely difficult to find at NIH, so if at all possible, we recommend taking public transportation. There is a Metro station on the NIH campus with shuttle bus service to the main buildings. If you plan to take the metro, please take the Red Line and get off at "Medical Center Metro Station". The NIH Gateway Visitor center is located directly across from the Metro Station. Please proceed to the Gateway Center to get a visitor badge and either walk or take one of the free campus shuttles to Building 10 South Entrance (see **Campus Shuttles** below).

Driving Directions Website (http://www.nih.gov/about-nih/visitor-information/driving-directions)

Directions for Parking OFF Campus in Visitor Parking Garage (MLP-11):

All visitors must enter through the NIH Gateway Center (Bldg. 66). If you are planning to drive to campus and park outside of campus, please drive to the Main Visitor Entrance at NIH Gateway Center (Bldg. 66) and park at the Multi-level Parking Garage (MLP-11). This garage is outside of the perimeter security and, thus, vehicles will not need to go through vehicle inspection, reducing the amount of time it takes to get on campus. The cost to park in MLP-11 is $2 per hour for the first three hours, $12 maximum for the entire day. Once you have parked, please proceed to the Gateway Center (Bldg. 66) to get a visitor badge and either walk or take one of the free campus shuttles to Building 10 South Entrance (see **Campus Shuttles** below).

Directions for Parking ON Campus:

Visitor parking on campus is limited, so we recommend parking off campus in MLP-11. If you are planning to drive and parking on campus, please allow at least 30 minutes to pass through campus security (Gateway Inspection Station, Bldg. 66A). You will be required to submit to a vehicle inspection. Visitors over 15 years of age must provide a form of government-issued ID such as a driver's license or passport. Once you are through the vehicle inspection, you may park in the green lot next to Bldg. 53 along South Drive and walk to Bldg. 10 (see NIH Visitor Map).

Campus Shuttles

NIH provides a number of free shuttle services throughout the day on the NIH Campus for employees, patients, and visitors. The following shuttles run from the Visitor Center to Building 10 **South** Entrance:

- "Metro/Building 10 South Express Shuttle – **Light Green** Line" – get on at Visitor Center ("Metro") and get off at "Bldg. 10 (South)" stop (1 stop).
    - Departs NIH Visitor Center roughly every 20 minutes.
- "Campus Limited – **Purple** Line" – get on at Visitor Center ("Metro") and get off at "Bldg. 10 (South)" stop (4 stops).
    - Departs NIH Visitor Center roughly every 25 minutes.
- "Campus Route – **Red** Line" – get on at Visitor Center ("Metro") and get off at "Bldg. 10 (South)" stop (6 stops).
    - Departs NIH Visitor Center roughly every 15 minutes.

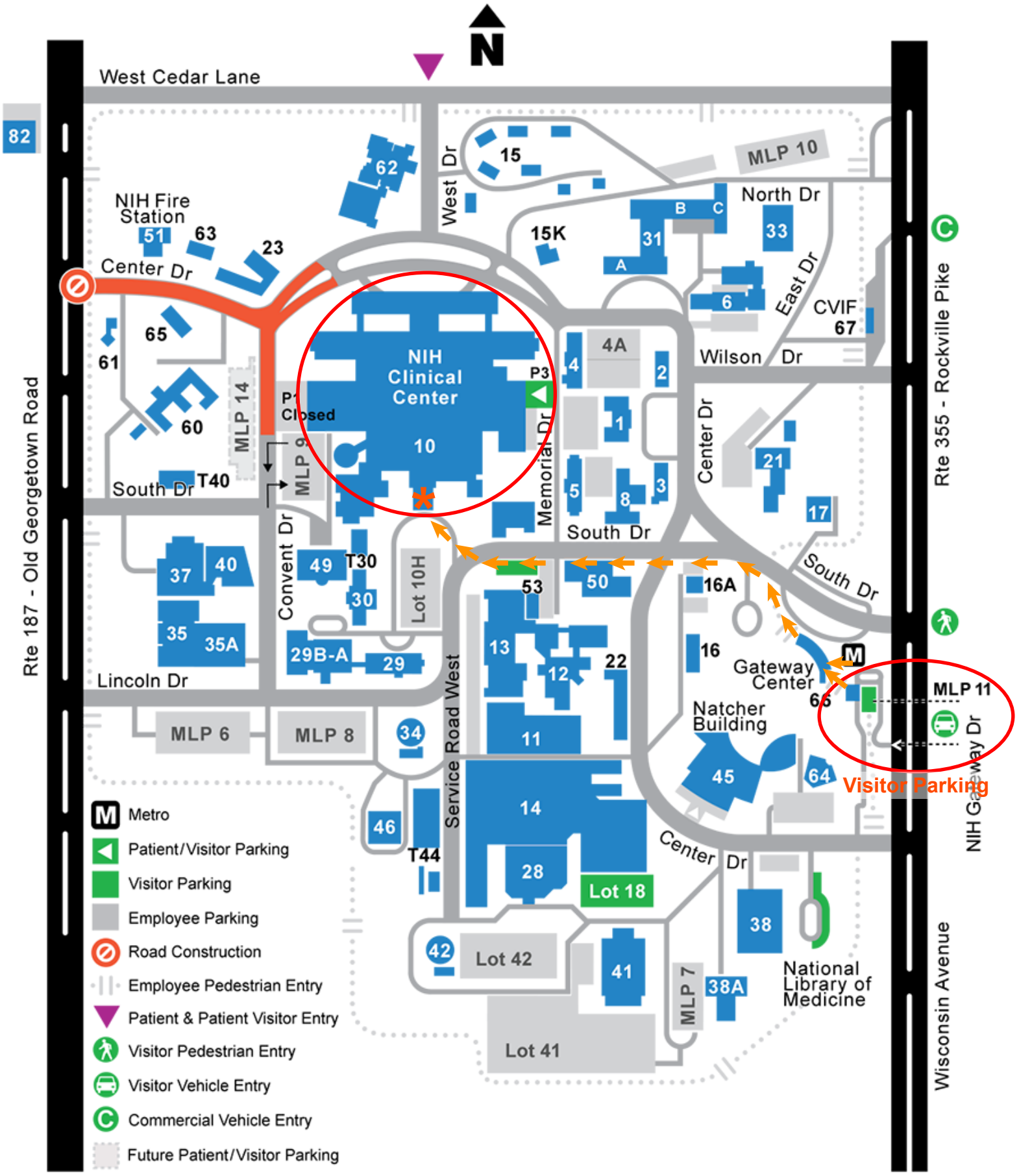
Directions

Directions to Masur Auditorium in Building 10:

- From the **North lobby** entrance: From the lobby, go down the right side, passing Admissions on your right. Continue straight through the sliding glass doors, following posted signs to the Masur. Continue following the "Detour" signs to the Masur. The auditorium is just past the main elevators.
- From the **South lobby** entrance: From the lobby, take either the left or right hallway up a slight incline until you come to the entrance of the Masur Auditorium. When the two hallways converge, you are standing in front of Masur Auditorium.

NIH Campus Map

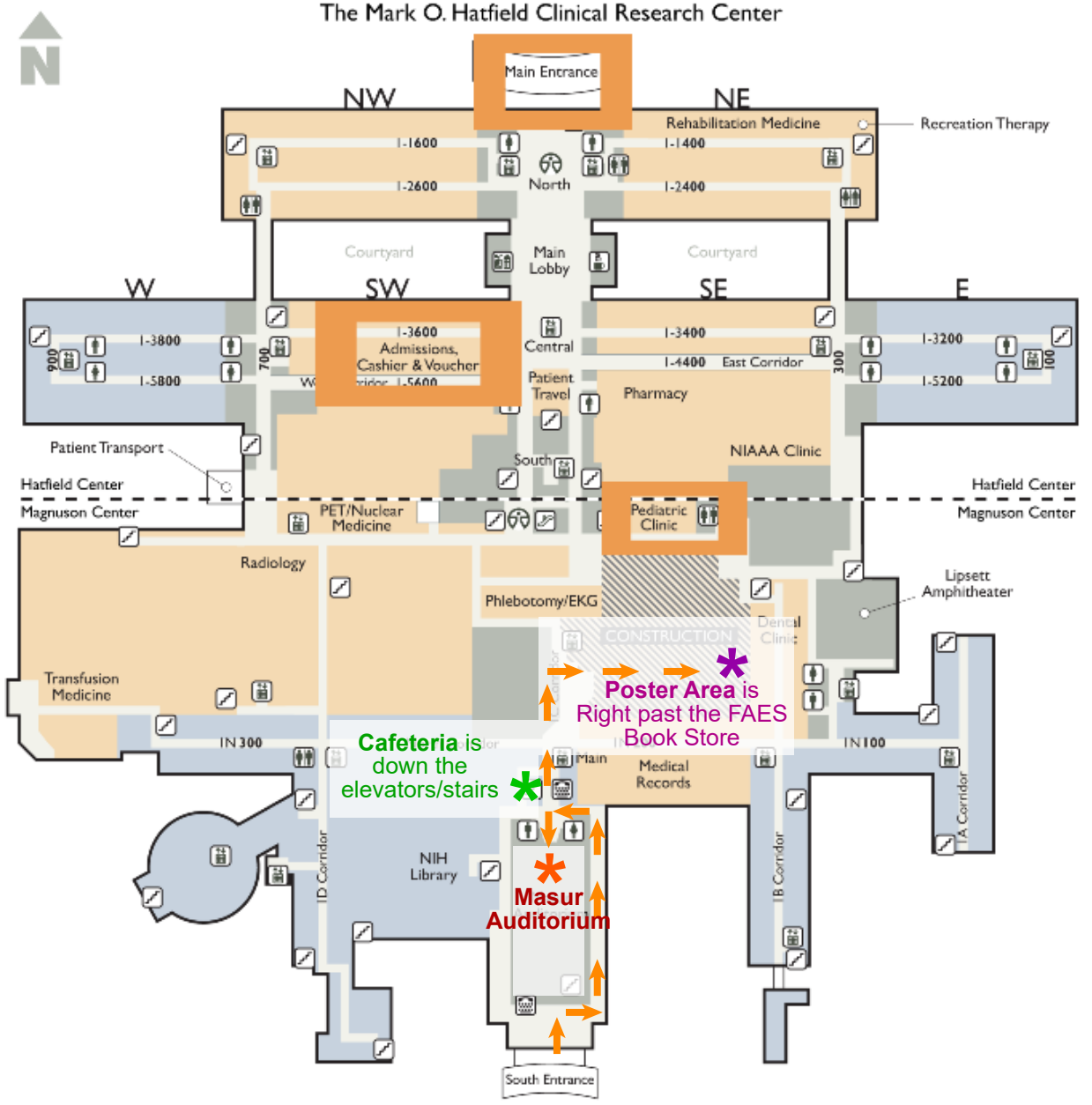
https://ors.od.nih.gov/maps/Pages/default.aspx



Building 10 Map

https://clinicalcenter.nih.gov/about/visitor1.html
https://ors.od.nih.gov/pes/emb/Documents/CC_MasurAudMap.pdf

Ist Floor



Meow