THE INSTITUTE FOR DATA-INTENSIVE ENGINEERING AND SCIENCE (IDIES)

2026

Annual Symposium

PROGRAM

THURSDAY, OCT. 22

02. Agenda
03. Keynote speakers
{ bios, titles and abstracts }
08. Update and plenary talk speakers
{ bios, titles, and abstracts }
12. Posters
35. Call for annual review content

{ IN DATA WE (AND AI) MUST TRUST }

idies The Institute for Data-Intensive Engineering and Science

WITH SUPPORT FROM:

JOHNS HOPKINS UNIVERSITY | Data Science and AI Institute

SYMPOSIUM AGENDA

8:30 to 9 a.m.	Breakfast	
9 to 9:15 a.m.	**Alex Szalay**	Opening remarks
9:15 to 9:35 a.m.	**Ani Thakar**	*Sloan Digital Sky Survey (SDSS) DR20 and SciServer*
9:35 to 9:55 a.m.	**Brian Mohr**	*IDIES Data Center Update*
9:55 to 10:30 a.m.	Keynote talk: **Clay Reid** (Johns Hopkins University)	*A Google Earth for the Human Brain: Petascale Imaging of Axonal Trajectories Throughout the Brain*
10:30 to 10:45 a.m.	Morning break	
10:45 to 11:05 a.m.	**Amy Binder** and **Mustafa Yavas**	*Visualizing Career Funneling in Top U.S. Universities*
11:05 to 11:25 a.m.	**Dennice Gayme**	*JHTDB-wind: Sharing High-Fidelity Wind Farm Simulation Data with the Research Community, and Insights Gained*
11:25 a.m. to 12 p.m.	Keynote talk: **Caroline Uhler** (Broad Institute of MIT)	*Causal Principles for Multi-Modal Data Integration and Translation*
12 to 12:30 p.m.	Poster lightning talks	
12:30 to 2 p.m.	**Lunch & poster gallery**	**Poster session 12:30 to 1:30 p.m.**
2 to 2:05 p.m.	Poster winner	
2:05 to 2:20 p.m.	**Elizabeth Smyth**	*Importance of Trusted Data and Data Science History at Hopkins*
2:20 to 2:40 p.m.	**Ting-Yi Lin**	Representing the Ippolito lab on the malaria computer vision project
2:40 to 3 p.m.	**Sue Werner**	*From Prototype to Pipeline—IDIES + Mapping the Modern Agora*
3 to 3:35 p.m.	Keynote talk: **Lyle Levine** (NIST)	*Supporting the Simulation Community with Rigorous Benchmark Measurements for Additive Manufacturing*
3:35 to 3:50 p.m.	Afternoon break	
3:50 to 4:10 p.m.	**Paulette Clancy**	*Advanced Research Computing at Hopkins (ARCH)*
4:10 to 4:45 p.m.	Keynote talk: **Roseanna Zia** (University of Missouri)	*Ribogenesis in Minimal Cells: Using AI to Accelerate Near-Atomistic Modeling to Discover the Matter-to Life Nexus and Enable Autonomous Synthetic Cells*
4:45 to 5 p.m.	**Alex Szalay** and **Charles Meneveau**	Closing remarks
5 to 6 p.m.	**Reception**	

KEYNOTE SPEAKERS (By aphabetical order))



Lyle Levine

AM Bench Program Founder and Co-Chair, National Institute of Standards and Technology (NIST)

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST)

AM BENCH

Talk title:

Supporting the Simulation Community with Rigorous Benchmark Measurements for Additive Manufacturing

Bio:

Dr. Lyle Levine is a physicist at the National Institute of Standards and Technology (NIST) in the USA. Dr. Levine founded and co-leads several prominent national and international efforts in the areas of additive manufacturing (AM) and qualification and certification (Q&C).

As one example, Dr. Levine founded and co-leads the Additive Manufacturing Benchmark Test Series (AM Bench), a NIST-led organization that has built collaborations with more than 30 US and European research organizations to provide rigorous AM benchmark measurement data and challenge problems to the AM modeling community.

AM Bench is the world's largest provider of AM validation data, with approximately 1670 unique user data downloads each month.

For further information on AM Bench, please visit their website at: **www.nist.gov/ambench**.

KEYNOTE SPEAKERS (By order of appearance)



R. Clay Reid, MD, PhD

Professor in the Departments of Neuroscience and Biomedical Engineering, Johns Hopkins University

Talk title:

A Google Earth for the Human Brain: Petascale Imaging of Axonal Trajectories Throughout the Brain

Bio:

Clay Reid recently joined the Hopkins faculty, hired as part of the Petascale Neuroscience and AI cluster in DSAI. He previously was an Investigator at the Allen Institute for Brain Science and Professor of Neuroscience at Harvard Medical School. Throughout his career, he has used high-throughput imaging approaches to investigate how the structure of neural circuits relates to their function. Starting 20 years ago, his group was one of the first to use volumetric electron microscopy (EM connectomics) to examine the interconnections of neurons in increasingly large networks, culminating in the MICrONS project (microns-explorer.org), which created a petascale dataset of visual cortex encompassing ~100,000 neurons and half a billion synapses. His research program at Hopkins will be to create multi-petascale atlases of the human brain, with the long-term goal of tracing the network of individual axons—comprising tens of thousands of kilometers—throughout the brain.

KEYNOTE SPEAKERS (By order of appearance)



Caroline Uhler

Director, Eric & Wendy Schmidt Center, Broad Institute of MIT and Harvard

Andrew (1956) and Erna Viterbi Professor of Engineering, MIT

Bio:

Caroline Uhler is a core institute member of the **Broad Institute of MIT and Harvard**, where she directs the **Eric and Wendy Schmidt Center** and is a member of the Scientific Leadership Team. She is also the Andrew (1956) and Erna Viterbi Professor of Engineering in the **Department of Electrical Engineering and Computer Science** and the **Institute for Data, Systems, and Society** at MIT.

Caroline's research lies at the intersection of machine learning, statistics, and genomics, with a particular focus on causal inference, representation learning, and gene regulation.

Caroline is recognized as a creative and innovative researcher and teacher at the intersection of machine learning, statistics, and biology. She is a SIAM Fellow, a Fellow of the IMS, a Sloan Research Fellow, and an elected member of the International Statistical Institute. In addition, she has received multiple awards including an NIH New Innovator Award, a Simons Investigator Award, and an NSF Career Award.

Caroline holds an MSc in mathematics, a BSc in biology, and an MEd all from the University of Zurich. She obtained her PhD in statistics from UC Berkeley in 2011 and then spent three years as an assistant professor at IST Austria before joining the faculty at MIT in 2015.

Talk title:

Causal Principles for Multi-Modal Data Integration and Translation

Abstract:

An exciting opportunity at the intersection of the biomedical sciences and machine learning stems from the growing availability of large-scale multi-modal data (imaging-based and sequencing-based, observational and perturbational, at the single-cell level, tissue-level, and organism-level). Traditional representation learning methods, although often highly successful in predictive tasks, do not generally elucidate underlying causal mechanisms. I will present a statistical and computational framework for causal representation learning and its applications towards identifying novel disease biomarkers as well as inferring gene regulation in different disease contexts.

KEYNOTE SPEAKERS (By order of appearance)



Roseanna N. Zia, PhD

Wollersheim Professor of Chemical Engineering, University of Missouri, Columbia

R. N. ZIA, V. SIVASANKAR, H. SACHAR, J. G. WANG, AND S. NARAYANAN

ZIA LAB

Bio:

Roseanna N. Zia is the Wollersheim Professor of Mechanical and Aerospace Engineering at the University of Missouri, Columbia. Her research combines computational biology with physics-based models that simulate entire bacterial cells with unprecedented molecular detail. To explicitly represent every protein, ribosome, and DNA molecule as a physical object in a whole-cell model, her lab is integrating AI-guided tools with atomistic-resolution physics. This new class of whole-cell models promises to expand synthetic biologists' ability to engineer bacteria for biotechnology, synthetic cells, and bespoke biomaterials. Current whole-cell models of E. coli and JCVI-Syn3A, the minimal cell capable of life, have revealed how colloidal-scale physics regulate protein synthesis rates and spatial organization. Her lab's new models predict design conditions for bottom-up, ultra-simple synthetic cells, which are subsequently made by a team of experimentalist collaborators.

Prof. Zia earned her Ph.D. at Caltech with John F. Brady studying colloidal hydrodynamics and completed postdoctoral work on colloidal gels at Princeton with William B. Russel before launching her faculty career at Cornell (2013–2017) and Stanford, where she earned tenure (2017–2022).

A Fellow of the American Physical Society, she has received two Presidential Early Career Awards (PECASE), ONR Director of Research Early Career Award, ONR Young Investigator Award, and NSF CAREER Award, among other honors. She has delivered over 100 invited lectures worldwide and served as Associate Editor of the Journal of Rheology (2017–2023). Her multidisciplinary work spans chemical engineering, physics, mechanical engineering, and biology.

Talk title:

Ribogenesis in Minimal Cells: Using AI to Accelerate Near-Atomistic Modeling to Discover the Matter-to Life Nexus and Enable Autonomous Synthetic Cells

Abstract:

A major goal of synthetic biology --- and understanding the matter-to-life transition --- is constructing cells that autonomously grow, divide, and produce bespoke...

[CONTINUED ON NEXT PAGE]

KEYNOTE SPEAKERS (By order of appearance)

Zia continued

[CONTINUED FROM PREVIOUS PAGE]

...proteins and materials. Ribosome biogenesis is central to this goal but remains notoriously difficult to recapitulate in cell-free and synthetic systems. One major contributor to this barrier is that the purposefully simplified spatial organization in ex vivo systems lacks the spatio-temporally coordinated features thought to foster ribosome assembly in natural cells. But identifying “what to leave in, what to leave out” to maintain simplicity and cell function requires Edisonian searching with current methods. Computational modeling of ribogenesis in synthetic cells promises a potential solution. But recapitulating near-atomistic interactions in whole-cell models using current approaches would require astronomical computational resources and time, presenting a second steep barrier to breakthroughs. My lab has developed a new approach to flatten both barriers: AI-guided, physics-based whole-cell models of bacteria that explicitly represent every biomolecule as a physical object. Using deep-learning neural networks trained on atomistic simulations, we bypass decades-old computational bottlenecks, enabling dynamic simulation of millions of biomolecules over biologically relevant timescales. We further employ AI-driven search engines to vastly reduce the combinatoric search space for optimal whole-cell composition. I'll discuss our progress and how we are using these methods to enable model-design-build-measure workflows for autonomous synthetic cells, and build a gateway for rational design of engineered microbes as living factories for next-generation protein-based materials.

PLENARY AND UPDATES SPEAKERS

(By order of appearance)



Alex Szalay and Charles Meneveau

Opening (Alex Szalay only) and closing remarks

Alex Szalay is a Bloomberg Distinguished Professor and Professor in the Department of Computer Science. He is the Director of the Institute for Data-Intensive Science (IDIES). He is a cosmologist, working on the statistical measures of the spatial distribution of galaxies and galaxy formation. He is a Corresponding Member of the Hungarian Academy of Sciences, and a Fellow of the American Academy of Arts and Sciences. In 2004 he received an Alexander Von Humboldt Award in Physical Sciences, in 2007 the Microsoft Jim Gray Award. In 2008 he became Doctor Honoris Causa of the Eotvos University, Budapest.



Charles Meneveau, the Louis M. Sardella Professor in Mechanical Engineering and associate director of the IDIES, is an expert in the multiscale aspects of turbulence, large-eddy simulations, and wind farm fluid dynamics. Meneveau has participated in efforts to democratize access to valuable “big data” in turbulence. As deputy director of JHU’s Institute for Data-Intensive Engineering and Science, he led the team of computer scientists, applied mathematicians, astrophysicists, and fluid dynamicists that built the **Johns Hopkins Turbulence Databases (JHTDB)**. This open numerical laboratory provides researchers from around the world with user-friendly access to large data sets arising from direct numerical simulations of various types of turbulent flows



Ani Thakar

SDSS DR20 and SciServer

Ani Thakar is a Principal Research Scientist in the Physics & Astronomy Department, and Associate Director of Operations for IDIES. He is also Catalog Archive Scientist for the Sloan Digital Sky Survey and Johns Hopkins University's Lead Scientist for the fifth phase of the Sloan Digital Sky Survey (SDSS-V).

PLENARY AND UPDATES SPEAKERS CONTINUED

(By order of appearance)



Brian Mohr

IDIES Data Center Update

Brian Mohr is the Senior Systems Engineer for the Institute for Data-Intensive Science (IDIES) within the Whiting School of Engineering.



Amy Binder and Mustafa Yavaş

Visualizing Career Funneling in Top U.S. Universities

Amy Binder is SNF Agora Professor of Sociology at Johns Hopkins University, where she studies education from a political, organizational, and cultural perspective. She is an award-winning author of three books, including, most recently, The Channels of Student Activism: How the Left and Right Are Winning (and Losing) in Campus Politics Today (University of Chicago Press, 2022) and numerous articles, including "Career Funneling: How Elite Students Learn to Define and Desire 'Prestigious' Jobs." Her research has been widely reviewed in the media and scholarly journals and is currently focused on how elite universities fail to prepare undergraduates for a wide range of creative careers for the public good.



Mustafa Yavaş is a postdoctoral fellow in the Center for Economy and Society at Johns Hopkins University. His scholarship focuses on economic sociology from a cultural, historical, and global perspective. He is the author of White-Collar Blues: The Making of the Transnational Turkish Middle Class (Columbia University Press, 2025) and several award-winning articles, including "White-Collar Opt-Out: How 'Good Jobs' Fail Elite Workers." His current agenda is centered on combining computational and historical inquiries to advance a critical economic sociology of elite labor by examining college-to-work transitions in elite higher education, the evolution of corporate work culture over the past century, and high-skill migration in a global context.

PLENARY AND UPDATES SPEAKERS CONTINUED

(By order of appearance)



Dennice Gayme

JHTDB-wind: Sharing High-Fidelity Wind Farm Simulation Data with the Research Community, and Insights Gained

Dennice F. Gayme is a professor of mechanical engineering with secondary appointments in the departments of Electrical and Computer Engineering and Applied Mathematics and Statistics. She is also a member of the Ralph O'Connor Sustainable Energy Institute (ROSEI) leadership council. Her research focuses on modeling, analysis, and control of spatially distributed and large-scale networked systems in applications such as wall-bounded turbulent and transitional flows, wind farms, and power grids. Her lab utilizes computational and theoretical methods from applied mathematics, dynamics, controls, optimization, and fluid mechanics.



Elizabeth Smyth

Importance of Trusted Data and Data Science History at Hopkins

Elizabeth Smyth is the Executive Director of Johns Hopkins Data Science and AI Instititue (DSAI).



Ting-Yi Lin

Representing the Ippolito Lab on the malaria computer vision project

Ting-Yi Lin is a master's student in Robotics at Johns Hopkins University and a student researcher in the Ippolito Lab. His work focuses on computer vision and domain adaptation for detecting Plasmodium parasites in blood smear images, aiming to improve model reliability across different staining and imaging conditions.

PLENARY AND UPDATES SPEAKERS CONTINUED

(By order of appearance)



Sue Werner

From Prototype to Pipeline - IDIES + Mapping the Modern Agora

Suzanne Werner is a senior developer and database administrator at IDIES, where she has spent nearly two decades building and maintaining the infrastructure behind some of Johns Hopkins' largest data resources, including the SDSS Catalog Archive Server, the Pan-STARRS PS1 archive, the Johns Hopkins Turbulence Database, and Mapping the Modern Agora. Her work sits where database engineering meets domain science: SQL Server at scale and the loading pipelines that turn raw data into something researchers can query. She holds degrees in history and computer science. Outside of IDIES, she plays bass in the Baltimore hardcore band War on Women.



Paulette Clancy

Advanced Research Computing at Hopkins (ARCH)

Paulette Clancy is the Edward J. Schaefer Professor of chemical and biomolecular engineering at Johns Hopkins University. She is the inaugural research director for Hopkins' Data Science and AI Institute She is a fellow of the Royal Society of Chemistry and of AIChE. Since 2019, she has also served as the Chair of Hopkins Research Computing Committee, the faculty oversight committee for high-performance computing at JHU. Prior to this, she spent over 30 years at Cornell, serving as the Samuel and Diane Bodman Professor of ChE before moving to Johns Hopkins in 2018 to become the inaugural department Head of ChE from 2018-2023. Clancy leads one of the top groups in the country studying atomic- and molecular-scale modeling and simulation of semiconductor materials. Over the past decade her work has turned strongly into algorithm development in machine learning, especially Bayesian optimization.

POSTERS

Rodrigo Araya Gracia	Toward Natural-Language Access to Turbulence Data: An LLM-Powered Agent for the JHTDB	13
Hajira Arooj	Association of Hormone Therapy Formulation With Retinal Vein Occlusion in US Women Aged 15-49 Years: A Case-Control Study	14
Aparna Bhattacharyya	SciDataReportR: Reusable Workflows for Interpretable Biomedical Data Analysis	15
Russell Calderon	Age-Relationship in Surgical Approach and Outcomes Following Retinal Detachment Repair	16
Minseo Choi	Adaptive KV Retention for LLM Agents at Human-Approval Timescales	17
Tamanna Ghosh	Longitudinal Laboratory Trajectories and Risk of Incident and Vision-Threatening Diabetic Retinopathy in Adults With Type 2 Diabetes	18
Leo Han	UnmixingMM: Unmixing in Whole-Slide Multiplex Immunofluorescence with Majorization–Minimization	19
Mashfique Hassan	Accelerating Markov Decision Process Policy Optimization on the GPU using a Simplification of the Bellman Equation	20
Jiayue Hu	CHAOS: Data for High-Entropy Oxide Discovery	21
Dibyajyoti Nayak	Learnable Time Integration for Stable Long-Term Extrapolation	22
Shivang Parmar	AI-Enabled Smartphone Ophthalmoscope for Automated Retinal Vessel Segmentation, Biomarker Extraction, and Disease Detection	23
Julian Posey	Emulating the Roman grism PSF across the focal plane for slitless flux-cube restoration	24
Dibakar Roy Sarkar	CINOC: Cardinality-Invariant Neural Operator Policies for Scalable PDE Control	25
Atharva Sathe	JHTDB-Wind: The Johns Hopkins Wind Farm Flow Database Providing Open Access to High-Fidelity Simulation Data	26
Jake Singh	A Differentiable Bayesian Framework for Celestial Navigation with Star and Satellite Imagery	27
Wei Wang	Solver in the Loop: Neural-Physics Hybrid Solver	28
Zeyu Wang	Fast urban CO2 emissions estimation using satellite data	29
Yongmeng Wu	Plug-in Neural Subdomains: FE–NO Coupling in 3D	30
Meltem Yashar	Surgical Success Calculator for Retinal Detachment	31
Geng Zhang	Estimating Hessian Log-Determinants without Forming the Hessian Matrix	32
Caroline Zhao	Investigating the Relationship Between Weather and Subarachnoid Hemorrhage Admissions	33
Lulu Zhao	Citation Grounding and Self Consistency for Large Language Model Extraction of Structured Data From Clinical Notes	34

Rodrigo Araya Gracia

Toward Natural-Language Access to Turbulence Data: An LLM-Powered Agent for the JHTDB

Large-scale turbulence simulations offer rich opportunities for studying complex flow phenomena. However, accessing and analyzing these datasets often requires substantial programming expertise. This project explores an LLM-powered scientific agent designed to provide natural-language access to the Johns Hopkins Turbulence Database (JHTDB). The agent combines a user's scientific request with dataset metadata, analysis instructions, and existing demonstration code to generate and adapt reproducible workflows for data querying, analysis, and visualization. In addition, a conversational interface within SciServer is envisioned, enabling users worldwide to request tasks ranging from extracting flow quantities and computing turbulence statistics to creating animations and performing more complex analyses. Preliminary tests demonstrate the ability of large language models to adapt JHTDB code for analyses across different datasets, including high-Reynolds-number channel-flow and magnetohydrodynamic cases. The proposed workflow aims to reduce implementation barriers while preserving access to the generated code, settings, and results for transparency and reproducibility. Future work will focus on applying this framework to support the systematic exploration of turbulence datasets and the discovery of previously unrecognized flow structures and physical phenomena.

Toward Natural-Language Access to Turbulence Data: An LLM-Powered Agent for the JHTDB

Rodrigo Araya Gracia[1] and Charles Meneveau[1]
[1]Department of Mechanical Engineering & IDIES, Johns Hopkins University

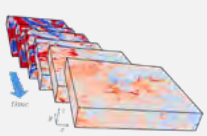
Introduction

Large-scale numerical simulations provide rich datasets for investigating turbulent flows. However, transforming these data into scientific insight requires programming expertise and substantial time to prepare queries, implement analyses, and generate visualizations. These demands create barriers for less experienced users and slow exploratory research even for specialists.

Large language models (LLMs) offer an opportunity to simplify this process through natural-language interaction. This project explores an LLM-powered agent for the Johns Hopkins Turbulence Database (JHTDB), aiming to reduce the human effort required to turn scientific questions into reproducible analyses.

JHTDB and Current Workflow

The Johns Hopkins Turbulence Database (JHTDB) is an open numerical turbulence laboratory that makes large-scale simulation data accessible to researchers worldwide. Its collection include direct numerical simulations (DNS) and large-eddy simulations (LES) of a wide range of turbulent flows. Among these is a snapshot from a record-breaking ($32{,}768^3$-grid-point DNS of isotropic turbulence at $Re_\lambda \sim 2{,}550$), occupying approx. 0.5 PB. The JHTDB infrastructure enables numerical experiments to be conducted remotely. Users typically write or adapt code to perform data analysis, while repeating or extending these analysises requires further implementation and verification.



Define scientific question
↓
Develop or adapt code
↓
Query JHTDB
↓
Analyze and visualize
↓
Interpret and refine

Proposed Agent Workflow

The proposed agent would combine a user's natural-language request with backend resources to adapt a DEMO code for turbulence analysis.

Frontend input – User prompt: Defines a desired task, such as computing turbulence statistics, extracting flow quantities, or generating visualizations.

Backend inputs

- **Agent instructions:** Specify how to use each resource, the requirements for code modification, and the expected output format.

Resources



Guided by these instructions and resources, the LLM modifies the DEMO to implement the user's task, adapting data queries, calculations, and visualizations as needed. The adapted code is intended for execution by the agent to produce the requested analysis. The backend instructions define the required deliverables and how they should be returned.

Envisioned User-Agent Interaction

A conversational interface within SciServer would allow users to request JHTDB analyses directly, with supporting code, dataset configuration, and physical metadata supplied automatically by the backend.

User: Using the MHD dataset, animate the x-component of the velocity on the z = pi plane from t = 0 to t = 1 in dataset time units.
Agent: I will use the full plane with the default spatial and temporal sampling, and keep the color scale fixed across frames. You can adjust these settings.

[The agent adapts the DEMO code, retrieves the data, and executes the visualization workflow.]

Agent: The animation is ready. You can view it below or access the generated Python script and analysis settings.

[View animation] [View code] [Analysis settings]

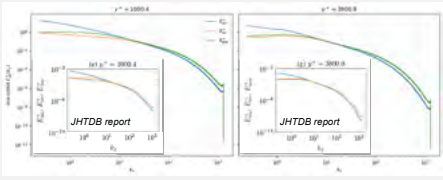
User: Now show the vorticity magnitude for the same plane and time interval.

Preliminary Evaluation of LLM Capabilities

Power Spectra for Channel5200

User prompt: "compute the power spectra E^+_{uu}, E^+_{vv}, and E^+_{ww}, at y+=1000.4 and y+=3800.9, as functions of k_x, for the channel5200 dataset."

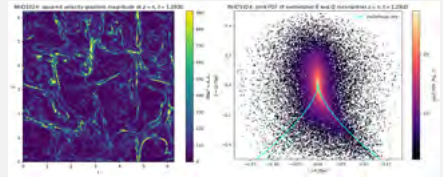
Results obtained from running GPT 5.6 Terra-generated code.



R-Q Joint PDF for MHD

User prompt: "(a) plot the contours of square velocity gradients on the plane z = pi, and (b) calculate the joint PDF of r and q (third and second order invariants of the velocity gradient tensor, respectively, normalized using the velocity gradient magnitude) for the magneto-hydrodynamic dataset. Include the Vieillefosse line."

Results obtained from running GPT 5.6 Terra-generated code.



Acknowledgements
The authors thank to the JHTDB team, specially to Michael Schnaubelt and Dmitry Medvedev for their support and technical coordination. The JHTDB is supported by NSF-grant 2103874.

Hajira Arooj

Association of Hormone Therapy Formulation With Retinal Vein Occlusion in US Women Aged 15-49 Years: A Case-Control Study

Association of Hormone Therapy Formulation With Retinal Vein Occlusion in US Women Aged 15-49 Years: A Case-Control Study

Hajira Arooj[1,2], Meltem Yashar, MD[2]; Luqman Munir[2]; Yannis M. Paulus, MD[2]

1. Department of Ophthalmology, Rawalpindi Medical University, Rawalpindi, Punjab, Pakistan
2. Wilmer Eye Institute, Department of Ophthalmology, Johns Hopkins University, Baltimore, Maryland, United States

Email contact: ypaulus1@jh.edu

Background

Retinal vein occlusion (RVO) is a **sight-threatening** retinal vascular disorder. Little is known about how RVO risk varies by hormone formulation in reproductive-age women

Purpose

To evaluate the association between specific hormone **formulations** (estrogen-only, progestin-only, and combined hormone therapy) and **risk of RVO**.

Methods

Design: Retrospective Case-Control study
Data Source:
TriNetX USA Federated Research Network
Accessed: October 18, 2026
Age and Gender: 15-49 years, female
Population and exposure groups
- Cases: with RVO (n = 3,240)
- Controls: without RVO (n = 11,417,290)

Exclusion: Pregnancy, malignancy, or thromboembolism (within 18 months of the index event)
Exposure groups: RxNorm and HCPCS codes (prescribed records) within **18 months** of the index event
- Estrogen-only
- Progestin-only
- Combined estrogen-progestin therapy

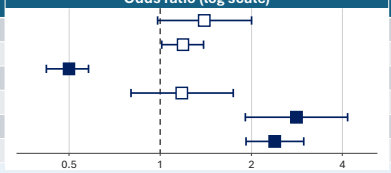
Reference group: no hormone exposure at any time.
Time Window for exposure:
18 months before index event
Index event:
- RVO [ICD-10-CM H34.81, H34.83] for cases
- Ophthalmology, Emergency department (ED), or Routine health visits for controls

Covariates:
Age, race, hypertension, diabetes mellitus, hyperlipidemia, obesity, tobacco use
(All assessed before the index date)
Statistical analysis:
- 2x2 contingency tables
- Crude odds ratios (ORs) with Woolf 95% confidence intervals, comparing each formulation with no exposure and each formulation with the others (6 comparisons)
- Two-sided Wald tests for P values
- Bonferroni correction for 6 comparisons (significance threshold P < 0.0083)
- Baseline balance assessed with standardized mean differences (SMD < 0.1 = adequate balance)

Results

Table 1: Crude odds ratios for retinal vein occlusion by hormone exposure group, with Bonferroni-adjusted P values

Comparison	Group A (Cases/ Controls)	Group B (Cases/ Controls)	Odds ratio (log scale)	OR (95% CI)	Bonferroni adjusted p-value
Estrogen vs No exposure	30 / 72,582	2,891 / 9,782,859		1.40 (0.98–2.00)	0.405
Progestin vs No exposure	155 / 442,383	2,891 / 9,782,859		1.19 (1.01–1.39)	0.234
Combined vs No exposure	164 / 1,119,466	2,891 / 9,782,859		0.50 (0.42–0.58)	**<0.0001**
Estrogen vs Progestin	30 / 72,582	155 / 442,383		1.18 (0.80–1.74)	1.000
Estrogen vs Combined	30 / 72,582	164 / 1,119,466		2.82 (1.91–4.16)	**<0.0001**
Progestin vs Combined	155 / 442,383	164 / 1,119,466		2.39 (1.92–2.98)	**<0.0001**



OR: odds ratio; CI: confidence interval. Bars show the 95% CI on a log scale; dashed line = no association (OR = 1).
Filled markers: significant after Bonferroni correction (adjusted P < 0.05); hollow markers: not significant. Adjusted P = P × 6 (six comparisons).

Table 2: Baseline characteristics and standardized mean differences among RVO cases, by hormone exposure group

Characteristic	Reference (no hormone) and RVO cases, n (%)	Estrogen exposure and RVO cases, n (%)	SMD	Progestin exposure and RVO cases, n (%)	SMD	Combined exposure and RVO cases, n (%)	SMD
Total N (analyzed)	2,719	30		151		162	
Mean age (SD)	39.6 (8.54)	43.4 (6.93)	0.4886	38.4 (9.33)	0.1403	37.2 (9.07)	**0.2751**
Race							
White	1,706 (62.7)	22 (73.3)	0.2286	100 (66.2)	**0.0728**	111 (68.5)	0.1218
Black or African American	436 (16.0)	10* (33.3)	0.4095	26 (17.2)	**0.0318**	17 (10.5)	0.1639
Asian	129 (4.7)	10* (33.3)	0.7819	10* (6.6)	**0.0812**	10* (6.2)	**0.0629**
Other	147 (5.4)	10* (33.3)	0.7554	11 (7.3)	**0.0771**	14 (8.6)	0.1269
Comorbidities							
Hypertension	675 (24.8)	13 (43.3)	0.3981	71 (47.0)	0.4755	59 (36.4)	0.2536
Diabetes mellitus	395 (14.5)	10* (33.3)	0.4519	36 (23.8)	0.2382	30 (18.5)	0.1076
Hyperlipidemia	307 (11.3)	12 (40.0)	0.6961	30 (19.9)	0.2382	28 (17.3)	0.1719
Obesity (overweight and obesity, E66)	468 (17.2)	13 (43.3)	0.5930	60 (39.7)	0.5154	53 (32.7)	0.3641
Tobacco use	52 (1.9)	0 (0.0)	0.1974	10* (6.6)	0.2346	10* (6.2)	0.2176

Characteristics are shown for RVO cases only. SMD: standardized mean difference of each hormone group versus the reference group (no hormone exposure); SMDs above 0.1 indicate meaningful imbalance. Groups were not matched.
**Counts under 10 are displayed as 10 by TriNetX, so percentages, SMDs and P values based on these cells are approximate.*

Conclusion

- Estrogen-only (OR 1.40, 95% CI 0.98-2.00) and progestin-only (OR 1.19, 1.01-1.39) therapy showed no significant crude association with RVO after Bonferroni correction.
- Combined therapy was associated with lower odds (OR 0.50, 0.42-0.58). Estrogen-only and progestin-only users had higher odds than combined users (ORs 2.82 and 2.39) but did not differ from each other (OR 1.18).
- Exposed groups differed from the reference in age and comorbidities (SMD > 0.1).

Limitations:
- The analysis was unmatched and crude, based on aggregate 2x2 tables, so residual confounding is likely.
- Exposure was based on prescription records, which do not confirm that medications were filled or taken, and could not distinguish contraceptive from therapeutic use.

Next Step:
- Stratified analysis by age, hypertension and diabetes, with Mantel-Haenszel pooled ORs across strata

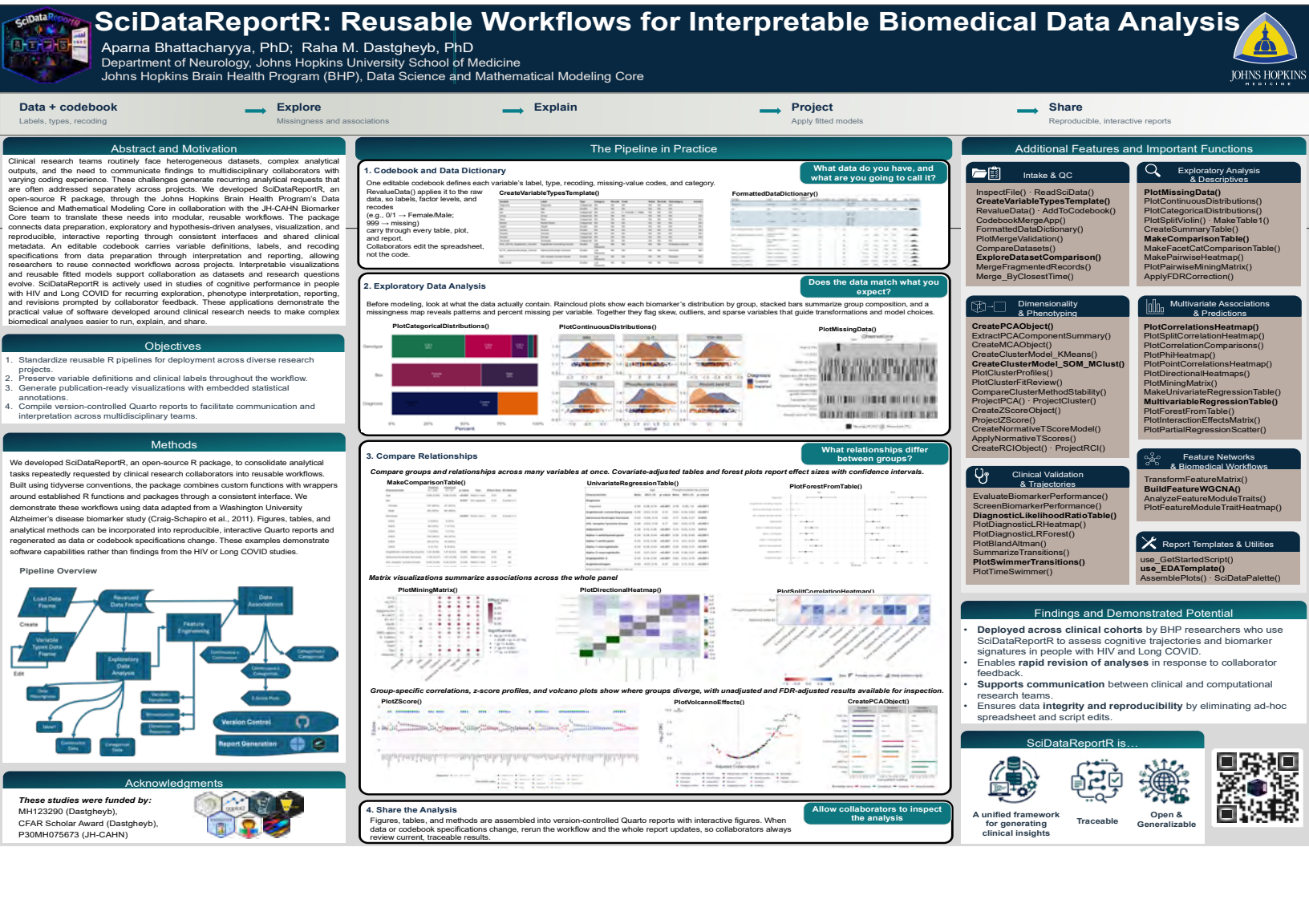
Funding

NIH 1R01EY033000, 1R01EY034325, 1R01EY029489, K08EY027458, 1R42EY035582, 1R41EB030875, University of Michigan MTRAC-FFMI grant, Alcon Research Institute Young Investigator Grant, Fight for Sight - International Retinal Research Foundation FFSGIA16002, Massey Foundation TBI Grand Challenge Award, Alliance for Vision Research, Research to Prevent Blindness. Foundation Fighting Blindness RC-CMM-0824-0899-JHU.

Aparna Bhattacharyya

SciDataReportR: Reusable Workflows for Interpretable Biomedical Data Analysis

Clinical research teams routinely face heterogeneous datasets, complex analytical outputs, and the need to communicate findings to multidisciplinary collaborators with varying coding experience. These challenges generate recurring analytical requests that are often addressed separately across projects. We developed SciDataReportR, an open-source R package, through the Johns Hopkins Brain Health Program's Data Science and Mathematical Modeling Core in collaboration with the JH-CAHN Biomarker Core team to translate these needs into modular, reusable workflows. The package connects data preparation, exploratory and hypothesis-driven analyses, visualization, and reproducible, interactive reporting through consistent interfaces and shared clinical metadata. An editable codebook carries variable definitions, labels, and recoding specifications from data preparation through interpretation and reporting, allowing researchers to reuse connected workflows across projects. Interpretable visualizations and reusable fitted models support collaboration as datasets and research questions evolve. SciDataReportR is actively used in studies of cognitive performance in people with HIV and Long COVID for recurring exploration, phenotype interpretation, reporting, and revisions prompted by collaborator feedback. These applications demonstrate the practical value of software developed around clinical research needs to make complex biomedical analyses easier to run, explain, and share.



Russell Calderon

Age-Relationship in Surgical Approach and Outcomes Following Retinal Detachment Repair

Purpose: To determine whether surgical approach and anatomical success following rhegmatogenous retinal detachment (RD) repair vary with age and whether age modifies the association between surgical approach and success.

Methods: We retrospectively studied 2,592 adults aged ≥50 years undergoing primary RD repair at the Wilmer Eye Institute from 2014–2026. Surgical approaches included pars plana vitrectomy (PPV), scleral buckle (SB), combined PPV/SB, and pneumatic retinopexy. The primary outcome was anatomical reattachment at 1 year. Associations were evaluated using Fisher's exact tests and multivariable logistic regression adjusted for age, sex, insurance, myopia, and calendar year.

Mean age was 64.6±9.1 years; 62.3% were men. Surgical approach differed across age groups (p<0.001). PPV use increased from 81.1% among patients aged 50–54 to 98.7% among those aged ≥80, while SB decreased from 15.0% to 1.3%. Overall anatomical success was 51.7% and differed by age (p=0.002) and surgical approach (p=0.010). After adjustment, surgical approach remained associated with success (p=0.014); pneumatic retinopexy had lower odds of success than PPV (adjusted OR=0.45, 95% CI 0.25–0.78; p=0.006). Age did not modify this association (interaction p=0.218).

Surgical management shifted toward PPV with advancing age, but age did not modify the relationship between surgical approach and anatomical success.

Age-Relationship in Surgical Approach and Outcomes Following Retinal Detachment Repair

Russell Calderon, BS;[1] Meltem Yashar, MD;[2] Luqman Munir, BS;[2] Yannis M. Paulus, MD, FACS, FARVO[2]

[1]Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

[2]Wilmer Eye Institute, Department of Ophthalmology, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States

Email contact: rcalde11@jh.edu

Lay abstract

Retinal detachment is a serious eye condition that becomes more common with age and often requires surgery to restore the retina to its normal position. We studied middle-aged and older adults to understand whether age was related to the type of surgery they received and their chances of successful retinal reattachment. We found that older patients were increasingly treated with pars plana vitrectomy (PPV), while scleral buckle surgery became less common with age. The likelihood of successful retinal reattachment varied by both age and type of surgery. However, we found no evidence that age changed the relationship between the type of surgery performed and the likelihood of success. These findings highlight important age-related differences in how retinal detachment is treated and may help guide future research on surgical outcomes in older adults.

Purpose

- Retinal detachment (RD) is a vision-threatening ophthalmic emergency that becomes increasingly common with age.[1,2]
- Multiple surgical approaches are used to repair RD, but it remains unclear how surgical management and outcomes differ across the aging spectrum.
- We aimed to determine whether surgical approach and anatomical success vary with age and whether age modifies the relationship between surgical approach and anatomical success.

Methods

- **Study design:** Retrospective cohort study conducted at the JHU Wilmer Eye Institute.
- **Participants:** 2,474 adults aged ≥50 years undergoing primary rhegmatogenous retinal detachment (RRD) repair.
- **Age groups:** 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, and ≥80 years.
- **Primary outcome:** Primary anatomical reattachment of the retina at 1 year.
- **Surgical approaches:** Pars plana vitrectomy (PPV), scleral buckle (SB), combined PPV/SB, and pneumatic retinopexy.

Statistical Analysis

- Surgical approach across age groups was compared using Fisher's exact test.
- Anatomical success across age groups and surgical approaches was compared using Pearson chi-square tests.
- Multivariable logistic regression evaluated the association between surgical approach and anatomical success, adjusting for age, sex, insurance status, myopia, and calendar year of surgery.
- An age × surgical-approach interaction tested whether age modified the association between surgical approach and anatomical success.
- Statistical significance was defined as p<0.05.

Results

Figure 1. Distribution of Primary Surgical Approach by Age Group Among Adults Undergoing Retinal Detachment Repair

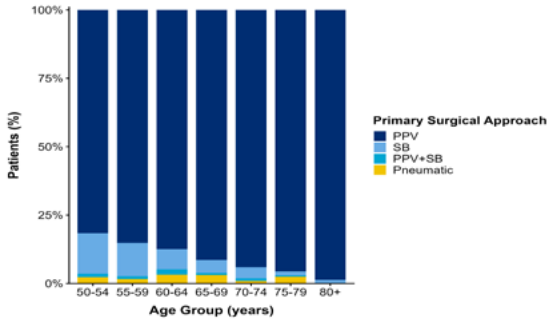


Figure 1. Bars represent the proportion of patients receiving each primary surgical approach within each age group. Surgical approach differed significantly across age groups (Fisher's exact test, p<0.001). PPV = pars plana vitrectomy; SB = scleral buckle; PPV+SB = combined pars plana vitrectomy and scleral buckle.

Figure 2. Primary Anatomical Success by Age Group Among Adults Undergoing Retinal Detachment Repair

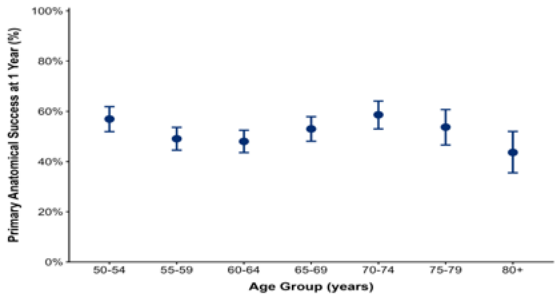


Figure 2. Proportion of patients achieving primary anatomical reattachment at 1 year; error bars represent 95% confidence intervals. Anatomical success differed significantly across age groups (Pearson chi-square test, p=0.004).

Table 1. Characteristics of the Study Cohort

Characteristic	Overall (N=2,474)
Age, years, mean ± SD	64.6 ± 9.1
Age group, n (%)	
50–54	397 (16.0)
55–59	485 (19.6)
60–64	502 (20.3)
65–69	419 (16.9)
70–74	319 (12.9)
75–79	203 (8.2)
≥80	149 (6.0)
Sex, n (%)	
Male	1,547 (62.5)
Female	927 (37.5)
Insurance status, n (%)	
Commercial	1,059 (42.8)
Government	1,088 (44.0)
Uninsured	327 (13.2)
Myopia, n (%)	
Yes	437 (17.7)
No	2,037 (82.3)
Primary surgical approach, n (%)	
Pars plana vitrectomy (PPV)	2,200 (88.9)
Scleral buckle (SB)	193 (7.8)
Combined PPV/SB	27 (1.1)
Pneumatic retinopexy	54 (2.2)

Conclusions

Among adults aged ≥50 years undergoing primary RD repair, age was strongly associated with RD surgical approach, but we found no evidence that age changes the relationship between procedure type and anatomical success. Future studies should investigate whether similar age-related patterns are observed in postoperative visual acuity to determine the functional consequences of different RD surgical approaches across the aging spectrum.

References

1. Ibrar A, Panayiotis M, Mohamed EA. Recognising and managing retinal detachments. Br J Hosp Med (Lond). 2021 Oct 2;82(10):1-11. doi: 10.12968/hmed.2021.0145. Epub 2021 Oct 26. PMID: 34726948.

2. Ferrara M, Al-Zubaidy M, Song A, Avery P, Laidlaw DA, Williamson TH, Yorston D, Steel DHW; BEAVRS and Euretina VR Retinal Detachment Outcomes Group. The effect of age on phenotype of primary rhegmatogenous retinal detachment. Eye (Lond). 2023 Apr;37(6):1114-1122. doi: 10.1038/s41433-022-02061-y. Epub 2022 Apr 26. PMID: 35473967; PMCID: PMC10102138.

Funding

NIH 1R01EY033000, 1R01EY034325, 1R01EY029489, K08EY027458, 1R42EY035582, 1R41EB030875, University of Michigan MTRAC-FFMI grant, Alcon Research Institute Young Investigator Grant, Fight for Sight - International Retinal Research Foundation FFSGIA16002, Massey Foundation TBI Grand Challenge Award, Alliance for Vision Research, Research to Prevent Blindness.

Minseo Choi

Adaptive KV Retention for LLM Agents at Human-Approval Timescales

LLM agents increasingly pause before consequential actions, such as sending email or running commands, to wait for human approval. These waits last minutes to hours, far longer than the seconds-scale tool-call pauses targeted by prior agent-serving systems. During the wait, the request's KV cache (the attention state for its prior tokens) can stay in GPU memory, move to host memory, or be discarded and recomputed at resume. We show that this choice is costly in both directions: keeping suspended KV in GPU memory reduces active-request goodput by 41%, while evicting it makes resumption roughly 10x slower. We introduce GPU opportunity cost, which places holding and reconstruction costs on a common GPU-second scale using three platform-calibrated coefficients, and apply Little's law to price scarce host memory under load. Our controller offloads KV at suspension and uses a held-out calibration sample to choose between indefinite host retention and load-dependent expiration, without predicting individual waits. Serving Llama-3.1-70B on four H100 GPUs with approval-gated workloads derived from tau2-bench, the controller improves active-request goodput by 23–51% over vLLM baselines, 22–29% over MORI, and 41–52% over Continuum.

Adaptive KV Retention for LLM Agents at Human-Approval Timescales

Minseo Choi[1], Ananya Joshi[1,2]
[1]Department of Computer Science, Johns Hopkins University
[2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine
arXiv:2608.30830 · AgenticOS Workshop @ NeurIPS 2026

Introduction

Human approval makes KV retention a capacity-allocation problem.

LLM agents pause for human approval before consequential actions. Prior agent-serving systems (MORI, Continuum) target 1–2 s tool-call pauses; human approvals take minutes to hours. **Where should suspended KV live meanwhile?**

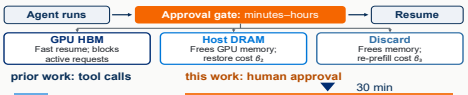


Figure 1. Approval waits and tool-call pauses (logarithmic axis).

Workload and production evidence

Production evidence: AgentSysBench [5]

35,037 production coding-agent sessions over 24 hours. Most idle gaps last 1–10 min, overlapping a 5-min cache TTL in that deployment.

59.4% Sessions with at least one eviction

55.9% Cache-create tokens attributed to eviction

31.5% Estimated token-billed cost from re-prefill after eviction

Cost uses Claude Opus 4.6 prices. External evidence motivates long-idle retention. Our experiments use **τ²-bench**; **AgentSysBench** evaluation remains future work.

Evaluation workload: τ²-bench [4]

From 2,624 released retail/airline runs, ToolType.WRITE selection and truncation before the first state-changing call yield 2,161 approval-gated prefixes. Median context: 8.4K tokens = 5.4K shared system + 3.0K private suffix. Prefix caching reuses shared state, leaving ≈0.98 GB per request. Simulated users approve 93–96% of gates.

(a) What retention costs

−41%

Active goodput with 72 suspended contexts in HBM

(b) What it buys: Resume TTFT

Resident in HBM: 66–75 ms

Evicted, then recomputed: 687–705 ms

Figure 2. ≈10× slower resume after eviction. Llama-3.1-70B, 4×H100 NVL, 48 active clients. Victim exclusion isolates residency cost and is disabled for stock-vLLM runs.

A 1-s resume stall is 0.06% of a 30-min wait. HBM holding cost accrues throughout suspension.

Limitations

Static, offline-calibrated policy. Synthetic waits and simulated approval outcomes. Live evaluation covers one model family. Higher goodput can trade off resume latency: mixture TTFT is 628 ms versus MORI's 606 ms. Production traces and broader model evaluation remain necessary.

Retention policy

Retention cost model: GPU opportunity cost

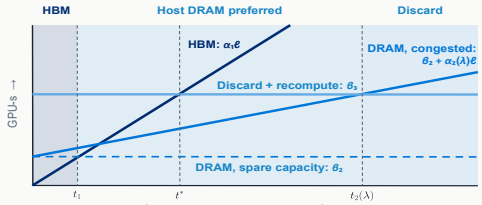
A common GPU-time scale compares the active-serving capacity consumed by HBM retention, host storage, and recomputation. The controller compares these costs without predicting each request's wait.

$$C_j(\ell) = \alpha_j \ell + \beta_j$$

$C_j(\ell)$: Total opportunity cost (GPU-s); ℓ: Wait length (s)
α_j: Holding cost per second (GPU-s/s)
β_j: One-time resume cost (GPU-s)

GPU HBM	Host DRAM	Discard
Hold: α_1 Resume: 0	Hold: 0, or $\alpha_2(\lambda)$ if full Resume: θ_2	Hold: 0 Resume: θ_3 (re-prefill)

Platform calibration: Llama-3.1-70B (BF16, TP4), 4×H100 NVL. The 3.0K-token private suffix occupies ≈0.98 GB. $\alpha_1 = 0.0176$ GPU-s/s derives from its KV-pool share. $\theta_2 = 0.02$ GPU-s is a conservative scheduling/synchronization allowance. $\theta_3 = 1.91$ GPU-s comes from suffix re-prefill. DMA can overlap GPU execution, so transfer time is not charged one-for-one.



$t_1 = \frac{\beta_2}{\alpha_1} \approx 1.1\,\text{s}$ and $t^* = \frac{\beta_3}{\alpha_1} \approx 109\,\text{s}$ Wait length ℓ →

Figure 3. Tier costs and break-even points on 4×H100 NVL. Offload to host DRAM at suspension because t_1 is about 1 s.

Host timeout: Capacity prices future recomputation

Host DRAM fits C_2 contexts. For mean wait W_{ref}, Little's law gives $\lambda_{\text{crit}} = C_2 / W_{\text{ref}}$. Above this load, overflow displaces host copies and causes future recomputation. Amortizing that cost across occupied slots gives:

$$\alpha_2(\lambda) = \max(0, \lambda - \lambda_{\text{crit}}) \cdot \frac{\beta_1}{C_2}$$
$$t_2(\lambda) = \frac{\beta_1 - \beta_2}{\alpha_2(\lambda)} \qquad (= \infty \text{ if } \lambda \le \lambda_{\text{crit}})$$

$\alpha_2(\lambda)$: Congestion price; $t_2(\lambda)$: Host expiration time; λ: Offered load; cpu_ttl(λ) discards at $t_2(\lambda)$. Runtime needs only λ. Finite-capacity replay evaluates the price.

Controller: Calibrate, then serve

Before serving	**Calibrate** Held-out waits	**Replay** Both host policies	**Freeze** Lower-cost branch
While serving	**Suspend** Offload KV to DRAM	**Retain or expire** $t_2 = \infty$ or $t_2(\lambda)$	**Resume** θ_2 present / θ_3 absent

Before serving, held-out waits select the lower-cost host policy. Calibration and serving traces are disjoint. Lognormal and exponential waits choose indefinite retention. The short/long mixture chooses expiration because elapsed age distinguishes near-term resumptions from long waits.

Live serving results

Llama-3.1-70B (BF16, TP4), 4×H100 NVL, vLLM [1], 375–380 GB host memory. All systems receive identical traces at $2\lambda_{\text{crit}}$. Waits: lognormal ($\sigma = 1$), exponential, or a 50/50 short/long mixture (means 60 s / 3,540 s; long-wait lognormal $\sigma = 0.7$). Each family has a 30-min mean.

Active-request goodput improves 23–51% over vLLM baselines, 22–29% over MORI [2], and 41–52% over Continuum [3].

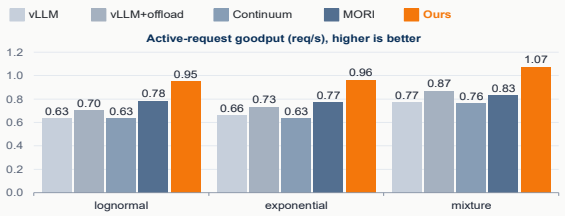


Figure 4. Completed active requests/s under identical traces and offered load. Higher is better.

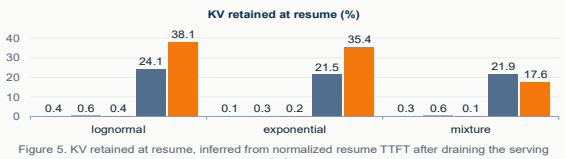


Figure 5. KV retained at resume, inferred from normalized resume TTFT after draining the serving window.

Even with 380 GB, vLLM+offload retains at most 0.6%: LRU evicts long-waiting KV first. On mixture waits, MORI retains 21.9% versus our 17.6%, yet serves 0.83 versus 1.07 req/s. Expiration improves active-serving capacity when retention costs exceed recomputation.

Read the paper

Future work

Online control and indexing

Online policy reselection as offered load and approval waits drift, with periodic platform-cost recalibration. An incremental index would find the next eviction candidate without rescanning suspended sessions.

Learned retention and evaluation

Investigate LHD-style retention that learns expected reuse per byte. Validate on production traces, including **AgentSysBench**, and additional model families.

Acknowledgments

This research was supported by the NVIDIA Academic Grant Program.

[1] Kwon et al., vLLM, SOSP 2023. [2] Xia et al., MORI, 2026. [3] Li et al., Continuum, 2025. [4] Barres et al., τ²-bench, 2025. [5] Chang et al., **AgentSysBench**, arXiv:2608.15127, 2026.
Baseline implementations: MORI is a policy port on vLLM. Continuum uses a public scheduling fork with fixed 2-s pin TTL, without its pause-duration estimator.

Tamanna Ghosh

Longitudinal Laboratory Trajectories and Risk of Incident and Vision-Threatening Diabetic Retinopathy in Adults With Type 2 Diabetes

Background: Diabetic retinopathy (DR) is a leading cause of preventable blindness. We examined whether time-updated trajectories of routinely collected laboratory tests, beyond HbA1c, identify patients at higher risk of incident and vision-threatening DR (VTDR).

(Methods section omitted from program for space)

Results: Over a median 5.6 years, 4,111 patients in the primary cohort developed DR (5-year cumulative incidence, 12.6%). Five-year VTDR incidence was 44.6% among patients with baseline mild or moderate DR versus 4.8% among those without. In fully adjusted models, higher mean HbA1c was the strongest correlate of incident DR (HR, 1.55; 95% CI, 1.50-1.60) and VTDR (HR, 1.29; 95% CI, 1.23-1.34). Lower eGFR (HR, 0.82), higher UACR (HR, 1.18), and higher NLR (HR, 1.08) were also associated with incident DR. Greater variability in HbA1c, eGFR, HDL cholesterol, and ALT was associated with higher risk of both outcomes. Lower mean ALT and AST were associated with higher risk. Associations for mean HDL cholesterol and triglycerides were null after adjustment for HbA1c, BMI, and blood pressure.Only higher HbA1c and UACR were associated with faster VA decline and 15-letter loss.

Conclusions: Glycemic control was the dominant laboratory correlate of DR, with kidney and inflammatory markers showing modest independent associations. HbA1c and UACR also tracked vision decline. Mean eGFR was associated with incident DR but not VTDR. These observational findings warrant external validation.

Longitudinal Laboratory Trajectories and Risk of Incident and Vision-Threatening Diabetic Retinopathy in Adults With Type 2 Diabetes

Tamanna Ghosh[1] • Luqman Munir[1] • Meltem Yashar, MD[1] • Yannis M. Paulus, MD[1,2]

1 Wilmer Eye Institute, Department of Ophthalmology, Johns Hopkins University, Baltimore, Maryland, United States

2 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States | Contact: ypaulus1@jh.edu

1 Background

Diabetic retinopathy (DR) is a leading cause of preventable blindness. Most studies use a single baseline value, such as one HbA1c, but diabetes unfolds over years.

Objective: To test whether time-updated trajectories of routine laboratory tests, beyond HbA1c, identify patients at higher risk of incident and vision-threatening DR (VTDR).

2 Methods

Design: Retrospective cohort (STROBE), Johns Hopkins EHR: adults with type 2 diabetes (E11 codes; other diabetes types excluded), aged 18+ at index, in continuing eye care (>2 ophthalmology visits after index). Index = first eye visit on or after first E11 code (Apr 2013–Jun 2026).

Primary cohort (n = 27,657): no DR or DME at baseline → incident DR.

Secondary cohort (n = 30,803): adds baseline mild or moderate DR (severe NPDR, PDR, DME excluded) → VTDR.

Exposures: 10 routine labs (HbA1c, LDL, HDL, triglycerides, eGFR, UACR, ALT, AST, NLR, PLR), each analyzed after 3+ draws; eGFR from creatinine (CKD-EPI 2021).

Outcomes: first code for DR (E11.31–E11.35) or VTDR (severe NPDR, PDR, DME).

Statistical analysis

- **Cox models:** time-updated running mean and variability of each lab (HR per 1 SD), adjusted for age, sex, race, smoking, running HbA1c, BMI, blood pressure, and baseline medications.
- **Disease subtype:** severe NPDR/PDR and DME analyzed as separate outcomes.
- **Dose-response:** spline curves of running mean HbA1c.
- **Absolute risk:** Kaplan–Meier, and cumulative incidence with death as a competing event.
- **Visual acuity:** labs vs VA decline in patients with DR.
- **Robustness:** FDR-adjusted; sensitivity to cataract surgery, index year, 365-day lag.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; DME, diabetic macular edema; DR, diabetic retinopathy; eGFR, estimated glomerular filtration rate; HR, hazard ratio; NLR, neutrophil-lymphocyte ratio; NPDR, nonproliferative DR; PDR, proliferative DR; PLR, platelet-lymphocyte ratio; UACR, urine albumin-creatinine ratio; VA, visual acuity; VTDR, vision-threatening DR.

3 Baseline Characteristics

Characteristic	Primary (n = 27,657)	Baseline mild/moderate DR (n = 3,146)	Secondary (n = 30,803)
Age, y, mean (SD)	64.1 (11.8)	64.2 (11.9)	64.2 (11.8)
Female sex	14,669 (53.0)	1,603 (51.0)	16,272 (52.8)
White race	14,222 (51.4)	1,359 (43.2)	15,581 (50.6)
Black race	9,655 (34.9)	1,296 (41.2)	10,951 (35.6)
Asian race	1,840 (6.7)	211 (6.7)	2,051 (6.7)
BMI, kg/m², mean (SD)	32.5 (7.2)	32.4 (7.3)	32.5 (7.2)
Systolic BP, mm Hg, mean (SD)	133.0 (18.6)	138.3 (22.0)	133.5 (19.0)
Insulin	5,028 (18.2)	1,285 (40.8)	6,313 (20.5)
Metformin	13,563 (49.0)	1,418 (45.1)	14,981 (48.6)
GLP-1 RA, SGLT2i, or DPP-4i	4,745 (17.2)	642 (20.4)	5,387 (17.5)
Statin	13,969 (50.5)	1,557 (49.5)	15,526 (50.4)
ACE inhibitor or ARB	13,422 (48.5)	1,622 (51.6)	15,044 (48.8)
Index year 2013–2014	8,793 (31.8)	1,249 (39.7)	10,042 (32.6)
Follow-up, y, median (IQR)	5.6 (3.0-9.3)	4.6 (2.2-8.1)	5.5 (3.0-9.2)
Eye visits after index, median (IQR)	8 (5-14)	9 (5-18)	8 (5-14)

Table 1. Data are n (%) unless noted. Medications: any order in the year before index. BMI and blood pressure were available for 16,897 and 17,331 primary-cohort patients.

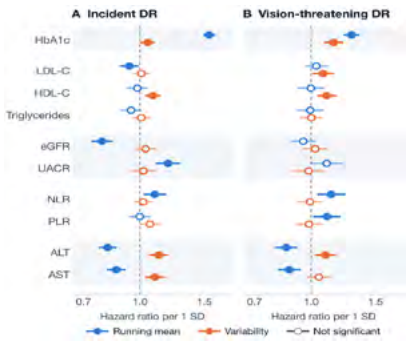
Key findings

- Beyond HbA1c: **lower eGFR** (HR 0.79), **higher UACR** (HR 1.19), and **higher NLR** (HR 1.10) were associated with incident DR (Figure 1).
- Mean HbA1c also predicted VTDR (HR 1.29); hazard rose steeply above about 7% (Figure 3).
- Greater **variability** in HbA1c, HDL cholesterol, and ALT was associated with higher risk of both outcomes.
- Lower mean ALT and AST were associated with higher risk; mean HDL and triglycerides were null after adjustment.
- Lower eGFR (HR 0.72) and higher UACR (HR 1.36) were associated with **severe NPDR/PDR**, not DME (Figure 3).
- Higher mean HbA1c and UACR tracked with faster VA decline, but effects were small (about 0.4 letters/year per SD of HbA1c; Figure 2).

4 Results

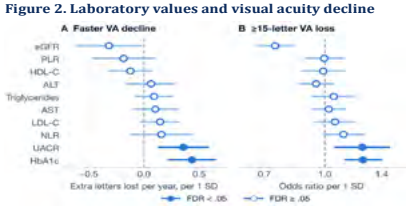
27,657	4,111	44.6% vs 4.8%	HR 1.55
adults, no baseline DR (primary cohort)	incident DR; 26.6 per 1,000 person-years	5-year VTDR risk with vs without baseline DR	per 1-SD higher mean HbA1c (95% CI, 1.50–1.60)

Figure 1. Laboratory trajectories and DR risk



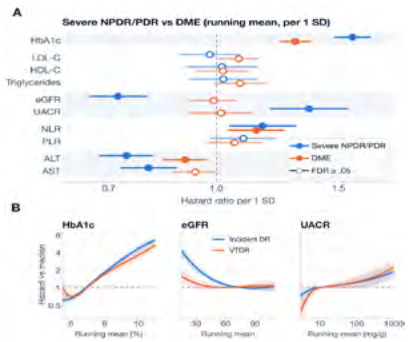
Hazard ratios per 1-SD higher running mean (blue) or variability (orange), fully adjusted for age, sex, race, smoking, running HbA1c, BMI, blood pressure, and baseline medications (baseline DR for VTDR). Bars are 95% CIs; hollow markers have FDR ≥ .05.

Figure 2. Laboratory values and visual acuity decline



Patients with DR and VA follow-up (3,797 patients, 6,979 eyes; 25.8% lost ≥15 letters). A, Adjusted change in VA decline per 1 SD higher mean lab (age, sex, race, baseline VA, DR group). B, Odds of ≥15-letter loss. Bars are 95% CIs; filled markers FDR < .05.

Figure 3. Disease subtype and dose-response



A, Fully adjusted HR per 1 SD of running mean for severe NPDR/PDR and DME; hollow markers have FDR ≥ .05. B, Hazard by running mean relative to the median for incident DR and VTDR; shaded bands are 95% CIs.

5 Conclusion

Glycemic control was the dominant laboratory correlate of DR. Lower eGFR and higher UACR were associated with incident DR and specifically with severe NPDR/PDR rather than DME, and inflammatory markers added modest signal. These observational findings warrant external validation.

Limitations

- Observational, single health system; exploratory analyses; DR outcomes come from diagnosis codes, not graded exams.
- Medications are baseline-only; variability may partly reflect testing frequency and treatment changes.
- About a third of index visits fall in 2013–2014; true diabetes duration is unknown. Death is a competing event for absolute risk only; HRs are cause-specific. The visual acuity analysis has no treatment data (anti-VEGF, laser).

Funding

NIH 1R01EY033000, 1R01EY034325, 1R01EY029489, K08EY027458, 1R42EY035582, 1R41EB030875; University of Michigan MTRAC-FFMI grant; Alcon Research Institute Young Investigator Grant; Fight for Sight – International Retinal Research Foundation FFSGIA16002; Massey Foundation TBI Grand Challenge Award; Alliance for Vision Research; Research to Prevent Blindness; Foundation Fighting Blindness RC-CMM-0824-0899-JHU.

Leo Han

UnmixingMM: Unmixing in Whole-Slide Multiplex Immunofluorescence with Majorization–Minimization

Multiplex immunofluorescence (mIF) imaging measures many biomarkers on a single tissue section, but the fluorophores have broadly overlapping emission spectra and tissue autofluorescence adds a strong, spatially varying background, so every measured band is a mixture. Spectral unmixing recovers per-pixel marker abundances. Standard pipelines solve an unconstrained least-squares problem whose solution contains physically infeasible negative abundances, then repair it with ad-hoc corrections: clipping negatives to zero or dropping the offending column of the spectral library and refitting. Neither solves a well-defined problem, and both introduce bias and visual artifacts.

We introduce UnmixingMM, which imposes strict nonnegativity as a constraint and solves the resulting problem with a majorization-minimization (MM) solver. Per-band noise is estimated from the data along a Hilbert-curve traversal, down-weighting the least reliable bands. The multiplicative updates keep every iterate nonnegative by construction and treat each pixel independently, so the solver runs on GPU at whole-slide, single-cell resolution.

On 43-band images of a melanoma panel, clipping leaves regions pinned to zero and drop-and-refit leaves pixel-level discontinuities. UnmixingMM recovers continuous abundance maps without such artifacts, which downstream cell segmentation and phenotyping depend on.

UnmixingMM: Unmixing in Whole-Slide Multiplex Immunofluorescence with Majorization–Minimization

Taehoon Han[1], Tamás Budavári[1,2], Alexander S. Szalay[2]
[1]Department of Applied Mathematics and Statistics, Johns Hopkins University
[2]Department of Physics and Astronomy, Johns Hopkins University

Introduction

Multiplex immunofluorescence (mIF) imaging measures many biomarkers on a single tissue section. The fluorophores labeling them have broadly overlapping emission spectra, and tissue autofluorescence (AF) contributes a strong, spatially varying background, making it hard to distinguish the different markers that are co-located.

Spectral unmixing recovers the per-pixel abundance of individual biomarkers. Standard pipelines solve an unconstrained least-squares problem whose solution contains physically infeasible negative abundances, which are then repaired with ad-hoc corrections such as clipping and dropping the offending column of the spectral library and refitting. Such repair introduces biases and visual artifacts.

We introduce UnmixingMM, a new mIF unmixing pipeline. It imposes strict nonnegativity as a constraint and solves the resulting problem with a majorization-minimization (MM) solver. Also, it estimates the noise level from the observation, down-weighting the noisier bands. The solver then updates each pixel independently, so it parallelizes directly on GPU and runs at whole-slide, single-cell resolution.

Linear Unmixing Model

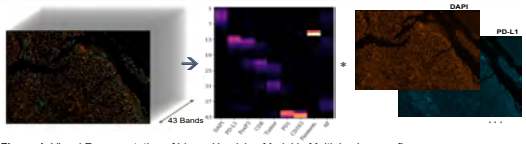
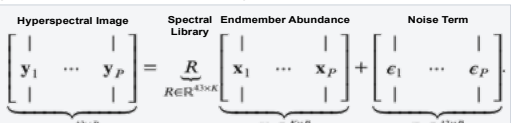


Figure 1. Visual Representation of Linear Unmixing Model in Multiplex Immunofluorescence.



Noise Estimation

Noise Model: zero-mean Gaussian, independent across pixels, with band-dependent variance σ_b^2

σ_b^2 estimated from difference of neighboring pairs of pixels along Hilbert-curve traversal, with only low-gradient (flat region) pixels selected.

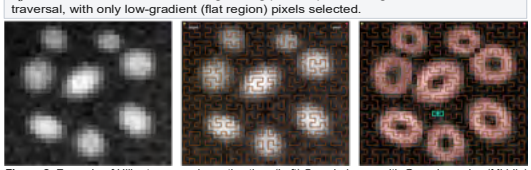


Figure 2. Example of Hilbert-curve noise estimation. (Left) Sample image with Gaussian noise (Middle) Image with the Hilbert-curve overlay (Right) Sobel gradient map with Hilbert-curve overlay and gradient-percentile-based masking. Highlighted region is a surviving pair whose difference $d \sim N(0, 2\sigma^2)$.

Least Squares Objective & Existing Solver

$$\hat{X} = \arg\min_{X} \sum_{b=1}^{43} \sum_{p=1}^{P} \frac{\left(Y_{b,p} - (RX)_{b,p}\right)^2}{\hat{\sigma}_b^2} \qquad w_b = \frac{1}{\hat{\sigma}_b^2},\ W = \mathrm{diag}(w_1, \ldots, w_{43}),$$

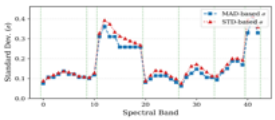


Figure 3. Estimation of band-dependent noise level. MAD is chosen for consistency to estimate σ_b

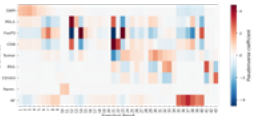


Figure 4. Heatmap of weighted pseudoinverse $R_W^\dagger$

$$X_{pinv} = R_W^\dagger Y = \left(R^T W R\right)^{-1} R^T W Y, \quad R_W^\dagger \equiv \left(R^T W R\right)^{-1} R^T W.$$

Unconstrained least-squares (LS) solution contains negative entries. Negative abundance does not make sense as these are physical photon counts. There are two approaches to repair such infeasibility: Clipping and Drop-and-Refit (DaR) methods.

Clipping

$$X_{clip} = \max\left(X_{pinv}, 0\right)$$

Negative entries are clipped to zero while positive entries are left untouched. The LS fit inflates some abundances to compensate for negative ones, clipping only removes negativity not the inflation. Also, negative abundances in X_{pinv} do not necessarily mean zero abundance in that pixel.

Drop-and-Refit

$$M_p = \mathrm{diag}\left(\mathbb{1}\left\{[X_{\mathrm{Piw}}]_{kp} \geq 0\right\}\right)_{k=1}^{K},$$

Repeat until no negative entries

$$R'_p \equiv R\, M_p,$$

$$x_p^{\mathrm{DaR}} = \left(R\, M_p\right)_W^\dagger y_p$$

Violating pixels are refit with a new weighted pseudoinverse, for all endmembers not only negative ones. But masking is still a guess, not a derivation.

Majorization–Minimization (MM) Solver

$$\hat{X} = \arg\min_{X \geq 0} \sum_{b=1}^{43} \sum_{p=1}^{P} \frac{\left(Y_{b,p} - (RX)_{b,p}\right)^2}{\hat{\sigma}_b^2}$$

To avoid the aforementioned artifacts and problems, we need to impose strict nonnegativity, solving a constrained optimization problem. We do so by setting up the MM multiplicative updates, which guarantee a monotonic decrease in the loss.

MM Update Rule (Update Factor)

$$X^{(t+1)} \leftarrow X^{(t)} \odot \frac{R^T W Y}{R^T W R X^{(t)}}$$

The MM update is derived from the Gaussian negative log-likelihood of the weighted LS objective. Nonnegativity holds by construction, since every term in the multiplicative update is nonnegative.

- $X^{(0)}$ initialized to ones, or any positive values
- W diagonal, entries $1/\hat{\sigma}_b^2$
- R the spectral library, with nonnegative coefficients
- Y the image, the measured photon counts, nonnegative

This strictly positive iterate multiplied by nonnegative terms always stays nonnegative. MM satisfies the constraint and guarantees the objective decreases monotonically.

Results: Multiplex IF Slides

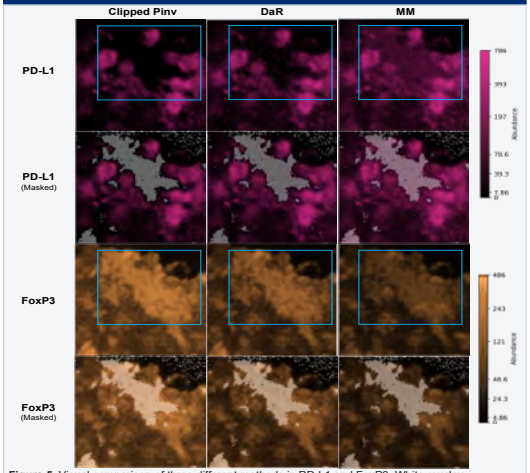


Figure 5. Visual comparison of three different methods in PD-L1 and FoxP3. White overlay highlights the pixels where the weighted pseudoinverse solution has negative abundance in PD-L1.

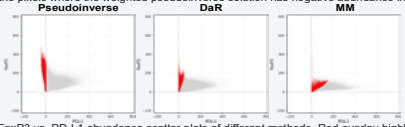


Figure 6. FoxP3 vs. PD-L1 abundance scatter plots of different methods. Red overlay highlights the pixels where the weighted pseudoinverse solution has negative abundance in PD-L1.

Conclusions & Future Work

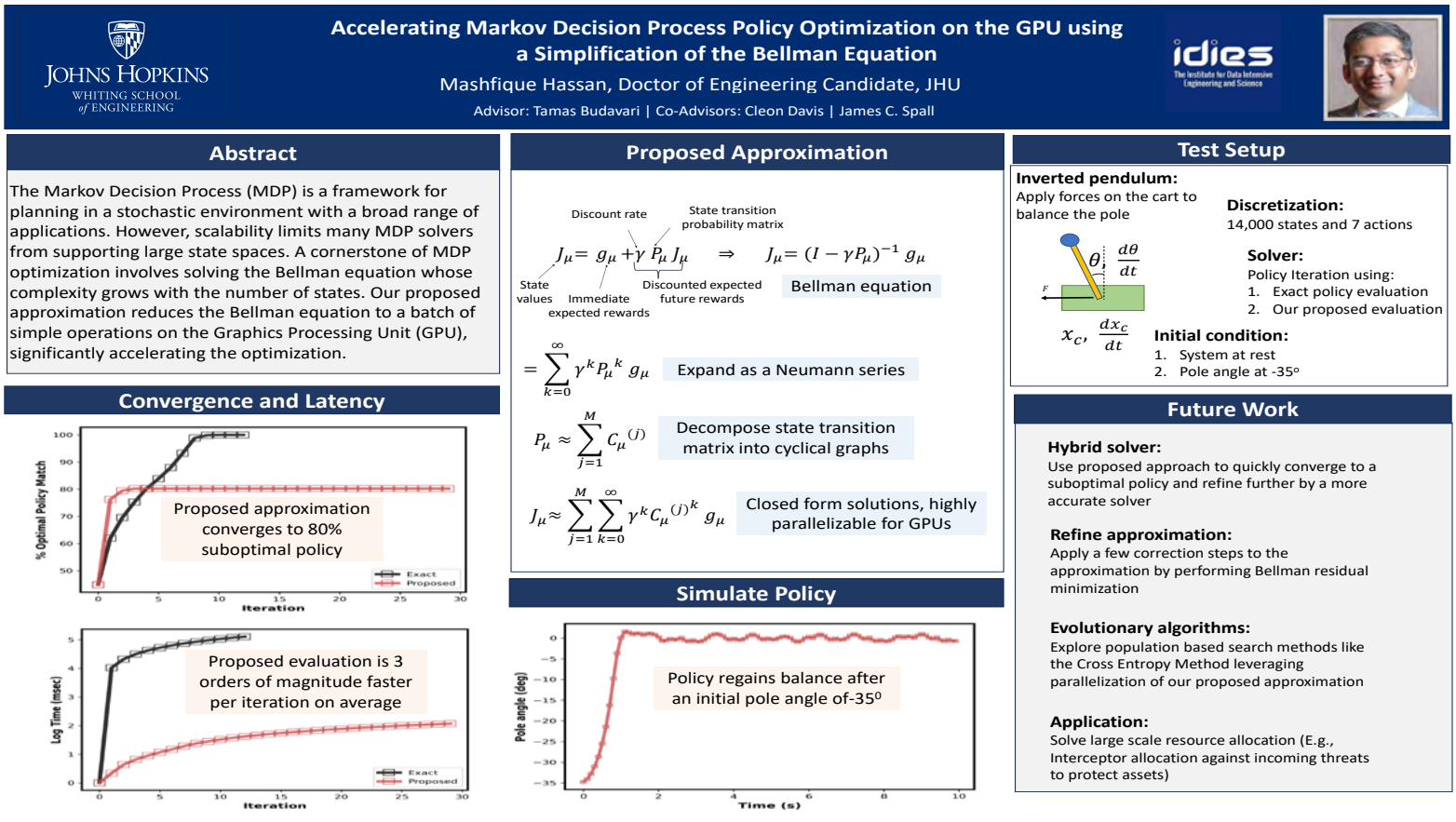
- **Constrained, not repaired.** Nonnegativity holds by construction, removing the bias, zeroed regions, and pixel discontinuities left by ad-hoc fixes.
- **Noise-aware.** Per-band σ_b estimated from the data down-weights the least reliable channels.
- **Scalable.** Monotone, tuning-free, per-pixel updates run on GPU at whole-slide, single-cell resolution for AstroPath.
- **Next:** Learn R from calibration data + Better AF model to reduce crosstalk into the endmembers

Sponsors: Bristol Myers Squibb, Emerson Collective, Harry J. Lloyd Charitable Trust, AKOYA Biosciences, The Mark Foundation for Cancer Research, Melanoma Research Alliance, NIH National Cancer Institute

Mashfique Hassan

Accelerating Markov Decision Process Policy Optimization on the GPU using a Simplification of the Bellman Equation

The Markov Decision Process (MDP) is a framework for planning in a stochastic environment with a broad range of applications. However, scalability limits many MDP solvers from supporting large state spaces. A cornerstone of MDP optimization involves solving the Bellman equation whose complexity grows with the number of states. Our proposed approximation reduces the Bellman equation to a batch of simple operations on the Graphics Processing Unit (GPU), significantly accelerating the optimization.



Jiayue Hu

CHAOS: Data for High-Entropy Oxide Discovery

High-entropy oxides combine several metals within a shared oxide lattice. Their chemical diversity offers opportunities in energy storage and catalysis. However, atomic arrangements, structural relaxation, and competing phases make material discovery difficult. We introduce CHAOS, a first-principles database that connects disordered oxide compositions with ordered reference oxides and their computational records. We represent chemical disorder with multiple atomic arrangements and calculate their relaxed structures and energies. Chemical and structural descriptors provide complementary views of how oxides accommodate mixed metals. Our current analysis focuses on rocksalt oxides and supports comparisons across compositions. CHAOS provides two routes to these data: direct database queries and natural-language searches through CHAOS-Agent. Both routes return records with identifiers, allowing researchers to inspect the calculations behind reported properties. This work provides a data foundation for studying complex oxides and selecting candidates for further investigation. Future directions include broader structural coverage, temperature and defect effects, and closer links between calculations and experimental feedback.

CHAOS: Data for High-Entropy Oxide Discovery

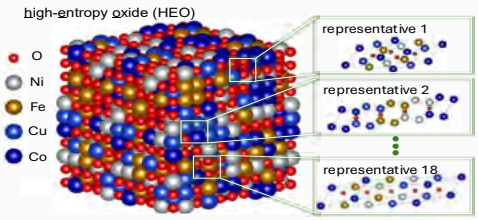
First-principles data connect chemical disorder to materials discovery.

Jiayue Hu, Shao-Yu Tseng, Matthew Brownrigg, Peter Boctor, Guangshuai Han, Tianhao Li, William Shiber, Xiao Xu, Bryan Lim, Maya Niyogi, Keith Clark, and Corey Oses (PI)
Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA

Why these oxides need better data

High-entropy oxides mix several metals in one lattice. Atomic disorder creates opportunities for energy materials and challenges for discovery.

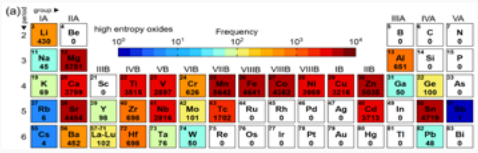
Representing atomic disorder



Schematic of the supercell representatives used to model disorder in (CoCuFeNi)O

Connected oxide records

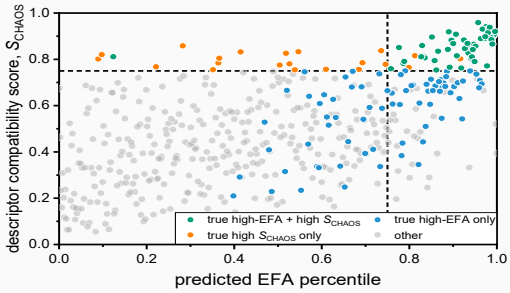
CHAOS links disordered compositions with ordered reference oxides. Its computational records help researchers compare materials and plan experiments.



Elemental coverage of CHAOS calculations. Color shows calculation frequency. White indicates no entries.

Data for materials prioritization

CHAOS combines chemical and structural descriptors to compare oxide compositions. The map identifies candidates with favorable predicted EFA rank and descriptor compatibility.



Each point represents one calculated rocksalt composition. Higher values on both axes identify candidates for further study.

52 candidates prioritized
48 have high calculated EFA

Entropy-forming ability (EFA) quantifies the tendency and ease with which a system forms high-entropy disordered phase.

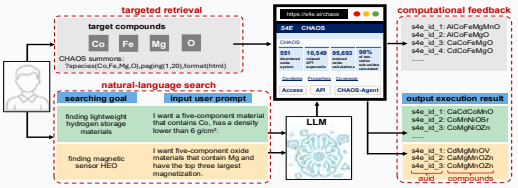
$$\mathrm{EFA} = \left(\sqrt{\frac{\sum_{i=1}^{N_s} g_i (H_i - H_{\mathrm{mix}})^2}{(\sum_{i=1}^{N_s} g_i) - 1}} \right)^{-1} \qquad H_{\mathrm{mix}} = \frac{\sum_{i=1}^{N_s} g_i H_i}{\sum_{i=1}^{N_s} g_i}$$

The compatibility score favors lower distortion and smaller chemical mismatch.

$$S_{\mathrm{CHAOS}} = \frac{1}{2} S_{\mathrm{distortion}} + \frac{1}{2} S_{\mathrm{chemistry}}$$

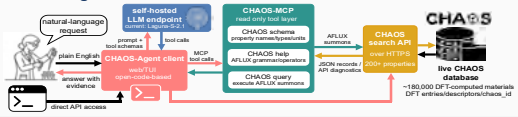
Ask in plain English

CHAOS supports direct API queries and natural-language questions through CHAOS-Agent. Both routes return records with IDs, so researchers can inspect the source calculations.



Direct queries and natural-language questions access the same database.

CHAOS-Agent architecture:



Conclusions and outlook

CHAOS provides a data foundation for complex oxide research. Future work can expand structure coverage and connect calculations with experiments. Candidate materials still need experimental tests.

Acknowledgments

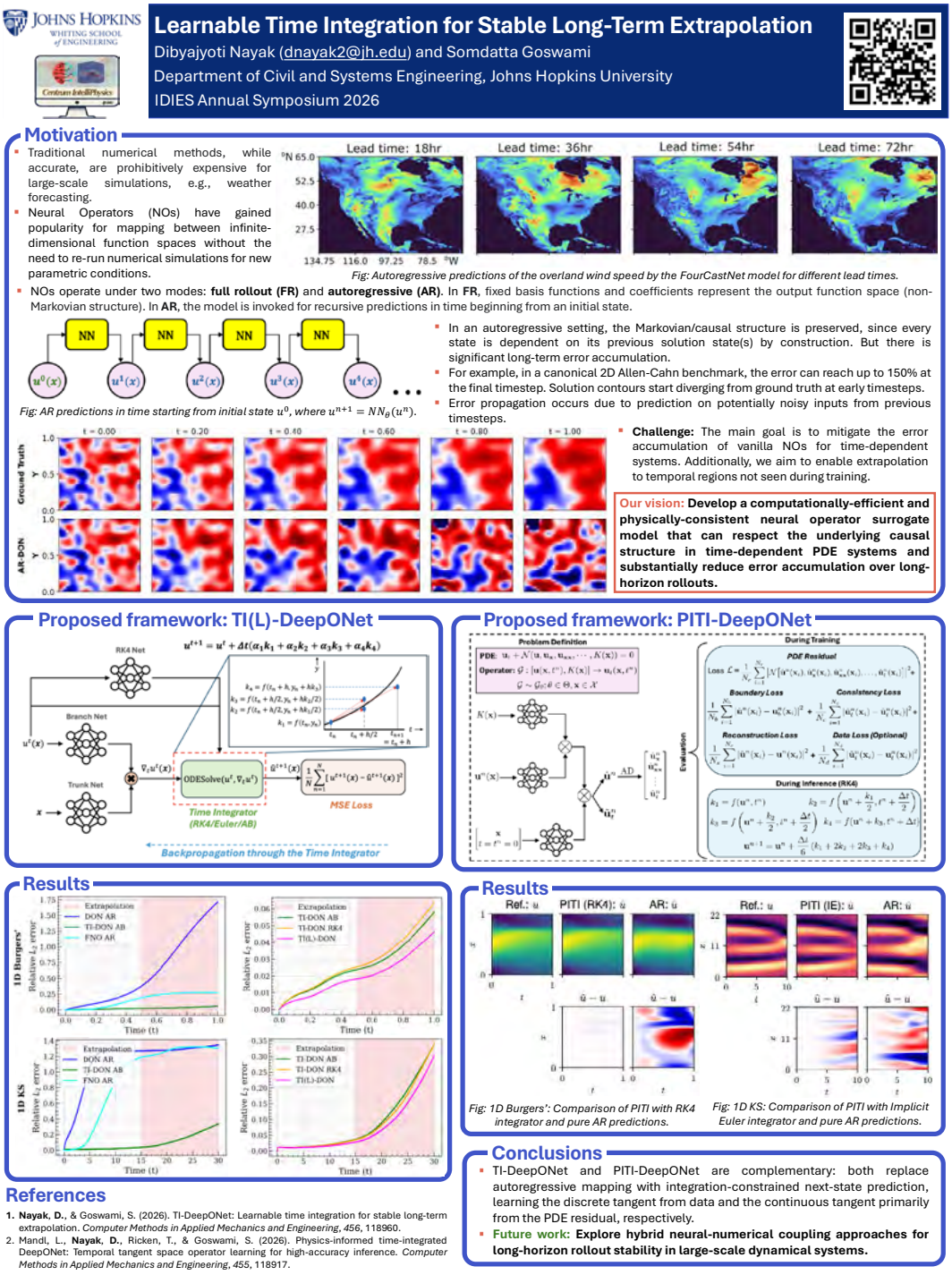
Support: ARPA-E (DE-AR0001787), JHU DSAI, ROSEI, and IDIES.
Computing: ARCH and TACC.

Explore CHAOS

Dibyajyoti Nayak

Learnable Time Integration for Stable Long-Term Extrapolation

Neural operator surrogates for time-dependent PDEs are typically trained either as full rollouts, which ignore temporal causality and generalize poorly beyond the training window, or autoregressively, which preserves causality but accumulates error over long horizons. We present two complementary frameworks that replace autoregressive state mapping with integration-constrained next-state prediction. TI-DeepONet learns the discrete time derivative from data and advances the solution through a classical integrator embedded in the training graph; its extension, TI(L)-DeepONet, makes the intermediate stage coefficients of a multi-stage scheme learnable, adapting the integrator to solution-specific variations. PITI-DeepONet instead recovers the continuous temporal tangent primarily from the PDE residual through a dual-output architecture, and monitors that residual at inference... *(Abstract truncated for space)*



Shivang Parmar

AI-Enabled Smartphone Ophthalmoscope for Automated Retinal Vessel Segmentation, Biomarker Extraction, and Disease Detection

Globally, over 2.2 billion individuals suffer from vision impairment, with more than 1 billion cases remaining preventable or untreated. The key reason behind this inequality is the fact that standard diagnostic imaging is expensive for low-resource and point-of-care settings. Beyond diagnosing eye diseases, retinal imaging also holds great potential for identifying neurological conditions. Specifically, retinal microvascular changes, such as altered blood vessel density and tortuosity, have been proven to serve as early, non-invasive biomarkers for neurodegenerative diseases, including dementia, where up to 75% of cases worldwide remain undiagnosed.

The engineering goal of this project is to design, build, and validate a low-cost, 3D-printed smartphone ophthalmoscope to automate retinal vessel segmentation and disease detection. These captured images are then processed through a multi-task Swin Transformer model with a UNet-style decoder and a separate disease classification model. Together, they segment the retinal vascular network, from which 6 retinal biomarkers are extracted, and classify retinal conditions across 8 disease categories. The prototype delivered an impressive DICE score of 0.86, which surpassed many SOTA models, while displaying an incredible AUC score of 0.92, with a score of 0.97+ for certain diseases.

INTRODUCTION

Eye Disease
- Over 2.2 billion people worldwide have some form of vision impairment, with more than 1 billion cases preventable or untreated.
- Some of these are
  - Diabetic Retinopathy (DR): Leading cause of preventable blindness that affects about 103 million people globally.
  - Pathological Myopia: Affects 1-3% of the general population. Children in low-income settings face higher risks of developing pathologic myopia due to lack of access to early diagnosis.
- 90% of people with unaddressed vision impairment live in low- and middle-income countries

Expensive Machinery
- Advanced retinal imaging systems such as clinical fundus cameras typically cost $30,000 to over $150,000 per device in clinical settings.
- These systems require:
  - Stable electricity and specialized computing hardware
  - Dedicated clinical space
- Because of these requirements, they are rarely available in rural clinics, mobile screening units, or low-resource healthcare systems, limiting early detection of diseases

Neurological Disease
- Neurological disease affect the quality of life for millions of people every year. For instance:
  - Dementia affects more than 55 million people globally, with Alzheimer's disease accounting for 60-70% of cases.
  - Multiple sclerosis (MS) affecting about 2.8 million people worldwide
- As the retina is so close to the brain, microvascular damage in the central nervous system also sometimes cause changes in the eye.
- For example:
  - Researchers have found that Alzheimer's is associated with retinal nerve fiber layer thinning and reduced vessel density
  - MS is strongly associated with thinning of the retinal nerve fiber layer.
- Apart from these 2 diseases, retinal biomarkers are there for countless other condition such as the Parkinson's disease.

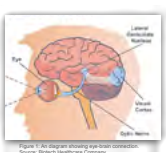


Figure 1: An diagram showing eye-brain connection. Source: Biotech Healthcare Company

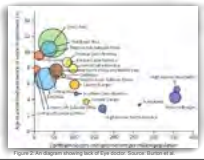


Figure 2: An diagram showing lack of eye doctor. Source: Author etc.

Engineering Goal

To design, build, and validate a **low-cost, 3D-printed smartphone ophthalmoscope headset** (with a headrest) featuring **novel cross-polarization and multispectral illumination**. It will also seamlessly integrate it with a SWIN Transformer for **automated retinal vessel segmentation** in order to calculate 6 **various biometrics** such as vessel density and tortuosity. It will also include a disease detection model that aims to identify most prevalent retinal diseases.

OTHER DESIGNS AND OUR NOVELTY

Passive Smartphone Adapters
- Other attempts at low-cost retinal imaging rely on passive, simple 3D printed clip-on adapters that just use the phones native camera flash. This fundamentally fails in a clinical setting because of how the blinding light induces miosis, requiring patients to use uncomfortable dilation drops. Our novel design creates a cohesive electronics-based system, with LEDs and cross polarization filters working in tandem with the phone via Bluetooth. This solves the critical issue of other designs in clinical setting.

Tabletop clinical cameras
- While tabletop slit-lamp fundus cameras remain the clinical gold standard for capturing high fidelity retinal data, they are massive, immobile machines that cost upwards of $10,000, this makes them inaccessible for point of care screening under rural areas. Our product and design allow for great amounts of modularity and different low-cost printable attachments that create a more portable and accessible experience, whether the patient is healthy or bedridden.

OCT Imaging Systems
- A recent development in ocular imaging that has been OCT scanning systems which allows for a better analysis on the retinal imaging of the brain and can create pathways into identifying neurological diseases. However, the main issue that is overlapping this is the reliance of such OCT scans, a system that is much more expensive and inaccessible than regular fundus imaging, as well as its lack of modularity that our project can provide. Our ML diagnostic is based around analyzing retinal paths and segments which is much more reliable economically long term than OCT scanning.

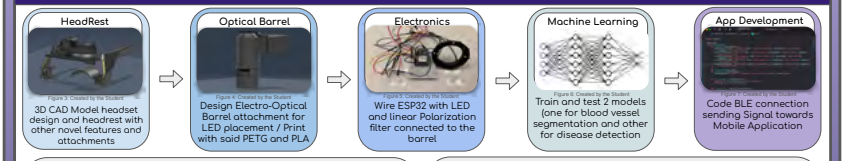
Machine learning structure
- Most of the disease detection models for retina diseases don't cover a wide spectrum of diseases, or fail to use more modern infrastructure. Furthermore, a lot of the models face the "black box" problem, where the users aren't able to see the logic behind the decision. Our model aims to address this issue. Through the use of an ensemble-based approach utilizing GradCAM, our model not only uses advance architecture but also illustrates its thinking behind the decision. The segmentation model, similarly, employs SWIN transformer to improve the DICE score of vessel segmentation and surpass current SOTA models.

Email: sparmar5@jh.edu

AI-Enabled Smartphone Ophthalmoscope for Automated Retinal Vessel Segmentation, Biomarker Extraction, and Disease Detection

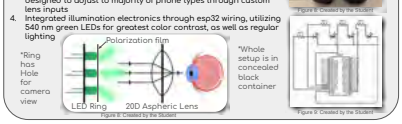
Shivang Parmar[1], Yahir Bernal[2], Yannis M. Paulus, M.D., FACS, FARVO, FASRS[1]
Johns Hopkins University, Baltimore, MD[1] Arizona State University, Tempe, AZ[2]

ENGINEERING METHODOLOGY - PROTOTYPE DESIGN AND PROCEDURE

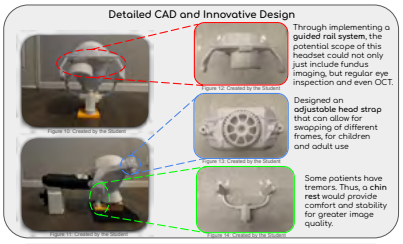


Process and Novel Engineering Imaging Barrel

1. Selected a 20D Aspheric condensing lens based on optical modeling of required field of view and working distance for smartphone couple fundoscopy.
2. Designed a cross-polarization system using two linear polarizers oriented at 90 degrees to each other, one over the illumination LED, an over the camera aperture to eliminate corneal specular reflection.
3. Engineered the barrel housing using Fusion 360 with an Adventurer flashforge 5m printer with PLA and PETG filament designed to adjust to majority of phone types through custom mms inputs
4. Integrated illumination electronics through esp32 wiring, utilizing 540 nm green LEDs for greatest color contrast, as well as regular lighting



Detailed CAD and Innovative Design

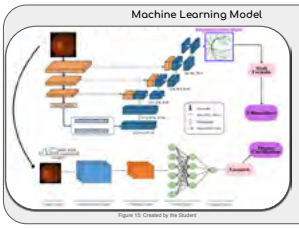


Through implementing a guided rail system, the potential scope of this headset could not only just include fundus imaging, but regular eye inspection and even OCT.

Designed an adjustable head strap that can allow for swapping of different frames, for children and adult use

Some patients have tremors. Thus, a chin rest would provide comfort and stability for greater image quality.

Machine Learning Model



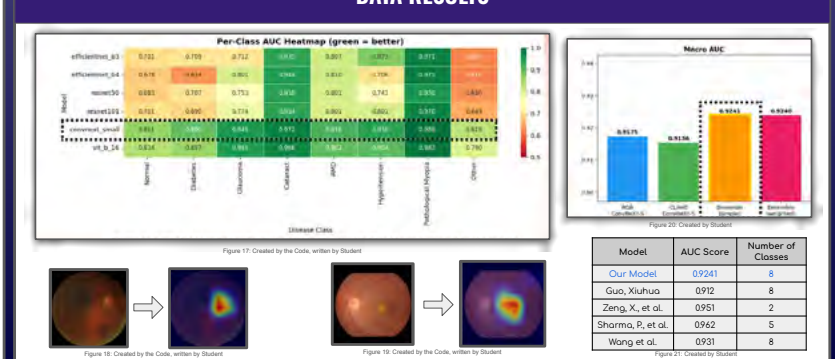
Explains the architecture behind the segmentation of blood vessel and disease detection of various diseases.

Mobile App Development Detail



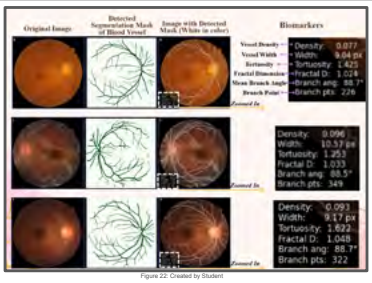
1. Developed a custom iOS application using Swift and Swift UI (soon to be adapted for Android use cases)
2. Integrated camera control for the fundus barrel: automated LED activation, exposure adjustment, and focus lock via bluetooth connection
3. Designed two user interface modes:
   a. Patient mode: simplified interface showing status and plain language guidance,
   b. Ophthalmology AI: A LLM which could be used to address the concerns of the patients in case they have any questions. Note: It is not suppose to substitute a physician's diagnosis. It is simply a tool for asking clarifying questions.

DATA RESULTS



Model	AUC Score	Number of Classes
Our Model	0.9241	8
Guo, Xiuhua	0912	8
Zeng, X., et al.	0951	2
Sharma, P., et al.	0962	5
Wang et al.	0931	8

RESULTS(Continued)



Model	Dice Score
Our Model	0.8675
Guo et al.	0.857
Ronneberger et al.	0.82
Alom et al.	0.823
Li et al.	0.847
Seo et al.	0.8302



DISCUSSION AND FUTURE WORK

Discussion
- The segmentation model's DICE score being greater than other state of the art published model points to how our novel model outperforms current segmentation AI system.
- The inference examples also emphasize the model's success at not only segmentation, but also calculating biomarkers (such as vessel density and mean branch angle)
- The Ensemble ConvNeXt's performance of 092 also illustrates the success of disease detection over a wide spectrum. It has achieved a score which is comparable to state of the art models.
- The physical components of the project successfully met its expectations. The physical prototype of the project had a flexible headrest. Thus, it could accommodate a large range of head sizes. The physical variables that typically ruin smartphone fundus photography were effectively mitigated. The modularity allows for different use cases, whether it is for a bed-ridden patient or someone with tremors. In future, multiple different attachments could be put on to increase ease-of-use or stability.

Why Matters
- Condensing a $30,000+ clinic machine into a sub-$150 smartphone kiosk. For instance, Optos Daytona, a leading retinal imaging system, cost ~$85,000. Thus, our design has price reduction of approximately 99.7%.
- Deliver clinically validated neurological biomarkers to doctors in low-resource environments
- Allows for easy screening for 8 of the most common eye diseases
- Provides an adjustable and intuitive experience with multiple accessibility options and a dynamic medical procedure

Future Work
- Improvements and additions that we could implement into our design include actual real-world validation through human testing / screening
- A means on integrating the rail system with other forms of retinal imaging, such as OCT scanning and slit lamp imaging
- Access to concrete neurodegenerative fundus images to further expand and train our model to better identify such diseases. This step would be possible with the UK BioBank.
- A more adjustable fundus barrel that allows the user to adjust the position of the Lens as well as the position of their phone camera
- An IR alignment system working in tandem with the LED ring and the mobile application to produce better and more centered images

KEY REFERENCES

- Cheung, Carol Y, et al. "Microvascular Abnormalities in the Eye as a Window to the Brain." Nature Reviews Neurology, vol. 13, no. 2, 2017, pp. 91-104.
- O'Bryant, Sid E., et al. "Retinal Microvascular and Structural Changes in Alzheimer's Disease." Alzheimer's & Dementia, various meta-analyses and review publications.
- Resnikoff, Serge, et al. "Estimated Number of Ophthalmologists Worldwide (International Council of Ophthalmology Update): Will We Meet the Needs?" British Journal of Ophthalmology, vol. 96, no. 6, 2012, pp. 783-787. Updated global workforce data, 2020.

Julian Posey

Emulating the Roman Grism PSF Across the Focal Plane for Slitless Flux-Cube Restoration

The Nancy Grace Roman Space Telescope will take slitless grism spectra over ~2,400 deg² as part of the High-Latitude Wide-Area Survey, yielding ~19 million galaxy redshifts. Analysis at that scale needs catalogue-free restoration methods, especially as neighboring traces overlap and can only be disentangled by combining several roll angles. To this aim, we are extending ImageMM, a multiplicative majorization–minimization image-restoration and super-resolution method, to produce latent image cubes (x, y, □) from multi-roll grism data.

As Roman's grism disperses each source into a ~900-pixel trace, one pixel per ~11 Å of the 1.0–1.93 μm band, the forward model will need the point-spread function in 846 wavelength bins at every source position. Given that the PSFs take ~1.75 s to calculate each with STScI's stpsf package and change by up to a third of their flux across a detector, we compared bilinear interpolation and PCA as emulators of the PSF across the focal plane. Training on 6,016 stpsf PSFs and testing at held-out positions and wavelengths on Roman's most variable detector, we found that a single 500-image linear basis reproduces stpsf to 0.28% of the flux (1.4% worst case), outperforming per-bin bilinear interpolation, which needs 27× more convolutions and misplaces 0.76%.

Emulating the Roman grism PSF across the focal plane

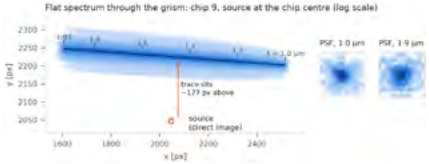
Using multi-frame deconvolution (ImageMM) for slitless flux-cube restoration



Julian Posey[1], Arik Mitschang[4], Tamás Budavári[1,2,3]
tposey2@jh.edu
[1]Department of Applied Mathematics and Statistics, Johns Hopkins University
[2]Department of Physics and Astronomy, Johns Hopkins University
[3]Department of Computer Science, Johns Hopkins University
[4]Data Science and AI Institute (DSAI), Johns Hopkins University

Roman's slitless spectra

Roman's High-Latitude Wide-Area Survey will take slitless grism spectra over ~2,400 deg², for ~19 million galaxy redshifts. The grism disperses *every* source into a ~900 px trace (1.0–1.93 μm, one pixel per ~11 Å), each wavelength with its own PSF.



Trace overlaps are disentangled by combining roll angles. **Goal: one non-negative flux cube $x(i, j, k)$ (sky pixel i, j; wavelength bin k) that explains every exposure,** with no source catalog or templates.

ImageMM for a grism

ImageMM (Sukurdeep, Budavári et al. 2025) restores one latent image from many exposures by majorization–minimization. For the grism, exposure t is

$$y^{(t)} = \sum_k f_{s,k} * \left(P_k^{(t)} x_k\right) + \eta^{(t)} = A^{(t)} x + \eta^{(t)}$$

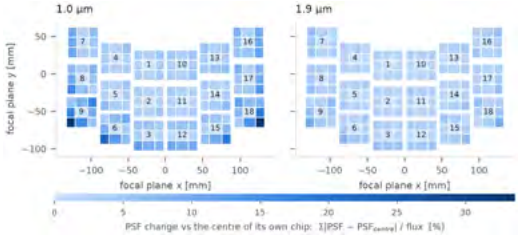
$$x_n = x_{n-1} \odot u_n, \qquad u_n = \frac{\sum_t A^{(t)\top} W^{(t)} y^{(t)}}{\sum_t A^{(t)\top} W^{(t)} A^{(t)} x_{n-1}}$$

$P_k^{(t)}$: rotates sky onto roll t and moves bin k to grism trace
$f_{s,k}$: bin k's PSF at source position s
$W^{(t)}$: inverse variances (and masks)
u_n: clipped to $[1/\kappa, \kappa]$, $\kappa = 2$ (ImageMM Eq. 9)

- **Matrix-free**: only A and $A^\top$ are applied; the cost is the PSF convolutions in A
- **Non-negative**: the update keeps $x \geq 0$ and lowers the loss every step if $A \geq 0$: non-negative PSFs

The bottleneck: the PSF

- The PSF $f_{s,k}$ changes with source position s and each of the 846 bins k (11 Å)
- STScI's **stpsf** (formerly WebbPSF) computes each from Roman's optical model in ~1.75 s: 20+ min per position
- One PSF per chip is not enough: up to 34% of the light misplaced at 1.0 μm



The way out: build every PSF from a few fixed images, so ImageMM needs a handful of convolutions per exposure instead of one per bin and source (next column).

A shared PSF basis

If every PSF is a fixed set of R images with weights that vary smoothly,

$$f_{s,k} = \sum_{r=1}^{R} a_r(s,k)\, b_r$$

then

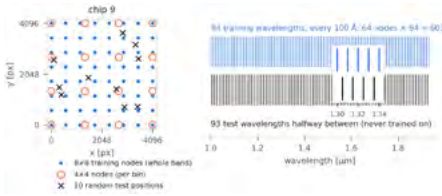
$$A^{(t)} x = \sum_{r=1}^{R} b_r * \left[\sum_k a_r(s,k)\, P_k^{(t)} x_k\right],$$

the sum over wavelengths (and sources) folds into a cheap placement step: **R convolutions per exposure for the whole band,** not one per bin and source.

Three candidate models

- **Bilinear** interpolation between PSFs on a 4×4 grid, per bin: non-negative, but 16 nodes × 846 bins = 13,536 images, and stpsf at every bin
- **Tangent PCA**, built for ImageMM's non-negativity: $\varphi = \sqrt{f}$ is a point on the unit sphere in L^2 ($\int \varphi^2 = \int f = 1$); PCA in its tangent space, then square back. Non-negative with unit flux by construction, but the square is quadratic in the images: $\sim K^2/2$ of them
- **Linear PCA**: one basis for the chip and the whole band; weights polynomial in position, spline in wavelength. Fewest images, but can dip below zero

Trained on 6016 stpsf PSFs (8×8 nodes, every 100 Å) on chip 9, Roman's most variable detector. **Tested where the model never saw the PSF**: 10 random positions × 187 wavelengths, half between training wavelengths.



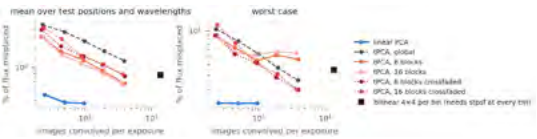
One wavelength bin at a time

Model	Images	1.03 μm	1.88 μm
chip-center PSF	1	22.1 / 50.9	5.77 / 13.2
bilinear 4×4	4	1.58 / 3.33	0.32 / 0.74
tangent PCA, K=5	21	**1.17 / 1.86**	**0.28 / 0.46**

a model per bin, trained at the bin's own wavelength; % of flux misplaced, mean / worst over 10 random positions

Per bin, tangent PCA is the most accurate: half of bilinear's worst case, and non-negative. But it still needs stpsf and its own images at every bin.

The whole band: one basis



With one basis for the whole band, tangent PCA falls behind: linear PCA with 500 images misplaces 0.28% of the flux on average (1.39% at worst), better than every tangent-PCA variant at 4000 images and per-bin bilinear at 27× the convolutions.

Why linear PCA can go negative



- On average 0.8% of pixels dip below zero, but they carry only **0.001% of the flux** (0.03% at worst), in the far wings
- Clipping them and renormalizing even improves the worst case (1.39% → 1.37%)
- In ImageMM on mock Roman frames, clipped, unclipped and the exact stpsf PSFs give **the same restored cube to 3 decimals**

So plain PCA looks safe for ImageMM, despite the non-negativity assumption.

Summary

Model	Images	stpsf	% misplaced
chip-center PSF	846	no	up to 51
bilinear, per bin	13,536	per bin	0.76 / 3.44
tangent PCA, K=87	3916	no	1.28 / 2.60
linear PCA	**501**	**no**	**0.28 / 1.39**

images convolved per exposure for the whole band; stpsf calls needed; mean / worst % of flux misplaced at positions and wavelengths not used in training

Linear PCA's basis alone is nearly exact (0.04%): **the error is in predicting the weights between nodes,** worst in the blue, where the PSF changes fastest.

Next: restoring a mock field



Mock chip-9 frames: five point sources, each dispersed with its own stpsf PSF along its own trace, at the four Data Challenge 3 roll angles.

The sky rotates between rolls while the traces stay along the detector's x axis: **traces overlap in one roll (A and B at 150°) and run off the chip in another (C at 0°).**

Further work

- Explore the implications of negative PCA components: look for edge cases (e.g. off-chip extrapolation) where they could break ImageMM's non-negativity
- Add more roll angles, regularization and error bars; then field-test with Roman's Data Challenge 3 (ESpRESSO) and the HLSS simulations

References

A collaboration with the JHU Data Science and AI Institute (DSAI), IDIES and NVIDIA.
Sukurdeep, Budavári, Connolly & Navarro 2025, arXiv:2501.03002 (ImageMM)
Perrin et al. 2014, Proc. SPIE 9143, 91433X (WebbPSF / stpsf)
Gabrielpillai et al. 2024, arXiv:2412.08883 (ESpRESSO, Data Challenge 3)
Wang et al. 2022, ApJ 928, 1 (HLSS simulations)

Dibakar Roy Sarkar

CINOC: Cardinality-Invariant Neural Operator Policies for Scalable PDE Control

Controlling heat transfer, fluid motion, and chemical reactions often requires networks of spatially distributed actuators. However, changing the number of actuators can force learned controllers to be redesigned and retrained. We introduce Cardinality-Invariant Neural Operator Control (CINOC), which learns a shared control policy that operates across different actuator populations. Each agent converts local field observations into actions, coordinating through the physical environment without direct communication. Policies are trained by differentiating through numerical solvers of partial differential equations. We demonstrate tracking of desired field patterns, chaos suppression, and density transport across linear, nonlinear, chaotic, and turbulent systems. In a two-dimensional density-transport task, a policy trained with 16 agents operates with 1,024 agents without retraining, a 64-fold increase, while achieving approximately 16% lower loss in matching the target density than deployment with 16 agents. A complementary mean-field analysis links learning in finite populations to a continuous-control model, while experiments reveal adjustment of individual control effort as populations change. These results offer a route to expanding or reconfiguring actuator networks without repeating expensive controller training, with potential applications in thermal management and distributed flow control.

Johns Hopkins Whiting School of Engineering

Energy at Hopkins

CINOC: Cardinality-Invariant Neural Operator Policies for Scalable PDE Control

Pietro Zanotta, Dibakar Roy Sarkar, Honghui Zheng, Somdatta Goswami, Ján Drgoňa
Civil and System Engineering, Johns Hopkins University

idies

Introduction

The Challenge: Controlling partial differential equations (PDEs) is critical, but existing learning-based multi-agent policies generally rely on fixed-dimensional representations.

Our Solution: We recast PDE policy synthesis as an operator learning problem, mapping state fields to continuous control functions.

The Result: End-to-end training via differentiable physics yields policies with cardinality invariance, allowing zero-shot transfer to significantly larger agent populations.

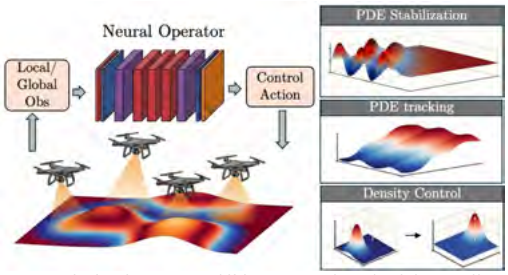


Figure 1: A shared neural operator maps PDE field observations to control actions for multiple agents, enabling zero-shot scalability across agent populations. Framework handles diverse tasks: (top right) chaos stabilization, (middle right) reference tracking, and (bottom right) density transport.

Contributions

Neural Operator Framework: Control policies parameterized as neural operators yield configuration-invariant behavior across variable sensing/actuation/population setups.

Differentiable-Physics Optimization: Policies trained end-to-end by differentiating through PDE solvers, providing scalable and efficient policy gradients.

Cardinality-Invariant Multi-Agent Control: An emergent self-normalization phenomenon enables consistent multi-agent behavior independent of agent count.

Mean-Field Grounding: Finite-population policy gradient learning is formally linked to a mean-field limit, theoretically establishing cardinality invariance.

Diverse PDE Benchmarks: Demonstrated across stabilization, tracking, and density transport on nonlinear, diffusion, and chaotic spatiotemporal dynamics.

Problem formulation and Methodology

$$\min_{u^M, v^M} \mathcal{J} = \mathbb{E}_{\psi\sim\mathcal{D}_{\text{prob}}}\left[\int_0^T \ell\big(z(\boldsymbol{x},t), u^M(t), v^M(t), \psi\big)dt\right]$$

$$\text{s.t.}\frac{\partial z}{\partial t} = \mathcal{F}\left(z(\boldsymbol{x},t);\psi\right) + \sum_{i=1}^{M}\mathcal{B}\big(\boldsymbol{x}, u_i(t), \boldsymbol{\xi}_i(t)\big),$$

$$\dot{\boldsymbol{\xi}}_i(t) = v_i(t), \quad i = 1,\dots,M,$$

$$\boldsymbol{c}\big(z(\boldsymbol{x},t), u^M(t), v^M(t), \boldsymbol{\xi}^M(t), \psi\big) \le 0.$$

This formulation defines the objective for our framework:

- ***Orange (Continuous Control Policy)**: the variables we are actively optimizing. The neural operator maps local observations to the control functions u^M and velocities v^M.*
- ***Purple (System Dynamics)**: represents the PDE state field (z).*
- ***Teal (Decentralized Actuation)**:. The spatial coordinates (ξ) define the position of each agent, whose effect on the field is given by $\mathcal{B}$.*
- ***Gray (Distributional Robustness)**: we evaluate the expected cost over a distribution of parameters (ψ, e.g. initial and reference shapes).*

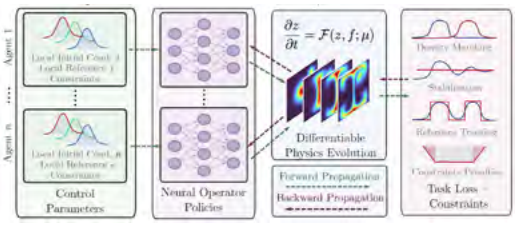


Figure 2: Schematic of the multi-agent operator learning for PDE control. The PDE solver acts as a differentiable layer. The operator policy (shared among the agents) takes error observations and outputs control actions. Gradients propagate through the solver to update

Theoretical Analysis

Mean-Field Gradient Consistency : As $M \to \infty$, discrete policy gradients $\nabla_\theta J_M$ converge to the continuous mean-field limit $\mathcal{D}_\theta J_\infty$ (Thm. 1), guaranteeing the learned operator captures true PDE dynamics rather than swarm-size artifacts.

Emergent Self-Normalization: Effort regularization implicitly drives each agent toward $u_i \approx \mathcal{O}(1/M)$, keeping aggregate forcing bounded without any explicit communication of swarm size M, directly producing cardinality-invariant behavior.

Results

PDE	Train	Deploy	Scale	Loss
FKPP 1D	30	20–150	5×	1.29×
Heat 1D	30	15–120	4×	2.08×
Heat 2D	16	16–256	16×	0.54×
KS 2D	196	144–256	1.3×	2.30×
KS 1D	30	15–90	3×	1.14×
Turbulence 2D	64	36–1,024	16×	0.76×
Density 2D	16	4–1,024	64×	0.84×

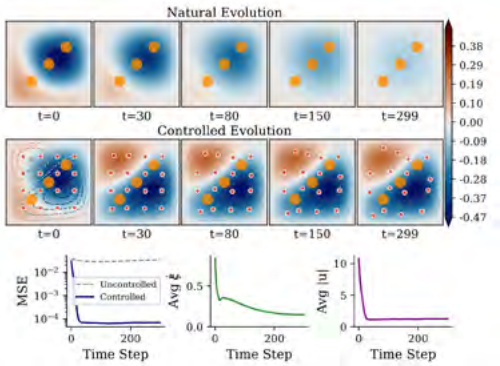


Figure 4: Reference tracking of 2D Heat equation. Top: Natural (uncontrolled) evolution of $tu(x,t)$ showing diffusion. Middle: Controlled evolution of the thermal field $u(x,t)$ where agents (red dots) regulate the temperature field toward dashed reference contours while avoiding static obstacles (orange circles). Bottom: Tracking MSE, average agent velocity (ξ) and average agent control intensity ($|u|$).

Conclusion

CINOC reformulates PDE control as operator learning, achieving zero-shot cardinality invariance across agent populations. Emergent self-normalization, grounded by mean-field gradient consistency, enables bounded aggregate forcing without explicit swarm-size communication. Future work targets parametric operators conditioned on physical coefficients and scaling to high-dimensional industrial environments.

Read the paper

Supported by U.S. DOE Office of Science ASCR (SciDAC LEADS) and ROSEI at JHU. Computing: JHU ARCH & Rockfish cluster (NSF OAC1920103). DRS and SG supported by NSF Grant 2436738.

View the code

Atharva Sathe

JHTDB-Wind: The Johns Hopkins Wind Farm Flow Database Providing Open Access to High-Fidelity Simulation Data

As offshore wind farms grow in size and complexity, understanding how turbines interact with one another and the surrounding atmosphere becomes increasingly important. High-fidelity computer simulations can reveal these interactions in remarkable detail, but their enormous computational demands have traditionally limited their use for large wind farms.

We have developed highly accurate analytical corrections to a widely used turbine model that fundamentally change this trade-off. Our approach captures the behavior of individual turbine blades with unprecedented precision, achieving agreement within 1% of reference results even on relatively coarse computational grids. By making predictions nearly independent of grid resolution, these advances enable accurate, large-scale wind farm simulations at a fraction of the computational cost.

Building on this breakthrough, we will simulate large offshore wind farms featuring modern 15-megawatt turbines. The resulting datasets will be made publicly accessible through JHTDB-Wind, the Johns Hopkins Wind Farm Flow Database, which already hosts two onshore datasets. Researchers worldwide will be able to explore these massive simulations without downloading terabytes of data, establishing a first-of-its-kind open resource for understanding offshore wind farm dynamics.

JHTDB-Wind: The Johns Hopkins Wind Farm Flow Database Providing Open Access to High-Fidelity Simulation Data

Atharva Sunil Sathe, Dennice F. Gayme, Charles Meneveau
The Ralph O'Connor Sustainable Energy Institute (ROSEI) | Johns Hopkins University

Introduction

As wind farms grow, their performance increasingly depends on interactions among turbine wakes and the atmosphere. Large-eddy simulations (LES) capture these interactions in detail but require specialized expertise and substantial computational resources. Each simulation also generates tens to hundreds of terabytes of data, making storage and sharing challenging.

Objective

JHTDB-wind opens these datasets to the whole community, with efficient public access and tools to analyze them without downloading terabytes of data. This poster introduces the first two datasets, new corrections that keep the actuator line model accurate on coarse grids, and plans for offshore wind farms.

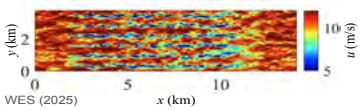
Simulation Approach

All datasets are generated using **JHU-LESGO**, a parallel LES solver employing a Cartesian grid, pseudo-spectral methods in the horizontal, second-order finite differences in the vertical, a Lagrangian scale-dependent dynamic model, an equilibrium wall model, and a filtered actuator-line model.

Wind Farm Datasets

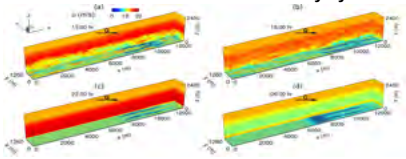
JHTDB-wind currently hosts two onshore wind farm datasets, both simulated with NREL 5 MW turbines.

60-turbine wind farm in a neutral boundary layer



Zhu et al., WES (2025)

8-turbine wind farm over a full daily cycle



Xiao et al., JRSE (2025)

Filtered Corrections for ALM

On coarse grids, the wide Gaussian kernel needed for numerical stability smears the blade forces and over-diffuses the trailed vorticity, so the ALM overpredicts loads and power. The FLL and FAM corrections restore the induced velocity that excessive smearing distorts, leading to accurate blade and turbine level data.

Generalized filtered lifting-line (FLL) correction

$$u_i'(z) = \frac{1}{2\pi}\int_0^S \frac{1}{V_{\text{rel}}(z')}\frac{G(z')}{\epsilon(z')^2}\left[e^{-(z'-z)^2/\epsilon(z')^2} + \frac{\epsilon(z')^2}{2(z'-z)^2}\left(e^{-(z'-z)^2/\epsilon(z')^2}-1\right)\right]\mathrm{d}z'$$

Martínez-Tossas et al., Wind Energy (2024)

Filtered annulus-momentum (FAM) correction

$$\Delta u_{\text{FAM}} = \underbrace{\underset{m}{\text{mean}}}_{\text{projection}} + \Delta u' - \frac{\langle \Delta u', \Delta u_{\text{FLL}}' \rangle}{\langle \Delta u_{\text{FLL}}, \Delta u_{\text{FLL}}' \rangle}\Delta u_{\text{FLL}}'$$

Sathe et al., in preparation

Highlights

- **Power within 1%** of BEM reference for all four turbines (Fig. 1).
- **Radial load distributions** are equally accurate (Fig. 2).
- **No parameter tuning** required for any turbine or grid.
- **Accurate on very coarse grids,** down to 3 LES grid points per blade.
- **Negligible cost,** under 0.5% of the solver time per step.

Results

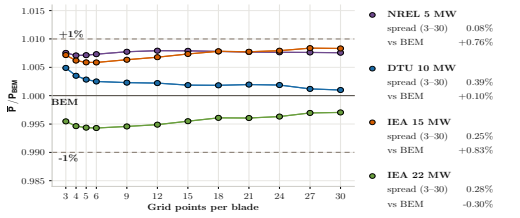


Figure 1. Rotor power relative to blade-element momentum (BEM) in uniform inflow. With both corrections, it varies by less than 0.4% from 3 to 30 grid points per blade, compared with more than 15% in uncorrected simulations.

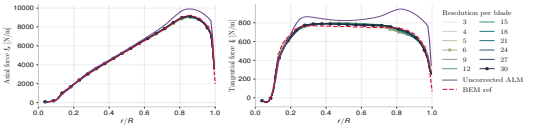
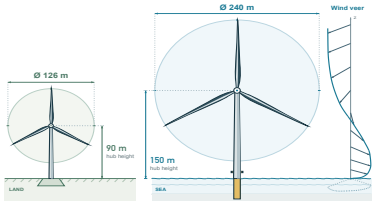


Figure 2. Axial and tangential force along the IEA 15 MW blade, corrected at every resolution, with the uncorrected ALM and the BEM reference.

Future Work

- **Next, offshore wind farms** will be simulated with the newly implemented IEA 15 MW reference turbine (right).
- **Offshore turbines are much larger** and experience strong wind veer across their rotors, so their physics differ from onshore turbines.
- **The very smooth sea surface** alters the incoming turbulence and thus wake meandering and recovery. High-fidelity simulations will be used to study these effects.



Acknowledgement: We thank the IDIES team for supporting JHTDB. We also gratefully acknowledge support from the Ralph O'Connor Sustainable Energy Institute and the Academic Center for Reliability and Resilience of Offshore Wind (ARROW). ARROW is funded by the U.S. Department of Energy (WETO DE-EE0011269), the Massachusetts Clean Energy Center and the Maryland Energy Administration.

Jake Singh

A Differentiable Bayesian Framework for Celestial Navigation with Star and Satellite Imagery

The proliferation of satellites in low Earth orbit enables a new form of celestial navigation: imaging satellites against the star field at known times and comparing the observations with reference catalogs and ephemerides. The method offers GNSS-independent, horizon-free position and attitude determination using low-cost CMOS cameras and embedded computers. It struggles, however, when prior position uncertainty is large: satellite identification becomes ambiguous, and a single misidentified star or satellite can badly corrupt the solution. An optimal estimator requires uncertainty quantification of all error sources, which to date has not been published.

To this end, we present the foundations of a Bayesian framework for statistically optimal celestial navigation with artificial satellites. We first address stationary Earth-surface positioning: estimating an observer's latitude, longitude, altitude, camera pointing, and clock error from one or multiple timestamped images taken at a single site. A forward model maps a prior on Earth-fixed position through the IAU 2006/2000A CIO-based transformation to the Geocentric Celestial Reference Frame, then to predicted star and satellite image coordinates in a wide-field camera. The framework is written end to end in JAX, so every component is auto-differentiable and compatible with gradient-based posterior sampling using the No-U-Turn Sampler in NumPyro.

A Differentiable Bayesian Framework for Celestial Navigation with Star and Satellite Imagery

Jake Singh[1] and Tamás Budavári[2]

[1]Doctor of Engineering Program, Whiting School of Engineering, Johns Hopkins University
[2]Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University

Introduction

The proliferation of satellites in low Earth orbit enables a new form of celestial navigation: imaging satellites against the star field and comparing the observations with reference ephemerides. The method offers GNSS-independent, horizon-free position and attitude determination using low-cost CMOS cameras and embedded computers. It struggles, however, when prior position uncertainty is large: satellite identification becomes ambiguous, and a single misidentified star or satellite can badly corrupt the solution. An optimal estimator requires uncertainty quantification of all error sources, which to date has not been published in prior art. This work is the first to apply Bayesian inference to the problem.

Objectives

1. Quantify uncertainty of all celestial navigation error sources
2. Develop forward model to predict star/satellite measurements
3. Solve the posterior using Bayesian inference
4. Develop reliable measurements for navigation filters

Materials and Methods

Python & uv
- Interpreted language with huge ecosystem
- Extremely fast package/project manger written in Rust

Numpyro & JAX
- Probabilistic programming package for inference
- JAX provides automatic differentiation for Hamiltonian Monte Carlo

Hipparcos & Space-Track.org Catalogs
- Hipparcos for star data
- Space-Track.org provides space object ephemerides

Field Experiments
- Nighttime data collection at known locations
- Camera, tripod, and laptop
- Previous experiments conducted with FLIR Blackfly

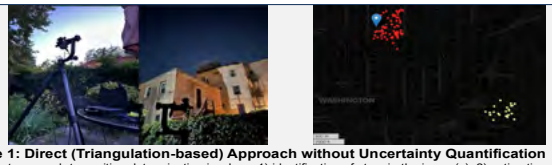


Figure 1: Direct (Triangulation-based) Approach without Uncertainty Quantification
The direct approach to position determination involves: 1) identification of stars in the image(s), 2) estimation of attitude from the known stars, 3) identification of satellites in the image(s), and 4) triangulation from the known satellites. Without uncertainty quantification of satellite ephemeris, image timestamp, and other errors, multimodal and correlated positioning errors arise as shown on the right. The experiment (left) was conducted in Washington, D.C.

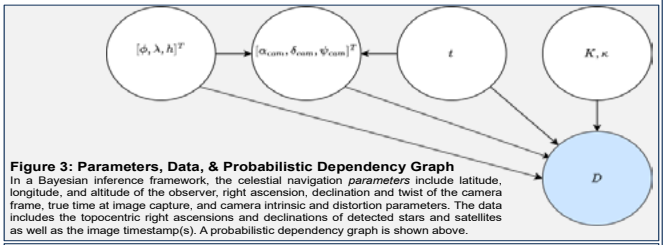


Figure 3: Parameters, Data, & Probabilistic Dependency Graph
In a Bayesian inference framework, the celestial navigation *parameters* include latitude, longitude, and altitude of the observer, right ascension, declination and twist of the camera frame, true time at image capture, and camera intrinsic and distortion parameters. The data includes the topocentric right ascensions and declinations of detected stars and satellites as well as the image timestamp(s). A probabilistic dependency graph is shown above.

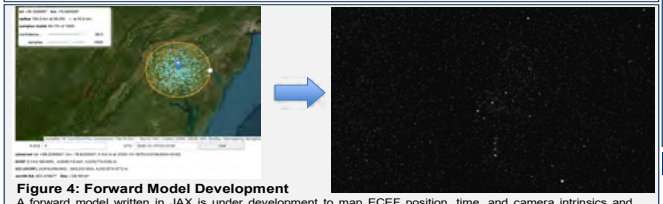


Figure 4: Forward Model Development
A forward model written in JAX is under development to map ECEF position, time, and camera intrinsics and distortion to predicted camera attitude and observed measurements of stars and satellites. The full model will implement an IAU 2000/2006 reduction, SGP4 satellite propagation, atmospheric refraction, stellar aberration, and light-time delay. The user can specify a position prior using a map point, geodesic distance and confidence level.

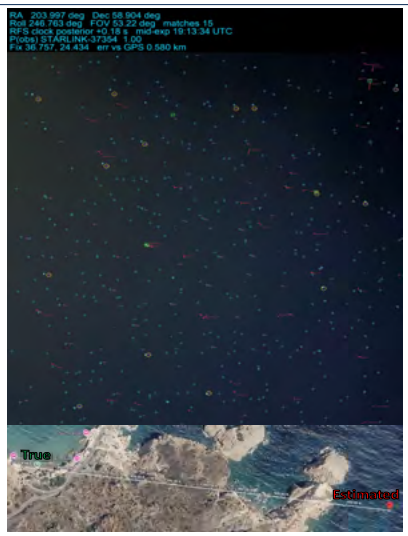


Figure 2: Direct Approach on an iPhone Image
This top image was taken with an iPhone 15 on the Greek island of Milos. The same direct approach from Figure 1 was used, except the timestamp error was selected to minimize true error. The best-fit yielded 580m error.

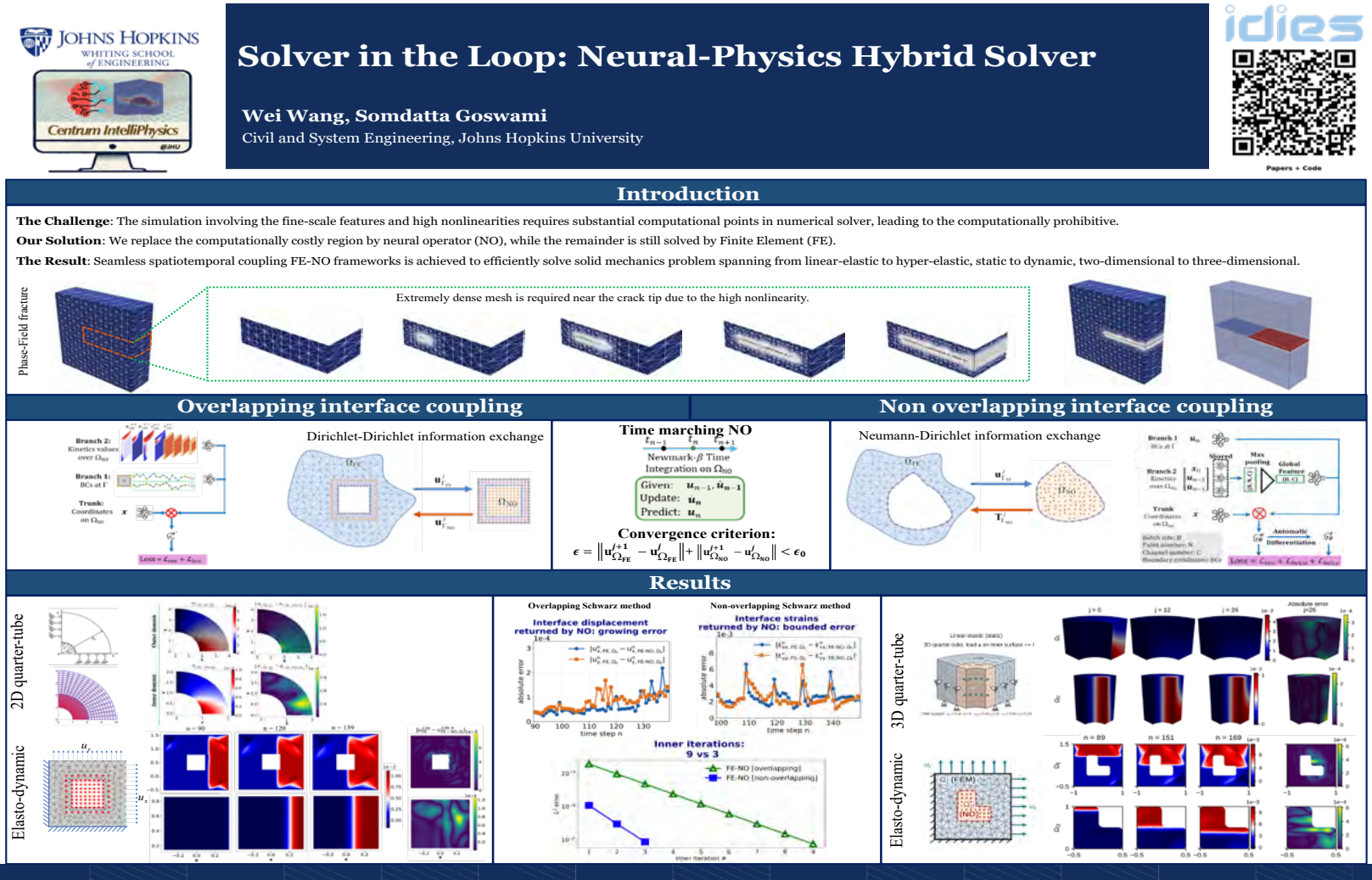
Upcoming Work

- Finish development of forward model
- Implement and test Hamiltonian Monte Carlo on real datasets
- Uncertainty quantification of TLE-based satellite propagation
- Celestial-inertial navigation with Kalman filtering

Wei Wang

Solver in the Loop: Neural-Physics Hybrid Solver

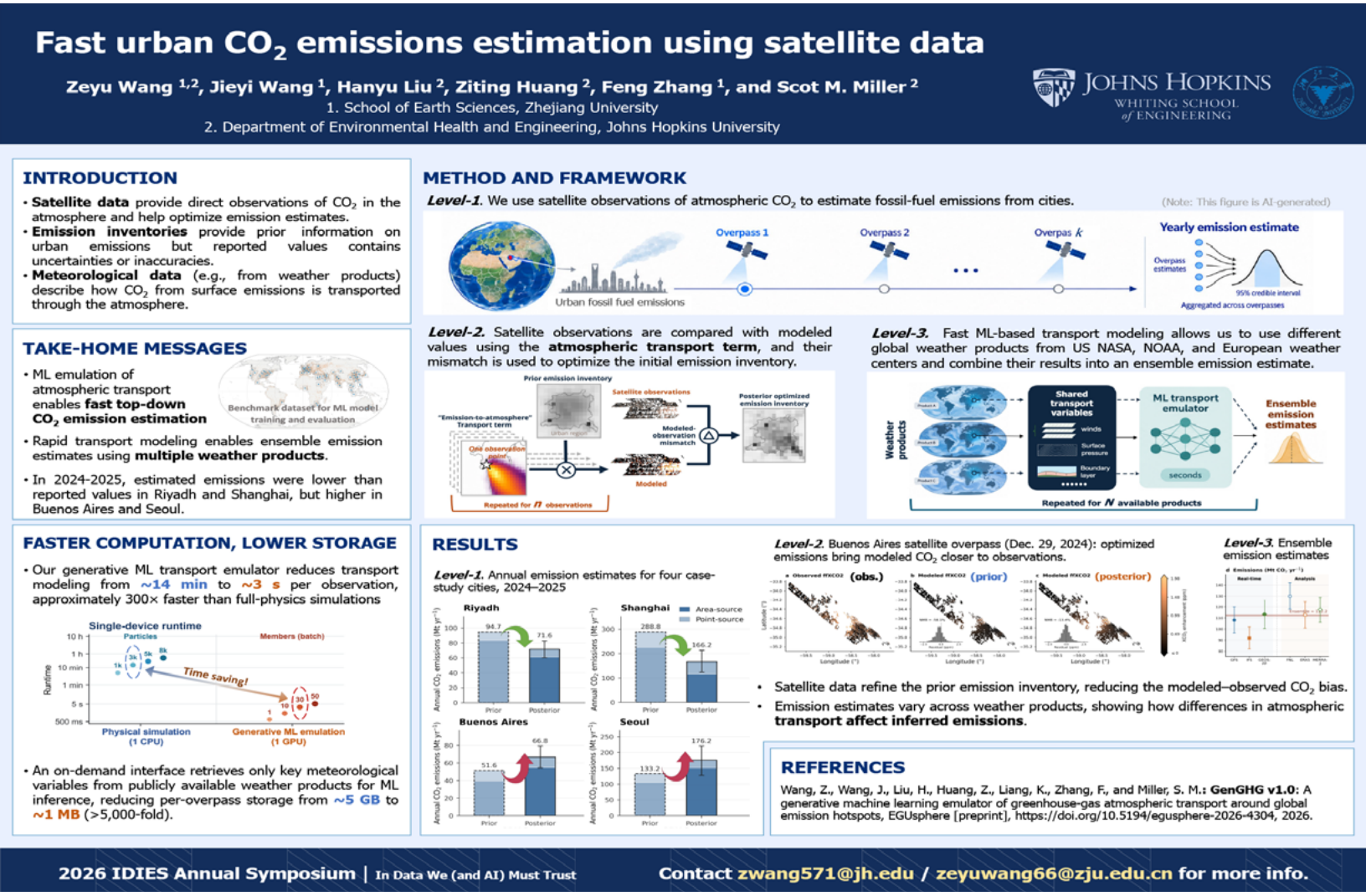
Finite element (FE) methods are the established tool for computational solid mechanics, but their cost grows substantially for problems with strong nonlinearities and fine-scale features. Neural operators (NOs) offer rapid inference, yet they require large volumes of training data, accumulate error in autoregressive prediction, and lose accuracy on large-scale problems. This talk presents a hybrid FE–NO solver based on Schwarz domain decomposition, in which a physics-informed DeepONet, pretrained on governing-equation residuals rather than labeled data, resolves the computationally intensive subdomain while FE handles the remainder. Overlapping and non-overlapping Schwarz schemes are compared on static and dynamic linear elasticity and quasi-static hyperelasticity. Embedding Newmark time integration in the NO mitigates error accumulation and synchronizes time stepping across subdomains. Neumann–Dirichlet interface exchange eliminates the overlap region and reduces inner Schwarz iterations from nine to three per time step. A Point-DeepONet operates directly on unstructured FE point clouds, admitting irregular and three-dimensional subdomains, while strain and stress are derived analytically from the learned displacement, reducing trained networks from five to two. Maximum relative displacement error remains below 2.5%, and for a three-dimensional quarter-tube with over one million degrees of freedom, the hybrid solver is 5.7× faster than FE–FE coupling.



Zeyu Wang

Fast Urban CO2 Emissions Estimation Using Satellite Data

Cities need independent estimates of carbon dioxide (CO2) emissions to evaluate reported values and track mitigation progress. Satellite measurements provide information about CO2 in the atmosphere, but translating these observations into emissions requires simulating how winds and weather conditions move CO2 through the air. These simulations are computationally expensive and remain a source of uncertainty. We developed a workflow that uses a machine-learning model to rapidly predict how CO2 moves through the atmosphere for each satellite snapshot, update an initial emission estimate, and combine estimates across snapshots into annual city totals. The workflow reduced calculation time from about 14 minutes to 3 seconds per observation and lowered weather-data storage from about 5 gigabytes to 1 megabyte per satellite snapshot.



Yongmeng Wu

Plug-in Neural Subdomains: FE–NO Coupling in 3D

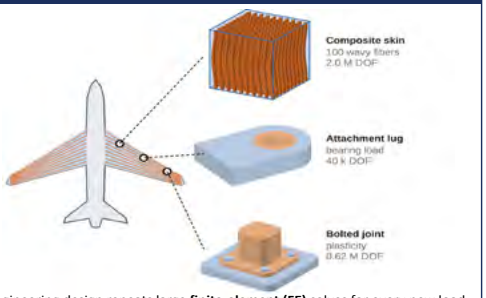
Engineering design repeats large finite-element (FE) solves for every new load. Neural operators (NOs) are fast but accuracy has no guarantees. We couple an NO subdomain to FE through Schwarz domain decomposition in JAX: the NO predicts stress from the interface displacement, and FE solves the rest. Every coupled answer is checked against monolithic FE with the relative L2 error. The coupled FE field reaches 9.2 × 10^{-4} error on a 3D lug and 5.7 × 10^{-3} on a 100-fiber composite with 2.0 M unknowns, where three accelerators agree on one fixed-point and IQN-ILS converges about 4× faster than constant relaxation. On an elastoplastic bolted joint the NO tracks five load cycles beyond its training set at 4.3 × 10^{-3} error.

Plug-in Neural Subdomains: FE–NO Coupling in 3D

A neural operator replaces a finite-element subdomain inside a Schwarz loop.

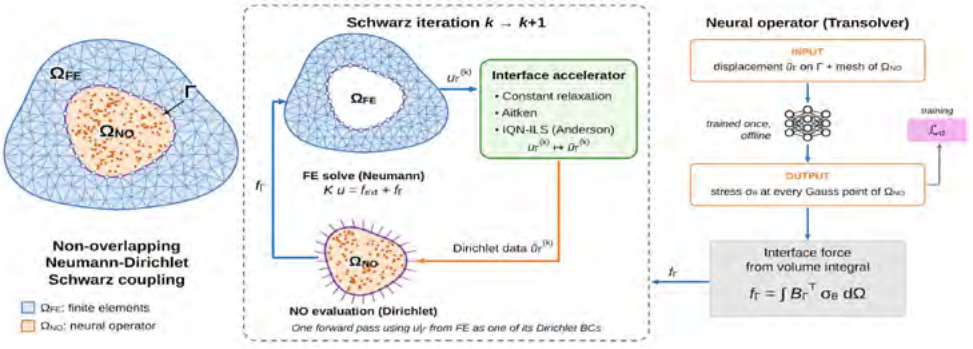
Yongmeng Wu (ywu383@jhu.edu), **Yanghao Chen, Wei Wang, Somdatta Goswami**
Department of Civil and Systems Engineering, Johns Hopkins University
IDIES Annual Symposium 2026

1 Motivation



Engineering design repeats large **finite-element (FE)** solves for every new load. **Neural operators (NOs)** are fast but carry no guarantees.
We couple an NO subdomain to FE through **non-overlapping Schwarz domain decomposition** in **JAX**. A hybrid solver can thus keep FE's checkability while delegating repeated subdomains to a learned operator.

2 Methodology: a neural operator in the Schwarz loop



3 Interface accelerators

fixed step → adaptive step → learned correction

Constant
$u^{k+1} = u^k + \omega\, r^k$
One fixed step in every Schwarz iteration.

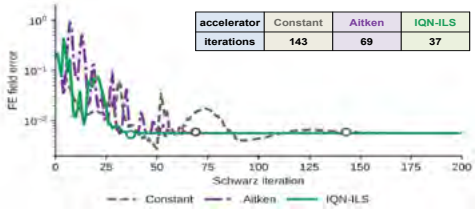
Aitken
$\omega_k = -\omega_{k-1}\,\frac{(r^{k-1}, \Delta r)}{\|\Delta r\|^2}$, $\Delta r = r^k - r^{k-1}$
Fit the step size to how the last two residuals changed.

IQN-ILS
$u^{k+1} = u^k + \omega\, r^k + \sum_i \alpha_i\,(\Delta u_i + \omega\,\Delta r_i)$, $\alpha = \arg\min_\alpha \left\| r^k + \sum_i \alpha_i\,\Delta r_i \right\|$
Combine past corrections in whatever mix best cancels the current residual.

r: residual (change of the interface displacement over one Schwarz iteration)
ω: relaxation factor ω_k :updated every iteration

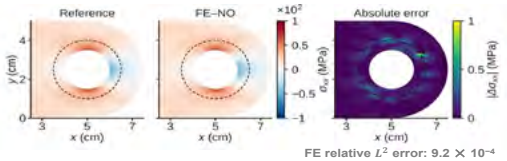
FE–NO for 100-fiber composite: all three accelerators reach the same answer.

accelerator	Constant	Aitken	IQN-ILS
iterations	143	69	37



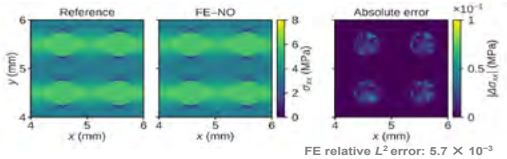
4 Results: linear 3D cases

3D lug: NO – inner ring | FE – outside domain



FE relative L^2 error: 9.2 × 10^{-4}

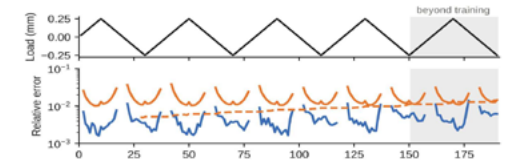
100-fiber composite: NO – fiber | FE – matrix



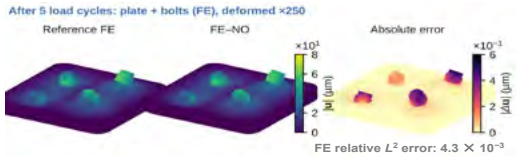
FE relative L^2 error: 5.7 × 10^{-3}

5 Result: plastic case with path-dependency

Bolted joint: NO – flange + housing | FE – plate + bolts



After 5 load cycles: plate + bolts (FE), deformed ×250



FE relative L^2 error: 4.3 × 10^{-3}

Meltem Yashar

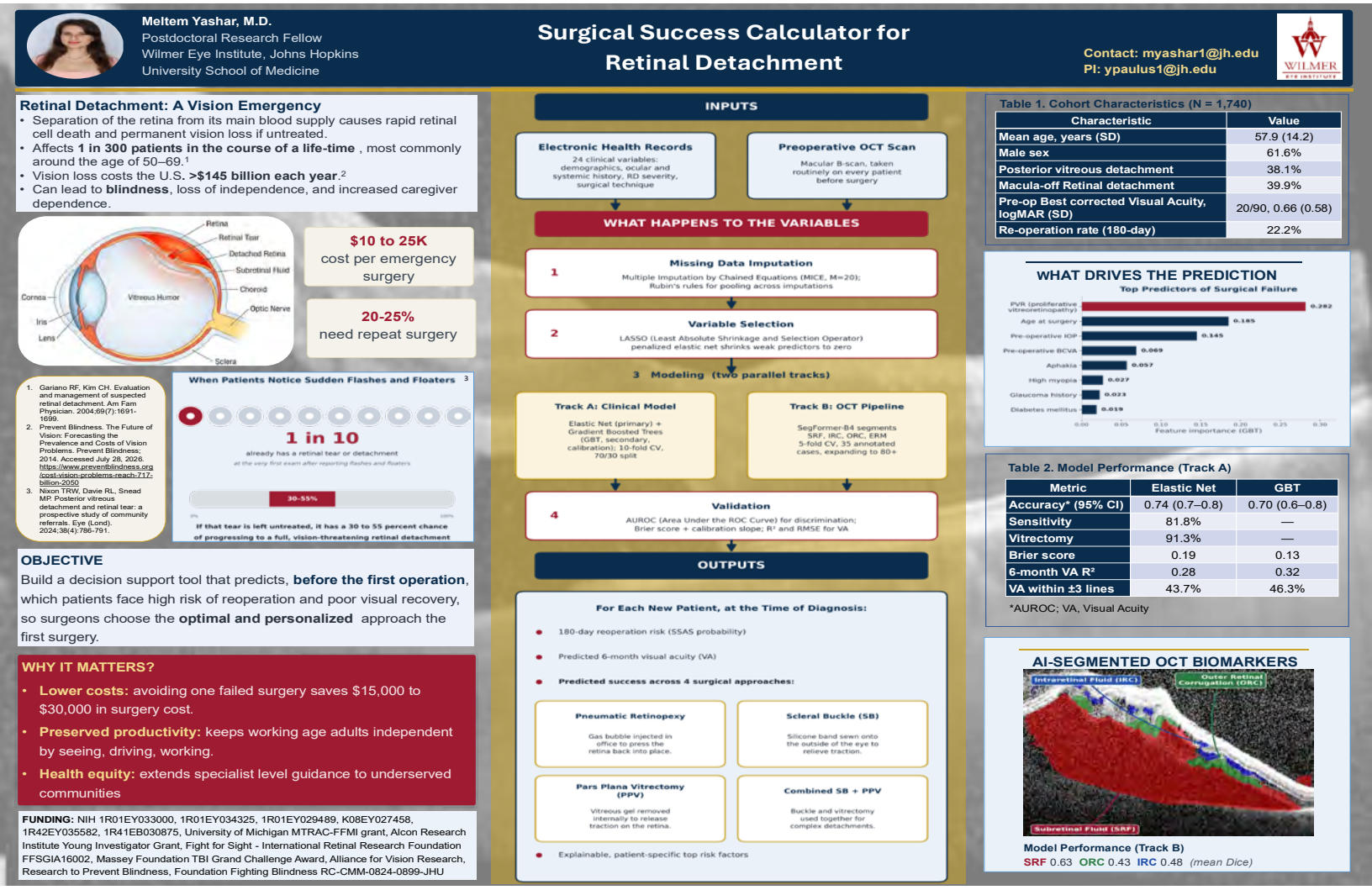
Surgical Success Calculator for Retinal Detachment

Purpose: To develop a decision-support tool that predicts 6-month reoperation risk and visual acuity (VA) before primary retinal detachment (RD) surgery.

Methods: Twenty-four electronic health record variables from 1,740 patients were imputed (MICE, M=20) and selected with LASSO. Elastic net and gradient boosted trees (GBT) were trained using 10-fold cross-validation and a 70/30 split. In parallel, a SegFormer-B4 model segmented subretinal fluid (SRF), intraretinal fluid (IRC) and outer retinal corrugation (ORC) on preoperative OCT B-scans.

Results: Mean age was 57.9 years, 39.9% were macula-off, and the 6-month reoperation rate was 22.2%. Elastic net predicted reoperation with an AUROC of 0.74 (95% CI 0.70–0.80) and 81.8% sensitivity. GBT reached an AUROC of 0.70 with better calibration (Brier 0.13 vs 0.19). For 6-month VA, GBT achieved an R^2 of 0.32, with 46.3% of predictions within ±3 lines. The top predictors were proliferative vitreoretinopathy, age, preoperative intraocular pressure and best corrected VA. Mean Dice was 0.63 for SRF, 0.48 for IRC and 0.43 for ORC.

Conclusions: Combining clinical data with AI-segmented OCT biomarkers may enable explainable, patient-specific outcome prediction, helping surgeons choose a more personalized surgical approach.



Geng Zhang

Estimating Hessian Log-Determinants without Forming the Hessian Matrix

The log-determinant of a Hessian matrix is used in uncertainty assessment, Bayesian model comparison, and experimental design. Computing it can be expensive when models have many parameters and derivatives are difficult to obtain. We combine simultaneous perturbation stochastic approximation (SPSA) with stochastic Lanczos quadrature (SLQ) to estimate the log-determinant of a positive-definite Hessian without forming the full matrix. SPSA estimates directional curvature from possibly noisy function evaluations, while SLQ uses small matrix projections to approximate the log-determinant. A local error analysis separates derivative estimation, projection, and random sampling errors. Numerical studies consider a Gaussian statistical model, simulated quantum-information measurements, and synthetic sensor placement. For the Gaussian model with 3,002 parameters, SPSA-SLQ achieves a relative root-mean-square error of 0.93% and uses 65% less peak GPU memory than a basic full-matrix estimator. The quantum-information study shows lower estimation errors at larger dimensions, while the sensor study compares improved full-matrix methods and evaluates layout selection. These results illustrate how SPSA-SLQ can reduce memory requirements for curvature-based calculations, including settings where only noisy function evaluations are available.

Estimating Hessian Log-Determinants without Forming the Hessian Matrix

Geng Zhang[1] and James C. Spall[1,2]
[1] Department of Applied Mathematics and Statistics, Johns Hopkins University
[2] Johns Hopkins University Applied Physics Laboratory
IDIES Annual Symposium 2026

Introduction

Log-determinant of Hessian matrix used in uncertainty assessment, Bayesian model comparison, and experimental design.

Hessian describes local curvature. Full-matrix computation and storage grow with parameter count.

Objective

Estimate positive-definite Hessian log-determinants from possibly noisy function values, without forming full matrices. Available derivatives can also be used.

SPSA-SLQ Method

SPSA estimates directional curvature from nearby function evaluations. SLQ combines information from small matrix projections [1,2].

1 Evaluate model
Obtain objective values at nearby parameter settings.

2 Perturb parameters together
Estimate Hessian-vector products: directional changes in gradient.

3 Build small projection
Keep curvature information along selected directions.

4 Compute in small matrix
Approximate one contribution to log-determinant.

5 Average scalar estimates
Combine contributions from random starting directions.

Local error analysis separates product-estimation, projection/quadrature, and random-probe errors. Target Hessian must be positive definite.

Conclusions

SPSA-SLQ reduces memory use and estimation error in reported large-parameter simulations. Function values suffice. Benefits depend on matrix structure and accuracy needed for each decision.

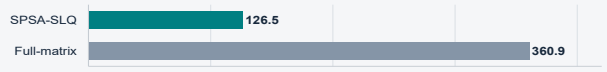
SPSA: simultaneous perturbation stochastic approximation
SLQ: stochastic Lanczos quadrature

[1] Spall, IEEE TAC 37, 332-341 (1992).
[2] Ubaru et al., SIAM J. Matrix Anal. Appl. 38, 1075-1099 (2017).
[3] Spall, JCGS 14, 889-909 (2005).
[4] Spall, ACC, 2395-2400 (2008).
[5] Gacon et al., Quantum 5, 567 (2021).
[6] Boukouvalas et al., CSDA 71, 1088-1102 (2014).

Geng Zhang gzhang37@jhu.edu

Numerical Results

1. Gaussian model: memory at 3,002 parameters



Peak allocated GPU tensor memory. MiB = 2^{20} bytes ≈ 1.049 MB.

Full-matrix baseline [3]: estimate complete Fisher information matrix from likelihood values using simultaneous perturbations, then compute log-determinant.

Three runs on RTX 3080 Ti. Nominal budgets match, but actual counts were not recorded. Diagonal reference has three distinct eigenvalues, favoring small projections.

1. Gaussian: accuracy and runtime

0.93% relative RMSE vs 127.80% full-matrix

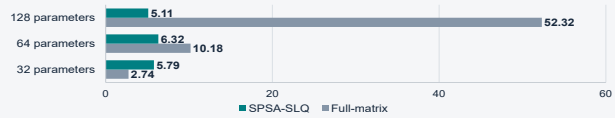
65% less peak GPU memory than full-matrix

Relative log-det RMSE (%)

Parameters	Full-matrix	SPSA-SLQ
209	12.94	**1.49**
989	75.62	**0.49**
3,002	127.80	**0.93**

GPU runtime at 3,002 parameters: 129.39 s (SPSA-SLQ) vs 140.09 s (full-matrix).

2. Noisy quantum-information measurements



Relative log-determinant RMSE (%). Lower is better.

Regularized information target in simulated quantum-state model [5]. Budget: 1,048,576 fidelity evaluations, 8,192 shots each. Thirty estimates per method and dimension. All measurements simulated.

2. Quantum: estimation accuracy

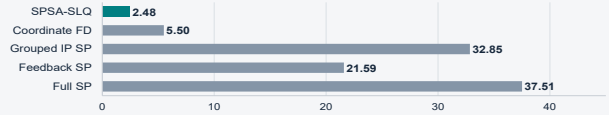
5.11% vs 52.32%
SPSA-SLQ vs full-matrix relative RMSE at 128 parameters

SPSA-SLQ has lower RMSE at 64 and 128 parameters. Full-matrix estimator is more accurate at 32 parameters.

Full-matrix baseline [5]: estimates complete information matrix from fidelity measurements before computing log-determinant.

RMSE: root-mean-square estimation error. Relative RMSE divides RMSE by mean absolute reference log-determinant. All studies use analytical reference values.

3. Synthetic sensor placement



Log-determinant RMSE. Lower is better.

18 layouts, 288 sensors per layout, 144 unknown spatial coefficients. SPSA-SLQ has lowest estimation RMSE among five tested methods [4,6].

SP = simultaneous perturbation. IP = independent perturbations. FD = finite differences.

3. Sensors: accuracy, choice, runtime

Lowest estimation error does not necessarily give best layout choice.

Best-of-six selection frequency
Coordinate FD: 77.8%
SPSA-SLQ: 54.4%
90 selections across three layout families.

CPU runtime/layout: 1.41 s (SPSA-SLQ), 1.48 s (Full SP), 0.97 s (Coordinate FD). Median of five serial runs on Ryzen 7 5800X.

Likelihood-term computation matches. Data generation and tuning costs differ: FD uses 4 simulated datasets per layout/run, compared with 41,472 for SPSA-SLQ.

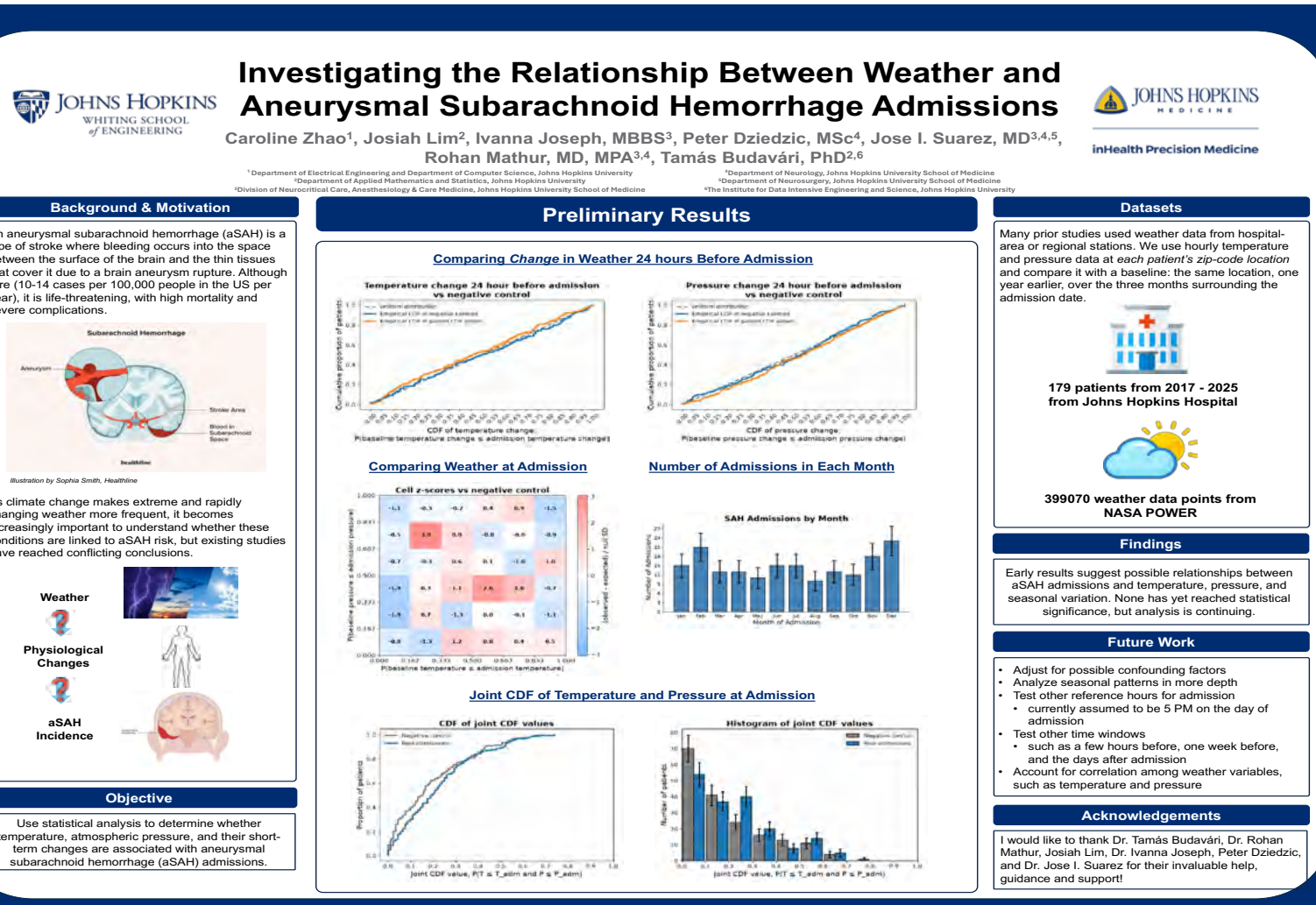
Caroline Zhao

Investigating the Relationship Between Weather and Subarachnoid Hemorrhage Admissions

A subarachnoid hemorrhage (SAH) is bleeding into the space between the surface of the brain and the thin tissues that cover it. Although rare (10-14 cases per 100,000 people in the US per year), it is life-threatening, with high mortality and severe complications.

As climate change makes extreme and rapidly changing weather more frequent, it becomes increasingly important to know whether these conditions are linked to SAH risk, but existing studies have been inconclusive and limited in scope. We are examining whether weather is a risk factor for SAH by linking admissions data from 179 patients at Johns Hopkins Hospital (2017 to 2025) to more than 386,000 hourly weather observations from NASA's open-source API, using statistical methods to test whether pre-admission weather differs from what would be expected by chance.

Preliminary results suggest possible relationships between SAH admissions and specific temperature and pressure conditions, as well as seasonal variation. However, no association has yet reached statistical significance, and analysis is ongoing. Understanding these relationships could help inform future preventative strategies and early warning for at-risk patients.



Johns Hopkins Whiting School of Engineering

Investigating the Relationship Between Weather and Aneurysmal Subarachnoid Hemorrhage Admissions

Johns Hopkins Medicine — inHealth Precision Medicine

Caroline Zhao[1], Josiah Lim[2], Ivanna Joseph, MBBS[3], Peter Dziedzic, MSc[4], Jose I. Suarez, MD[3,4,5], Rohan Mathur, MD, MPA[3,4], Tamás Budavári, PhD[2,6]

[1]Department of Electrical Engineering and Department of Computer Science, Johns Hopkins University
[2]Department of Applied Mathematics and Statistics, Johns Hopkins University
[3]Division of Neurocritical Care, Anesthesiology & Care Medicine, Johns Hopkins University School of Medicine
[4]Department of Neurology, Johns Hopkins University School of Medicine
[5]Department of Neurosurgery, Johns Hopkins University School of Medicine
[6]The Institute for Data Intensive Engineering and Science, Johns Hopkins University

Background & Motivation

An aneurysmal subarachnoid hemorrhage (aSAH) is a type of stroke where bleeding occurs into the space between the surface of the brain and the thin tissues that cover it due to a brain aneurysm rupture. Although rare (10-14 cases per 100,000 people in the US per year), it is life-threatening, with high mortality and severe complications.

Illustration by Sophia Smith, Healthline

As climate change makes extreme and rapidly changing weather more frequent, it becomes increasingly important to understand whether these conditions are linked to aSAH risk, but existing studies have reached conflicting conclusions.

Objective

Use statistical analysis to determine whether temperature, atmospheric pressure, and their short-term changes are associated with aneurysmal subarachnoid hemorrhage (aSAH) admissions.

Preliminary Results

Comparing *Change* in Weather 24 hours Before Admission

Comparing Weather at Admission

Number of Admissions in Each Month

Joint CDF of Temperature and Pressure at Admission

Datasets

Many prior studies used weather data from hospital-area or regional stations. We use hourly temperature and pressure data at *each patient's zip-code location* and compare it with a baseline: the same location, one year earlier, over the three months surrounding the admission date.

179 patients from 2017 - 2025 from Johns Hopkins Hospital

399070 weather data points from NASA POWER

Findings

Early results suggest possible relationships between aSAH admissions and temperature, pressure, and seasonal variation. None has yet reached statistical significance, but analysis is continuing.

Future Work

- Adjust for possible confounding factors
- Analyze seasonal patterns in more depth
- Test other reference hours for admission
  - currently assumed to be 5 PM on the day of admission
- Test other time windows
  - such as a few hours before, one week before, and the days after admission
- Account for correlation among weather variables, such as temperature and pressure

Acknowledgements

I would like to thank Dr. Tamás Budavári, Dr. Rohan Mathur, Josiah Lim, Dr. Ivanna Joseph, Peter Dziedzic, and Dr. Jose I. Suarez for their invaluable help, guidance and support!

Lulu Zhao

Citation Grounding and Self Consistency for Large Language Model Extraction of Structured Data From Clinical Notes

Large language models (LLMs) can convert unstructured clinical notes into structured research data, but unsupported assertions, confusion between related categories, and misattribution of historical information to clinical episodes can compromise extraction accuracy. We describe a citation-grounded LLM extraction method and compare its performance with single-pass extraction against an independent human reference standard.

Clinical notes from a rhegmatogenous retinal detachment cohort were processed in independent segments. Patient history and surgical variables were extracted in separate passes. The method combined verbatim supporting quotations verified against source text, worked-example prompting, secondary verification, rule-based relevance filtering, repeated independent extractions with majority voting, rule-based temporal classification, and numeric calculations performed by code. Outputs were evaluated against a reconciled reference standard derived from 2 independent human reviewers.

For history variables showing the largest improvements, agreement increased from 7%–36% to 76%–81%. Structured surgical findings that were not extracted at baseline had reported accuracy of 94%–100% when present using the final method. The proportion of extracted surgical findings discarded because they lacked a valid supporting citation decreased from 65% to 4.8%.

These findings support further evaluation of the method for clinical research data curation. Validation across other clinical settings and variable types is needed to establish generalizability.only noisy function evaluations are available.

Citation Grounding and Self Consistency for Large Language Model Extraction of Structured Data From Clinical Notes

Lulu Zhao, Luqman Munir[1] , Meltem Yashar, M.D.[1] , Lana Hakma, Varun Bobba, Russel Calderon, Yannis M. Paulus, M.D.[1,2]

[1] Wilmer Eye Institute, Department of Ophthalmology, Johns Hopkins University, Baltimore, MD | [2] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

Contact: ypaulus1@jh.edu

Background

- Ophthalmology clinical notes contain detailed patient histories and surgical information that support clinical research.
- Converting these complex narratives into structured datasets requires time-consuming manual review.
- LLMs offer automated extraction, but unsupported findings, confusion between categories, and misattribution of historical information can limit reliability.
- Combining citation grounding, independent verification, repeated extraction, and explicit processing rules may improve reliability and make automated clinical data extraction more practical and reproducible.

Purpose

To evaluate whether a citation-grounded LLM workflow incorporating verification, repeated extraction, and rule-based processing improves accuracy over single-pass extraction, using an independent human reference.

Methods

Study Design and Data

- We used 13,695 clinical notes from 50 patients with rhegmatogenous retinal detachment (RRD), averaging 273.9 ± 307.3 notes per patient. This includes progress notes, assessment and plan documentation, history and physical notes, and treatment plans.
- Notes were processed using HopGPT, Johns Hopkins' secure generative AI platform approved for protected health information (PHI).
- Models used: GPT-5.4 mini for all extraction and verification passes.
- Notes were processed independently without carrying outputs forward. Patient-history and surgical variables were extracted from distinct note subsets.

LLM Extraction Workflow

- Prompting: Worked examples illustrated expected output
- Citation grounding: Supporting quotations were verified against source text.
- Secondary verification: Quotations were checked for support of extracted values.
- Repeated extraction: Independent extractions were combined using majority voting.
- Rule-based processing: Rules filtered irrelevant information and classified event timing.
- Numeric calculations: Code performed calculations.

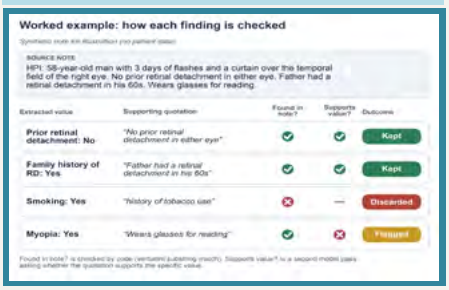
Reference Standard

- Two reviewers independently extracted data and reconciled their findings to establish a reference standard. Variables included: detachment characteristics, retinal findings, surgical details, and postoperative positioning.

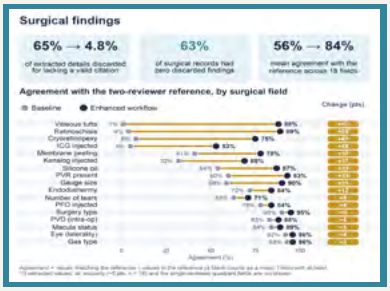
Performance Evaluation

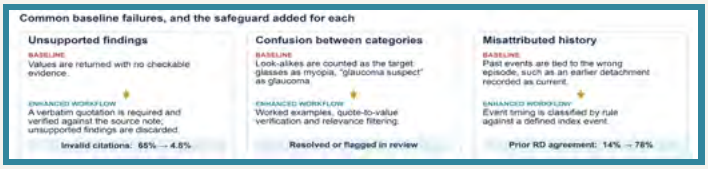
- The enhanced workflow was compared with single-pass extraction without safeguards. Outcomes included agreement with the reference standard for each variable and the proportion of surgical findings discarded for lacking valid citations.











Results

History Variable Agreement

- Largest improvements: 7%-36% at baseline vs. 76%-81% with the enhanced workflow

Surgical Findings

- Findings missed at baseline achieved 94%-100% accuracy when present with the enhanced workflow
- Findings discarded for invalid supporting citations decreased from 65% to 4.8%

Conclusion

Research Application

- Citation grounding, verification, repeated extraction, and processing rules improved the reliability of structured data extraction.
- Supporting quotations make extracted findings traceable to the clinical record and easier to audit.
- This workflow shows promise for converting ophthalmology notes into datasets for clinical research.

Limitations and Future Directions

- Evaluation was limited to a single retinal-detachment cohort with manual review of 50 patients.
- Further validation across institutions, clinical settings, and variable types is needed to establish generalizability.

Funding

NIH 1R01EY033000, 1R01EY034325, 1R01EY029489, K08EY027458, 1R42EY035582, 1R41EB030875, University of Michigan MTRAC-FFMI grant, Alcon Research Institute Young Investigator Grant, Fight for Sight - International Retinal Research Foundation FFSGIA16002, Massey Foundation TBI Grand Challenge Award, Alliance for Vision Research, Research to Prevent Blindness, Foundation Fighting Blindness RC-CMM-0824-0899-JHU

Call for Annual Review content

We're busy putting together the 2026 IDIES Annual Review covering all data-intensive research from our members. Let us share your professional achievements! Please send all announcements to Kate Stevens at **kstevens@jhu.edu** no later than October 31, 2026.

Example content includes any of the following from **October 1, 2025 to September 30th**: Publications, grants, appointments, honors received, speaking engagements and media appearances, event recaps, organizational firsts, theses defended, new partnerships, and anything else of interest to colleagues working with big data.