Proceedings of the 2023 Delaware Data Science Symposium

Benjamin E. Bagozzi, Hanan Abou Ali, Michael Blaustein, Daria Blinova, Jeffrey Buler, Lynette Carney, Sunita Chandrasekaran, Adam Davey, Adam Fleischhacker, Mina Ostovari, Daniel Peart, Sam Smith, Nii Adjetey Tawiah, Cathy H. Wu, and the University of Delaware Data Science Institute

2023

The 2023 Delaware Data Science Symposium was held on September 22nd with a primary focus on the role of data science in financial technology (FinTech) and health equity. The Symposium was organized by the University of Delaware’s (UD’s) Data Science Institute (DSI) with support from Tech Impact, Dupont, Kendal Corporation, Intellitec Solutions, UD’s Library, Museums, & Press, the UD Career Center, the UD Graduate College, the UD Master of Science in Data Science Program, UD’s Artificial Intelligence Center of Excellence (AICOE), and the DSI. It represented the fourth Delaware Data Science Symposium hosted at the University of Delaware, and the third such Symposium since the DSI’s inception. Altogether, the Symposium saw over 280 registered attendees from the University of Delaware and partner institutions across the Mid-Atlantic and beyond. The 2023 Delaware Data Science Symposium included multiple keynote speakers, a series of initiative & lightning talks, a poster session, a panel on data science-driven equity from healthcare, FinTech, community, and educational perspectives, and a session on UD’s summer 2023 Data Science (DS) + Artificial Intelligence (AI) Hackathon. Alongside these sessions, the Symposium also facilitated two associated satellite events. The first was a September 21st Data Science and Analytics Open House for UD graduate programs focused on data science and analytics. The second was a September 25th workshop on the use of MATLAB for low-code AI.
Symposium Overview

The 2023 Delaware Data Science Symposium had over 280 registered attendees. Approximately 83% were affiliated in some manner with the University of Delaware (UD). The remaining 17% were solely affiliated with UD partner institutions. Attendees with partner institution affiliations included faculty and/or students from Delaware State University, Lincoln University, Georgia Tech, the LSST Discovery Alliance, and Rowan College of South Jersey, among others. The 2023 Delaware Data Science Symposium also saw wide attendance from non-academic partner organizations. This included over 60 attendees from approximately 30 distinct partner organizations such as American Fintech Council, Central Intelligence Agency, ChristianaCare, Cledar, Cross River, the Data Innovation Lab at Tech Impact, the Delaware Technology Park, DuPont, Esri, FMC Corporation, IFF, Innovative Precision Health (IPH), Intellitec Solutions, IPQ Analytics LLC, Kendal Corporation, MathWorks, Nemours Children’s Health System, Newark City Council, Teamwave Tech Inc., Tech Impact, The Chemours Company, United Way of Delaware, Vanguard, Waters Corporation, and Winterthur Museum, among others. A visual presentation of the non-academic collaborators’ organizations, taken directly from the 2023 Delaware Data Science Symposium’s registration form, appears immediately below.
The 2023 Delaware Data Science Symposium’s attendees held a wide variety of positions across the institutions and organizations outlined above. A high-level summary of attendees’ primary-reported positions appears immediately below. Therein, one can observe significant representation among (undergraduate and graduate) students, external collaborators, university faculty, and university staff, as well as others.

**Primary Occupation of Attendees**

Among University of Delaware attendees, approximately 65 distinct units were represented, spanning a majority of UD’s colleges. The Symposium also saw a total of 122 registered student participants, including 32 master’s students, 70 PhD students, and 20 undergraduate students. In total, 31 posters were presented by undergraduate students, graduate students, postdoctoral researchers, or faculty that were either affiliated with the University of Delaware or with Lincoln University. Among the posters presented at the Symposium, the following three best poster awards were awarded to Lincoln University- and UD-affiliated students:

- The University of Delaware Master of Science in Data Science-sponsored **Best Undergraduate Student Poster Award** was awarded to **Iteoluwa Ibitoye, Undergraduate Student at Lincoln University** for their poster “Data Alignment Techniques in Railway Track Geometry Inspection: A Comparative Study.”

- The University of Delaware Library, Museums & Press-sponsored **Best Graduate Student Poster Award** was awarded to **Rachel Keown, PhD Student, University of Delaware, Delaware Biotechnology Institute** for their poster “Unwinding the World of Viral Genetic Diversity: Helicases as Phenotypic Marker Genes.”

- The University of Delaware Data Science Institute-sponsored **Best Overall Data Science Poster Award** was awarded to **Huma Rasheed, PhD Student, University of Delaware, Communication** for their poster “Abortion Salience After the Roe v. Wade Leak.”
In addition to the student poster awards outlined above, several teams from UD’s summer 2023 DS+AI Hackathon presented their hackathon results and received Tech Impact-sponsored awards. These hackathon teams, awards, and projects, which are discussed in detail further below, included:

- The DS+AI Hackathon’s **Technical Complexity Award**, awarded to the **Kendal Transformers** for their project “Predicting the Need for Long Term Care Among Senior Citizens Using Machine Learning.”

- The DS+AI Hackathon’s **Impact Award**, awarded to the **Medical Marauders** for their project, “Towards Interpretable Machine Learning for U.S. Hospitals’ CMS Ranking.”

- The DS+AI Hackathon’s **Statement Execution Award**, awarded to the **Chemours Innovators** for their project “A Self-optimized Impedance Fitter from Chemours Innovators.”

- The DS+AI Hackathon’s **Creativity Award**, awarded to the **Rain Makers** for their project, “Generating a High-quality, Fine-scale Precipitation Dataset for the Great Lakes Region Building upon Existing Datasets”

The Symposium also saw lightning talks from UD students, faculty and staff, as well as from external collaborators such as DuPont, ChristianaCare, IPQ Analytics LLC, Innovative Precision Health, and others. These were complimented by four initiative talks highlighting UD’s recently established Institute for Engineering Driven Health, Gerard J. Mangone Climate Change Science and Policy Hub, and Data Science-FinTech Consortium, as well as Nemours’ Research Expanding Access to Child Health (REACH) initiative.

The Symposium’s event webpage can be found here¹ and a Udaily article on the Symposium can be found here². In what follows, we present a formal proceedings for the 2023 Delaware Data Science Symposium. This proceedings includes all associated 2023 Delaware Data Science Symposium Program items, as well as abstracts and titles for each keynote, lightning talk, and presented poster. We then provide a series of photographs from the Symposium. Finally, the Symposium’s planning committee and sponsors are listed.

---

¹ [https://dsi.udel.edu/events/dsi-symposium-2023/](https://dsi.udel.edu/events/dsi-symposium-2023/)
#DataScienceUD23
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 AM</td>
<td>Registration &amp; Poster Setup</td>
</tr>
<tr>
<td>9:00 AM</td>
<td>Welcome &amp; Introductions</td>
</tr>
<tr>
<td>9:05 AM</td>
<td>Remarks from Tracy Shickel</td>
</tr>
<tr>
<td>9:15 AM</td>
<td>Keynote 1: Dr. Andrea Parker</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Panel: Data Science Driven Equity</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Lightning Talks Part I</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>12:40 PM</td>
<td>Poster Lightning Talks</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>Poster Session</td>
</tr>
<tr>
<td>1:50 PM</td>
<td>Initiative Showcase</td>
</tr>
<tr>
<td>2:15 PM</td>
<td>Keynote 2: Hubert Niewiadomski</td>
</tr>
<tr>
<td>3:00 PM</td>
<td>Lightning Talks Part II</td>
</tr>
<tr>
<td>3:45 PM</td>
<td>Coffee &amp; Snack Break</td>
</tr>
<tr>
<td>4:00 PM</td>
<td>Hackathon Presentation &amp; Awards Ceremony</td>
</tr>
<tr>
<td>4:50 PM</td>
<td>Poster Awards &amp; Closing Remarks</td>
</tr>
<tr>
<td>5:00 PM</td>
<td>End Symposium</td>
</tr>
</tbody>
</table>
WELCOME REMARKS & INTRODUCTIONS

Benjamin Bagozzi
Chair, Data Science Symposium Planning Committee
Department of Political Science & International Relations
University of Delaware
https://www.poscir.udel.edu/people/faculty/Bagozzib

Tracy Shickel
Associate Vice President
Corporate Engagement
University of Delaware
Andrea Grimes Parker is an Associate Professor in the School of Interactive Computing at Georgia Tech. She is also an Adjunct Associate Professor in the Rollins School of Public Health at Emory University and at Morehouse School of Medicine. Dr. Parker holds a Ph.D. in Human-Centered Computing from Georgia Tech and a B.S. in Computer Science from Northeastern University. She is the founder and director of the Wellness Technology Lab at Georgia Tech. Her interdisciplinary research spans the domains of human-computer interaction and public health, as she examines how social and interactive computing systems can be designed to address health inequities.

Dr. Parker has published widely in the space of digital health equity and received several best paper and honorable mention paper awards for her research. Her research has been funded through awards from the National Science Foundation, the National Institutes of Health, the Aetna Foundation, Google, and Johnson & Johnson. Additionally, she is a recipient of the 2023 Georgia Tech College of Computing Outstanding Mid-Career Research Award, and the 2020 Georgia Clinical & Translational Science Alliance Team Science Award. Dr. Parker has held various leadership roles, including serving as co-chair for Workgroup on Interactive Systems in Healthcare (WISH) and as a member of the Johnson & Johnson / Morehouse School of Medicine Georgia Maternal Health Research for Action Steering Committee.
Transforming the Health of Communities through Innovations in Social Computing

Despite years of digital health research and commercial innovation, profound health disparities persist. In this talk, I will argue that to reduce health disparities, ICTs must address social determinants of health. Intelligent interfaces have much to offer in this regard, and yet their affordances—such as the ability to deliver personalized health interventions—can also act as pitfalls. For example, a focus on personalized health interventions can lead to the design of interfaces that help individuals engage in behavioral change. While such innovations are important, to achieve health equity there is also a need for complimentary systems that address social relationships. Social ties are a crucial point of focus for digital health research as they can provide meaningful supports for positive health, especially in populations that disproportionately experience barriers to wellbeing.

I will offer a vision for digital health equity research in which interactive and intelligent systems are designed to help people build, enrich, and engage social relationships that support wellbeing. By expanding the focus from individual to social change, there is tremendous opportunity to create disruptive interventions that catalyze and sustain population health improvements.

Abstract: Digital health research—the investigation of how technology can be designed to support wellbeing—has exploded in recent years. Much of this innovation has stemmed from advances in the fields of human-computer interaction and artificial intelligence. A growing segment of this work is examining how information and communication technologies (ICTs) can be used to achieve health equity, that is, fair opportunities for all people to live a healthy life. Such advances are sorely needed, as there exist large disparities in morbidity and mortality across population groups. These disparities are due in large part to social determinants of health, that is, social, physical, and economic conditions that disproportionately inhibit wellbeing in populations such as low-socioeconomic status and racial and ethnic minority groups.
Currently serving as the CEO of Cledar, Hubert Niewiadomski moved into this role after previously acting as the company's CTO. Equipped with a foundation in machine learning, artificial intelligence, software engineering and particle physics, Hubert boasts over twenty years of hands-on experience in data analytics across both academic and business landscapes. At Cledar, he has steered the team through diverse data science projects, addressing challenges in areas like macroeconomic forecasting, credit rating, and network monitoring. His expertise extends to image and video processing, with particular attention given to medical imaging applications.

Of late, he has been spearheading the development of LLM-based algorithms tailored for legal contexts. Alongside these responsibilities, Hubert collaboratively explores research on Large Language Models in conjunction with ETH Zurich and Warsaw University. His portfolio features a stint as the co-Coordinator of Data Analytics within the TOTEM Collaboration at CERN (European Laboratory for Particle Physics), Geneva, Switzerland. At CERN, Hubert's introduction of two machine learning strategies considerably improved the accuracy of measurements related to diffractive phenomena observed in proton collisions at the Large Hadron Collider. Such advancements were instrumental in the joint discovery of the odderon particle in partnership with Fermilab. Consistently, throughout his tenure at CERN, Hubert promoted the adoption of machine learning methodologies to analyze large, multi-dimensional high-energy physics datasets.

Academically, Hubert is an alumnus of the University of Manchester, where he earned his Ph.D., and the Warsaw University of Technology, from which he obtained his MSc. Beyond his technological pursuits, Hubert nurtures a deep-seated love for music. His musical journey led him to the Geneva University of Music, where he delved into classical music studies, concentrating on opera singing and pipe organ.
Transformative Role of AI & Large Language Models in Economy and Fintech Industry

The evolution of artificial intelligence (AI) has brought transformative shifts, with the financial services sector standing at the forefront of this change. This presentation delves into the escalating role of AI in economy, with a specific focus on the advancements in Natural Language Processing (NLP) and the capabilities of Large Language Models (LLMs).

Journeying back, we'll touch upon the foundational mathematical concepts that gave birth to artificial neural networks in the mid-20th century, and spotlight the surprising rate at which these models have grown, even outpacing Moore's Law in the recent decade. Within this context, the diverse landscape of LLMs will be explored, illuminating their unique competencies and their notable achievements in various benchmarks.

In the realm of application, we’ll examine the increasing investments pouring into AI and LLM research and development. We’ll also delve into the innovative techniques of LLM training, highlighting how the models themselves optimize outcomes, even with constrained datasets. Finally, we’ll look ahead, discussing the potential disruptions and a few illustrative applications of these technologies within the data industry.
Panel: Data Science-Driven Equity

Federica Bianco (**Moderator**)

Associate Professor
Sharp Lab
University of Delaware
*Federica Bianco (udel.edu)*

Federica Bianco splits her time as a professor in the Department of Physics and Astrophysics (where she runs a lab focusing on light curves), the Biden School of Public Policy and Administration and the Urban Observatory, where she uses her astrophysics skills to study urban problems. She is also the coordinator of the Large Synoptic Survey Telescope (LSST) Science Collaboration, a network of more than 1,500 scientists. She has published more than 100 peer-reviewed papers, was a Smithsonian predoctoral and James Arthur postdoctoral fellow and was the recipient of a Department of Energy “Innovative Development in Energy-Related Applied Science” grant.

Emily Kauffman

Senior Manager
Strategic Initiatives and Impact
United Way of Delaware

Emily Kauffman is a Senior Manager for Strategic Initiatives and Impact at the United Way of Delaware and serves as a Program Manager for the Equity Counts Data Center Initiative. She is currently completing her Master of Public Health degree at George Washington University. After growing up in Delaware and forging a post-college career in health promotion and wellness, Emily is now passionate about using evidence-based programs to address health inequities at the local level.
Dr. Conaty-Buck is a University of Delaware Asst. Prof of Nursing, Family Nurse Practitioner & Informatics Researcher. She educates and precepts DNP/MSN/MBA/BSN students in Interprofessional Healthcare Informatics, Integrated Care Delivery, and Transition to Practice. She co-developed UD’s DNP program and is a member of the SON Graduate Committee. She is a member of the CHS IRB committee and co-chair of UD’s Data Science Institute’s Training Working Group. She began service on the Faculty Senate Graduate Curriculum Committee this fall. Degrees include UVA (MSN and DNP), JMU (BSN), Virginia Tech (MFA Arts Administration), Rowan/Glassboro (BA Music Education) & UD Lerner COB (Graduate Certificate in Entrepreneurship & Innovation). She co-chairs the American Association of Nurse Practitioner’s (AANP) Health Informatics & Telehealth Community representing over 500 NPs across the nation. She is on the HIMSS committee for NP Education serving as a panelist for national webchats on AI and Nanotechnology and is a planning team member for SONSIEL’s upcoming 2023 Innovation Conference. Dr. Conaty-Buck is State Advocacy Director for the Delaware Coalition of Nurse Practitioners and Treasurer for the DELMARVA region of the American Nurses Informatics Association. Her research interests include Patient Technology & Innovation, AI and Virtual Reality for Patient Care and Practitioner/Student Learning, Healthcare Cybersecurity and NP Advocacy & Leadership. Named an AANP Fellow in 2019, she was Delaware’s 2021 Top Nurse for Advancing and Leading the Profession and AANP’s 2022 State Advocate for Excellence.
Phil Goldfeder  
CEO  
American Fintech Council  
*Who We Are — American Fintech Council*

With nearly two decades of experience at the intersection of the public and private sectors, Phil Goldfeder currently serves as Chief Executive Officer of the American Fintech Council (AFC), a leading industry association representing responsible financial technology (fintech) companies creating critical access to safe and affordable financial services. AFC, comprised of the nation’s largest fintech companies, fosters innovative, transparent, and responsible products that promote competition, consumer protection, and financial health, inclusion, and equity. AFC is committed to robust industry standards, with a focus on consumer protection and regulatory compliance, in addition to advocating for and embracing appropriate government regulation.

Before joining AFC, Goldfeder served as Senior Vice President of Global Public Affairs at Cross River, a financial institution and technology infrastructure provider that offers embedded financial solutions. In this role, Goldfeder founded the Online Lending Policy Institute (OLPI) and has been a leader in shaping the new financial services landscape since early 2015. At the onset of the COVID-19 pandemic and the passage of the CARES Act in March 2020, Goldfeder helped lead the Cross River team to mobilize internally and offer a streamlined and automated system to provide more than $12 billion in PPP funding to the most vulnerable small businesses in every state, as well as additional short-term relief efforts to communities in the company’s footprint and beyond.

He previously served as an elected member of the New York State Assembly representing diverse neighborhoods of Queens, N.Y. After most of his district was devastated during Superstorm Sandy, Goldfeder lead recovery efforts with a specific focus on partnering with the banking and insurance industry to help rebuild his community and reform outdated policies. He was the author of transformative banking and insurance modernization legislation and was also a key legislative leader on a diverse array of issues. Prior to his election, he served as a senior advisor to Senate Majority Leader Chuck Schumer (D-NY) and New York City Mayor Michael Bloomberg.
Dr. Claude Tameze is a distinguished academic leader with a profound impact on program management, innovative research, and STEM education. With a Ph.D. in Applied Mathematics from Delaware State University, Dr. Tameze stands out for his exceptional work in image processing, cutting-edge pedagogy to foster problem-solving, and securing substantial multi-million-dollar NSF grants, including the Mid-Atlantic Data Science Corps NSF funded grant with the University of Delaware to advanced data science education at Lincoln University. As a former chair, and currently professor & founding Director of Mathematics Learning Center at Lincoln University, his transformative leadership elevated departmental enrollment by a remarkable 133% and boosted graduation rates by over 500%, with 100% placement of the graduates accepted in prominent universities for advanced studies or competitive professional employments. Dr. Tameze is renowned for forging strategic collaborations with prestigious institutions such as the University of Delaware, Princeton University, University of Michigan, and others, reinforcing his commitment to fostering academic excellence. He has been instrumental in pioneering pivotal courses like MAT-115- Foundation of Data Science, co-taught with prominent data science faculty from the University of Delaware; and leading initiatives like the Mathematics Learning Center (MLC) using innovative pedagogy, including ALEKS & Mobius. His dedication to diversity and equity in STEM education extends beyond research, as evidenced by invited talks at high-profile events and workshops on Equity, Diversity & Inclusion in STEM as he has delivered invited talks at prestigious events, including the Congressional Black Caucus, ASU-hosted events, and EduTech Europe. Dr. Tameze was also invited to give lectures on Probability in several prominent universities in China, but he is also a relentless advocate for advancing education and promoting equity, making an indelible mark on the academic landscape.
Ryan Harrington is the Director of Strategy and Operations for the Data Innovation Lab (DIL) at Tech Impact where he focuses on developing data science projects for social good. Previously, Ryan was the lead data scientist at CompassRed Data Labs, a data science and analytics consulting firm where his focus was on developing models using a combination of statistical methods and machine learning techniques. Outside of work, he co-founded and co-organized Open Data Delaware, an organization dedicated to improving data transparency within government institutions and applying data and technology to solve civic problems.
Lightning Talks Part I

Daniel Peart (Moderator)
Senior Assistant Librarian
Library
University of Delaware

Arthur Andrews
DuPont
“Data Science Applications for Process Improvement at DuPont”
Data science and modern software tools hold great promise for improving manufacturing process efficiency and product quality. However, identifying the specific opportunities can be a vast and open-ended problem. In this talk, we overview the database system and software tools available at modern manufacturing plants and offer general principles for identifying and implementing data science solutions for process improvement.

Kristen Castell
Managing Director
Center for Accelerating Financial Equity (CAFE)
“Fintech for Financial Equity”
How technologies can help individuals that have been traditionally underserved by the financial industry to attain financial health & wellness.

David Chen
ChristianaCare
“Mechanism Matters: Changes in Assault Patterns Among Delaware Medicaid Beneficiaries from 2018-2020”
Compared to 2019, in 2020 the United States experienced an increase in firearm homicide across all age groups, sexes, and census divisions. This surge occurred intercurrently with a worldwide pandemic which sharply affected behavioral norms and healthcare access. However, it is unclear whether other forms of
violence – including non-fatal and non-penetrating assault – similarly increased. We examined the Delaware Medicaid claims and encounters database for assault injuries identified by ICD-10 code from 2018-2020 using a taxonomy based on injury, intent, and penetration of assault. Overall, those with new non-penetrating assaults decreased in 2020 compared to prior years, but this was offset by the increase in new penetrating injuries, especially for those younger than 25 years of age. As assault-related Medicaid healthcare utilization patterns are mechanism dependent with differing ambulatory sensitivity, intervention and prevention efforts require different adaptive strategies.

**Toby Driscoll**

*Unidel Chaired Professor of Mathematical Sciences*

*Director, Master of Science program in Data Science*

*Mathematical Sciences*

*University of Delaware*

*“Data-driven Model Discovery Using Universal Differential Equations”*

For systems whose dynamics are only partially understood, a neural network can be used within a differential equations model and trained by fitting to observations of the system. The resulting model can be used as a black-box simulator or be subjected to further model discovery to look for simple closed-form expressions that capture the same dynamics.

**Susan Conaty-Buck**

*Assistant Professor*

*School of Nursing*

*University of Delaware*

*“Large Language Models for Healthcare - Can we Trust Them?”*

Patients and clinicians alike report great interest in Natural Language processing model ChatGPT adapted for healthcare including generation of human-like text to perform tasks like answering medical questions, creating clinical documentation and recommending care measures. As more practitioners are retiring/resigning, ChatGPT Medical could create chatbot-based apps for patient engagement including real time question response and supporting patient efforts to self-manage their health. ChatGPT’s ability to analyze large volumes of medical data is theorized to increase medical discoveries. Despite the vision of ChatGPT Medical as a well-trained healthcare assistant, concerns abound include data privacy and security, need for regulatory oversight, data quality, quantity and equity, and integration into current healthcare systems. ChatGPT reports it does not always tell the truth - can health systems and providers ethically trust this model to provide safe and effective healthcare?
Ruiqi Wu  
Assistant Professor  
Business Administration  
University of Delaware  
“Platform Information Design and Competitive Price Targeting”
As the rule-maker of its marketplace, a platform shapes sellers' information and, thereby, their competitive outcomes. What information about consumers should sellers be allowed to use for price-targeting? We develop an empirical framework for this information design question, combining flexible demand estimation and equilibrium pricing modeling. Applying it to Steam, we first estimate consumer preferences using high-frequency microdata and a two-stage decision model. Then, we project sellers' equilibrium prices and profits under a comprehensive set of information designs. Finally, counterfactual comparisons reveal that the platform can bring sellers 75% of the potential incremental profits with information on consumer price sensitivities alone. By contrast, information on consumer feature tastes leads to only a marginal profit increase due to seller competition, which suggests that coarsened targeting does not necessarily come at the expense of profitability.

Greg Dobler  
Associate Professor  
Public Policy & Administration (Biden School)  
University of Delaware  
“Are 20 Million Pictures Worth 20 Billion Words? Urban Dynamics Through Persistent Remote Imaging”
I will discuss the concept of the "Urban Observatory" (UO) a long-standing facility in NYC soon to have a partner site in Wilmington, DE that uses persistent synoptic imaging of city skylines from a distance of over a mile and with temporal granularity as high as 10 seconds per image. The massive data set on urban dynamics that the UO is generating is being used to answer questions in a variety of domains in urban science including lighting, vegetation health, heating/cooling use, behavioral dynamics, and public policy assessment.

Adam Fleischhacker  
Associate Professor  
Business Administration  
University of Delaware  
“The Art of Prediction: Interpretable Models with Black-Box Assistance”
This research intends to combine the transparency and interpretability of hierarchical Bayesian models with the predictive power of black-box models. We propose an ensemble of diverse Bayesian models, each capturing specific data aspects, and leverage a black-box model to assign dynamic weights to these interpretable models, optimizing predictive accuracy. This fusion of transparency and accuracy allows
decision-makers to not only understand the final predictions but also comprehend why they were made, enabling more informed and confident decisions.

**Michal Herzenstein**  
*Associate Professor  
Business Administration  
University of Delaware*  

“The Language of (Non)Replicable Social Science”

The language used in academic publications in psychology and behavioral economics is predictive of whether their findings were successfully replicated by other researchers. To understand how and why, we focus on uncovering the textual differences between replicable and nonreplicable papers. We find that the language in replicable papers alludes to the authors’ transparency and confidence in the research, and generally exhibits markers of truthful narratives. Nonreplicable research, however, is shorter, linguistically obfuscated, written with more clout and has the structure of an interesting, immersive story. Thus, authors of nonreplicable papers are more likely to make an effort, through their writing, to persuade readers of their results. Moreover, since the differences in texts are mostly around context-free words, and since we trained our model on a large set of papers from many disciplines, our results can be relevant beyond the social sciences.

**Karen Hoober**  
*Associate Director, Graduate Education & Outreach  
Center for Bioinformatics & Computational Biology (CBCB)  
University of Delaware*  

“The Center of Bioinformatics and Computational Biology Graduate Programs”

The Center of Bioinformatics and Computational Biology (CBCB) graduate program is built on the core curriculum of Bioinformatics Data Science (BDS). CBCB offers two graduate certificates in Applied Bioinformatic and Biomedical Informatics & Data Science (online or hybrid), an MS degree in Bioinformatics and Computational Biology, and a PhD in Bioinformatics Data Science. CBCB is offering a T32 predoctoral training program to enhance graduate research training in *Computational Biology, Bioinformatics, and Biomedical Data Science* (CBB). CBCB and the CBB training program educates the next-generation of researchers/professionals to play key roles in multi- and inter-disciplinary teams, bridging life- and computational-sciences. Experts in the CBCB fields are housed in several Colleges: Engineering, Arts & Sciences, Agriculture & Natural Resources, Health Sciences and Earth, Ocean & Environment: the BDS degrees are university-wide interdisciplinary programs with emphasis on professional skills and immersive internship opportunities (e.g., Christiana Care, Delaware Health Information Network, Nemours Children’s Health, and Delaware State University), preparing graduates for careers in industry, government agencies, or non-profits.
1. Huong Le, Postdoc, Plant & Soil Sciences, University of Delaware *
“Optimal Monitoring Sites Based on suLHS to Improve the Representativeness of Global Soil Properties”
We present a data-driven method for identifying optimized samples from spatial information of global soil data. The spatial univariate Latin Hypercube Sampling (suLHS), combines a Latin Hypercube to obtain a representative sample of the univariate probability distribution function and an autocorrelation model to ensure a reproducible spatial dependency function. The suLHS is tested with a case study using data of global soil respiration map that are relevant for carbon cycle science. The results show that the suLHS is able to reproduce the univariate probability distribution and the spatial variability of the global soil respiration.

2. Guna Gurazada, PhD Student, Data Science Institute, University of Delaware *
“The Promise of Multi-omics and Machine Learning in Modeling CHO Cell Productivity”
The large-scale production of monoclonal antibodies (mAbs) has been an important focus for the biopharmaceutical industry, with Chinese hamster ovarian (CHO) cells being the preferred host cell factories for their manufacture. Persistent improvements in process conditions have led to significant increase in mAb productivity. However, the molecular mechanisms of the host cell that have led to these improvements are not well understood. Here, we present our efforts in implementing a machine learning-based computational framework for the integration of multi-omics data with the goal of answering two questions: a) Can multimodal analysis of the CHO cell system accurately predict mAb productivity, and b) What are the biological markers correlating to desirable productivity traits. The multi-omics framework that we establish can be
generalized and have broader applications to other cellular systems including HEK293 (used to produce viral vectors for gene therapy), plants and even bacteria.

3. Luis Herrera, PhD Student, Physics & Astronomy, University of Delaware *
“A Comparative Study of Different Machine Learning Methods for Dissipative Quantum Dynamics”
It has been recently shown that supervised machine learning (ML) algorithms can accurately and efficiently predict long-time population dynamics of dissipative quantum systems given only short-time population dynamics. Here, we benchmarked 22 ML models on their ability to predict long-time dynamics of a two-level quantum system linearly coupled to harmonic bath. The models include uni- and bidirectional recurrent, convolutional, and fully connected feedforward artificial neural networks (ANNs) and kernel ridge regression (KRR) with linear and most commonly used nonlinear kernels. Our results suggest that KRR with nonlinear kernels can serve as inexpensive yet accurate way to simulate long-time dynamics.

4. Kris Holton, PhD Student, Data Science Institute, University of Delaware *
“Uncovering Key Predictive Channels & Clinical Variables in 40Hz ASSR EEG in Early Stage Psychosis”
Psychotic disorders are characterized by abnormalities in synchronization of neuronal responses. The 40Hz gamma band deficit measured by auditory steady-state response (ASSR) electroencephalogram (EEG) is a robust observation in psychosis, and is associated with symptoms and functional deficits. However, most scalp EEG studies report only one or a few electrodes to classify cases and controls, nor do they investigate if clinical variables are correlated with 40Hz ASSR signals in patients with early stage psychosis (ESP). There is also a lack of longitudinal 40Hz ASSR ESP studies. Here, we examine ESP 40Hz ASSR deficits using all channels, classify ESP status using machine learning on all channels, correlate EEG channels with modules of clinical variables to find associations, and assess whether baseline 40Hz ASSR deficits predict short-term functional outcome.

5. Yalin Liao, PhD Student, Electrical & Computer Engineering, University of Delaware *
“SWAR: Sliced Wasserstein Autoencoding Representation for Anomaly Detection”
Unsupervised anomaly detection is a fundamental problem in machine learning. We propose a two-step method, SWAR, for anomaly detection, which learns more informative representations via a sliced Wasserstein autoencoder (SWAE) and then estimates the probability density function from the new representations.

6. Huma Rasheed, PhD Student, Communication, University of Delaware *
“Abortion Salience After the Roe v. Wade Leak”
The Roe v. Wade leaked draft opinion, and the subsequent Supreme Court decision overturning Roe, drew a wave of media attention to the issue of abortion. This study sought to examine the topical and attribute salience of abortion in the tweets of CNN, Fox News, and ABC News during this critical timeframe (N = 20,144). Findings reveal a gap in the salience of abortion, with Fox News more likely to tweet about abortion than CNN and ABC News. In terms of attributes, Fox News and CNN were more likely to frame abortion politically and use language that had a more negative sentiment. The study concludes with a discussion on the implications of the findings on public opinion and policymaking.

7. Bilal Riaz, PhD Student, Electrical & Computer Engineering, University of Delaware *
“Applications of Optimal Transport in Data Science and Machine Learning”
Data hygiene is a common issue in machine learning. While it is desirable to have more data for increased diversity and improved downstream learning tasks, mislabeled data can be detrimental to the learning process. To address this issue, we propose using a modified form of the Kantorovich-Wasserstein distance that employs carefully curated clean data to filter out incoming data with label noise. We further extended our approach to the applications in distributions alignment and shape registration. We further extended our approach to the Gromov-Wasserstein Distances for distributions alignment and shape registration across different domains.

8. Tatiana Acero Cuellar, PhD Student, Physics & Astronomy, University of Delaware *
“Detangling the Mysteries of a CNN Used to Separate Astrophysical Transients from Artifacts”
Moving from raw images taken by a telescope to astronomical data ready for scientific discovery requires powerful techniques and computational resources. Current techniques generate artifacts that contaminate the data. I created a Convolutional Neural Network (CNN) for the separation of astrophysical transients from image artifacts that does not rely on the computationally expensive Difference Image Analysis technique, or template subtraction. Compare the efficiency of two CNNs, one trained with the templates, search, and difference images and one that takes as input the template and search only. Testing accuracy is reduced from 96% to 91%. We are investigating the performance drop observed in the model by addressing the latent space characteristics of each CNN. We are implementing unsupervised machine learning
techniques (UMAP) and searching for features in the latent space that give insight of the reason behind the performance drop as well as in the physical nature of the transients.

9. Kyungmin Lee, PhD Student, Public Policy & Administration (Biden School), University of Delaware *
“Detection of Building Energy End-Use Via Proximal Infrared Remote Sensing and Computer Vision”
We propose a novel, non-intrusive method of energy use monitoring that uses proximal infrared remote sensing of building envelopes to find patterns of heating and cooling use. We have tested two cases of residential buildings in NYC, one in downtown Brooklyn and a public housing project in Gowanus, North Brooklyn. We collected ~160,000 infrared images at approximately 10-second intervals of the buildings’ facades taken from June 2018. We have applied a maximum Gaussian difference of pixels from that sequence of images to identify exterior-venting HVAC units and generate their infrared time series. Finally, we determine the on/off transitions of each AC unit using a one-dimensional edge detection algorithm and identify aggregated and disaggregated patterns of end-user behavior. Our results show that the method of infrared image processing can discern on-and-off patterns of exterior ACs in a dense urban scene at distances up to approximately 2 miles (3.2 km).

10. Riley Clarke, PhD Student, Physics & Astronomy, University of Delaware *
“Stellar Flares as a Case Study of Atmosphere-Aided Studies of Transients in the Rubin LSST Era”
Stellar flares are short-duration, stochastic brightening events that occur on the surfaces of stars. Space telescopes designed to find planets outside our solar system have measured the brightness evolution of stellar flares in great detail, but have made few accurate measurements of their temperatures. The Vera C. Rubin Legacy Survey of Space and Time (LSST) is the premier ground-based survey of the 2020s, which will observe millions of stellar flares over its 10 year mission, beginning in 2025. While each one of these flares will only be observed by LSST once, so that we will not be able to characterize their evolution in time, we are developing a unique and innovative technique to use the Earth’s atmosphere, which splits light into its color components in the same way water droplets produce a rainbow, to measure the color, and thus the temperature, of the flares observed by Rubin LSST. This project is supported by NSF Award 2308016.

11. Lan Yu, PhD Student, Public Policy & Administration (Biden School), University of Delaware *
“Empirical Measurement of Lighting Technology Changeover in New York City”
In 2013, NYC planned to replace 250,000 streetlights with LEDs by 2017, but no post-assessment was conducted. In this work, we identified light sources and the associated lighting technology type in two Hyperspectral images collected by Urban Observatory, one from 2013 and one from 2018, to empirically measure the efficacy of this policy. For each source in the scene, we used its brightness to determine the lighting technology type by comparing this color “spectrum” with NOAA and LSPDD Templates. We also used 1-D Convolutional Neural Networks to determine the source’s type, and have a model accuracy of roughly 94%. When comparing the 2013 and 2018 results from the above methods, we find new sources in 2018 compared to 2013 are dominated by LEDs. However, our results also provide empirical evidence that the policy intervention was not fully realized as we find HPS streetlighting sources in the scene that were targeted for changeover by 2017 but have not been upgraded by 2018.

12. Elizabeth Smith, PhD Student, Plant & Soil Sciences, University of Delaware *
“Spatiotemporal Variability and Uncertainty of Soil Respiration in the Conterminous United States”
We used digital soil mapping and machine learning techniques to estimate the spatiotemporal variability of Rs across the Conterminous United States (CONUS) from 2000-2020. The Soil Respiration Database was used to train a quantile regression forest model and Shapley values, a concept derived in game theory, were used to determine the contribution of each covariate. We found positive correlations between Rs and soil texture (sand, silt, clay) and soil organic carbon across National Ecological Network (NEON) regions. Our bottom-up approach estimates the annual Rs as 8.72±4.49 Pg C yr-1 (mean ± standard deviation) with associated uncertainty of 16.61 Pg C yr-1 (90% confidence interval), which is consistent with other bottom-up estimates of Rs for this region. These results improve current estimates, further our understanding of spatiotemporal trends of Rs, and highlight the knowledge gap in our understanding of the regional and global carbon budget.

13. Shar Daniels, PhD Student, Physics & Astronomy, University of Delaware *
“Astronomical Discovery with Neural Networks: Exploring the Sub-Second Transient Sky”
With the novel application of deep learning models to continuously exposed astronomical data, we are creating tools to discover and reveal the nature of rapidly-evolving optical astrophysical phenomena. Traditional observational modes require seconds-to-minutes for each image, so the evolution of optical astronomical phenomena at sub-second timescales is barely explored. However, the “continuous-readout” nontraditional observing modality enables resolution at sub-second timescales by integrating the images of astrophysical objects along one spatial dimension. Analyses of these data require custom-made reduction pipelines. We are developing neural networks for the analysis of 450GB continuous-readout astronomical dataset sampled at 300 Hz from the Zwicky Transient Facility (ZTF). This poster will show the performance of CNN and
transformer models under development on the ZTF data and outline the potential of our analysis tools for detecting and analyzing rapid transients at scale.

14. Vineeth Gutta, PhD Student, Computer & Information Sciences, University of Delaware *
“Comparison of Neural Networks with Tree-based Machine Learning Approaches”
Using deep learning in cancer research to tackle scientific challenges is becoming an increasingly popular technique. Advances in enhanced data generation, machine learning algorithms, and compute infrastructure have led to an acceleration in the use of deep learning in various domains of cancer research such as drug response problems. In this study, we explored tree-based models to improve the accuracy of a single drug response model and demonstrate that tree-based models such as XGBoost have advantages over deep learning models, such as CNN, for single drug response problems. The case studies we present focus on cancer drug response datasets however the work is relevant to tabular data in most domains.

15. Sri Vishnuvardhan Reddy Akepati, Master’s Student, Chemical & Biomolecular Engineering, University of Delaware *
[Title and abstract omitted from Symposium Proceedings]

16. Adam Kenney, Master’s Student, Data Science Institute, University of Delaware
“Egg Shape Morphology Accurately Predicts Dam-origin Using Machine Learning Tools”
Zebra shark (Stegostoma tigrinum) females are commonly co-housed in aquariums, making it difficult to monitor individual reproductive cycles externally if eggs are laid concurrently. The aim of this project was to determine if eggs should be accurately ascribed to the correct female based on their external morphology (i.e., shape and size). Yolked and non-yolked (wind) eggs were collected and twenty-two external measurements were taken. Genetic testing of yolked eggs was used to confidently assign eggs to the correct female. Using a random forest model trained on 80% of the data, there was a 96% ability to correctly assign a yolked egg to its correct female when institution was included in the model. Preliminary analysis indicates certain variables may be useful to determine female origin of wind eggs. Although power decreased when institution was excluded (roughly 60%), egg shape morphology was found to be a useful and accurate tool to track individual female reproductive cycles.

17. Shayla Sharmin, PhD Student, Computer & Information Sciences, University of Delaware
“Insights into Cognitive Engagement: Comparing the Effectiveness of Game & Video Based Learning”
This study aims to examine the effects of GBL and VBL on cognitive processes by measuring oxygenated hemoglobin levels using fNIRS. A user study of twelve participants divided into experimental (game) and control groups (video) was conducted, and fNIRS was employed to acquire cerebral signals, pre- and post-tests were administered. Although there were no statistically significant changes, the mean levels of oxygenated hemoglobin (ΔHbO) were higher in the GBL group, suggesting the possibility of cognitive improvement. Both experimental groups exhibited increased neural activation in the Lateral Prefrontal Cortex (Lateral PFC), with the GBL group showing 2.33 times greater hemodynamic activity in this specific brain region. Examination of knowledge acquisition indicated a significant increase of 47.74% in the GBL group as compared to the VBL group.

18. Siddharth Chaini, PhD Student, Physics & Astronomy, University of Delaware
“Light Curve Classification using Distance Metrics”
The rise of synoptic sky surveys has ushered in an era of big data in time-domain astrophysics, making data science and machine learning essential tools for studying celestial objects. Here, we present an investigation of distance metrics for classification of astrophysical time series, or light curves. While tree-based models (e.g., Random Forest and Gradient Boosted trees) are the standard in astrophysical classification, the direct use of distance metrics is an approach that has not been explored in time-domain astrophysics. With a comprehensive study of 18 distance metrics applied to a catalog of 700,000 variable stars in 10 classes, we make recommendations for efficient and physically interpretable classification algorithms.

19. Kennet Rueda, PhD Student, Physics & Astronomy, University of Delaware
“Simulating Astatide in Water: From Basis Sets to Force Fields Using Particle Swarm Optimization”
Isotope At-211 has been identified as a promising agent for treating certain types of cancer. Here we propose new methods for large-scale simulations of astatide At- in water. In particular, we developed a simple yet accurate force field for At- in water based on reliable relativistic DFT calculations. We employ the evolutionary algorithm called particle swarm optimization for the force field and basis set optimizations. Accounting for strong relativistic effects, via our redesigned infrastructure, we elucidate a noticeable dissimilarity between At- and I- in halide-water force field parameters, radial distribution functions, diffusion coefficients, and hydration energies. In general, this work establishes a
framework for the systematic development basis sets for relativistic DFT and accurate force fields for molecular dynamics simulations to be used in large-scale models of complex molecular systems with elements from the bottom of the periodic table.

20. Rachel Keown, PhD Student, Delaware Biotechnology Institute, University of Delaware *
“Unwinding the World of Viral Genetic Diversity: Helicases as Phenotypic Marker Genes”
Helicases are a large and diverse class of motor proteins required for many cellular functions and ubiquitous to every form of life on earth. Phylogenetic analysis and historical literature show that specific sub-types of helicases encoded by phage are structurally and functionally distinct. HeliBase, an in-house database of virus isolate helicases, was manually curated and used to query a viral metagenome assembly to investigate helicase–polymerase relationships in natural viral populations. Analysis of helicase–polA replication modules from metagenomic and reference databases shows a 77% co-occurrence of fast ring shaped helicases with fast polymerase types. Additionally, SNF2 helicase-like protein annotations were found often associated with slower polymerase types. A multitude of helicase sub-types investigated here show great promise as markers for genome to phenotype studies in environmental viruses.

21. Elena Diez Pastor, PhD Student, Plant & Soil Sciences, University of Delaware
“Optimizing the National Forest Inventory of Mexico with the Conditioned Latin Hypercube Sampling”
The optimization of environmental sampling is now more accessible thanks to the development of computational techniques. The National Forest Inventory of Mexico monitors and characterizes the country’s forest resources, despite experiencing a decrease in the funding necessary to carry out this activity. It is essential to help redesign their sampling strategy and reduce their monitoring effort, saving costs but recovering enough information on the measured variables. For this specific objective, we have applied a novel data-driven method for identifying optimized samples from spatial information of environmental data: the conditioned Latin Hypercube Sampling. We prove that it is possible to significantly reduce the sample size, still representing the behavior of the variables of interest and saving a relevant amount of resources. These results allow the survival of a project that helps to quantify how the ecosystem services that exist in Mexico contribute to alleviate climate change.

22. Iteoluwa Ibitoye, Undergraduate Student, Mathematical Sciences, Computer Science and Music, Lincoln University *
“Data Alignment Techniques in Railway Track Geometry Inspection: A Comparative Study”
Track geometry defects are major cause of train derailment incidents in the United States, accounting for a significant portion roughly 40% of them. Track geometry is characterized by parameters such as alignment, profile, gage, cross level and twist/warp. Railroads collect large size data on the condition of these track geometry parameters on a constant interval, near constant interval or temporally using track inspection vehicles. They analyze numerous inspection runs conducted on the same track over time to assess the degradation rates. The collected inspection data frequently suffer from longitudinal misalignment, sometimes spanning several hundred feet mainly due to operator errors in providing locations events. The goal of this research is to explore different data alignment techniques and compare their performance and applicability on a geometry data from specific track segment.

23. Ibrahim Wilson, Undergraduate Student, University Studies, University of Delaware *
“Gaussian Processes in the Legacy Survey of Space and Time” (NOTE: This poster has multiple authors: Authors: Ibrahim Wilson, Siddarth Chaini, Dwight Higgins, Federica Bianco)
Light curves, measures of time-dependent variations of the brightness of celestial objects, are crucial for studying objects like supernovas and variable stars. Astrophysical time series are typically irregular and sparse. The nightly and seasonal rotation of the sky prevents continuous monitoring of astrophysical sources, observations suffer from weather-related and telescope maintenance-related gaps as well as observing choices in the context of limited telescope resources available to monitor large samples of objects. However, most statistical methods for the analysis of time-series rely on regular and continuously sampled datasets. Gaussian Processes (GPs) are flexible probabilistic models that allow us to incorporate priors and estimate uncertainties on our interpolated values. By applying GPs, we aim to fill in the gaps in the data and enable the application of machine and deep learning methods for the investigation of the nature of celestial objects. We showcase their application of GPs to sparsely sampled simulated datasets for the upcoming Vera C. Rubin Observatory Legacy Survey of Space and Time. This research is partially supported by the NASA Space Grant and NSF HDR Data Science Corps Award Number 2123264.

24. Brandon Foreman, Undergraduate Student, Physics & Astronomy, University of Delaware *
“Lipid chain dynamics are chemistry dependent (Authors: Brandon Foreman, Shea Fitzgerald, and Edward Lyman)”
The dynamics of cellular membranes are important for many biological functions, but it is difficult to generalize from measurements on complex membranes as are found in living cells. Instead, we take a biophysical approach, and study the dynamics of simplified models, in order to understand how lipid chemistry controls membrane dynamics. Here, we present an analysis of classical molecular dynamics simulations of two different lipid membranes — a sphingolipid and a glycerolipid. Based on structural parameters (like hydrocarbon chain order and area per molecule) these two lipids appear interchangeable — and indeed have been treated as such for several decades. To our surprise, we find
significant differences in the dynamics of the hydrocarbon chains, as measured by the autocorrelations of chain isomerization. This observation offers a possible explanation for the significantly higher viscosity of sphingolipids compared to their glycerolipid analogues. This research is partially supported by NSF HDF DSC Award 2123264.

25. Alex Mulrooney, Undergraduate Student, Electrical & Computer Engineering, University of Delaware

“fMRI Encoding of Visual Cortex with Contrastive Learning”

Advances in computing power and neuroimaging data collection allow for new avenues of research into building computational models of the brain. Using the extensive Natural Scenes Dataset, we introduce a method of fine-tuning a pretrained convolutional neural network (CNN) using a contrastive loss function such to more closely align the CNN’s image representations to the brain’s representations. We train models with this approach that are specific to subjects and regions of interest (ROI). This method achieves significantly higher encoding accuracy as compared to both using the features from the untuned CNN as well as a CNN that is fine-tuned with a simple regression approach. We also show that the learned representations are moderately transferable across subjects. Furthermore, the CL-tuned CNNs suffer from less of a performance drop on the original training classification task as compared to the regression approach, as well as on other classification tasks not seen during training.

26. Spencer Toth, Undergraduate Student, Plant & Soil Sciences, University of Delaware

“Genomic Similarities Reflect Infectivity Patterns In Bradyrhizobium Rhizobacteriophage”

Bradyrhizobium is a nitrogen-fixing bacteria that plays a fundamental role in the nitrogen cycle by providing the host plant with a source of fixed nitrogen through symbiotic root nodules. However, Bradyrhizobium lytic phages infect the cells of its host by overtaking the host machinery for viral replication causing the cell to lyse. Whole genome analyses of Brady lytic phage isolates were conducted to show connections between their genotypic and phenotypic characteristics. Phage DNA was extracted and sequenced and genomic contigs were assembled and annotated. Genome visualization and pangenome analyses showed that viruses that infect B. elkanii were more similar to one another than the viruses infecting B. diazoefficiens. Further research will include investigating the evolutionary history of the annotated replication proteins shared amongst phages infecting the same host and the connection between structural proteins and the morphology of the viruses.

27. Feranmi Adepoju, Undergraduate Student, Mathematical Sciences, Computer Science and Music, Lincoln University

“Rubin Rhapsodies: Sonification of Astronomical Data for Enhanced Exploration and Accessibility”

While visualizations are the traditional means of data representation, sonifications are a less common but powerful alternative in which data features are mapped to sound characteristics such as pitch or volume. Sonification has been previously applied to astronomical images and time series data, but the Vera C. Rubin Legacy Survey of Space and Time (LSST) presents a new opportunity to sonify highly multidimensional astronomical data and accompanying metadata. We additionally represent moon brightness with a “drone,” a musical technique that establishes tonality and provides a foundation to the sound. By incorporating sonifications alongside traditional modes of data analysis, we propose that sonification can foster inclusivity and enable complementary methods of exploring the rich, complex, and multidimensional Rubin LSST dataset. This project is supported by NSF Award 2123264 and by the Heising Simons Foundation through grant 2021-2975, and administered by Las Cumbres Observatory.

28. Mitchel Igolimah, Undergraduate Student, Mathematical Sciences, Computer Science and Music, Lincoln University

“Comparative Study of Anomaly Detection Algorithms for Network Intrusion Detection”

The primary objective of this work is to evaluate unsupervised anomaly detection algorithms using distinct attack datasets. This endeavor sheds light on their performance, robustness, and applicability to real-world scenarios. We employed three well-known anomaly detection algorithms: Isolation Forest (iForest), k-Nearest Neighbors (kNN), and Local Outlier Factor (LOF). These algorithms were applied to an abridged version of an open network packet dataset from Darknet. Due to limitations of computational power, we used only 3000 records of network packet with 79 columns. Each row represents various attributes related to source and destination IP addresses, ports, protocols, and more. The 3000x79 data shape in our data is transformed into 3000x5641 during the preprocessing due to 6 categorical features. The original Darknet dataset encompasses 150K records.

29. Arthi Jayaraman, Professor, Chemical & Biomolecular Engineering, University of Delaware

[Title and abstract omitted from Symposium Proceedings]

30. Nii Tawiah, Assistant Professor, Sociology & Criminal Justice, Delaware State University

“Predicting Analysis of Opioid Consumption Patterns Among Diverse Ethnic Cohorts Using Machine Learning”
The opioid crisis continues to evolve with recent shifts in increasing opioid-related overdose deaths among Americans. Strikingly, opioid overdose has not only increased for all Americans but has become more pronounced among African Americans. More research is needed to better understand how this subgroup is using opioids so prevention and treatment efforts can be tailored to reach this population.

31. Rodiat Ayinde, Professor, Physics & Astronomy, University of Delaware
“Using MetaAI’s Insight Toward the Automated Detection of Rare Echoes of Stellar Explosions”
Light echoes are the reflection of stellar explosions on interstellar dust. They offer valuable information in the study of dust and transients, but their detection is challenging as they appear as extended, morphologically diverse, time-evolving transients often at the limit of surveys’ signal-to-noise ratio. Our group is testing and developing AI models to automate the detection and study of light echoes in the Rubin LSST data. The Large Synoptic Survey Telescope at the Rubin Observatory is expected to acquire an unprecedented 20TB of information-rich images each night, straining the astrophysical resources and imposing a need for the automated detection of anomalies and rare phenomena such as light echoes. We tested a recent MetaAI model “Segment Anything”, or SAM (Kirillov et al. 2023), to assess its suitability in the segmentation of template-subtracted astronomical images containing light echoes using the ATLAS dataset as a precursor of LSST.

*Denotes Poster Lightning Talk Participant
Institute for Engineering Driven Health

The UD Institute for Engineering Driven Health is a cluster of innovators who develop and translate new technologies to advance human health. This talk will introduce our activities that aim to catalyze idea generation and technology commercialization and to educate a new genre of trainees in translational research so that they can thrive in entrepreneurial or translational research-oriented careers.

Emily Day

Associate Professor, Biomedical Engineering
Associate Director,
Institute for Engineering Driven Health
University of Delaware

Emily Day is Associate Professor of Biomedical Engineering and Associate Director of the Institute for Engineering Driven Health at UD. She also holds a Joint Associate Professor appointment in Materials Science & Engineering at UD. Her research focuses on developing nanoparticle-based delivery vehicles that can transport medicines to diseased tissues with high precision to combat aggressive cancers, blood disorders, and reproductive health conditions. Prior to joining UD, Dr. Day earned her B.S. in Physics with a minor in Mathematics from the University of Oklahoma and her Ph.D. in Bioengineering from Rice University. She also
received postdoctoral training in the Department of Chemistry at Northwestern University. Since joining UD, Dr. Day has received notable awards and honors, including the Rita Schaffer Award from the Biomedical Engineering Society, the Gerard Mangone Young Scholars Award, and the Mid-Career Faculty Excellence in Scholarship Award from UD. She was named a Fellow of the American Institute for Medical and Biological Engineering in 2022.

The Research Expanding Access to Child Health (REACH)

The Research Expanding Access to Child Health (REACH) Center is an NIH Center for Biomedical Research Excellence (COBRE P20GM144270) at Nemours Children’s Hospital, Delaware. REACH focuses on developing, evaluating, and implementing innovative stakeholder-engaged technology-enhanced interventions and models of care to overcome the complex socioecological factors responsible for inadequate access to healthcare and child health. The REACH Center includes two Research Cores and funds multiple research and pilot projects. Three programs to develop investigators are based in the Administrative Core: the Mentoring Programs, the Project Funding Program, and the Research Management Program. The IMPACT Core provides resources including expertise, tools, training, and technical assistance to investigators developing and conducting programs of technology-enhanced intervention research to reduce disparities and improve the health and wellbeing of children and families across Delaware. The PROMISE Core creates an institutional infrastructure to support and empower communities and investigators through mutual engagement by removing barriers to authentic partnership and community-participatory research.

Kimberly Canter

Kimberly Canter, Ph.D., is a Pediatric Psychologist, Senior Research Scientist in the Nemours Center for Healthcare Delivery Science, and Associate Professor of Pediatrics at Thomas Jefferson University. She is also Core Director of the Intervention Methodology: Provision and Connection through Technology (IMPACT) Research Core of the Research Expanding Access to Child Health (REACH) Center, a newly funded Center of Biomedical Research Excellence (COBRE) based at Nemours Children’s Health in Delaware. Dr. Canter received her Ph.D. from the Clinical Child Psychology Program at the University of Delaware.
Kansas in 2015, and completed her clinical internship and postdoctoral research fellowship at Nemours Children’s Health. Her research focuses on developing and implementing evidence-based, technology-enabled psychosocial interventions to decrease traumatic medical stress and enhance coping skills for children with chronic illnesses and their families. She has a specific interest in and commitment to interventions that are wide-reaching and helpful in terms of increasing access and reducing healthcare disparities among underserved populations. Dr. Canter has received funding from a number of organizations, including the National Cancer Institute, the American Cancer Society, Alex’s Lemonade Stand Foundation, to support this work.

The Gerard J. Mangone Climate Change Science and Policy Hub
Global climate change touches every discipline in the University. Collaboration is crucial to fully understand climate change impacts and develop solutions for society and the planet. The Gerard J. Mangone Climate Change Science and Policy Hub promotes collaboration and brings together students, faculty, stakeholders and specialists to position the University of Delaware as a national leader in climate change education, research and innovation. Together we discover and implement solutions for a more sustainable tomorrow.

Dana Veron
Director
Gerard J. Mangone Climate Change Science and Policy Hub
University of Delaware

Dr. Dana Veron is an Associate Provost for Faculty Development, Director of the Gerard J. Mangone Climate Change Science and Policy Hub, and Professor of Geography and Spatial Sciences and Marine Science and Policy. Her research interests include climate change, off-shore wind resource assessment, sea breeze circulation, land surface-atmospheric interactions, stochastic radiative transfer, cloud-aerosol-radiation interactions, Arctic energy balance, and surface optical properties.
Data Science-FinTech Consortium
The Data Science Institute, in partnership with the Fintech Innovation Hub, the AI Center of Excellence, and the Lerner College, is launching a consortium open to for-profit and not-for-profit organizations, with the goal of facilitating access to the University’s data science resources and capabilities in support of member workforce and training needs and research opportunities.

Michael Blaustein
Industry Liaison, Data Science Institute
University of Delaware

Michael Blaustein serves as Industry Liaison of the Data Science Institute. He is Director of Business Intelligence and Commercialization Strategy in UD’s Office of Economic Innovation and Partnerships. Prior to joining the University, he held a variety of business and technology roles at DuPont, with a strong focus on strategy, innovation, and technology commercialization.
**Lightning Talks Part II**

**Nii Tawiah (Moderator)**  
_Assistant Professor_  
_Sociology and Criminal Justice_  
_Delaware State University_

---

**David Hong**  
_Assistant Professor_  
_Electrical & Computer Engineering_  
_University of Delaware_

**“Making Sense of Big Data with Matrix and Tensor Decompositions”**

This lightning talk gives a high-level introduction to matrix and tensor decompositions, which are unsupervised and general-purpose methods for making sense of big data. Numerous big datasets across science and engineering are in the form of matrices (i.e., 2-D arrays) and tensors (i.e., N-D arrays). We will show a demo on crime data to illustrate how matrix and tensor decompositions can be used to uncover underlying patterns (i.e., signals) and make sense of the data.

---

**Michael Liebman**  
_IPQ Analytics LLC_

**“Precision Medicine is Dependent on Accurate Medicine”**

Clinical diagnosis connects patient care/disease management, drug and diagnostic development and payer reimbursement, yet accuracy remains elusive. Misdiagnosis impacts approximately 12M patients annually in the US, causing 40k-80k deaths and costs the US economy approximately $750 B annually. There are numerous causes but a critical component is lack of critical thinking to understand and evaluate disease itself. Disease is a process, not a state, requiring re-evaluation of how to apply technologies to address critical questions not adequately asked because of real world complexity. Clinical guidelines and risk scores attempt to simplify decision-making in diagnosis and treatment but do not recognize the complexity of a real-world patient, of the disease process and factors that affect it, nor of influences affecting the practice of medicine. Critical thinking is being applied in several disease areas, e.g., multiple sclerosis, and infant/maternal morbidity and mortality.
Neil Mathur
Innovative Precision Health (IPH)
“Innovative Precision Health”
Innovative Precision Health is a data aggregator and analytics company operating in the healthcare sector. The company collaborates with both digital device vendors and clinics in order to mine patient population data and glean insights that could be used for optimizing patient care.

Vishal Saxena
Associate Professor
Electrical & Computer Engineering
University of Delaware
“Energy-Efficient Neuromorphic System-on-a-Chip (NeuSoC) for Artificial Intelligence at the Edge”
In this talk, the speaker will provide an overview of his ongoing research in brain-inspired integrated circuits (ICs) or chips for realizing pervasive AI on the edge devices. Such architecture is known as Neuromorphic Computers and utilize non-volatile memory (NVM) arrays in a way similar to the synapses in the biological brain. These architectures leverage analog computing as opposed to digital and perform deep neural network (DNN) computations within the memory array. Thus, the elements that perform compute and storage are one and the same and leads to energy savings that would otherwise be lost in shuttling data back and forth between the processor and external memory. This talk will provide a quick overview of analog neuromorphic computing hardware, the challenges due to the non-ideality of the memory devices, and their future prospects in solving the energy and carbon footprint challenge of widespread AI.

Ulf Schiller
Associate Professor
Computer & Information Sciences
University of Delaware
“Data-driven Materials Discovery and Design”
The quest for discovering and designing materials with desired properties has traditionally been a labor-intensive and time-consuming endeavor. However, the rapid advances of ML/AI approaches has ushered in the new paradigm of harnessing the power of materials data for accelerated materials discovery and design. In this lightning talk, I will outline how the integration of high-throughput experiments, computational simulations, and materials data collection paves the way to solving inverse problems, where the goal is to predict the material composition and processing conditions that yield specific desired properties. As a specific example, I will showcase our NSF DMREF project that targets the development of experimentally guided, direct and inverse sintering models taking into account specifics of micro- and macro-structure in sintering assisted additively manufacturing (SAAM) processes.

Puranjit Singh
PhD Student
Plant & Soil Sciences
University of Delaware
“A Deep Learning-Based Smartphone App for Field-Based Blueberry Yield Prediction”
Blueberry production has experienced a significant surge, with the US leading as the top producer with an impressive revenue of $932 million in 2020. The benefits of blueberries, including their antioxidant-rich,
nutrient-rich, and anti-inflammatory properties, contribute to this growth. Accurate blueberry yield and maturity estimation in field using traditional methods is a labor-intensive task. We developed a smartphone android application leveraging deep learning techniques to automate yield prediction for different blueberry cultivars. Lightweight DL models of EfficientNet were fine-tuned on a dataset of various southern highbush and rabbiteye cultivars. The trained model achieved an AP of 0.681 and 0.344 under 0.5 and 0.75 IoU thresholds during testing. Blueberry yield was predicted by ML methods using mature berry count and user-defined average berry weight with satisfactory accuracy. The android app can enable blueberry growers and breeders to assist management decision making.

**Troy Wilford**  
*Director of UD Commercialization Programs*  
*University of Delaware*  
*“The Innovation Delaware (InDE) Fellowship”*

As the creative engine for entrepreneurship education and advancement at UD, Horn Entrepreneurship empowers aspiring innovators from every discipline as they pursue new ideas for a better world. During this session, you will learn about a new program that provides inventors, researchers, and innovators the opportunity to develop critical knowledge and entrepreneurial skillsets through experiential learning.
**HACKATHON PRESENTATION/AWARDS CEREMONY**

**Ben Bagozzi (Moderator)**
Chair Data Science Symposium Planning Committee  
Department of Political Science & International Relations  
University of Delaware  
https://www.poscir.udel.edu/people/faculty/Bagozzib

**Kendal Transformers**
Technical Complexity Award  
**Title:** Predicting the Need for Long Term Care Among Senior Citizens Using Machine Learning

**Team Members**
Nikhil Rao, Master’s Student, Electrical and Computer Engineering  
Mauricio H. Ferrato, PhD Student, Computer and Information Sciences  
César Claros-Olivares, PhD Student, Electrical and Computer Engineering Department  
Jason Eldridge, BSN, RN, Business Systems Analyst, Kendal Corporation  
Kurt Rahner, MBA, MCS, VP of Technology, Kendal Corporation

**Abstract:** Understanding the needs of seniors as they progress through the aging process can significantly enhance healthcare providers’ ability to cater to the appropriate services. Knowing if they will require ongoing long-term care in the future and the variables that substantiate this prediction can potentially improve the quality of the service. In this regard, machine learning could help providers better understand the populations they serve to discern the likelihood of those individuals needing long-term care, which is challenging given the unbalanced nature of the problem. We analyzed survey data from the RAND Corp. Health and Retirement Study, considering cognitive, physical and financial information, among others, and developed a model that classifies older adults as high risk/low risk of requiring long-term care in the future.
We also analyzed the specific domain areas identified by the model that increase the risk. Some of the techniques applied in this project include supervised feature transformation (weight of evidence), feature selection (LASSO), logistic regression, and analysis of precision and recall metrics for performance. Preliminary results show promising predictive performance with a weighted average precision of 0.5 in the precision-recall curve.

Medical Marauders
Impact Award
Title: Towards Interpretable Machine Learning for U.S. Hospitals' CMS Ranking

Team Members
Mina Ostovari, Team Lead, DSI
Andrew Kallai, Undergraduate Student, Computer and Information Sciences
Talha Mahmood, Undergraduate Student, Computer and Information Sciences
Shreya Pamulapati, Undergraduate Student, Computer and Information Sciences
Shaquana Smith, Undergraduate Student, Department of Linguistics & Cognitive Science

Abstract: In this study, we aimed to predict U.S. hospital rankings using Centers for Medicare and Medicaid (CMS) Hospital Compare Data. Hospital rankings, derived from five measure groups (Mortality, Safety of Care, Readmission, Patient Experience, and Timely & Effective Care), summarize care quality and performance. CMS's current ranking methodology involves standardizing measures, aggregating scores, applying weights, and utilizing clustering for star ratings. Hospitals need a minimum of three measure groups with three measures each, including Mortality or Safety of Care, for ranking. Given the spatial distribution of hospitals across the US, as well as the number of available measures, we leveraged non-spatial and spatial random forests (RFs) for prediction, considering hospital zip codes for our spatial units. After removing missing data, the RF model outperformed its spatial counterpart in predictive accuracy. The models highlighted death rate for pneumonia patients and readmission for heart failure patients as the most important variables. By comparison, the ratio of unplanned hospital visits after outpatient surgery and the percentage of outpatients with cardiac imaging stress tests before low-risk surgery were identified as the least important among variables considered. Our study advances hospital quality assessment by developing interpretable machine learning techniques for the modeling of hospital star ratings, providing a comprehensive understanding of influential predictors, and enhancing the accuracy of predictive models for hospital rankings.
Chemours Innovators
Problem Statement Execution Award
Title: A Self-optimized Impedance Fitter from Chemours Innovators

Team Members:
Maria van Venrooy, Data Scientist, The Chemours Company; UD CS ’22
Austin Plymill, Senior Staff Scientist, The Chemours Company; UD CS ’22
Yiqin Cao, Master’s Student, School of Education
Hang Chen, PhD Student, Computer and Information Sciences
Nihaal Chowdary Surpani, Master’s Student, Data science
Lalith Teja Nagidi, Master’s Student, Data Science

Abstract: Our project extends the National Renewable Energy Laboratory’s Open-Source Impedance Fitter (OSIF), enabling it to autonomously optimize model parameters for electrochemical impedance spectra from Proton-exchange membrane fuel cells in hydrogen/nitrogen environments. These parameters align with quasi-transmission line, one-dimensional linear diffusion, or spherical diffusion models. The result is an upgraded software version featuring a user-friendly interface (UI) for streamlined functionality. It excels at automatically applying optimal model parameters to additional datasets, significantly reducing the need for manual trial and error. This innovation stands to save valuable time for chemical engineers and researchers. Our software empowers swift modeling and analysis of Proton-exchange membrane fuel cell impedance data, enhancing research and development in electrochemistry. Its accessibility through a user-friendly interface broadens its utility, facilitating broader exploration and adoption of these critical technologies. Our project represents a crucial advancement in fuel cell technology research and engineering efforts.
Rain Makers
Creativity Award

Title: Generating a High-quality, Fine-scale Precipitation Dataset for the Great Lakes Region Building upon Existing Datasets

Team Members:
Siamak Malakpour Estalaki, PhD Student, Department of Geography and Spatial Sciences
Amirreza Meydani, PhD Student, Department of Geography and Spatial Sciences
Mahadev Maitri, Master’s Student, Computer Science
Aman Jot Singh, Undergraduate Student, Computer Science
Christian Munley, Undergraduate Student, Computer Science and Physics
Jianbo Zhang, Master’s Student, Data Science
Yao Hu, PhD, Assistant Professor, Department of Geography and Spatial Sciences (Primary); Department of Civil and Environmental Engineering

Abstract: The changing climate has significant effects on the patterns of rainfall, causing widespread impacts on water resources, flood risks, agricultural practices, and the health of ecosystems around the world. This impact is particularly noticeable in regions such as the Great Lakes, where alterations in precipitation patterns contribute to an elevated risk of flooding and a decline in water quality. For a comprehensive understanding and prediction of these environmental hazards, it is crucial to have access to precise and dependable precipitation data. However, current statistical and machine learning algorithms used to generate such datasets have limitations when it comes to dealing with various challenges, including spatial accuracy, complex non-linear precipitation patterns, distant connections between data points, and uncertainties in predictions. To tackle these constraints, we adopted generative artificial intelligence models to create new precipitation data specifically for the Great Lakes area, building upon existing precipitation datasets. Preliminary findings indicate that the newly generated dataset of precipitation performs better than the pre-existing ones. This success serves as an illustration of the potential of generative AI models in producing finely detailed, high-quality precipitation data across expansive geographical regions. Additionally, these models offer valuable insights into the dependability of predictions. Ultimately, the enhancement of precipitation data accuracy, which stands as a foundational element for various physics-driven numerical models, will contribute to heightened precision in forecasting the diverse array of physical processes influenced by rainfall.
POSTER AWARDS/CLOSING CEREMONY

Cathy Wu
Director, Data Science Institute
University of Delaware
https://dsi.udel.edu/

Cathy Wu and Ben Bagozzi, along with Poster Committee Co-Chairs Daria Blinova and Hanan Abou Ali, Confer the 2023 Symposium’s Best Poster Awards to Award Winners and/or their Faculty Representatives
Symposium Chair Ben Bagozzi gives introductory remarks (left) and Associate Vice President of Corporate Engagement, Tracy Shickel gives opening remarks (right)

Dr. Andrea Parker offers her keynote talk to the Symposium’s Audience
Federica Bianco introduces the Data Science-driven Equity from Healthcare, FinTech, Community, & Educational Perspectives panel (left), and participants Ryan Harrington, Phil Goldfeder, Susan Conaty-Buck, and Claude Temeze offer perspectives (right)

Susan Conaty-Buck offers her perspective alongside Data Science-driven Equity from Healthcare, FinTech, Community, & Educational Perspectives panel participants Phil Goldfeder, Claude Temeze, and Emily Kauffman (left), and Emily Kauffman offers her perspective alongside panelists Susan Conaty-Buck and Claude Temeze (right)
Dr. Hubert Niewiadomski offers his keynote talk to the Symposium’s Audience

Emily Day discusses the University of Delaware’s Institute for Engineering Driven Health (IEDH) (left), while Kimberley Canter introduces Nemours’ Research Core of the Research Expanding Access to Child Health (REACH) Center (right)
POSTER SESSION PHOTOS
DATA SCIENCE SYMPOSIUM PLANNING COMMITTEE

Benjamin Bagozzi
Committee Chair

Cathy Wu
Director, DSI

Hanan Abou Ali

Michael Blaustein

Daria Blinova

Jeffrey Buler

Lynette Carney

Sunita Chandrasekaran
DATA SCIENCE SYMPOSIUM PLANNING COMMITTEE

(Cont’d)

Adam Davey
Adam Fleischhacker
Mina Ostovari

Daniel Peart
Sam Smith
Nii Tawiah
THANK YOU TO OUR SPONSORS

Visit the Data Science Institute at:
https://dsi.udel.edu/

JOIN US ON TWITTER

@dsiudel

twitter

#DataScienceUD23