UNIVERSITY OF DELAWARE
DATA SCIENCE INSTITUTE

DELAWARE Data Science
DARWIN COMPUTING SYMPOSIUM

FEB 12, 2024

#DARWINcomputing2024

Delaware Data Science DARWIN Computing Symposium FEB 12, 2024

University of Delaware Data Science Institute

Agenda

Registration and Poster Setup	**8:30 – 9:00 AM**
Introduction – Welcome Remarks: Tom Hsu, Committee Chair	**9:00AM**
Remarks from Angela Chen, Chief Information Officer UD IT	**9:10AM**
Remarks from Bill Miller, NSF-AOC	**9:20AM**
Keynote – Tanya Berger-Wolf, Ohio State University	**9:40AM**
Coffee Break	**10:25AM**
Session I – Broadening Participation: Regional/National Resources	**10:40AM**
Lunch	**12:15 PM**
Poster Lightning Talks	**12:50PM**
Poster Session	**1:15PM**
Session II – Science on DARWIN	**2:15PM**
Coffee Break	**3:15PM**
Panel: Beyond DARWIN: The Next Generation	**3:30PM**
Best Poster Award & Closing Remarks	**4:30PM**

WELCOME REMARKS

Tian-Jian "Tom" Hsu

Chair DARWIN Computing Symposium Planning Committee
Professor, Civil & Environmental Engineering
Director of Center for Applied Coastal Research
University of Delaware

REMARKS FROM UD IT

Angela Chen

Vice President of Information Technologies, CIO
University of Delaware

REMARKS FROM THE NSF

William Miller

Senior Advisor, Office of Advanced Cyberinfrastructure (OAC) , National Science Foundation

"Update on the National AI Research Resource (NAIRR) Pilot"

KEYNOTE SPEAKER

Rudi Eigenmann *(Moderator)*

DARWIN Principal Investigator (PI)
Director, Data Intensive & Computational Science (DiCoS) Core
Distinguished Professor, Electrical and Computer Engineering
University of Delaware
https://www.ece.udel.edu/people/faculty/eigenman/

Tanya Berger-Wolf

Professor, Computer Science and Engineering, Electrical and Computer Engineering, Evolution, Ecology and Organismal Biology
Director, Translational Data Analytics Institute
The Ohio State University

"AI for Biodiversity: Combatting Extinction Together"

Abstract: We are in the middle of the 6th extinction, losing the planet's biodiversity at an unprecedented rate and scale. In many cases, we do not even have the basic numbers of what species we are losing and how many. New data collection technology, such as GPS, high-definition cameras, UAVs, genotyping, and crowdsourcing, are generating data about the living planet that are orders of magnitude richer than any previously collected. AI can turn these data into high resolution information source about living organisms, enabling scientific inquiry, conservation, and policy decisions. The talk will present a vision and examples of trustworthy AI for biodiversity, discussing opportunities and challenges.

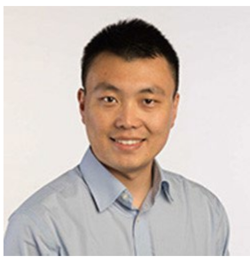
Bio: Dr. Tanya Berger-Wolf is a Professor of Computer Science Engineering, Electrical and Computer Engineering, and Evolution, Ecology, and Organismal Biology at the Ohio State University, where she is also the Director of the Translational Data Analytics Institute. She is leading the US National Science Foundation funded Imageomics Institute and the newly funded AI for Biodiversity Change Global Climate Center.
Berger-Wolf is a member of the US National Academies Board on Life Sciences and CNRS International Scientific Advisory Board, Artificial Intelligence for Science, Science for Artificial Intelligence (AISSA) Centre. She served on the Global Partnership on AI (GPAI) AI on Biodiversity working group, WWF working group on

AI Collaboration to End Wildlife Trafficking, AAAS-FBI Big Data in the Life Sciences and National Security Working Group, and the organizing committee of the National Academies First U.S.-Africa Frontiers of Science, Engineering, and Medicine Symposium, among many others. She has received numerous awards for her research and mentoring and given hundreds of talks.

Berger-Wolf is also a director and co-founder of the AI for conservation non-profit Wild Me, home of the Wildbook project, which has been chosen by UNSECO as one of the 100 AI projects worldwide supporting the UN Sustainable Development Goals.

SESSION I – HPC AND DATA SCIENCE

Marianna Safronova *(Moderator)*
Professor, Physics & Astronomy
University of Delaware



Ruo-Qian "Roger" Wang
Assistant Professor, Civil & Environmental Engineering
Rutgers University
https://whirlab.onrender.com.

"AI-Powered Coastal Flooding Analysis: Leveraging Computer Vision and Large Language Models for Informed Decision-Making"

In the dynamic realm of coastal research, the convergence of high-performance computing (HPC), artificial intelligence (AI), and environmental data analysis promises a groundbreaking transformation in our ability to confront pressing challenges, foremost among them being coastal flooding. This presentation unveils the pioneering applications of computer vision, AI, and Large Language Models (LLM) that empower us to harness and decipher extensive flood-related datasets. In this presentation, I will present our recent efforts in using social media and monitoring cameras to recognize and quantify coastal flooding. Through the integration of monitoring cameras and AI, we demonstrate how large-volume and unstructured data can be used to contribute to a comprehensive understanding of coastal flooding dynamics. Furthermore, we explored the frontier to apply Large Language Models such as ChatGPT in support of decision-making. The developed framework has the potential to enhance the communication and data digestion involved in intricate flooding decisions, bridging the gap between technical experts, stakeholders, policymakers, and the wider public. Our findings showcase the power of language models to distill complex data into actionable knowledge, fostering greater awareness and informed decision-making.

Jing Gao

Assistant Professor, Geography & Spatial Science, Data Science, Computer Science, Environmental Engineering
University of Delaware

“Data-driven Spatiotemporal Modelling for Long-Term Large-Scale Human-Environment Interaction Studies”

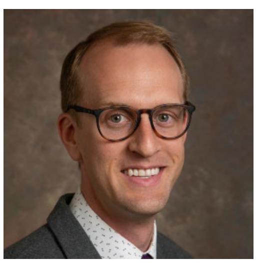
Jing Gao is a transdisciplinary scholar integrating data science, urban sustainability, and climate change. Her research examines and models long-term, large-scale spatiotemporal dynamics between human population and climate change, using new, diverse data through computation. She is the inaugural chair of the External Advisory Board to the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) Program of NSF (successor of XSEDE), a chapter author of the U.S. Fifth National Climate Assessment (NCA5), an expert reviewer for the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), and an NSF CAREER awardee. Dr. Gao is an Assistant Professor at the University of Delaware (UD), in Geography, Data Science, Computer Science, and Environmental Engineering. Before joining UD, she was a research scientist at the National Center for Atmospheric Research (NCAR), following a postdoctoral fellowship at the National Center for Supercomputing Applications (NCSA), and a PhD from the University of Wisconsin – Madison.

Gulnihal Ozbay

Associate Dean for Extension and Applied Research,
Professor, Natural Resources
Delaware State University

“Investigating Turbidity/Total Suspended Relationships and Aragonite Saturation State in Rehoboth Bay Aquatic Systems for Sustainable Shellfish Aquaculture Industry”

Turbidity resulting from high concentrations of suspended solids acts as an important indicator for potentially undesirable conditions in aquatic habitats that support ecologically and economically significant species like oysters and blue crabs. Quantitative data for suspended solids is difficult to directly measure in the field and thus often discussed synonymously with turbidity. Aragonite Saturation State is also an important indicator for calcifying species because it is a measure of the calcium carbonate concentration in the water which oysters and blue crabs use to form their outer shells. The objectives of this study were to 1) investigate the correlation between turbidity and total suspended solid metrics within the Rehoboth Bay region of Delaware, 2) improve methodology to allow for greater quantities of total suspended solid data to be collected in the field, and 3) assess aragonite saturation state values to determine its impacts to calcifying species. Field data for turbidity, salinity, total dissolved solids, dissolved oxygen, chlorophyll, and pH were collected at multiple sites in Rehoboth Bay through use of a YSI 556 Multiprobe, ProDSS, YSI 6 series sonde, YSI Exo2 sonde, and Aquafluor fluorometer. Water samples were also collected from each site. These field samples were then processed in Delaware State University's ONE Health Lab for total suspended solids through the standard method, total alkalinity, total calcium hardness, and nutrient content using a 9500 YSI photometer. Regression analysis and analysis of variance (ANOVA) were then conducted on the data collected. These analyses show a more accurate correlative relationship between total suspended solids and turbidity; thus, allowing for quicker identification of changes to and around aquaculture habitat. These same water quality parameters were also inputted into R programming to calculate the aragonite saturation state at the different site locations. With the new oyster aquaculture blooming in Delaware waters, water quality data collection has become more important to sustain and grow shellfish farming and help the resource managers make informed decisions on policies supporting shellfish farming.



Kyle Davis
Assistant Professor, Environmental Data Science
University of Delaware

"AI Approaches for Achieving Sustainable and Resilient Food Systems"

Sustainable food systems aim to provide sufficient and nutritious food, while maximizing climate resilience and minimizing environmental impacts. Yet historical practices, notably the Green Revolution, prioritized the single objective to maximize production over other nutritional and environmental dimensions. With a focus on the food-water-climate nexus, my research explores AI-enabled solutions to maintain humanity's achievements while overcoming past shortcomings of global food systems.

SESSION II

SCIENCE ON DARWIN

Benjamin Bagozzi *(Moderator)*
DARWIN Co-PI
Associate Director, Data Science Institute (DSI)
Associate Professor, Political Science & International Relations
University of Delaware

Carolina Perez Segura
PhD Candidate, Chemistry and Biochemistry
Hadden-Perilla Research Laboratory
University of Delaware

"Atomistic Simulations as a Tool to Optimize Hepatitis B Assembly Inhibitors"

The Hepatitis B virus (HBV) poses a significant global health threat, affecting over 296 million individuals, despite the availability of a vaccine. Chronic HBV infection lacks a definitive cure, emphasizing the need for new therapeutic approaches. The HBV capsid, a protein shell comprising 120 dimeric core protein (Cp) subunits, represents a promising drug target. During infection, the capsid self-assembles to encapsulate the viral genome, leading to the production of infectious virions. Consequently, disrupting capsid assembly offers an opportunity to impede viral progression. Certain small molecules, called CAMs, have shown potential in misdirecting, or disrupting HBV capsid assembly by binding between Cp subunits. Here, we performed microsecond-long molecular dynamics simulations of the HBV capsid bound to CAMs and extracted 48 million snapshots of the CAM binding sites. Our results – based on processing 10 TB of data for eight different capsid simulations – reveal that the CAM binding sites are far more dynamic than anticipated. We found that the volume of the binding sites can range from being completely closed to twice the size captured by experimental structures. Further, we found that the binding sites can adopt a wide range of 3D shapes, and our extensive dataset enabled us to determine which shapes are most probable. Characterizing the topology of the CAM binding site is valuable for designing and optimizing new HBV capsid assembly inhibitors.

Based upon joint work with Boon Chong Goh and Jodi A. Hadden-Perilla.

Vineeth Gutta
PhD Candidate, Computer & Information Science
Computational Research and Programming Lab
University of Delaware

"DARWIN-driven Innovation: A Showcase of HPC Research at CRPL"

High Performance Computing (HPC) is a common element of all the research and innovation at the Computational Research and Programming Lab (CRPL) lead by Dr. Sunita Chandrasekaran. DARWIN is crucial for the various research projects including building testsuites to validate and verify OpenMP and OpenACC programming model features, exploring open-source compiler ecosystem such as LLVM, the MLIR framework, experimenting the migration of real-world applications such as sequence alignment algorithms, improving modeling and portability of drug discovery ML models, accelerating plasma physics code, and exploring LLMs for OpenACC and OpenMP test generation. Hardware accelerators such as GPUs are important for many of the projects above and optimizing codes and models for GPUs is a priority. DARWIN's NVIDIA V100 GPUs have been vital for such tasks. In addition to DARWIN's compute resources, software support in particular is an important enabler of cutting edge research at CRPL. In this talk I discuss in depth the role DARWIN plays in each of the major research initiatives in our group

Based upon joint work with Sunita Chandrasekaran.

Nitant Gupta
Postdoctoral Researcher, Chemical and Biomolecular Engineering
Jayaraman Lab
University of Delaware

"Computational Reverse Engineering Analysis of Scattering Experiments Method for Analyzing 2D Small-Angle Scattering Profiles (CREASE-2D)"

Characterization of structural diversity within soft materials is needed for engineering new materials for a variety of applications. Small-angle scattering (SAS) is a widely used characterization technique that provides structural information of soft materials at varying length scales, by recording the scattered intensity I(q,θ) as a 2D profile which is a function of the scattered wave vector represented by its magnitude q and azimuthal angle θ. While isotropic structures can be interpreted from azimuthally averaged 1D SAS profile, to understand anisotropic spatial arrangements, one has to interpret the 2D SAS profile, I(q,θ). In this talk, I will present our recently developed method called CREASE-2D that interprets I(q,θ) as is and outputs the relevant structural features. CREASE-2D is an extension of the 'computational reverse engineering analysis for scatting experiments' (CREASE) method that has been used successfully to analyze 1D SAS profiles for a variety of soft materials. CREASE uses a genetic algorithm for optimization and a surrogate machine learning (ML) model for fast calculation of 1D 'computed' scattering profiles that are then compared to the experimental 1D scattering profiles during optimization. In CREASE-2D, which goes beyond CREASE in interpreting 2D scattering profiles, we use XGBoost as the surrogate ML model to relate structural features to the I(q,θ) profile. The CREASE-2D workflow identifies the structural features whose computed I(q,θ) profiles match the input experimental I(q,θ). We test the performance of CREASE-2D by using as input a variety of in silico 2D SAS profiles with known structural features and demonstrate that CREASE-2D converges towards their correct structural features. We expect this method will be valuable for materials' researchers who need direct interpretation of 2D scattering profiles to explore structural anisotropy.

Based upon joint work with Sri Vishnu Vardhan Akepati and Arthi Jayaraman.

PANEL DISCUSSION: Beyond DARWIN

John Huffman *(Moderator)*

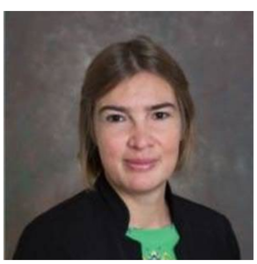
Director
IT Research Cyberinfrastructure
University of Delaware

Jing Gao

Assistant Professor, Geography and Spatial Science
University of Delaware

Sunita Chandrasekaran

Co-Director, AI Center of Excellence
Associate Professor, Computer and Information Sciences
University of Delaware



Marianna Safronova

Professor, Physics & Astronomy
University of Delaware

Robert Henschel

Project Director, Research Engagement, Research Technologies
Pervasive Technology Institute
Indiana University

POSTER SESSION

1. **Sri Vishnuvardhan Reddy Akepati, Master's Student, Chemical & Biomolecular Engineering, University of Delaware**
   *Title***:** Machine Learning Enhanced CREASE Method for Analyzing 2D Small Angle Scattering Profiles
   Abstract: Understanding structural diversity in polymeric materials is a key step towards engineering new materials. One way to probe structures at varying length scales is through small angle scattering; where a typical measurement provides as output the scattered intensity (I) vs. the magnitude of the wavevector (q) and angle (ϴ). To overcome some challenges in interpreting this 2D I(q, ϴ), especially for structures exhibiting anisotropy, we present a machine learning boosted 'computational reverse engineering analysis for scatting experiments' (CREASE) method. The chosen ML model, XGBoost, is trained to relate structural features to the I(q, ϴ) profile. Using the trained model within the CREASE workflow, we accelerate the identification of structural features whose computed I(q, ϴ) matches the experimental I(q, ϴ). This streamlined XGBoost-CREASE methodology eliminates complexities in manual interpretation and provides an efficient way to understand structural diversity in soft materials.
2. **Willow Fortino, PhD Student, Physics & Astronomy, University of Delaware**
   Title: How Long Can We Go? Minimum Sprectroscopic Requirements for Supernova Subtype Classification
   Abstract: Millions of supernovae will be discovered with the Rubin Observatory Legacy Survey of Space and Time (LSST). As a result, spectrographs around the world will have to make difficult decisions about which supernova candidates receive spectroscopic follow-ups. This work identifies the minimum spectral resolving power, R, as a function of signal-to-noise-ratio at which spectral classification of supernova subtypes becomes impossible. We developed a novel method for simulating future low-resolution spectra by degrading existing high-resolution data. Our dataset includes 16 different supernova types including subtypes of Ia, Core Collapse and interacting supernovae. We hope that target observation managers associated with the LSST alert stream, as well as designers of future observatories, will benefit from knowing what spectral resolution is necessary to classify a supernova for arbitrary SNR.
3. **Federico Garcia, PhD Student, Physics & Astronomy, University of Delaware**
   Title: Nonclassical Dynamics and Ensuing Radiation in Light-Driven of Strongly Correleated NiO
   Abstract: Ultrafast light-induced spin and charge excitation and resulting magnetization dynamics studies for antiferromagnetic materials like NiO with strongly-correlated electrons are still in its infancy. In NiO, the magnetic moments can no longer be treated classically. Here we use multiscale quantum-classical formalism where conduction electrons are described by the quantum mechanics, while outgoing electromagnetic radiation is computed using Jefimenko equations for retarded electric and magnetic fields. We apply this formalism to Mott insulator NiO with Rashba spin-orbit coupling to simulate proximity effects of a heavy metal like Pt to compute the time-dependent magnetization and entanglement of the system and from the pumped charge currents we compute the ensuing THz radiation.
4. **Hrishikesh Kamath, Master's Student, Electrical & Computer Engineering, University of Delaware**
   Title: Performance Evaluation of Machine Learning Models for Drug Discovery on DARWIN
   Abstract: In this project, we validate and port machine learning models used for drug discovery. We also assess the computational time of these models on DARWIN. This involved examining the enhancements in performance across different compute modes available on DARWIN.
5. **Jae Kim, PhD Student, Chemical & Biomolecular Engineering, University of Delaware**
   Title: Application of Similarity-enabled Machine Learning (ML) on Small Datasets
   Abstract: Machine learning (ML) has been successfully applied to learn patterns in experimentally generated chemical data to predict molecular properties. However, experimental measurements can be expensive and, as a result, experimental data for several properties is scarce. Several ML methods face challenges when trained with limited data. Here, we introduce a similarity-based ML methodology to efficiently train ML models on small datasets. We group molecules with similar structures, represented by molecular fingerprints, and use these groups to train separate ML models. We apply the methodology to predict viscosities of biomass-derived molecules. Our method shows noticeable improvement in model performance compared to the model trained on the entire dataset. Our methodology provides a robust framework for scenarios with limited data can be readily generalized to a diverse range of molecular datasets.
6. **Christian Munley, Undergraduate Student, Computer & Information Sciences, University of Delaware**
   Title: LLM4VV: Developing LLM-Driven Testsuite for Compiler Validation
   Abstract: Large language models (LLMs) are a new and powerful tool for a wide span of applications involving natural language and demonstrate impressive code generation abilities. In this paper, we explore the capabilitity of state-of-the-art LLMs, including closed-source options like OpenAI GPT-4 and open-source alternatives like Meta AI Codellama, to automatically generate tests and use these tests to validate and verify compiler implementations of a directive-based programming paradigm, OpenACC. Our approach entails exploring various prompt engineering techniques including a code template, retrieval-augmented generation (RAG) with code template, expressive prompt using RAG with code template, one-shot example, and RAG with one-shot example. This paper focusses on (a) exploring the capabilities of the latest LLMs for code generation, (b) investigating prompt and fine tuning methods, and (c) analyzing the outcome of LLMs generated tests.

7. **Mohammadsadegh Nouri, PhD Student, Civil & Environmental Engineering, University of Delaware**
Title: Novel Eulerian Two-phase Numerical Simulation Tool for Scour Burial of Munitions
Abstract: We present research highlights for the development and validation of the high fidelity Eulerian two-phase model, SedFoam, to simulate mobility and burial dynamics of munition, modeled here as a 3D short cylinder, under waves and currents. We carried out model validation for the mobility/burial of a short cylinder driven by an accelerating current, which is similar to a comprehensive laboratory experiment reported by Rennie et al. (2017, Ocean Engineering). The model is able to reproduce the rolling (no burial), mobile and buried, and locally buried scenarios for a given flow acceleration and cylinder densities. The accuracy of predicted critical velocity for cylinder motion is within 10% of observed value. We also apply the model to simulate highly energetic sediment transport events (sheet flow) where rapid burial of high-density short cylinders is observed. Preliminary results suggest the model is able to predict the rapid burial and the sensitivity to munition density.

8. **Mateo Mezic, PhD Student, Civil & Environmental Engineering, University of Delaware**
Title: MateoTsunami Prediction for the Eastern United States
Abstract: Meteorological tsunamis, or MeteoTsunamis for short, pose a significant natural hazard risk to coastal regions. Unlike seismic tsunamis, MeteoTsunamis result from atmospheric disturbances resonating with water surfaces under specific conditions. Proposed is a model that uses weather radar data to predict MeteoTsunami events, focusing on the eastern coast of the United States. This work uses real-time radar images to predict MeteoTsunamis. Although typically overshadowed by large scale seismic Tsunami events, MeteoTsunamis under conditions of high tide, storm surge and high mean sea level can create record water levels. Tsunami warning systems have increased significantly in recent years, however systems for predicting MeteoTsunami events and their severity are in the early stages of development. Modeling the risk of these weather events can not only improve our understanding of the physical processes responsible, but can also help protect human lives and property from damage.

9. **Jay Shah, PhD Student, Chemical & Biomolecular Engineering, University of Delaware**
Title: Development of Machine Learning Approaches for Analyzing Data from Experimental Measurements
Abstract: Semi-crystalline polymers hold significant importance in materials science due to their unique combination of crystalline and amorphous regions, providing mechanical strength, thermal stability, and processing versatility. Characterizing new semi-crystalline polymeric materials poses several challenges particularly when 100 percent crystalline polymer structure has not yet been observed. Some of the widely used characterization methods for quantifying percentage crystallinity in semi-crystalline polymers are atomic force microscopy (AFM), polarized optical microscopy (PLOM), wide-angle X-ray scattering, infrared spectroscopy and differential scanning calorimetry (DSC). This poster introduces our development of machine learning (ML) models to extract information from AFM images, PLOM images and DSC curves. These ML models identify domains and texture features in the images and establish their relationships with percentage crystallinity calculated from DSC curves.

10. **Alfred Worrad, PhD Student, Chemical & Biomoleculuar Engineering, University of Delaware**
Title: Analysis of the Electronic Properties for Different Coverages of Vanadium Oxide on Titania
Abstract: Modeling the vast configurational space of supported vanadium oxide on titania using ab initio methods such as density functional theory (DFT) is quite computationally expensive when considering structures beyond monomers and dimers. To explore larger structures, we perform high throughput calculations using machine learning interatomic potentials to obtain a family of low-energy structures for clusters up to monolayer coverage. These structures are then optimized using DFT to find the lowest energy structures for each coverage. We see that vanadium oxide's surface coverage impacts the materials' electronic properties and that isolated V_2O_5 dimers are not necessarily representative of larger oligomers, as was previously thought.

11. **Puranjit Singh, PhD Student, Plant & Soil Sciences, University of Delaware**
Title: A Deep Learning-based Smartphone App for Field-Based Blueberry Yield Prediction
Abstract: Accurate anticipation of blueberry yield and maturity before the harvesting is paramount for growers seeking to optimize production and refine marketing strategies. Pre-harvest estimation methods, relying on visual assessment or manual sampling, have proven to be prone to inaccuracies, labor-intensive, and time-consuming. The prospect of predicting yield and maturity with greater precision before the actual harvest represents a significant advantage for growers as it would help them analyze informed decisions that could in turn lead to enhance profitability on harvesting. Latest advancements in the field of artificial intelligence using Deep Learning strategies, we propose an efficient and user-friendly strategy. This strategy involves leveraging cell phone images of unharvested blueberry plants to generate accurate predictions of yield and maturity. This app aims to empower growers by enabling them to monitor the yield and maturity of individual blueberry plants in near real-time.

12. **Ryan Schanta, Master's Student, Civil & Environmental Engineering, University of Delaware**
Title: Prediction of Nearshore Wave Statistics Using ML Models Trained on Wave-Resolving Boussinesq Models
Abstract: Modeling coastal processes necessitates extraordinary computational power. Models thus make tradeoffs in the processes they resolve. One tradeoff is wave-averaging, whereby wave spectra are modeled instead of individual waves. However, modeling individual waves in the nearshore is crucial to understand complex processes. To this end, a series of 1000 wave-resolving simulations using the FUNWAVE model were run to generate a training dataset to predict nearshore wave skewness and asymmetry. Regression and neural network models

were developed to predict the resulting skewness and asymmetry. Preliminary results show that even relatively simple ML models can provide drastic improvements over simplified parameterizations. Furthermore, the results prove robust to overfitting. This suggests that the spatial distribution of temporal statistics of nonlinear wave dynamics in the nearshore environment may be adequately parametrized in a meaningful way via a machine learning approach.

13. **Dhana Lakshmi Kankanala, Master's Student, Data Science, University of Delaware**
Title: VN-Solver: Vision-based Neural Solver for Combinatorial Optimization Over Graphs Using PointNet
Abstract: Data-driven approaches have been proven effective in solving combinatorial optimization problems over graphs such as the traveling salesman problems and the vehicle routing problem. The rationale behind such methods is that the input instances may follow distributions with salient patterns that can be leveraged to overcome the worst-case computational hardness. For optimization problems over graphs, the common practice of neural combinatorial solvers consumes the inputs in the form of adjacency matrices. In this paper, we explore a vision-based method that is conceptually novel: can neural models solve graph optimization problems by taking a look at the graph pattern in 2d/3d space? Our results suggest that the performance of such vision-based methods is not only non-trivial but also comparable to the state-of-the-art matrix-based methods, which opens a new avenue for developing data-driven optimization solvers.

14. **Alexander Bryer, PhD Student, Chemistry & Biochemistry, University of Delaware**
Title: The HIV-1 Capsid Mechanoelastic Properties Regulate Nuclear Import and Uncoating
Abstract: The mature HIV-1 capsid must remain stable to facilitate nuclear trafficking and engender the chemistry of reverse transcription, while allowing the capsid to uncoat in a timely manner within the nuclei of infected cells. The balance of stability and propensity for timely uncoating is now recognized as a hallmark of HIV-1 infection. Despite this, the molecular determinants of capsid stability and a causal understanding of uncoating remain unrealized. We present in silico AFM simulations that probe different regions of full-scale conical HIV-1 capsids. Analyses from our simulations reveal that the HIV-1 capsid strongly resists positive-curvature deformation while allowing negative-curvature deformations with a nearly fivefold reduction in loading force. We identify regions of the capsid that are more susceptible to external forces and characterize the out of equilibrium strength of different CA-CA interfaces. Further, we characterize the mechano-elastic properties of two HIV-1 capsid mutants: one that is hyper stable and noninfectious, and a compensatory mutant which is known to rescue infectivity. The latter two systems offer a valuable opportunity to elucidate the role of capsid stiffness relevant to successful nuclear import and infection. We additionally subjected full-scale HIV-1 capsids to isotropic growth simulations, mimicking an expansive model where disassembly is driven by the build-up of reverse transcription products, where we observe failure of the capsid consistent with empirical observations. Our results show that the HIV-1 capsid is a robust container with finely tuned viscoelastic properties that allow it to adapt to a range of constrained geometries during trafficking and nuclear entry.

15. **Amirreza Meydani, PhD Student, Chemistry & Biochemistry, University of Delaware**
Title: Improving Runoff Prediction for Sustainable Angriculture and Enhanced Water Quality
Abstract: The Great Lakes are impacted by runoff fertilizer, leading to water quality issues like oxygen depletion and harmful algal blooms. This study aims to enhance runoff predictions using NOAA's National Water Model (NWM), aiding farmers in optimal fertilizer application timing. We developed a daily runoff prediction framework integrating statistical post-processing models with NOAA's NWM outputs at a 1km scale, using two models: tree-based machine learning (TML) and Quantile Mapping Bias Correction (QMBC). Results show TML's superiority in predicting runoff occurrences and magnitudes. This framework assists farmers in timing fertilizer application to reduce nutrient runoff, improving agricultural practices and water quality.

16. **Zach Schreiber, PhD Student, Center for Bioimformatics and Computational Biology, University of Delaware**
Title: An Embedding-based Approach for Profiling Replication Protein Associations in Bacteriophage Genomes
Abstract: Selenoprotein S (selenos) is a membrane protein involved in cellular stress management and signaling pathways. It assists protein homeostasis through the endoplasmic reticulum-associated degradation (ERAD) pathway, a mechanism from the stress response machinery to degrade misfolded proteins. In the ERAD pathway, selenos recognizes the AAA ATPase p97. This ATPase p97 is the main energy source to extract misfolded protein for degradation. The function of selenos is crucial for p97 function. How these proteins associate remains unclear, limiting the understanding of the role of selenos in assisting p97 functions. To investigate this question, we use cryo-electron microscopy (Cryo-EM) to reconstruct 3D structures of the selenos/p97 complex. The 3D reconstructions of the selenos/p97 complex reveal insights into the assembly mechanism of complexes in the ERAD. Understanding how these proteins associate will depict the principles governing protein degradation.

17. **Fabio Gonzalez, PhD Student, Chemistry & Biochemistry, University of Delaware**
Title: Using Cryo-Electron Microscopy to Understand the Physiological Roles of Selenoprotein S
Abstract: Bacterial viruses (phage), play crucial roles in ecosystem dynamics and are a significant source of genetic diversity. Previous studies have demonstrated that replication proteins, in particular DNA polymerase (PolA), have biochemical characteristics that can be key predictors of phage biology and infection strategy. However, we lack frameworks to identify and interpret modules of multiple proteins that would enable predictions of the impact of unknown virus populations on ecosystems. We hypothesize that an embedding-based approach leveraging protein structural prediction and KNN classification can effectively predict and classify conserved replication modules across phage genomes. Results show promise for defining distinct PolA groups and identifying modules of neighboring replication genes, with potential for implementing a replication module-based framework for predicting viral phenotypes.

18. **Alex Mulrooney, Undergraduate Student, Electrical & Computer Engineering, University of Delaware**
Title: Accelerating Tensor Decompositions in Julia
Abstract: Tensor-valued data appears in a variety of applications, including signal processing, neuroscience, and computer vision. A commonly used tensor decomposition, CP decomposition, aims to express a data tensor as sum of simpler rank-1 tensors, which can reveal underlying patterns in the data or be used for compression. Previously, Dr. Hong has introduced generalized CP (GCP) decompositions, which generalizes CP decomposition to arbitrary loss functions other than least squares, and created a open-source software package, GCPDecompositions.jl, that implements GCP in the Julia programming language. In this work, we present several improvements to this package, including an implementation of a previously proposed faster algorithm for the matricized tensor times Khatri-Rao product (MTTKRP), a key component of fitting the decomposition, and support for GPU computation via interfacing with CUDA.jl, which we show can provide significant speed-ups for larger tensors.

19. **Kyungmin Lee, PhD Student, Energy & Environmental Policy, University of Delaware**
Title: Detection of Building Energy End-Use Via Proximal Infrared Remote Sensing and Computer Vision
Abstract: 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).

20. **Lan Yu, PhD Student, Public Policy & Administration (Biden School), University of Delaware**
Title: Empirical Measurement of Lighting Technology Changeover in New York City
Abstract: In 2013, the Bloomberg administration in New York City (NYC) announced that 250,000 streetlights were to be replaced by LEDs by 2017. However, an assessment of this policy's progress has not been possible due to the unavailability of publicly accessible data. Recently, Dobler et al. (2015, 2021) deployed an "Urban Observatory" (UO) in NYC that uses cameras mounted atop tall buildings far from the city to take images that have sufficient color sensitivity and spatial resolution to determine the lighting type of individual sources in the city skyline (Dobler et al., 2016). In this work we identified light sources and the associated lighting technology type in two UO images, one from 2013 and one from 2018, to empirically measure the efficacy of the 2017 NYC LED retrofit policy. Our most complex models incorporate 1-D Convolutional Neural Networks to determine the type for each source, and we find that these have a model accuracy of roughly 92%.

21. **Shar Daniels, PhD Student, Physics & Astronomy, University of Delaware**
Title: Neural Networks for Sub-Second Astronomical Transient Discovery
Abstract: 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.

22. **Arjun Valiya Parambathu, Undergraduate Student, Behavioral Health and Nutrition, Delaware State University**
Title: Exascale Computing Enables Better Accuracy in Modeling Electrostatistics in Protein Interactions
Abstract: We present two cases of contrasting length scales where the conventional description of electrostatics is lacking in describing protein interactions. In the first case, we examine the self-association of a monoclonal antibody (mAb) using atomistic molecular mechanics calculations. We find that the electrostatic contribution in the associated state, computed using Poisson-Boltzmann (PB) solvers, varies widely with key parameters, questioning the reliability of these methods. In the second case, we examine the rank order of ion-pair association relevant to cation-exchange (CEX) chromatography using all-atom molecular dynamics simulations. We find that explicitly accounting for the solvent predicts the correct rank order for the association constant among the cations but not the anions. We find that explicit polarization, which allows the charge distribution to change based on local environment, is required to predict the rank order correctly for both cations and anions.

23. **Hassan Baker, PhD Student, Electrical & Computer Engineering, Delaware State University**
Title: Patch2Loc: Learning to Localize Neuroimage Patches to Detect Brain Lesions
Abstract: Detecting brain lesions as abnormalities observed in magnetic resonance imaging (MRI) is essential for diagnosis and treatment. In the search of abnormalities, such as tumors and malformations, radiologists may benefit from computer-aided daignostics that use computer vision systems trained with machine learning to segment normal tissue from abnormal brain tissue. While supervised learning methods require annotated lesions, we propose a new unsupervised approach (Path2Loc) that learns from normal patches taken from structural MRI: %2. We

train a neural network model to map a patch back to its relative spatial location within the brain volume. During inference, abnormal patches are detected the relatively higher error and/or variance of the location prediction. A threshold for the abnormality detection can be made using the distribution across normal pateches (i.e., an unsupervised threshold), or additionally a small set of annotated data (i.e., supervised threshhold). We demonstrate the ability of our model to detect abnormal brain tissues by applying our approach to the detection of tumor tissues in MRI on the BraTS2021 dataset, significantly outperforming state-of-the-art unsupervised methods.

24. **Parinaz Barakhshan, PhD Student, Electrical and Computer Engineering, University of Delaware**
    Title: An Automated Tool for Capturing, Replaying, and Validating Results of Experimental Section in Science and Engineering Applications
    Abstract: This poster presents a new methodology and tool that speeds up the process of optimizing science and engineering programs. The tool, called CaRV (Capture, Replay, and Validate), enables users to experiment quickly with large applications, comparing individual program sections before and after optimizations in terms of efficiency and accuracy. Using language-level checkpointing techniques, CaRV captures the necessary data for replaying the experimental section as a separate execution unit after the code optimization and validating the optimization against the original program. The tool reduces the amount of time and resources spent on experimentation with long-running programs by up to two orders of magnitude, making program optimization more efficient and cost-effective.
25. **Juan S. Rey, PhD Student, Chemistry & Biochemistry, University of Delaware**
    Title: Data-driven Computational Analysis of ALLINIs at the IN CCD-CTF Interface
    Abstract: Allosteric HIV-1 integrase inhibitors (ALLINIs) are small molecules that disrupt the viral lifecycle by inducing aberrant integrase (IN) multimerization and eliciting eccentric viral morphologies in which the viral genome is localized outside the capsid. Here, we perform molecular dynamics simulations of IN in complex with four ALLINIs: BID, Pirmitegravir (PIR), gs9822 and EKC110 to study the role of these molecules in multimerization. We find that the ALLINIs work as a "molecular glue" that induces and stabilizes an interface between the IN catalytic core domain (CCD) and the carboxyl terminal domain (CTD) via electrostatic and hydrophobic interactions with conserved residues in the CTD and CCD. Interestingly, we find that to accommodate extended ALLINI gs9822, the CTD can be displaced and still maintain the CCD-CTD interface via π-π stacking interactions with the ALLINI. All ALLINIs share the same chemical design principles: an aromatic scaffold with tertbutoxyl and carboxylate sidechains and a bulky hydrophobic group; however, small changes in these moieties lead to increased potency against HIV-1 mutants. Thus, we perform free energy perturbation (FEP) in silico mutation simulations, to quantify the energetic favorability of interactions with residues in the CCD-CTD interface to different ALLINIs. In addition, we tested the effect of resistance mutations on the CCD-CTD interface conformation and their energetic contribution to ALLINI binding. Overall, our computational pipeline allows the study of ALLINI induced IN condensation and provides insights for the rational development of this series of antiretrovirals through the optimization of their key contacts with the viral target.
26. **Zaid Alhusban, Postdoctoral Researcher, Civil & Environmental Engineering , University of Delaware**
    Title: Studying the Implementation of Hybrid Living Shorelines to Reduce Sea Level Rise Effects
    Abstract: Due to the anticipated sea level rise brought on by climate change, Lewes beech in Delaware Bay might experience flooding. Nature-based shoreline protection options are suggested to dampen the waves and reduce the effect of sea level rise. Numerical modeling was used to determine the suitability of implementing living shorelines close to Lewes beech and select the best living shoreline options and designs. FUNWAVE-TVD Boussinesq model software was used for smaller-scale modeling by using a rectangular grid with 1 m resolution and forcing offshore data to determine wave hydrodynamics, including wave height and wave-induced currents near the shoreline around the living shorelines. The performance of living shorelines will be assessed based on the structure's effects on significant wave heights, water velocity, salinity, reduction of erosion, and potential for habitat creation.
27. **Jiaye Zhang, PhD Student, Civil & Environmental Engineering, University of Delaware**
    Title: XBeach Modeling of Cross-shore Hydrodynamics and Morphodynamics in a Shallow Surf Zone
    Abstract: XBeach (Surfbeat) is used widely to predict storm-induced beach erosion. However, previous studies showed that XBeach often over-predicts erosion near the shoreline. This study focuses on investigating the skill of XBeach in predicting the offshore-directed mean currents (undertow) and the wave-orbital-velocity third moments (skewness and asymmetry) by validation with near-prototype large wave flume experiment data (DUNE3). Combining with high-resolution simulation data obtained from OpenFOAM via 3D large-eddy simulation (LES), the parameter values for the roller energy dissipation parameter β and the threshold water depth h_min are suggested to be sensitive to improve the over-predicted undertow in the shallow surf zone. Moreover, the existing wave shape parameterizations in XBeach are shown to remarkably overestimate wave asymmetry in the nearshore. The coefficients of γ_ua and T_bfac have been demonstrated as effective parameters to further improve the morphodynamic prediction.
28. **Daniel Golbaz, Staff, Civil & Environmental Engineering, University of Delaware**
    Title: Rainfall Interaction on Waves, Tides and River Discharge on Varying Beach Slope – Sensitivity Analysis on the Temporal Distribution of Precipitation
    Abstract: Coastal communities are at high risk of various hazards from tropical cyclones, including storm surges, heavy precipitation, tornadoes, and flooding. These hazards can have severe consequences for communities, such as damage to property and infrastructure, loss of life, and economic disruption. Accurately estimating water levels during coastal flooding is crucial for evaluating the vulnerability and effective emergency management planning at the community level. However, this remains a challenging task due to the complex interactions between various factors, such as wind, waves, tides, currents, and precipitation. Rainfall intensity and duration are important factors that impact water

levels, particularly when storms induce heavy precipitation for extended periods. Finer scale models are to return accurate details of ocean characteristics with the use of numerical modeling and ability to account for smaller scale features, complex processes, and interactions. Therefore, adding extra contributive variables to a high-fidelity model improves surge height estimation and physical processes involved in flood events aiding in developing effective strategies to mitigate the risks to coastal communities. In this study, we utilized fully nonlinear Boussinesq wave model (FUNWAVE), to evaluate the effects of rainfall on water level estimation. A Sensitivity analysis was applied to simulate the impacts of rain effects on a coastal area on varying beach slopes and evaluating the interaction of rain rates to other flood derivers. Our primary result demonstrated that the incorporation of rainfall dynamics into fine-scale, physics-based model can improve the prediction accuracy of water level during tropical cyclones, leading to better quantification of flooded area and improve the accuracy of exchanging information between coupled models at boundaries. Overall, our study highlights the importance of considering the impacts of rainfall on flood related studies and predictions for better understanding coastal resilience and informing emergency management strategies.

29. Maria Alejandra Carattini-Colon & Maxine Robinson, Undergraduate Students, Physics & Engineering, Delaware State University

Title: Rainfall Interaction on Waves, Tides and River Discharge on Varying Beach Slope – Sensitivity Analysis on the Temporal Distribution of Precipitation

Abstract: Coastal communities are at high risk of various hazards from tropical cyclones, including storm surges, heavy precipitation, tornadoes, and flooding. These hazards can have severe consequences for communities, such as damage to property and infrastructure, loss of life, and economic disruption. Accurately estimating water levels during coastal flooding is crucial for evaluating the vulnerability and effective emergency management planning at the community level. However, this remains a challenging task due to the complex interactions between various factors, such as wind, waves, tides, currents, and precipitation. Rainfall intensity and duration are important factors that impact water levels, particularly when storms induce heavy precipitation for extended periods. Finer scale models are to return accurate details of ocean characteristics with the use of numerical modeling and ability to account for smaller scale features, complex processes, and interactions. Therefore, adding extra contributive variables to a high-fidelity model improves surge height estimation and physical processes involved in flood events aiding in developing effective strategies to mitigate the risks to coastal communities. In this study, we utilized fully nonlinear Boussinesq wave model (FUNWAVE), to evaluate the effects of rainfall on water level estimation. A Sensitivity analysis was applied to simulate the impacts of rain effects on a coastal area on varying beach slopes and evaluating the interaction of rain rates to other flood derivers. Our primary result demonstrated that the incorporation of rainfall dynamics into fine-scale, physics-based model can improve the prediction accuracy of water level during tropical cyclones, leading to better quantification of flooded area and improve the accuracy of exchanging information between coupled models at boundaries. Overall, our study highlights the importance of considering the impacts of rainfall on flood related studies and predictions for better understanding coastal resilience and informing emergency management strategies.

30. Leonardo Pierre, PhD Student, Physics & Engineering, Mathematics & Computer Science, Delaware State University

Title: K-Means Clustering Analysis of the Membrane Structure of Giant Unilamellar Vesicles from the Hyperspectral Dark-Field Microscopy

Abstract: Giant Unilamellar Vesicles (GUVs) serve as simplified models mimicking the biological membranes. To investigate the membrane integrity of GUVs, we introduced cholesterol (CHOL) in concentrations ranging from 0 to 40 mol%. The reflectance spectra of GUVs, resulting from the absorption or scattering of photons upon their vesicular interactions, were captured spatially via the hyperspectral dark-field microscopy. To analyze the spectral variation inside the lipid membranes, we utilized the K-Means clustering technique, a machine-learning algorithm that categorizes data into distinct "clusters" based on similarities in their characteristics. We used K-Means from the Spectral Python (SPy) Module to section the membrane structure of GUVs based on the similarities in their spectral profiles. The K-Means algorithm was found to be very effective in segmenting the membrane structure with submolecular resolution. Our results are promising in revealing the patterns of GUVs' microstructure in relation to the CHOL concentrations based on the spectral attributes. Integrating machine learning to the analysis of hyperspectral microscopy data leads to the exceptional submolecular resolution otherwise unavailable in the live sample imaging. We acknowledge support by the Delaware INBRE program, with a grant from the National Institute of General Medical Sciences - NIGMS (P20 GM103446) from the National Institutes of Health. Leonardo Pierre acknowledges support from NRT-MIDAS (Computing and Data Science Training for Materials Innovation, Discovery, AnalyticS) at University of Delaware, an NSF award (2125703) of Research Traineeship (NRT) program.

DARWIN COMPUTING SYMPOSIUM PLANNING COMMITTEE

Rudolf Eigenmann, PI

Arthi Jayaraman, co-PI

Benjamin Bagozzi, co-PI

William Totten, co-PI

Cathy Wu, co-PI

Tian-jian Hsu (Tom), Committee Chair

DARWIN COMPUTING SYMPOSIUM PLANNING COMMITTEE (CON'T)

Daria Blinova

Lynette Carney

Michael Blaustein

Sam Smith

Jiaye Zhang

THANK YOU TO OUR SPONSORS!

TECHIMPACT®

University of Delaware
Delaware Environmental Institute

University of Delaware
Faculty Senate

University of Delaware
Information Technologies

CACR
Center for Applied Coastal Research
University of Delaware

VISIT THE Data Science Institute AT:

https://dsi.udel.edu/

JOIN US ON TWITTER

twitter **@dsiudel**

hashtag: **#DARWINcomputing2024**