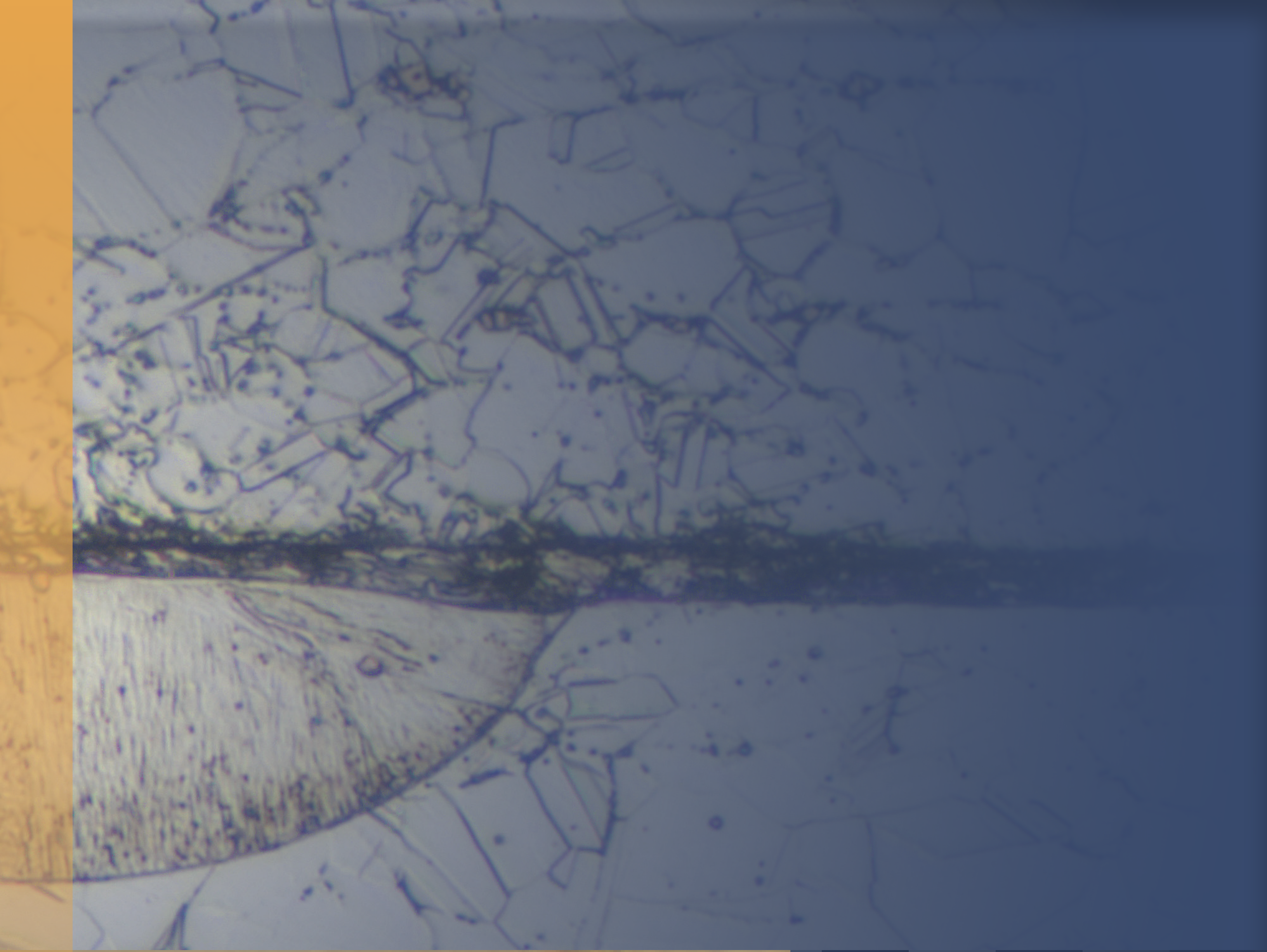


idies

The Institute for Data Intensive Engineering and Science

'22 ANNUAL REVIEW

OCTOBER 1 – SEPTEMBER 30
2021 – 2022

MESSAGE FROM THE

DIRECTOR

At IDIES, faculty and students work together to solve amazing data-intensive problems, from genes to galaxies, including new projects in materials science, and urban planning. Over the last few years, our members have successfully collaborated on many proposals related to Big Data, and we have hired new faculty members, all working on different aspects of data-driven discoveries. Together, we are successful in building a collection of unique, large data sets, giving JHU a strategic advantage in making new discoveries.

We want to share exciting news with the IDIES community about some ground-breaking developments that we have been eagerly waiting for.

Earlier this year, the Schmidt Futures foundation established the Virtual Institute of Scientific Software (VISS) comprised of four world-renowned universities—Cambridge University (UK), University of Washington, Georgia Institute of Technology, and Johns Hopkins—each chosen to host their own scientific software engineering center (SSEC). VISS will address the growing demand for high-quality professional software engineers who can build dynamic, scalable, open software to facilitate accelerated scientific discovery across fields.

Work on the NSF-funded National Science Data Fabric project has begun. Building on the success of the Open Storage Network project the NSDF will provide distributed storage hubs across the country for large science data sets and provide support for containerized compute near the data.

ON THE COVER: Material Sciences—Close-up of a laser meltpool
IMAGE CREDIT: National Institute of Standards and Technology

AstroPath - the cell imaging analysis platform that was developed by IDIES scientists in collaboration with the School of Medicine to apply astronomy image analysis techniques to clinical data consisting mostly of multispectral immunofluorescence microscopy images - has revolutionized cell pathology. AstrpPath's high-precision characterization of tumor microenvironments has had a critical impact on cancer inmmunotherapies. To date, more than 100k cells have been annotated for nuclear features, and "ground truth" dataset annotations will soon be used to train a deep learning algorithm to accurately identify the nuclei and membranes.

Over the past year, IDIES has expanded its involvement in projects with a number of different Hopkins departments and external partners across multiple science, health and engineering disciplines on solving challenging data-intensive problems, while continuing to provide essential support to the Hopkins community with data science resources, grant proposal support and funding opportunities.

SciServer, the generic science platform that IDIES developed and unveiled in 2015, continues to increase in popularity with the Hopkins and worldwide community. SciServer installations outside of Hopkins now include NIST, NASA Goddard, Max Planck (Germany) and the National Astronomical Observatory of Japan. A new engagement is with Brookhaven National Labs.

IDIES has initiated a collaboration with the Sheridan Libraries Data Services Center (DSC) aimed at promoting each other's activities and sharing resources. Dr Pete Lawson, who leads the DSC, has replaced Dr Sayeed Choudhury on the IDIES Executive Committee. Sayeed, who had been a devoted friend to IDIES since its inception, left Hopkins to join Carnegie Mellon University this past summer. IDIES is grateful for his friendship and support for all these years, and wishes him the very best at CMU.

IDIES also thanks Dr Jeff Leek for several years of very fruitful collaborations on multiple projects, and Tara Engel for her able leadership as IDIES admin for the past few years. We welcome Elliot Rosen as the new IDIES admin and Kate Stevens as IDIES communications specialist.

IDIES continues to increase its footprint and relevance across the University by fostering and sustaining data intensive projects in all disciplines. We welcome your ideas, big or small, on how we can improve our engagement with your research community.

— Alex Szalay

OCTOBER 20TH, 2022 | WELCH LIBRARY

2022 IDIES ANNUAL SYMPOSIUM

KEYNOTE SPEAKERS

BY SPEAKING ORDER

STEPHANIE HICKS PhD

Associate Professor, Department of Biostatistics,
Johns Hopkins Bloomberg School of Public Health

PRESENTING:

Data-efficient and Multimodal Computational Pathology

Dr. Stephanie Hicks is an Associate Professor in the Department of Biostatistics at Johns Hopkins BSPH where she is also a faculty member of the Johns Hopkins Data Science Lab, and has affiliations with the Malone Center for Engineering in Healthcare, Center for Computational Biology, the Department of Genetic Medicine, and the Department of Biochemistry and Molecular Biology. She is an expert in developing scalable computational methods and open-source software for biomedical data analysis, in particular single-cell and spatial transcriptomics genomics data, leading to an improved understanding of human health and disease.

She serves on a variety boards including the Bioconductor Technical Advisory Board, and the Editorial Board at Genome Biology and the Journal of American Statistical Association. Locally, she co-founded and co-organizes the R-Ladies Baltimore chapter to promote gender diversity in the R programming language community. She is a recipient of several professional awards including a K99/R00 Pathway to Independence Award, a High-Impact Project Award from the Bloomberg American Health Initiative, Teaching in the Health Sciences Young Investigator Award from the American Statistical Association (ASA), and the the COPSS Emerging Leader Award from the ASA, arguably the statistical profession's most prestigious award for early career leaders in Statistics and Data Science.

FAISAL MAHMOOD PhD

Associate Professor, Pathology, HarvardMedical School;
Division of Computational Pathology, BWM

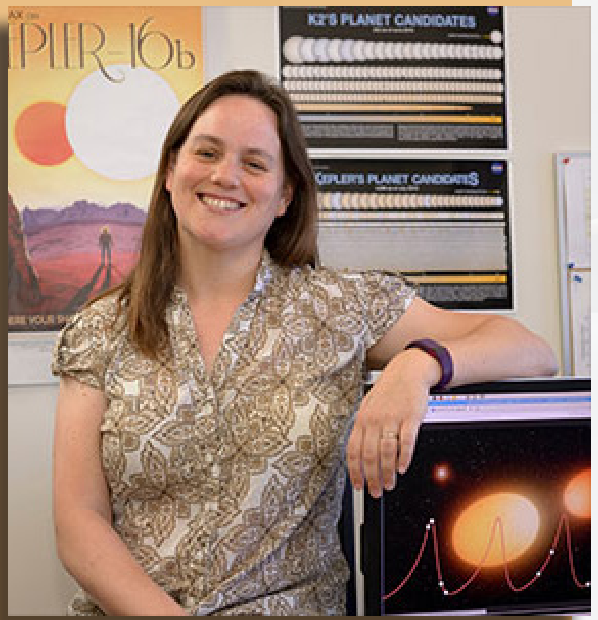
PRESENTING:

Scaling up Single-cell and Spatial Data Analyses

Dr. Mahmood is an Associate Professor of Pathology at Harvard Medical School and the Division of Computational Pathology at the Brigham and Women's Hospital (BWM). He received his Ph.D. in Biomedical Imaging from the Okinawa Institute of Science and Technology, Japan and was a postdoctoral fellow at the department of biomedical engineering at Johns Hopkins University. His research interests include pathology image analysis, morphological feature, and biomarker discovery using data fusion and multimodal analysis. Dr. Mahmood is a full member of the Dana-Farber Cancer Institute / Harvard Cancer Center; an Associate Member of the Broad Institute of Harvard and MIT, and a member of the Harvard Bioinformatics and Integrative Genomics (BIG) faculty.

KEYNOTE SPEAKERS CONTINUED

BY SPEAKING ORDER



SUSAN MULLALLY PhD

Associate Scientist,
Space Telescope Science Institute

THE WEBB TELESCOPE: THE DATA BEHIND THE IMAGES

Dr. Susan E. Mullally is an expert in astronomical time series data collection and analysis and has used these skills to study the population of exoplanets in our Galaxy and the details of stellar evolution. She is currently the Deputy Project Scientist for JWST and works to ensure the scientific productivity of the JWST mission. Previously she worked in the MAST as the lead for archiving the data from NASA's Transiting Exoplanet Survey Satellite, to help develop the exo.MAST interface, and to improve the archive for the TESS and Kepler missions.

Prior to her employment at STScI, Dr. Mullally was a scientist for the Kepler Science Office. As part of that team, Dr. Mullally led the creation of the final Kepler catalog of exoplanet candidates in such a way that it can be used to reliably determine the frequency of small planets in our Galaxy. As part of that work she developed a machine learning technique to automatically vet signals found from the Kepler spacecraft. She has also worked to find and understand a new class of tidally-distorted, eccentric binary stars known as heartbeat stars. Earlier in her career she developed tools and organized the data collection efforts of the Whole Earth Telescope, a collaboration of ground-based telescopes that work to measure the pulsations of white dwarf stars. Dr. Mullally started her career as an assistant professor at Colorado College and enjoyed teaching students the art of astronomical observing.

Dr. Mullally enjoys giving outreach talks and most recently has been found teaching local pre-schoolers about planets and hikers on top of a mountain in Ireland about unusual exoplanets.

PAT HELLAND

Software Architect,
Salesforce

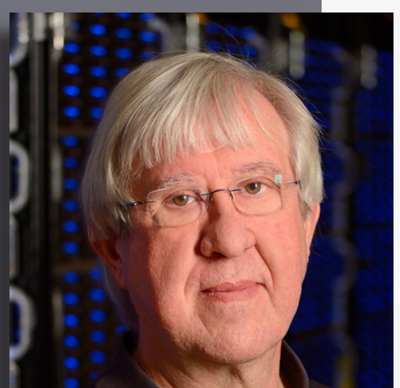
STANDING ON THE DISTRIBUTED SHOULDERS OF

Pat Helland has been building distributed systems, database systems, high-performance messaging systems, and multiprocessors since 1978, shortly after dropping out of UC Irvine without a bachelor's degree. That hasn't stopped him from having a passion for academics and publication. From 1982 to 1990, Pat was the chief architect for TMF (Transaction Monitoring Facility), the transaction logging and recovery systems for NonStop SQL, a message-based fault-tolerant system providing high-availability solutions for business critical solutions. In 1991, he moved to HaL Computers where he was chief architect for the Mercury Interconnect Architecture, a cache-coherent non-uniform memory architecture multiprocessor. In 1994, Pat moved to Microsoft to help the company develop a business providing enterprise software solutions. He was chief architect for MTS (Microsoft Transaction Server) and DTC (Distributed Transaction Coordinator). Starting in 2000, Pat began the SQL Service Broker project, a high-performance transactional exactly-once in-order message processing and app execution engine built deeply into Microsoft SQL Server 2005. From 2005-2007, he worked at Amazon on scalable enterprise solutions, scale-out user facing services, integrating product catalog feeds from millions of sellers, and highly-available eventually consistent storage. From 2007 to 2011, Pat was back at Microsoft working on a number of projects including Structured Streams in Cosmos. Structured streams kept metadata within the "big data" streams that were typically 10s of terabytes in size. This metadata allowed affinitized placement within the cluster as well as efficient joins across multiple streams. On launch, this doubled the work performed within the 250PB store. Pat also did the initial design for Baja, the distributed transaction support for a distributed event-processing engine implemented as an LSM atop structured streams providing transactional updates targeting the ingestion of "the entire web in one table" with changes visible in seconds. Starting in 2012, Pat has worked at Salesforce on database technology running within cloud environments.

His current interests include latency bounding of online enterprise-grade transaction systems in the face of jitter, the management of metastability in complex environments, and zero-downtime upgrades to databases and stateful applications. In his spare time, Pat regularly writes for ACM Queue, Communications of the ACM, and various conferences. He has been deeply involved in the organization of the HPTS (High Performance Transactions Systems - www.hpts.ws) workshop since 1985. His blog is at pathelland.substack.com and he parsimoniously tweets with the handle @pathelland.

2022 IDIES ANNUAL SYMPOSIUM

OCTOBER 20TH, 2022 | WELCH LIBRARY

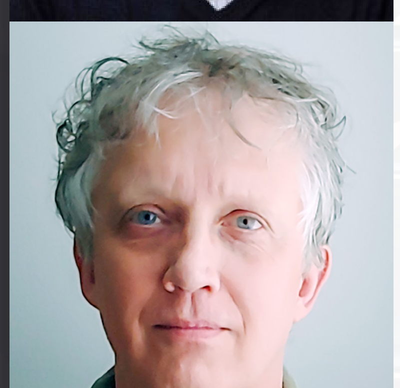


IDIES

ALEX SZALAY, PhD is the founding director of IDIES and the SSEC, a Bloomberg Distinguished Professor, Alumni Centennial Professor of Astronomy, and a professor of Computer Science. As a cosmologist, he works on the use of big data in advancing scientists understanding of astronomy, physical sciences, and life sciences.

IDIES, Moderator

CHARLES MENEVEAU, PhD, the Louis M. Sardella Professor in Mechanical Engineering, and associate director of IDIES, is an expert in the multiscale aspects of turbulence, large-eddy simulations, and wind farm fluid dynamics. Meneveau has participated in efforts to democratize access to valuable "big data" in turbulenceand led the team that built the Johns Hopkins Turbulence Databases (JHTDB).



Scientific Software Engineering Center (SSEC)

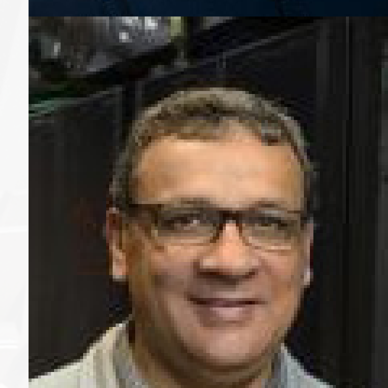
GERARD LEMSON has his PhD in theoretical cosmology and is currently a research scientist at IDIES and the associate director of the new SSEC. He is also associate director for science coordination in the NSF -unded SciServer project (www.sciserver.org) and assists in code development of that platform.

MODERATORS & UPDATES SPEAKERS

BY ORDER OF APPEARANCE

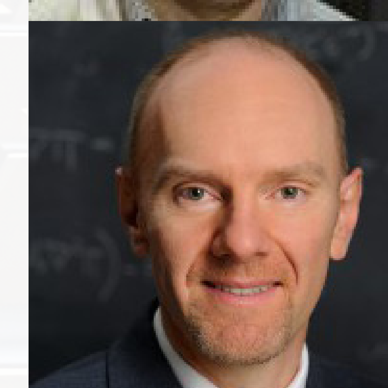
Data Services at Sheridan Libraries & Museums

PETER LAWSON, PhD, the Data and Visualization Librarian at Sheridan Libraries' Data Service and the newest member of the IDIES executive board. He holds a PhD in Bioinnovation from Tulane University. As a computational scientist and librarian, Lawson leverages his research background to help students and researchers analyze and visualize their data.



Advanced Research Computing at Hopkins & Maryland Advanced Research Computing Center (ARCH & MARCC)

JAIME COMBARIZA, PhD, is the director of the Maryland Advanced Research Computing Center (MARCC), a shared high-performance computing facility for John Hopkins University and the University of Maryland.



Moderator

THOMAS HAINE, PhD, is a professor of Earth & Planetary Sciences, with a research interest in the fundamental understanding of the physics of the basin-scale ocean and its role in Earth's climate. An IDIES member and former IDIES SEED Awardee, Haine leads the NSF-funded Poseidon project for ocean circulation.

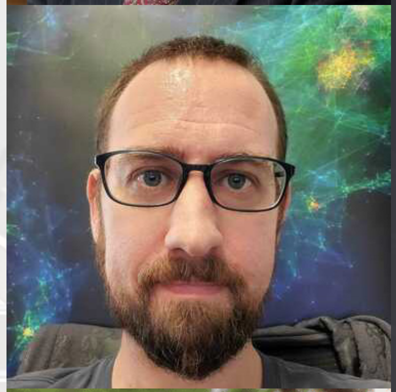
MODERATORS & UPDATES SPEAKERS CONTINUED

BY ORDER OF APPEARANCE

AI-X

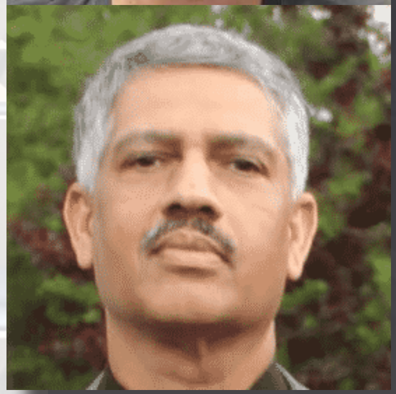
KT RAMESH, the Alonzo G. Decker, Jr., Professor of Science & Engineering, is the director of the Hopkins Extreme Materials Institute (HEMI) and serves as senior advisor to the university president. He is a professor of mechanical engineering, with joint appointments in department of Materials Science and Engineering and Earth and Planetary Sciences. He is the founding director of the multidisciplinary HEMI, which seeks to develop the science and technology to protect people, structures, and the planet.

SCISERVER



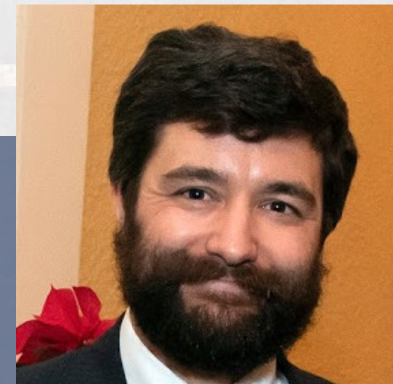
ARIK MITSCHANG earned his PhD in Astronomy at Macquarie University in Sydney, AUS, then worked as a software engineer in the Search Engine division of Amazon in moved to Tokyo, Japan. He joined the IDIES faculty in 2019, where he works on development of the SciServer platform (he's also a heavy user of the system for data science tasks within the institute).

Moderator



ANI THAKAR, PhD is a principal research scientist in the Department of Physics & Astronomy (PHA) and at the Center for Astrophysical Sciences (CAS). Thakar serves as the associate director of operations for IDIES.

TOM LIPPINCOTT PHD



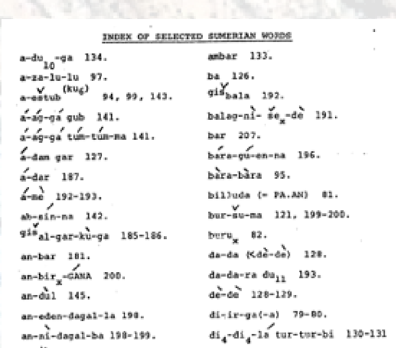
Assistant Research Professor, Director of Center for Digital Humanities, Alexander Grass Humanities;

Institute Assistant Research Professor, Department of Computer Science;

Research Scientist, Human Language Technology Center of Excellence, Johns Hopkins Krieger School of Arts and Sciences

Tom Lippincott is an Assistant Research Professor in Computer Science and the Humanities Institute and director of Krieger's Center for Digital Humanities, where research focuses on boundaries between the practices of traditional humanistic scholarship and the evolving capacities of computational intelligence. Prior to joining Johns Hopkins he received his PhD from Cambridge University in computational linguistics, and is affiliated with the Center for Language and Speech Processing and the Human Language Technology Center of Excellence.

CO-IS: Patesede Makonnen, Richard Essam, and Ben Allsoppe



INDEX OF SELECTED SUMERIAN WORDS

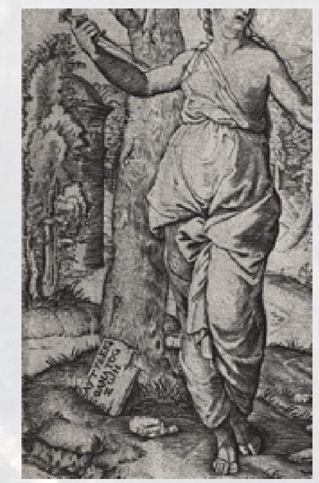
a-du$_{10}$-ga 134.	ambar 133.
a-za-lu-lu 97.	ba 126.
a-ešub (kug) 94, 99, 143.	giš bala 192.
á-ág-gá gub 141.	balag-nì- še$_x$-dè 191.
á-ág-gá tùn-tùn-ma 141.	bar 207.
á-dam gar 127.	bára-gú-en-na 196.
á-dar 187.	bàra-bàra 95.
á-mè 192-193.	bil₃uda (= PA.AN) 81.
ab-sín-na 142.	bur-šu-ma 121, 199-200.
giš al-gar-kù-ga 185-186.	buru$_x$ 82.
an-bar 181.	da-da (<dè-dè) 128.
an-bir$_x$-GANA 200.	da-da-ra du$_{11}$ 193.
an-dùl 145.	dè-dè 128-129.
an-eden-dagal-la 198.	di-ir-ga(-a) 79-80.
an-nì-dagal-ba 198-199.	di$_4$-di$_4$-lá tur-tur-bi 130-131
an-šub-ba 121-122.	du-bu-ul da-ba-al za/du$_{11}$ 204.

(left) An example of the format of the transliterated Sumerian word index. Daniel Reisman, "Two Neo-Sumerian Hymns". A Dissertation in Oriental Studies, University of Pennsylvania, 1969.

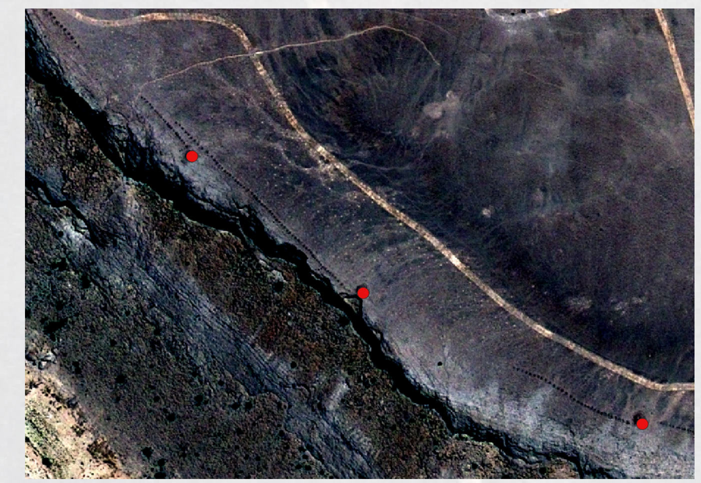
(right) Handwritten correspondence of Maria Edgeworth, early to mid-19th-century.

Exploring Computer Vision Models and Developing Infrastructure for OCR and Image Clustering

The Center for Digital Humanities (CDH) is exploring the current state of the art for broad classes of computer vision models in terms of their utility for humanistic scholarship. At this stage, we have assembled a set of image corpora representing heterogeneous interests and tasks. These include the handwritten correspondence of poet Percy Shelley (1792-1822) and Maria Edgeworth (1768-1849), early Italian Renaissance paintings with embedded symbologies, especially Dürer, satellite imagery of terrain with dense archaeological remains, and rasterized images of secondary scholarship related to the cuneiform writing system that employs field-specific transliterations to describe an ancient syllabic orthography.



Marcantonio Raimondi, "Dido." Detail from early 16th c. engraving.



Ancient tomb structures marked on a satellite image of southern Yemen

At the same time we are selecting recent pre-trained computer vision models that represent the current state of the art for supervised tasks like character recognition, to unsupervised tasks like image segmentation and edge detection. With these in hand, our next goal is to arrange the application of the models to the corpora using the general infrastructure being refined at the CDH, and present the resulting annotated images back to the domain scholars for scrutiny and discussion. This process additionally serves to advance our research into and development of the computational mechanisms thatoperate at the boundaries between the humanities and computation (as scholars translate their primary sources into a form amenable to machine learning models) and computation and the humanities (as the potential insights from those models are translated back and made legible to the scholars).

IDIES SEED AWARD UPDATE

TOM WOOLF PHD

Research Professor, Department of Physiology
Johns Hopkins School of Medicine

Tom Woolf started development of the Daily24 project when Apple released HealthKit/ResearchKit. This was collaborative work within computer science and the initial App was called Metabolic Compass. The ideas led to an active collaboration across multiple departments, most recently within General Internal Medicine. In particular, Daily24 was part of AHA funded research into the timing-of-eating. Dr. Woolf's team brings together researchers within the School of Medicine with expertise in Covid and researchers from the Applied Physics Lab with expertise in

Co-Is: Paul Nagy, Brian Garibaldi, Scott Pilla, Jared Zook, Harold Lehmann, Jane Valentine, Daniel Berman, and James Howard

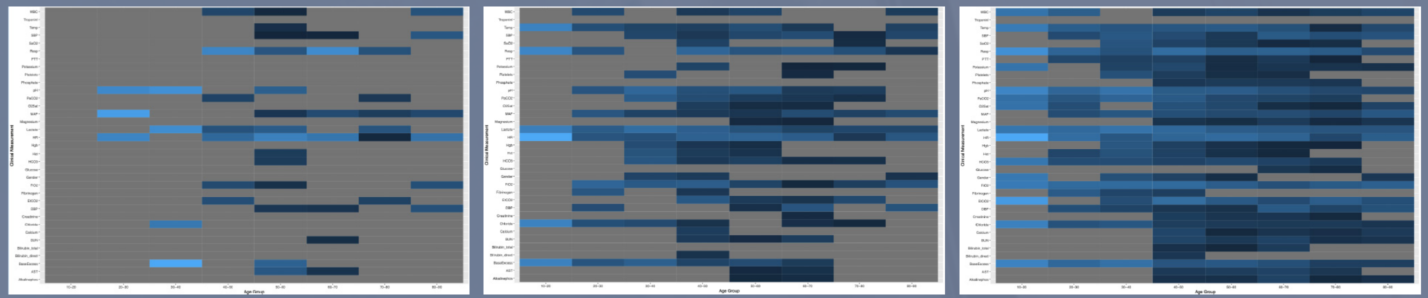


Figure: Our COVID24 App aims at predicting real-time risk based on features from each patient. In this case we build from the analysis of patients at 3, 6, and 12 hours in the ICU and the color code indicates the relative importance of each feature for predicting the outcome.

COVID24 Now a Three-tiered App

Our COVID24 App has grown from its beginning as a project aimed at predicting patient risk prior to infection. Our original IDIES proposal was to expand our circadian rhythm based Daily24 App to enable a real-time, data-driven analysis of local risk. We have since added two more research directions to our App development. In these two additional aims we address Long Covid and Acute Respiratory Distress Syndrome (ARDS). With Long Covid we aim to add further understanding to the analysis of risk for major symptoms that have not cleared after an initial COVID infection. Our third tier is to enable faster data-driven analysis of lung function to support a more rapid response to the pre-ICU stage of ARDS.

Based on the IDIES funding we have submitted two NIH proposals focused on growing the App from the patient risk tier into the analysis of Long COVID (grant-1) and into the lung function analysis (grant-2). Initially, for grant-1, the proposal is for a small clinical trial at Green Spring Station. The proposal will recruit from General Internal Medicine patients, and w–e have proposed to integrate the patient generated data into Epic. In the grant-2, we have proposed to build a lung structure/function model based on CT scans. From the scan a model for lung structure is defined and simulations (on AWS) can estimate lung function. Our goal in this third tier is to enable patients to see the locations of the lung infection and how that has impacted their lung function.

PAULETTE CLANCY PHD

Head, Professor, Dept. of Chemical and Biomolecular Engineering
Johns Hopkins University Whiting School of Engineering

Paulette Clancy is a Professor and the inaugural Head of the Department of Chemical and Biomolecular Engineering at Johns Hopkins University. And she holds the Samuel and Diane Bodman Professor Emerita of Chemical Engineering at Cornell.

She was the inaugural Director of the Cornell Institute for Computational Science and Engineering for almost 10 years and is reprising a similar role at Hopkins, chairing the faculty oversight committee for our petascale research computing resources, ARCH.

Her research group is recognized as one of the country's leading computational groups in atomic-scale modeling of materials and algorithm development. Her current thrust is to develop machinelearning algorithms to accelerate the search for optimal materialsprocessing protocols. Her group has always been focused onnovel electronic materials and materials processing but it also includes more esoteric projects likexenobiology ('Life' on Titan). She has won numerous awards for mentoring, service learning and civic engagement, and promoting the professional advancement of those from under-represented groups.

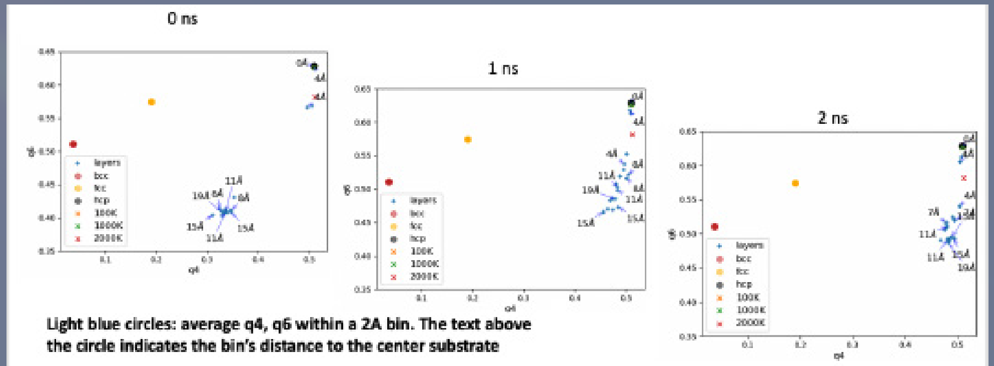


Figure 1.

Active Learning Captures a Reaction-Diffusion Simulation of Novel Additive Manufacturing of Gallium Nitride

We used an "on-the-fly" active learning approach, a machine learning technique, to simulate a novel additive manufacturing (AM) process to create high-value gallium nitride thin films. Being able to essentially 'print' this material should have implications for improved energy transmission and efficiency. Such an approach can provide a level of atomic-scale detail on the crystallization process, e.g., identifying the fastest growing facets, that is unavailable by experimental means. This kind of information would be invaluable for providing a level of control of the AM process that doesn't currently exist.

In principle, Molecular Dynamics (MD) is a suitable way to obtain this kind of atomic-scale information. But its accuracy is limited either by the inherent inadequacies of using an empirical force field to solve the central F=ma equation of motion in MD, especially for a system involving a chemical reaction like the AM process. Or it is limited by the extreme computational cost of using an ab initio approach to obtain those forces more accurately. The practical value of an "on-the-fly" approach is that it combines the accuracy of an ab initio, quantum mechanical, method within an MD framework. This is essentially the best of both worlds: Active learning allows the system to "learn" where next to move the particles without invoking the high cost of performing an ab initio calculation unless the algorithm detects that it is deviates sufficiently from its training set that it needs to make an ab initio calculation to stay accurately on track. In this way, we can capture the details of the formation of gallium nitride by a chemical reaction of gaseous nitrogen atoms diffusing through liquid gallium to form crystalline GaN.

In this seed award, we are focusing on following the ability of the active learning algorithm to capture this chemical reaction process and ensure that the solid that it produces is both crystalline and has an expected crystal structure. The algorithm we are using is known as FLARE, Fast Learning of Atomistic Rare Events, developed by Kozinsky (Harvard) in 2020. It is a Bayesian approach based on a Gaussian process regression technique. FLARE is not a "turn-key" system and it has yet to be tested on a system that involves a reaction-diffusion system and three phases of matter (solid, liquid, gas). So far in this seed award, we have used well-known Steinhardt Order Parameters that provide a unique "fingerprint" for known crystal structures (e.g., FCC, BCC, etc.) based on the angles between nearest neighbors of a given atom. This allows you to see the evolving crystal structure as the liquid crystallizes, which may not follow a simple monolithic path. Figure 1 (a-c) shows how the order parameters, q4 and q6, evolve from values representative of a random liquid (shown in 1a at time t=0) towards that of pure HCP (hexagonal close packed structure) shown as a black dot at the top RH corner of the plot. The crosses show order parameters from each layer. These get closer to pure HCP as time progresses from 1 ns I(Fig. 1b) to 2 ns (Fig.1c). The simulations are continuing for longer times to bring the structure to completion as an HCP crystal characteristic of GaN.

Development of a Searchable Database for Human Temporal Bone Otopathology Education and Research

AMANDA M. LAUER PHD

Associate Professor, Dept. of Otolaryngology-HNS, Department of Neuroscience, Johns Hopkins University School of Medicine

Amanda Lauer is an Associate Professor in Otolaryngology-HNS and Neuroscience at Johns Hopkins. Research in the Lauer Lab focuses on comparative models of hearing loss, the efferent feedback pathways between the ear and brain, and human otopathology. Along with Dr. Bryan Ward, she leads a team of scientists, clinicians, engineers, and medical illustrators building a human temporal bone resource for the scientific community. Dr. Lauer is also active in mentoring programs aimed at increasing diversity and inclusion in science and supporting early career scientists.

Co-Is: Bryan Ward, John Carey, Depts. of Otolaryngology-Head and Neck Surgery; Tilak Ratnanather, Dept. of Biomedical Engineering

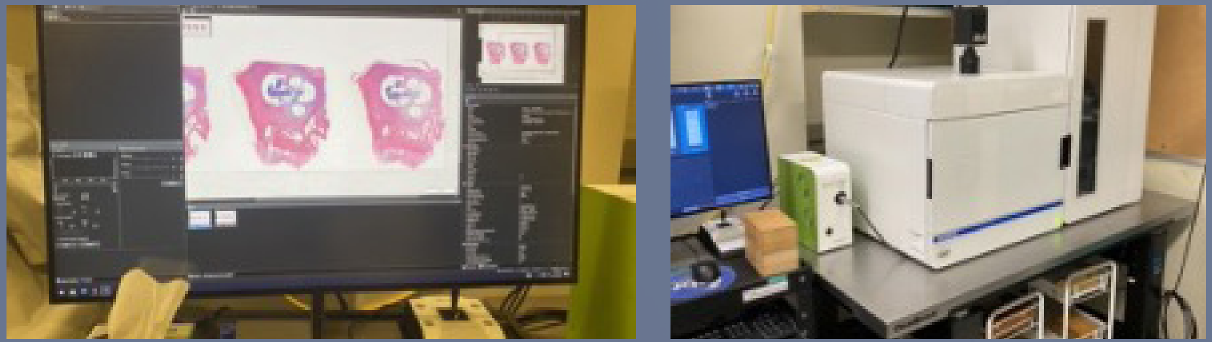


Figure 1. Human temporal bone specimen images (left) captured using a digital slide scanner with microscope and robotic loading arm (right).

The goal our IDIES Seed Grant project is to establish proof-of-concept that we can build a searchable database of digitized sections from human temporal bone (inner ear) specimens that can be accessed by any scientist with an interest in human hearing and balance research and reconstructed in 3D for segmentation and analysis of different soft tissue structures housed within the bone. Our long-term plan is to 'democratize' human temporal bone research to accelerate the pace of discovery of the causes of human inner ear diseases affecting hearing and balance. Research by our colleagues in the Cochlear Center for Public Health and the Department of Otolaryngology at Johns Hopkins and elsewhere has linked hearing loss and balance dysfunction with increased risk of dementia, cognitive decline, and other negative health outcomes. Inner ear disorders are potentially modifiable risk factors, but there are no cures for many etiologies of hearing and balance dysfunction. However, we still don't fully understand the structural and cellular abnormalities contributing to symptoms associated with many conditions.

Our group has revived a century-old tradition of human temporal bone research at Johns Hopkins, which houses one of the most extensive collections in the world, including many specimens from Black/African American individuals--a group that has historically been underrepresented in hearing research. Our collection also contains specimens from people with enigmatic inner ear disorders like Meniere's disease. We have begun to digitize images of the specimens (Figure 1), which take up about 1-2 TB per ear. These large data volumes are difficult for most researchers' personal computers to handle, so we have been developing a workflow to upload the images to SciServer so our collaborators can work with the specimens to build 3D reconstructions of various structures and fluid-filled spaces in the inner ear.

We are also assessing the challenges of working with datasets in which we only have one out of every ten sections cut through the ears (traditional histology) compared to having every section, every other section, every 5th section, and so on. Once this process has been optimized for use with our collaborators' AI analysis algorithms, it will inform our future tissue processing efforts as we begin to collect new specimens from human donors and animal models and provide scientists with valuable tools for the efficient characterization of inner ear structures. These tools can be applied to a broad range of research questions, including those regarding specific inner ear disorders and more general changes in inner ear structures across the lifespan. Increasing the accessibility of these specimens to the wider scientific community will catalyze temporal bone research and provide opportunities for researchers across the career spectrum. Particularly important to us is that we already have been able to involve undergraduates, medical students, residents, and graduate students in this work. Success in this endeavor with managing large micro-anatomic datasets could help democratize temporal bone research; investigators not located at the few institutions with these rare collections could fully participate in investigations to drive our field forward much faster.

Development of an Artificial Intelligence System for Phenotyping of Patients with Acute Stroke

RAMA CHELLAPPA PHD

Associate Professor, xxxx

Co-Is: Soumyajit Ray; Robert D. Stevens

Rama Chellappa is a Bloomberg Distinguished Professor in the Department of Electrical and Computer Engineering in the Whiting School of Engineering and in the Department of Biomedical Engineering in the School of Medicine. His research interests are Computer Vision, Artificial Intelligence, Biomedical Data Sciences, and Machine Learning.

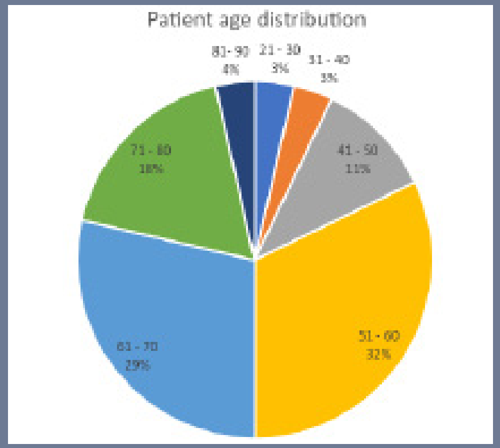
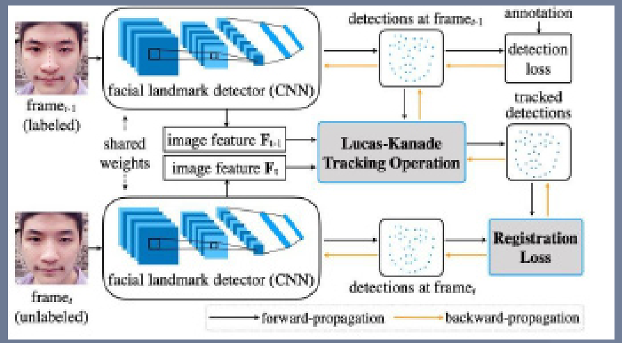
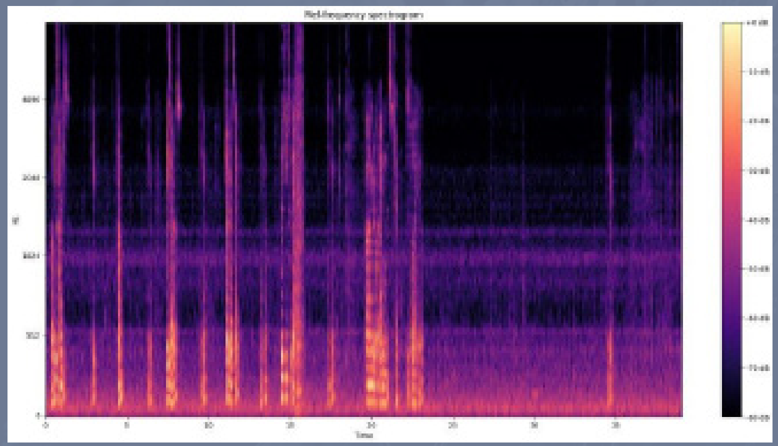


Figure 1. (left) Patient age

Figure 2. (lower left) Architecture for stabilizing facial landmark extraction using optical flow-based tracking. Adapted from: Dong, Xuanyi, et al. "Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

Figure 3. (lower right) Mel frequency spectrogram extracted from a study participant's (Sub ID: C0001) speech.





To advance our aim of developing a smartphone-based AI system to detect and quantify the severity of ischemic stroke, we have received approval from the Johns Hopkins Medicine Institutional Review Board for our research protocol (study protocol #IRB00296535). In accordance with this protocol, we have started collecting data from patients and control group participants.

Our study is a prospective observational cohort. We are recruiting healthy participants through advertisements in the JHU Hub which reaches out to the Johns Hopkins University community. We are also recruiting eligible patients who are being evaluated for the clinical disorders (ischemic stroke, hemorrhagic stroke and other underlying conditions presenting similar symptoms) by the Brain Attack Team (BAT) in Johns Hopkins Bayview Medical Center (JHBMC) and Johns Hopkins Hospital (JHH). We have as of now collected data from 28 subjects (this includes 5 stroke patients, the remaining being patients who were evaluated by the BAT team, but the eventual diagnosis was a non-stroke condition) and 2 control group subjects.

Starting from the NIH Stroke Scale (NIHSS), we have iteratively developed a subject data recording protocol to enable a standardized and easily replicable set of tasks which are highly informative of the underlying condition of the subject. These tasks have been specially designed to be algorithmically interpreted using our custom developed AI modules. This subject data recording protocol incorporates evaluation schemes / task paradigms for non-responsive and aphasic patients (as per the Glasgow coma scale) apart from responsive patients. All our data collection is being done after the primary therapeutic intervention has been completed.

While we still do not have a sufficient amount of data to evaluate an end-to-end deep learning AI platform, we are working on individual components of the AI/ML data analysis pipeline. Our overall architecture is based on multimodal early fusion of the two primary data streams - image based information of facial movement and the speech patterns. For the analysis of the image stream, we have developed two pipelines. The first pipeline consists of a convolutional neural network (CNN) based architecture to extract 68 facial landmarks from the face in every frame of the video. Given the subtle nature of the information in the facial movements, we are using optical flow-based tracking to stabilize the facial landmarks thus ensuring a high signal to noise ratio (Figure 2). These landmarks are used to compute standard facial metrics for every frame. The time series of these metrics is the input for both the classifier for stroke detection and the regressor for severity quantification. The second pipeline for analyzing the image data is segmenting the face into regions (such as forehead, eye region, mouth etc.) and using a CNN with separate filters for each region to extract the appropriate regional features. In both the image-based analysis pipelines we are also extracting a CNN based attention score for each frame which will inform the classifier/regressor to weigh the data from that frame appropriately.

For the audio analysis, we are extracting a mel frequency spectrogram (Figure 3) from the recorded speech of the patient. To enable this, we are manually annotating the regions of the complete video where only the patient's voice is audible. These sections are used to create the spectrogram. The spectrogram will be passed through a CNN to extract features which will be fused with the features from the video stream at multiple stages.

While we have built/trained the individual components (stabilized landmarks, calculation of facial metrics, face parsing/segmentation, mel frequency spectrogram), the final classifier / regressor will be jointly trained on the all the data collected once we have reached a critical mass.

KISHORE V. KUCHIBHOTLA PHD

Assistant Professor, Dept. of Psychological & Brain Sciences, Neuroscience and Biomedical Engineering at Johns Hopkins University

Kishore Kuchibhotla is an Assistant Professor in Psychological & Brain Sciences, Neuroscience and Biomedical Engineering at Johns Hopkins University. He is an expert in Alzheimer's disease and the neuroscience of learning and memory. In addition, he has extensive experience in industry working on developing novel solutions for healthcare-related challenges.

A Technology Platform to Monitor Cognitive Fluctuations and Lucid Intervals in Dementia at Home

In our original IDIES seed proposal, we sought to provide a robust framework for quantifying cognitive fluctuations and lucid intervals in individuals with Alzheimer's disease (AD) using a mobile technology platform applied within one's home or local environment. After receiving support from the IDIES Seed Fund, we worked with a team of software developers at Valere Labs LLC to fully restructure and re-design our prototype Lucidity app to standardize the UI/UX based on current best practices. Efforts included integrating the core front-end components of the mobile application (an onboarding sequence, game-based psychometric tests administered in the form of ecological momentary assessments (EMAs), and a diary-based catalog of notable events), back-end architecture planning/data modeling, and API integration with wearable devices.

We are now in the development phase of the application and ecosystem with Valere Labs LLC to build a robust, reliable, and extensible codebase initially for Android OS and subsequently for iOS (Apple). Key features of our Lucidity technology platform include: 1) Detailed onboarding workflow to collect caregiver and patient information; 2) Visual psychophysical 'games' to quantify performance across cognitive domains: spatial memory, attention, language, and executive functions; 3) Novel episodic memory tests in which caregivers pre-load photos of family members (Face Memory game) and notable events (Event Memory game) that are tagged with relevant information (names, location, event type, etc.); 4) Intuitively designed caregiver diaries that allow for structured subjective evaluation by caregivers at regular time-points and during periods where they observed a cognitive change; 5) Minimal integration with Fitbit sensor to allow passive collection of biomarkers that can be time-stamped with game performance and caregiver diaries; 6) Integrated back-end in which all three dimensions of data (games, caregiver diaries, and health sensor information) are time-stamped and saved in the cloud.

Following the completion of the development of the Lucidity platform and ecosystem, we will move forward with a 2-month pilot study, enrolling caregiver-AD patient dyads, caregiver-MCI patient dyads, and cognitively-healthy elderly subjects to use and interact with the Lucidity application and ecosystem. We will be collecting preliminary data (timestamped health sensor recordings, cognitive test metrics, and caregiver

Laying the Foundation for Large-scale Precision Stellar Parameter Inference in the Field of Exoplanets

Keyi Ding
Majors: Computer Science, Physics
Faculty Mentor: Kevin C. Schlaufman
William H. Miller III Department of Physics and Astronomy
Project Date: June 1, 2022 - August 15, 2022 (with a week's intermission for summer travel)

Laying the Foundation for Large-scale Precision Stellar Parameter Inference in the Field of Exoplanets

We propose to determine the stellar physical properties of stars in open clusters using the Isochrones Python package to execute with MultiNest, a Bayesian fit of the MESA Isochrones Stellar Tracks (MIST) isochrone grid to parallax and multiwavelength photometry data for open cluster stars of known metallicities and ages. Stars in open clusters are stars with similar ages and metallicities, and their physical properties provide valuable insight into the study of stellar evolution and exoplanets. We will apply our data science approach of determining stellar properties of stars in open clusters with photometric data and verify that our methodology provides accurate and precise results. Once verified, the result of this project will be valuable to researchers studying galactic astronomy and exoplanets in determining the physical parameters of stars through a data science approach.

Proposal Title
Optimizing Routes in the Operating Room

Student
Ryan Chou
JHU Pre-medical Biomedical Engineering Undergrad

Faculty Mentor
Gregory D Hager
Mandell Bellmore Professor of Computer Science

Project Dates
06/01/2022 - 08/10/2022

Optimizing Routes in the Operating Room

An ongoing effort by the healthcare community is directed towards understanding how novice surgeons become experts. Knowing what expert surgeons do that novices do not has the potential to greatly influence surgical training and assessment methods, and technologies. This project looks at septoplasty, a surgical procedure to remove asymmetrical parts of the nasal septum (which separates the two nostrils), specifically a phase where a surgeon must separate the mucosal flap (tissue) from the nasal septum (cartilage and bone). The objective of this study is to determine whether expert septoplasty surgeons can be differentiated from novices by their ability to move their instrument (cottle elevator) efficiently (in an optimal path) during the separation of the mucosa and the septum.

IDIES MISSION STATEMENT

To further the JHU mission of “Knowledge for the world” by providing intellectual leadership in data-driven science:

- Research and Education in adaptive disruptive technologies
- Technical and domain expert guidance via collaboration and consultation
- Open and sustainable long-term access to high-value datasets

THANK YOU TO OUR GENEROUS SPONSORS

The IDIES Executive Committee would like to extend our heartfelt gratitude to our affiliates, collaborators, contributors, editors, and staff, without whose continued support and cooperation IDIES would not be possible.

IDIES is always accepting members. For more inforamation about our membership categories, please visit idies.jhu.edu join for more information, and to join today!