MFS Symposium on AI – Now and Beyond (AINBe) as part of Institute Research Day (IRD 2023)

Venue: BSBE Auditorium, New Bio-Technology Building

March 14, 10:00 AM

Session 1
Speaker : Shankar Subramaniam

Shankar Subramaniam, Distinguished Professor Joan and Irwin Jacobs Endowed Chair in Bioengineering and Systems Biology, University of California San Diego

TITLE: Is big data the sine qua non of human future? Does it define the future of AI?

ABSTRACT:

In today's world, data is everywhere, and the ability to analyze and make sense of it has become increasingly important. This has led to the rise of "big data" and the development of powerful tools for processing and analyzing it. At the same time, the field of artificial intelligence (AI) has seen tremendous growth, with applications ranging from self-driving cars to virtual assistants. But what is the relationship between big data and AI, and what role does big data play in defining the future of AI?

In this talk, we'll explore the interplay between big data and AI, and consider whether big data is truly the sine qua non of the human future. We'll examine the ways in which big data has enabled breakthroughs in AI, from natural language processing to image recognition. We'll also discuss the challenges and limitations of big data, such as issues of privacy and bias. Finally, we'll consider alternative approaches to AI that don't rely on big data and discuss the potential implications for the future of AI.

Whether you're a data scientist, AI researcher, or simply interested in the intersection of technology and society, this talk will provide valuable insights into the role of big data in shaping the future of AI.

BIO: Shankar Subramaniam is a Distinguished Professor of Bioengineering, Computer Science and Engineering, Cellular and Molecular Medicine, Computer Science and Engineering and Nano Engineering. He was the Chair of the Bioengineering Department at the University of California at San Diego (2008-13). He holds the inaugural Joan and Irwin Jacobs Endowed Chair in Bioengineering and Systems Biology. He was the Founding Director of the Bioinformatics Graduate Program at the University of California at San Diego. Prior to moving to UC San Diego, Dr. Subramaniam was a Professor of Biophysics, Biochemistry, Molecular and Integrative Physiology, Chemical Engineering and Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign (UIUC).

In 2020 he was elected as a Fellow if IAMBE and in 2013 he was elected as a Fellow of AAAS. In 2002 he received the Genome Technology All Star Award. He is a fellow of the AIMBE and is a recipient of Smithsonian Foundation and Association of Laboratory Automation

Awards and his research work is described below. In 2019 he was elected as the diamond jubilee distinguished alumni by the Indian Institute of Technology Kanpur. In 2008 he was awarded the Faculty Excellence in Research Award at the University of California at San Diego. In 2011 he was appointed as a Distinguished Scientist at the San Diego Supercomputer Center. In 2019 he was awarded the diamond jubilee Distinguished Alumni Award by the Indian Institute of Technology Kanpur. He has served on the External Advisory Boards for several Bio/Biomedical Engineering Departments including Johns Hopkins U., Case Western Reserve U., U. Penn, Rice U., and UT Austin. He is currently the Chair of the Wellcome Trust-DBT India Alliance Team Science Grant Panel. In 2012, he was elected as the Chair of the College of Fellows of AIMBE. He also serves on the Scientific Advisory Board of Janssen Pharmaceuticals (the research arm of Johnson and Johnson). He has served on the Scientific Councils of NIGMS and NHGRI (NIH Institutes) and as a Chair of three distinct study sections at the National Institutes of Health. Subramaniam has graduated over 80 Ph.D. students who occupy leading academic and industrial positions. He has trained over 100 postdoctoral researchers. His research is funded by the National Institutes of Health, National Science Foundation, the Wellcome Trust, and the Chen Foundation.

Subramaniam's innovative work has major impact on research and development in academia and industry by allowing the synthesis of complex biological and medical information from genes and molecules into integrated knowledge at cellular and system levels, thus providing important basis for drug discovery and innovation. He was a pioneer in bioinformatics with his development of the Biology Workbench, the first of its kind in web-based infrastructures. He has fostered training and research in systems biology and bioinformatics at the national level, serving on the NIH Director's Advisory Committee on Bioinformatics and played a key role in the formulation of the NIH Director's Roadmap which places a major emphasis on the use of quantitative approaches of engineering to biomedical research in health and disease. He has been instrumental in raising national awareness of the roles of these engineering approaches to biomedical research. He founded the UCSD Bioinformatics program and was Chair of the nationally top-ranked bioengineering program from 2008-2013. Subramaniam has collaborated with colleagues in clinical medicine to elucidate the molecular and genomic basis of the pathogenesis of diabetes, inflammation, atherosclerosis, and myopathies by using modern approaches of systems biology and bioinformatics to analyze physiological and pathophysiological data, leading to the development of novel therapeutic measures and drug discovery.

Subramaniam has made innovative contributions at the interface of engineering and medicine. In addition to inventing new methods for analysis of complex systems, he pioneered a novel technology for RNA sequencing with the smallest quantities of RNA leading to our ability to analyze human tissues at the microscale. His contributions to models of human disease are wide and profound and have strong implications for precision and personalized systems medicine.

Session 2
Speaker : Noel Buckley

TITLE: Using Deep Learning to Identify Neurodegenerative Cellular Phenotypes

Noel J Buckley, *Dept. Psychiatry & Kavli Institute for Nanoscience Discovery & Wellcome Centre for Human Genetics, University of Oxford*

ABSTRACT: Dementia, particular Alzheimer's Disease (AD), is the greatest unmet medical need of the 21st century. Despite huge effort from both Pharma and academia, there are currently no disease-modifying treatments. Human cellular models of AD are essential to understanding disease mechanism and to identifying interventions that can slow, block or reverse disease pathophysiology. Human iPSC-derived neural cells have emerged as a powerful tool in this arena and, in combination with gene editing, are leading to an increasing suite of cells derived from defined genetic backgrounds and disease risk that can be applied in forward or reverse screens. Thus far, the vast majority of work has focussed on the rare familial form of AD (representing less than 1% of AD) using iPSCs derived from disease causing APP or PSEN backgrounds, but the relevance of these studies to the prevalent sporadic form of AD is unclear. In contrast, attempts to study sAD have largely used iPSCs derived from AD patients and non-AD "controls". Typically, the outcomes used to assess changes in (disease) phenotype in neurons have used cellular readouts such as neurite length and branching, synapse number, excitability, cell death. However, this pre-selection of cellular features represents a low sampling of available information and an overfocus on 'late-stage' phenotype. We are circumventing these roadblocks by generating neurons and glia from iPSCs gene-edited to bear different alleles of the biggest sAD risk gene, APOE, followed by use of Deep Learning (DL) approaches to rapidly and robustly identify phenotypic signatures that distinguish risk from neutral or protected backgrounds. One of the important advantages of DL lies in its ability to learn characteristic features from raw input data. This contrasts with 'shallow' machine learning methods that require features (nuclear shape, size, neurite length) to be engineered through human-suggested data transformations - including commercial imaging software, such as Harmony and CellProfiler. Our approach requires DL to accomplish three metatasks of classifying cell phenotype according to (i) cell type (ii) underlying genetics (iii) external perturbation. I will present our initial studies that show all these three meta-tasks can be accomplished.

BIO: The human brain contains around 100 billion neurons. Their genesis, development and degeneration are all governed by underlying gene regulatory networks (GRNs). These fundamental biological processes represent a nexus where basic and translational neurobiology converge; the same processes that orchestrate normal neuronal development, maturation and death are those that malfunction in neurodevelopmental and neurodegenerative disorders. We aim to identify these networks and uncover how they vary across individuals, thereby causing errors of neurodevelopment or susceptibility to neurodegeneration. This overarching goal is as much about developing a basic understanding of how neuronal phenotype emerges from interactions among the genes of the network as it is about translating this understanding into identification of novel therapeutic targets. All of our work is guided by a credo that understanding basic biological mechanisms and identifying targets for therapeutic intervention are inextricably intertwined. To implement this vision, we work collaboratively with basic, clinical and translational neuroscientists, molecular biologists, stem cell biologists and computational biologists.

This mission can be broken down into three overlapping phases, First, we need to define the networks. We do this by using induced pluripotent stem cells (iPSCs) to examine neuronal development and degeneration on specific genetic backgrounds (our disease focus is increasingly on ASD and Alzheimer's Disease). We use a variety of genomic tools to harvest transcriptional and epigenetic data and then interrogate these data to infer the interactions among the genes and to define how the network topology changes during neuronal development and neuronal death. Second, we need to attribute changes in network topology to

specific genes. This we do by using genome editing to introduce specific genetic changes and then reanalyzing the network topology. Third, we use a range of genetic and small molecule manipulations to target specific pathways inferred from the network analyses. Our hope is to identify candidate pathways and genes that may represent novel therapeutic targets to delay, abrogate or rescue aberrant neuronal development or degeneration.

As well as this systems approach, we also focus on transcription factors that are known regulators of neurodevelopment and neurodegeneration. Much of our attention has been focused on REST, a key transcriptional regulator, that we have shown plays a critical role in embryonic neurogenesis and in maintaining adult neuronal phenotype. Furthermore, REST is a critical component in mediating cell death in Huntington's Disease and potentially in rescuing cell death in Alzheimer's disease.

We are part of a broad consortium of scientists at Oxford under the ARUK Drug Development Institute that will expedite rapid translation of our findings to develop target-enabling packages including assays, screens, probe compounds and other reagents for target development.

Session 3
Speaker : Rajesh Gupta

Rajesh Gupta, Professor and Qualcomm Endowed Chair, University of California San Diego

TITLE: New Services and Programming Infrastructure for Causally-Connected Cyber-Physical Systems

ABSTRACT: Over the past decade, the embedded and control systems community have vigorously pursued a vision of coupled feedback-controlled systems with a broad range of real-life applications from transportation, smart buildings to human health. These efforts have continued to push intelligent processing to edge and near-edge devices, provide new capabilities for improved sensing with high quality timing information, establish limits on the quality of time and its impact on the stability of control algorithms etc. In this talk, I will outline our vision of how we can treat physical spaces and built environments as consisting of sensing, actuation, processing and communication resources that are dynamically discovered and put to use through emerging meta-data schema and methods. Technical challenges and progress in creating a software stack to enable new class of applications.

BIO: Dr. Rajesh Gupta serves as the founding director of the Halıcıoğlu Data Science Institute and as a distinguished professor of Computer Science and Engineering at UC San Diego. Professor Gupta's research is in embedded and cyber-physical systems with a focus on sensor data organization and its use in optimization and analytics. He currently leads NSF project MetroInsight and a co-PI on DARPA/SRC Center on Computing on Network Infrastructure (CONIX) with the goal to build a new generation of distributed cyber-physical systems that use city-scale sensing data for improved services and autonomy.
Dr. Gupta is a recipient of IEEE Computer Society W. Wallace McDowell Award, IEEE TCCPS Distinguished Leadership Award, the National Science Foundation CAREER Award, two Departmental Achievement Awards, and a Components Research Team Award at Intel. Dr. Gupta and his students have received a best demonstration paper award at ACM BuildSys'16, best paper award at IEEE/ACM DCOSS'08, and a best demonstration award at

IEEE/ACM IPSN/SPOTS'05. He served as an advisor to Tallwood Venture Capital, RealIntent, Calypto and Packet Digital Corporation.

He also served as editor-in-chief of IEEE Design & Test of Computers and founding editor-in-chief of IEEE Embedded Systems Letters. He currently serves as editor-in-chief of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. Prof. Gupta holds the Qualcomm Endowed Chair in Embedded Microsystems at UC San Diego and the INRIA International Chair at the French International Research Institute in Rennes, Bretagne Atlantique. He is a Fellow of the IEEE, the ACM, and the American Association for the Advancement of Science (AAAS).

Session 4
Speaker : Ananth Grama

Ananth K. Grama, Samuel D. Conte Professor of Computer Science, Purdue University

TITLE: Novel Matrix Decompositions and Applications in Clinical and Biological Data Analysis

ABSTRACT: Matrix decompositions are commonly used in various forms (e.g., clustering, dimensionality reduction) in diverse applications. In this talk, I will present a novel variant of matrix decomposition based on Archetypal Analysis. Archetypal Analysis aims to find extremal points in high-dimensional data, with the benefit of representing all data points as convex combinations of these extremal samples. We extend Archetypal Analysis to supervised settings by regularizing on coherence of label-data associated
with samples. We present methods for computing supervised archetypes and demonstrate their excellent performance characteristics.

We demonstrate the power of our methods on two important applications. Our first application deals with identification of cell types from single cell transcriptomes. Methods for automatically characterizing the functional identity of cells, and their associated properties, can be used to uncover processes involved in lineage differentiation as well as sub-typing cancer cells. They can also be used to suggest personalized therapies based on molecular signatures associated with pathology. Our new method
is used to infer the functional identity of cells from their transcriptional profile, classify them based on their dominant function, and reconstruct regulatory networks that are responsible for mediating their identity. We identify novel Melanoma subtypes with differential survival rates and therapeutic responses, for which we provide biomarkers along with their underlying regulatory networks.

Our second application focuses on inferring types, etiology, and interventions associated with clinical data on Sepsis. We identify disease states in sepsis and model disease progression using clinical variables and patient samples in the MIMIC-III database. We identify six distinct patient states in sepsis, each associated with different manifestations of organ dysfunction. We find that patients in different sepsis states are statistically significantly composed of distinct populations with disparate demographic and comorbidity profiles. Our progression model accurately characterizes the severity level of each pathological trajectory and identifies significant changes in clinical variables and treatment actions during sepsis state

transitions. Finally, we demonstrate how our method can be used to construct powerful personalized intervention models for Sepsis through an example of ventilator use recommendation. Collectively, our framework provides a holistic view of sepsis, and our findings provide the basis for future development of clinical trials, prevention, and therapeutic strategies.

BIO: Ananth Grama's research interests span parallel and distributed computing architectures, algorithms, and applications. His work on distributed infrastructure includes software support for dynamic clustered and multi-clustered environments. His recent work focuses on resource location and allocation mechanisms in peer-to-peer networks. His research on applications has focused on particle dynamics methods, their applications to dense linear system solvers, and fast algorithms for data compression and analysis.

Dr. Grama has authored several papers and co-authored a text book, *Introduction to Parallel Computing: Design and Analysis of Algorithms*, with Vipin Kumar, Anshul Gupta, and George Karypis. He is a member of the American Association for Advancement of Sciences and Sigma Xi.

Dr. Grama is the Samuel Conte Professor of Computer Science at Purdue University. He also serves as Associate Director of the Center for Science of Information, a Science and Technology Center of the National Science Foundation. He joined Purdue in 1996 as an Assistant Professor and has been there since. He received a PhD in Computer Science from the University of Minnesota (1996), an MS in Computer Engineering from Wayne State University (1990), and a B. Egg. in Computer Science from the Indian Institute of Technology Roorkee (1989). His primary research interests lie in broad areas of parallel and distributed computing, large-scale data analytics, design and simulation, and applications. Dr. Grama's work has been recognized through a number of awards, including the NSF CAREER award (1998), Purdue University Faculty Scholar (2002), Fellow of the American Association for Advancement of Science (AAAS) (2014), Distinguished Alumnus of the University of Minnesota (2015), and Amazon Research Award (2021).