ARBIC

ARKANSAS BIOINFORMATICS CONSORTIUM

Artificial Intelligence (AI) in Arkansas (AR)

AR-BIC 2020: AI in AR – 6th Annual Conference

February 10 –11, 2020

Conference Center
Wyndham Riverfront Little Rock
North Little Rock, Arkansas

www.aralliance.org/ar-bic

THANK YOU

The AR-BIC Governance Board wishes to thank all of the Scientific Program Committee, our visiting guests, student-scholars, researchers, mentors, and the gifted administrators from universities and laboratories across Arkansas who helped organize and are participating in AR-BIC 2020. YOU ARE THE FUEL OF THE AR-BIC COMMUNITY. Our sincere gratitude also goes to the conference sponsors for their continued support and investment in providing the opportunities for training up-and-coming scientific talent and showcasing some of the most exciting research in Arkansas.

CONFERENCE SPONSORS

- Arkansas Research Alliance (ARA)
- Arkansas Biosciences Institute (ABI)
- University of Arkansas (UA)
- Arkansas State University (A-State)
- University of Arkansas for Medical Sciences (UAMS)
- University of Arkansas at Little Rock (UA Little Rock)
- University of Arkansas at Pine Bluff (UAPB)
- Arkansas Economic Development Commission (AEDC)
- National Centre for Toxicological Research (NCTR), US- Food and Drug Administration (US FDA)

ACKNOWLEDGEMENT

Funding for this conference was made possible, in part, by the U.S. FDA through grant 1R13FD006690-01. Views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the government of the United States.

ABOUT AR-BIC 2020:
ARTIFICIAL INTELLIGENCE (AI) IN ARKANSAS (AR)

The Arkansas Bioinformatics Consortium (AR-BIC) is an Arkansas-centric community that stimulates communication, collaboration, training, and entrepreneurial activities to leverage expertise and resources in data science and bioinformatics. Established in 2014, AR-BIC began as an annual conference to identify and take stock of the scientific research activities and assets across the Natural State. The annual meeting evolved over the next four years to explore bioinformatics in precision medicine, genomics, microbiome, food and agriculture, and data analytics and data science. This year, AR-BIC's sixth annual conference focuses on Artificial Intelligence (AI) in Arkansas (AR). The one-and-half-day conference is being held at the Wyndham Riverfront Little Rock, February 10-11, 2020.

AI is a broad concept of training machines to think and behave like humans. It encompasses a wide range of statistical and machine learning (ML) approaches with a specific emphasis on learning from existing data and information to predict future outcomes. This conference will discuss the concepts and methodologies of AI applied in the fields of biomedical research, food safety, drug discovery and development, and clinical settings. In addition to providing a platform for networking and collaboration, AR-BIC will showcase AI expertise and experience in Arkansas.

AR-BIC GOVERNANCE BOARD

- Arkansas Biosciences Institute (ABI): **Bobby McGehee**
- Arkansas Economy Development Commission (AEDC): **Tom Chilton**
- Arkansas Research Alliance (ARA): **Jerry Adams (Chair), Bryan Barnhouse,** and **Julie LaRue**
- Arkansas State University (A-State): **Thomas Risch**
- National Center for Toxicological Research (NCTR): **William Slikker, Weida Tong, Shraddha Thakkar**
- University of Arkansas (UA): **Dan Sui**
- University of Arkansas at Little Rock (UA Little Rock): **Brian Berry**
- University of Arkansas for Medical Sciences (UAMS): **Shuk-Mei Ho**
- University of Arkansas at Pine Bluff (UAPB): **Mansour Mortazavi**

SCIENTIFIC PROGRAM COMMITTEE

- Arkansas Research Alliance (ARA): **Bryan Barnhouse** and **Julie LaRue**
- Arkansas State University (A-State): **Jason Causey, Jake Qualls, Xiuzhen Huang, Emily Bellis**
- National Center for Toxicological Research (NCTR): **Weida Tong (Chair), Shraddha Thakkar, Tanmay Bera, Syed Ali**
- University of Arkansas (UA): **Justin Zhan, Douglas Rhoads**
- University of Arkansas at Little Rock (UA Little Rock): **Mary Yang** and **Jerry Darsey**
- University of Arkansas for Medical Sciences (UAMS): **David Ussery, Timothy Ashby, Michael Bauer**
- University of Arkansas at Pine Bluff (UAPB): **Karl Walker**

CONFERENCE POINTS OF CONTACT

- Logistics: Julie LaRue (jlarue@aralliance.org)
- Scientific Programs: Weida Tong (Weida.Tong@fda.hhs.gov)

AR-BIC CONTACT INFORMATION

Jerry B. Adams, Chair, AR-BIC Governance Board
President & CEO
jadams@aralliance.org
Bryan J. Barnhouse, Chief Operating Officer, ARA
bbarnhouse@aralliance.org
Julie LaRue, Senior Project Manager, ARA
jlarue@aralliance.org

Arkansas Research Alliance
1125 Oak Street, Suite 301
Conway, Arkansas 72032
(501) 450-7818
www.aralliance.org
www.aralliance.org/ar-bic

CONFERENCE PROGRAM

Day 1 (Monday, February 10)
10:00 am – 1:00 pm: Registration and Poster Set-up
1:00 pm – 1:10 pm: Welcome Remarks • Governor Asa Hutchinson (video) • Jerry Adams (ARA) • Stephanie Gardner (UAMS)
1:10 pm – 4:00 pm: Introduction to Artificial Intelligence (AI) Co-Chaired by Weida Tong (NCTR) and Brian Berry (UA Little Rock) • **AI – An Introduction**, May Wang (Georgia Tech) • **AI for Text Mining**, Xiaowei Xu (UA Little Rock) • **AI for Biomedical Image Analysis**, Mariofanna Milanova (UA Little Rock)
4:00 pm – 5:00 pm: AI – A Perspective from NSF, Balakrishnan Prabhakaran (National Science Foundation, USA)
5:00 pm – 7:30 pm: Poster Presentation and Reception

Day 2 (Tuesday, February 11)
8:00 am – 9:00 am: Breakfast and Networking
9:00 am – 10:00 am: An Introduction to Transfer Learning for Predictive Medicine, with Cautionary Tales, Cesare Furlanello, Fondazione Bruno Kessler (FBK), Trento, Italy & HK3 Lab, Milan, Italy
10:00 pm – 10:20 am: Coffee Break
10:20 am – 12:00 pm: Session 1: ARA Academy on AI (25 min each) Chaired by Bryan Barnhouse and Doug Hutchings (ARA) • **AI for Brain,** Keith Bush (UAMS) • **FDA Efforts on ML for Genomic Biomarkers,** Weida Tong (NCTR) • **Biomedical Imaging Segmentation,** Xiuzhen Huang (A-State) • **Data Bridge and Deep Learning for AI,** Justin Zhan (UA)
12:00 pm – 1:00 pm: Lunch

1:00 pm – 2:20 pm: Session 2: Imaging and Radiomics (20 min each)
Co-Chaired by Douglas Rhoads (UA) and Fred Prior (UAMS)

- **FIDE: An AI-based Recognition System to Advance ORA Regulatory Operations at Importing Mail Facilities,** Leihong Wu (NCTR)
- **Advanced Approaches in Deep Learning and Computer Vision to 3D Face Emotion Recognition in Food Eating,** Khoa Luu (UA)
- **Deep-Learning for Automated Segmentation of Skin Wound Features,** Kyle Quinn (UA)
- **Radiomics in Cancer Imaging,** Fred Prior (UAMS)

2:20 pm – 2:50 pm: Coffee Break

2:50 pm – 4:30 pm: Session 3: AI/ML in biomedical research (20min each)
Co-Chaired by Emily Bellis (A-State) and Shraddha Thakkar (NCTR)

- **Advances in Neural AI and Applications to Drug Discovery,** Andreu Vall, Institute for Machine Learning, Johannes Kepler University, Linz, Austria (JKU)
- **AI for Drug Design and Drug Discovery,** Jerry Darsey (UA Little Rock)
- **Deep Learning Inferential Methods for Protein Structure Predictions,** Mahmoud Moradi (UA)
- **Drug Repositioning in the Era of AI,** Zhichao Liu (NCTR)
- **AI/ML for Predicting Drug-Induced Liver Injury,** Shraddha Thakkar (NCTR)

4:30 pm – 5:00 pm: Poster Awards and Concluding Remarks

AR-BIC 2020 EVALUATION LINK

www.surveymonkey.com/r/AR-BIC2020-Evaluation

SAVE THE DATE FOR AR-BIC 2021: FEBRUARY 8-9

LIST OF POSTER PRESENTERS

#	Name	Institute	Title	Pg. No.
1	**Hong Fang**	**NCTR**	FDALabel: Amazon Cloud Database for Drug Labeling information	28
2	**Wei Vivian Zhang**	**NCTR**	Statistical Methods to Analyze Uncertain Liquid Biopsy qPCR Data	29
3	**Trevis Huggins**	**USDA**	Efficient Accession Management and Characterization of the USDA-ARS Rice Germplasm Collection through Phenotyping and Genotyping	30
4	**Shanzhi Wang**	**UA Little Rock**	Deep sequencing for dissecting the roles of the enzymatic and scaffolding functions of EXO1 *in vivo*	31
5	**Stefan Graw**	**UAMS**	proteiNorm – A user-friendly tool for normalization and analysis of TMT and label-free protein quantification	32
6	**Nesreen Aljahdali**	**NCTR**	Genomic characterization of plasmids containing antimicrobial resistance and virulence genes of the susceptible recipient strain	33
7	**Santosh Sharma**	**USDA**	Genomic Prediction and Bayesian Network Analysis of Multiple Root Architecture Traits in Rice	34
8	**Tanmay Bera**	**NCTR**	Developing imaging methods for automated species level identification of food contaminating beetles	35
9	**Dongying Li**	**NCTR**	Global Profiling of miRNA Expression for Candidates Associated with Ketoconazole-induced Liver Injury	36
10	**Kapil Khadka**	**NCTR**	Understanding sex-specific drug properties and drug-induced liver toxicity related adverse outcome pathways in humans	37
11	**Amit Kumar**	**NCTR**	Microbiome shift and immune responses: differential effects of corn oil on different rodent models	38
12	**Manoj K. Shah**	**UAPB**	Bandwidth measurement of GeSn based high speed photodetectors	39
13	**Kristin Ashby**	**NCTR**	Risk Factors Associated with Severe Drug Induced Liver Injury	40
14	**Kevin Chappell**	**UA Little Rock**	PTMViz: A tool for analyzing and visualizing histone post translational modification data	41
15	**Ryan Sanders**	**UA**	Extracting Patterns in Medical Claims Data for Opioid Overdose	42
16	**Nnamdi. S. Ekesi**	**UA**	Genomic, and Virulence Comparisons of Different Bacterial Isolates from BCO Lesions in Broilers	43
17	**Thy A. Dao**	**UA**	A Distance Based Multisample Test for High-Dimensional Compositional Data with an Application to the Human Microbiome	44
18	**Saly Abouelenein**	**UAMS**	Identification of Risk Factors of Lung Cancer Mortality in the American Poultry Workers Using Random Forest Algorithm	45

19	**Ting Li**	**NCTR**	Genomics model to predict Drug-Induced Liver Injury with deep learning: DeepDILI+	46
20	**Abdulkarim Shwani**	**UA**	Directed Genome Evolution Identifies Deoxyribose Phosphate Aldolase as a Macrophage Survival Factor in Staphylococcus agnetis	47
21	**Yheni Dwiningsih**	**UA**	Identification of QTLs Associated with Drought Resistance Traits at Reproductive Stage in K/Z RILs Rice Population	48
22	**Xinming Li**	**UA Little Rock**	Reference2Vec: Teach Machine to Read People References	49
23	**Hunter L. Tiner**	**UA Little Rock**	Fall 2019 Cheminformatics OLCC	50
24	**Aaron S. Kemp**	**UAMS**	Semantic Representations of Multi-Modal Data, NeuroInformatic Processing Pipelines, and Derived Neuroimaging Results in the Arkansas Image Enterprise System (ARIES)	51
25	**Abdulla Said Ali**	**UAPB**	Design and Simulation of GeSn based Waveguide for Infrared Photonic at $\lambda = 2\mu m - 3\mu m$ from 1% to 10% of Sn Composition	53
26	**Abass A. Oduola**	**UA**	Comparison of the Impacts of Broadband and Selected Infrared Wavelengths on Inactivation of Microbes on Rough Rice	54
27	**Xixi Wang**	**UA Little Rock**	Industry Payment to Vascular Neurologists: A Six-year Analysis of the Open Payments Program from 2013 through 2018	55
28	**Khulud Alotaibi**	**UA Little Rock**	Isolation and Molecular Characterization of Shiga-Toxin producing *Escherichia coli* from Food and Clinical samples	56
29	**Kate Sanders**	**Hendrix College**	Improving the Phylogeny of *Arceuthobium* (Dwarf Mistletoe)	57
30	**Anjala Jiji**	**A-State**	Evaluation of lightweight alignment tools for RNA-seq quantification in plants	58
31	**Gabrielle A. Agcaoili**	**UA**	Differences in learning and gene expression in brains of male and female *Bicyclus anynana*	59
32	**Jessica Gilmore, Di' Granger**	**UA**	Using Procalcitonin Levels: Benefits to Antimicrobial Stewardship and Aspiration Diagnosis	60
33	**Benjamin Ingle**	**UA**	Identification of Individual and Regional Features Impacting Childhood Nutrition in Tanzania	61
34	**Stetson Ledbetter, Emily Bell**	**UA**	Exploring the Relationship between Maternal Health Characteristics and Preterm Birth using Classification and Regression Tree (CART) Methods	62
35	**Abigeal Omolewu**	**UA**	Forecasting the Number of Illnesses Related to Salmonella Outbreaks	63
36	**Andrew Palmer**	**UA**	Regional Differences in Factors Influencing Arkansas Teenage Pregnancy Rates	64

BIOGRAPHIES OF THE SPEAKERS

KEYNOTE SPEAKERS

Name	*Institute*	*Topic*	*Page No.*
Balakrishnan Prabhakaran	NSF	Artificial Intelligence: A Perspective from the National Science Foundation (NSF)	10
Cesare Furlanello	HK3 Lab & FBK, Italy	NSF An introduction to Transfer Learning for Predictive Medicine, with cautionary tales	11
May Wang	Georgia Tech	AI – an introduction	12

SPEAKERS

Name	*Institute*	*Topic*	*Page No.*
Andreu Vall	JKU, Austria	Advances in neural AI and applications to drug discovery	13
Fred Prior	UAMS	Radiomics in cancer imaging	14
Jerry Darsey	UA Little Rock	AI for drug design and drug discovery	15
Justin Zhan	UA	Data Bridge and Deep Learning For AI	16
Keith Bush	UAMS	AI for brain	17
Khoa Luu	UA	Advanced Approaches in Deep Learning and Computer Vision to 3D Face Emotion Recognition in Food Eating	18
Kyle Quinn	UA	Deep-learning for automated segmentation of skin wound features	19
Leihong Wu	NCTR	FIDE: an AI-based Recognition System to advance ORA regulatory operations at Importing Mail Facilities	20
Mahmoud Moradi	UA	Deep learning inferential methods for protein structure predictions	21
Mariofanna Milanova	UA Little Rock	AI for imaging analysis	22
Shraddha Thakkar	NCTR	AI/ML for predicting drug-induced liver injury	23
Weida Tong	NCTR	FDA efforts on ML for genomic biomarker	24
Xiaowei Xu	UA Little Rock	AI for text mining	25
Xiuzhen Huang	A-State	Biomedical imaging segmentation	26
Zhichao Liu	NCTR	Drug Repositioning in the Era of AI	27

Balakrishnan Prabhakaran, Ph.D.
Program Director
Information and Intelligent Systems (IIS)
National Science Foundation (NSF)

Dr. Prabha (Balakrishnan) Prabhakaran is currently a Program Director in the Cyber Human Systems (CHS) program of Information and Intelligent Systems (IIS) Division of the CISE (Computer and Information Science and Engineering) Directorate of the National Science Foundation (NSF). He is also involved with other programs such as Fairness in Artificial Intelligence, Future of Work, Secure and Trustworthy Computing.

Prabhakaran is a Professor in the faculty of Computer Science Department, University of Texas at Dallas. Dr. Prabhakaran received the prestigious NSF CAREER Award FY 2003 for his proposal on Animation Databases. He was selected as an ACM Distinguished Scientist in 2011 and is currently an IEEE Senior Member. He is an Associate Editor of IEEE Transactions on Multimedia. He is Member of the Editorial board of Multimedia Systems Journal (Springer), Multimedia Tools and Applications journal (Springer), and other multimedia systems journals. He received the Best Associate Editor for 2015, from Springer's Multimedia Systems Journal. Dr Prabhakaran is a Member of the Executive Council of the ACM Special Interest Group on Multimedia (SIGMM) and is the Co-Chair of IEEE Technical Committee on Multimedia Computing (TCMC) Special Interest Group on Video Analytics (SIGVA). Dr. Prabhakaran served the General Co-Chair of the IEEE International Conference on Health Informatics (ICHI) 2015. He was also a General Co-Chair of ACM International Conference on Multimedia Retrieval 2013 (ICMR 2013), IEEE Haptic, Audio, and Visual Environments (HAVE) 2014, a General Co-chair of ACM Multimedia 2011, and ACM Multimedia and Security (MM&Sec) 2007. Prof Prabhakaran's research has been funded by Federal Agencies such as the National Science Foundation (NSF), USA Army Research Office (ARO), and the US-IGNITE Program, apart from industries and consortiums. Full CV: http://utdallas.edu/~praba/cv.pdf.

Artificial Intelligence: A Perspective from the National Science Foundation (NSF)

This talk starts off with a brief background on Prabhakaran's personal research in health-care domain – primarily, building novel mixed-reality systems for evaluation as well as therapeutic use in Physical Medicine and Rehabilitation, targeting stroke and amputee patients. Pilot studies with these systems revealed the lack of health-care data (in medical research areas such as physical medicine and rehabilitation) that would be sufficient for the data hungry algorithms used in deep learning. Hence, Prabhakaran's research addressed motif discovery as well as synthetic data generation for increasing the data size to be used for training deep learning algorithms.

With this research background, Prabhakaran has been involved National Science Foundation (NSF) programs such as Cyber Human Systems, Fairness in Artificial Intelligence (AI), Future of Work, Secure and Trustworthy Computing, Smart and Connected Communities. The talk will provide an overview of NSF's leadership in AI and NSF CISE (Computer and Information Science and Engineering) Directorate's AI related activities. The discussion will then highlight the National initiative on AI Research Institutes and then some of the NSF's Big Ideas (such as Future of Work, Harnessing Data Revolution, and Mid-scale Research Infrastructure). Following that, the presentation will be on Information and Intelligent Systems (IIS) Division's programs such as Robust Intelligence, Information Integration and Informatics, Cyber Human Systems, Smart and Connected Health.

Cesare Furlanello, Ph.D.
Senior Scientific Advisor
Fondazione Bruno Kessler (FBK), Trento Italy &
HK3 Lab, Milan, Italy

Dr. Cesare Furlanello is an expert in Machine Learning (ML), predictive data analytics for massive bioinformatics data, and reproducibility of AI. He has more than 20 years of experience in ML applied to massive data in healthcare and biology, as head of the Neural Networks for Complex Data, now FBK/MPBA (Predictive Models for Biomedicine & Environment) research unit of the Kessler Foundation. He has served as Head of Data Science at FBK and as President of the MAQC international Society (2018-19). He chaired the MGED11 Workshop on Microarray Gene Expression Databases and the 3rd Workshop or the MAQC Society on Reproducibility of Artificial Intelligence for Predictive Medicine (2019). He has also coordinated two dedicated events on Deep Learning (DL) for Precision Medicine (ECML 2016 and 2017). His expertise is the accurate implementation of ML methodologies in infrastructures that enable reproducibility and applicability of AI solutions, with more than 80 projects funded by public agencies and industry. His goal is the design of reproducible AI algorithms for health, with a focus on multimodal integration of bioimaging (including digital pathology), omics and clinical data, supported by efficient technological upscaling in cloud computing frameworks. He is the recipient of two MS Azure Awards, including one of Deep Learning for Precision Medicine (2017). Furlanello has published more than 160 peer-reviewed scientific papers, including major journals such as Nature, Nature Genetics, Nature Biotechnology, Bioinformatics. He is a founding member of the MSc in Data Science of Trento University and supervised > 70 theses (BSc, MSc, including 11 PhDs) on machine learning projects. He is Adjunct Faculty of the Wistar Institute since 2012 and received the Full Professor National Habilitation in Biomedical Engineering in Dec 2017. He is the founder of three startups, including HK3 Lab (2019) on Deep Learning and its application to deliver reproducible AI for predictive medicine.

An introduction to Transfer Learning for Predictive Medicine, with cautionary tales

Learning Artificial Intelligence (AI) models that can be robustly applied on the same tasks across datasets or even on different tasks by minimal domain shift adaption is a topical issue in all fields of application of algorithms emulating human cognitive abilities. All starts with the success in creating artificial experts that can play multiple videogames, managing increasing levels of complexity. Almost every clinical specialty will use AI someway, sometime in the future; however, we do now yet if even for the same task within one specialty we will be employing a large population or a small team of models, or eventually a highly versatile machine. Transfer learning (and meta-learning more in general) is currently the most useful approach to train models on health data sets of small-moderate size by tuning with further training from larger data collections, even from non-medical domains. In some sense, models are used to train each other, distilling some characteristics of the teacher model together with extra data of interest. However there is a transversal and urgent need to improve the machine learning user experience on biomedical data with more robust model selection protocols, without compromising to selection bias or lack of interpretability. This talk will review the existing geography of transfer learning methods, with examples in different cases of interest for precision medicine. We will discuss how to deal with selection bias and hidden stratification by means of valid Data Analysis Plans applied in the context of transfer learning, showcasing studies on Deep Learning applied to genomics and bioimaging. Cautionary tales on information leakage in digital pathology will enforce the importance of Data Analysis Plans such as those developed within the MAQC initiative.

May Dongmei Wang, Ph.D.
Professor, Kavli Fellow
Georgia Research Alliance Distinguished Cancer Scholar
Georgia Institute of Technology and Emory University

Dr. May Dongmei Wang is a full professor in the Joint Biomedical Engineering Department of Georgia Tech and Emory University, a Kavli Fellow, a Georgia Cancer Coalition Scholar, and a Fellow of the American Institute for Biological and Medical Engineering (**AIMBE**). Her research interest is in Biomedical Big Data Analytics with a focus on Biomedical and Health Informatics (**BHI**) for predictive, personalized, and precision health (**pHealth**), including high throughput next-generation-sequencing and -omic data mining to identify clinical biomarkers, bionanoinformatics, pathological imaging informatics to assist clinical diagnosis, critical and chronic care health informatics for evidence-based health decision support, and predictive systems modeling to improve health outcome. In the FDA-organized MAQC international consortium, Dr. Wang led RNA-Seq data analysis pipeline study. She has published over 200 papers in peer-reviewed journals and conference proceedings. She delivered more than 200 invited and keynote lectures in professional conferences, academic institutions, healthcare systems, government agencies, and industry. Dr. Wang is a recipient of an Outstanding Faculty Mentor Award for Undergraduate Research at Georgia Tech, and a MilliPub Award from Emory University School of Medicine.

In addition, Dr. Wang is Georgia Tech Biomedical Informatics Program Co-Director in Atlanta Clinical and Translational Science Institute (**ACTSI**), Co-Director of Georgia-Tech Center of Bio-Imaging Mass Spectrometry, and Biocomputing and Bioinformatics Core Director in Emory-Georgia-Tech Cancer Nanotechnology Center. Her research has been supported by NIH, NSF, CDC, Georgia Research Alliance (GRA), Georgia Cancer Coalition (GCC), Emory-Georgia Tech Cancer Nanotechnology Center, Children's Health Care of Atlanta (CHOA), Atlanta Clinical and Translational Science Institute (ACTSI), and industrial partners such as Microsoft Research and HP.

Andreu Vall, Ph.D.
Postdoctoral researcher
Institute for Machine Learning
Johannes Kepler University
Linz, Austria

Dr. Andreu Vall received his Licentiate Degree (MSc) in Mathematics from the Technical University of Catalonia in 2008. From 2009 to 2014 he worked as a research assistant, first in the field of geomatics, and later on operations research applied to the fashion retail industry. He received his doctorate in 2019 from the Johannes Kepler University Linz, with a body of work focused on machine learning approaches to music recommender systems and specializing in the generation of personalized music playlists. Since 2019 he is a postdoctoral researcher at the Institute of Machine Learning of the Johannes Kepler University Linz, where he works on deep learning methods for drug discovery, with a particular interest for uncertainty estimation methods.

Advances in neural AI and applications to drug discovery

New methods of artificial intelligence, especially those based on deep neural networks, attracted much attention in drug discovery research after wining the Merck Molecular Activity Challenge in 2014 and the Tox21 Data Challenge in 2015. Since then, many areas of drug development (e.g., virtual screening, bioactivity prediction, toxicity prediction, QSAR modeling, but also synergy models, generative models, and chemical synthesis) completely changed the analysis of pharma-related microscopy data.

Neural AIs in the form of Deep Learning suffer from the fundamental problem of vanishing and exploding gradients, but despite this fundamental problem, yield overwhelming successes. Recently developed techniques, such as normalization techniques, residual networks, or recurrent networks with memory, have contributed to mitigate the vanishing gradient problem and enabled scientific progress in many areas, notably in computer vision and speech, but also in drug discovery.

We give an overview of recent developments in the area of neural AIs, the vanishing gradient problem, how to mitigate it, and successful application areas. We focus on drug discovery as a especially successful application area.

Fred Prior, Ph.D.
Professor and Chair
Department of Biomedical Informatics
University of Arkansas for Medical Sciences

Dr. Fred Prior, is Professor and Inaugural Chair of the Department of Biomedical Informatics as the University of Arkansas for Medical Sciences (UAMS). Dr. Prior has over 30 years of R&D experience in industry and academia, focused on the design of advanced medical information management and imaging technologies. He has held senior management positions in a variety of R&D environments ranging from Silicon Valley startups to major multi-national corporations in the US and Europe.

Dr. Prior's research interests include cancer informatics, radiomics, advanced imaging technologies and big data analytics. He serves as principle investigator and director of the US National Cancer Institute's Cancer Imaging Archive project and is the lead PI of an NCI funded Quantitative Imaging Network team developing tools for radiomic analysis and an NCI ITCR team exploring the integration of radiomics and pathomics. He serves as a fellow of the International Cancer Imaging Society and with European colleagues created an international conference, "The Wizardry of AI and Machine Learning in Cancer Imaging" which is held annually in Lisbon Portugal. Dr. Prior teaches Machine Learning at UAMS and lectures internationally.

Dr. Prior also directs informatics efforts for the UAMS Translational Research Institute. He is an associate editor of several leading information technology journals, and a reviewer for numerous other scientific and engineering journals as well as U.S. and European funding agencies. He is the author of over 100 scientific publications and holds 7 US and international patents.

Radiomics in Cancer Imaging

Imaging based cancer research is undergoing a transition from analyses based on human observers to the use of advanced computing platforms and software to automatically extract quantitative features relevant to prognosis or treatment response. Artificial Intelligence (AI) and machine learning (ML) are not new to medicine or medical imaging. There is a substantial literature dealing with the application of machine learning techniques in medical imaging beginning in the 1980s. Known as Computer Aided Detection/Diagnosis (CAD), research in this field led to the development of key deep learning algorithms in the 1990s and the application of these technologies in commercial products. In spite of years of research and development, the number of clinically successful CAD products with FDA approval has been rather limited, until recently.

Identifying quantitative imaging phenotypes across scales through the use of robust image analysis and deep learning methods is an evolving approach to improving our understanding of cancer biology. Prediction of response to treatment and development of optimized therapies require integration of Radiology and Pathology imaging information, molecular and genomics data, and clinical information. High quality imaging feature sets from Radiology, (radiomics) and Pathology images, (pathomics) play a critical role in the development of robust and predictive models for characterization of cancer onset and progression.

As basic and translational investigators dig deeper into the complexity of underlying disease processes, larger sample populations are required for research studies and increasing quantities of data are being generated. In addition, well-curated research and trial data are being pooled in public data repositories to support large-scale analyses. The ability to find new patterns in data, to identify new image features of significance to cancer diagnosis, and precise phenotypes to inform precise therapies will exert an increasing influence on cancer research.

Jerry A. Darsey, Ph.D.
Director, Center for Molecular Design and Development &
Department of Chemistry
University of Arkansas at Little Rock
Little Rock, AR USA

Dr. Jerry A. Darsey is a professor in the Department of Chemistry & Director for the Center of Molecular Design and Development at the University of Arkansas at Little Rock and is an adjunct professor in the Department of Biopharmaceutical Sciences, the University of Arkansas for Medical Sciences in Little Rock, Arkansas. He also consults for the FDA at the National Center for Toxicological Research. His B.S. degree is in Physics and his PhD is in Physical Chemistry both from Louisiana State University. Dr. Darsey is the author or co-author of approximately 150 manuscripts in the fields of molecular modeling, neural network simulations, conformational modeling of polymers and proteins, Monte-Carlo modeling; and a book in final revision on applications of neural networks to molecular and biomolecular systems. He is also involved in drug design He has made over 300 presentations at regional, national and international meetings. He also is inventor or co-inventor on 7 patents on file with in the Unites States Patent and Trademark office with three patent pending. Dr. Darsey's research interests are primarily in computer modeling techniques of atomic and molecular systems for the elucidation of their chemical and physical properties. He is also working in the area of nanotechnology to model atomic and molecular nanoclusters for applications to enhanced hydrogen storage. As either a PI or Co-PI, he has received approximately $2 million in funding from NASA, Department of Energy, Arkansas Science and Technology Authority and American Chemical Society-Petroleum Research Fund.

AI for drug design and drug discovery

The field of drug design is a rapidly growing area in which many successes have occurred in recent years. The tremendous growth of genomics, proteomics, and other structural information on proteins, has provided thousands of new targets and opportunities for future drug discovery. The most fundamental goal in drug design is to predict whether a given molecule will bind to an active target region of a protein and if so how strongly. Molecular mechanics or molecular dynamics is most often used to estimate the strength of the intermolecular interaction between the small molecule and its biological target (usually a protein). These methods are also used to predict the conformation of the small molecule and to model conformational changes in the target that may occur when the small molecule binds to it. Semi-empirical, *ab initio* quantum chemistry methods, or density functional theory (DFT) are often used to provide optimized parameters for the molecular mechanics calculations and also provide an estimate of the electronic properties (electrostatic potential, polarizability, etc.) of the drug candidate that will impact binding affinity. This talk provides two new types of approaches to drug discovery; an approach based on a combination of artificial intelligence [AI] coupled with high-level quantum mechanical [QM] methods; a second method using functional correlations. These processes of drug design includes the choice of a target (usually a protein), the evaluation of the structure of that target protein, data on the binding affinities to active cites of the target protein, which are all pivotal in considering this method for drug discovery.

Justin Zhan, Ph.D.
Professor of Data Science
Department of Computer Science and Computer Engineering
University of Arkansas
Arkansas, USA

Dr. Justin Zhan is an ARA Scholar and Professor of Data Science at the Department of Computer Science and Computer Engineering, College of Engineering, University of Arkansas. He has been the director of the Big Data Hub and a professor in the Department of Computer Science, College of Engineering, University of Nevada, Las Vegas. His research interests include data science, biomedical informatics, artificial intelligence, blockchain technologies, information assurance, and social computing. He was a steering chair of the IEEE International Conference on Social Computing, IEEE International Conference on Privacy, Security, Risk and Trust, and IEEE International Conference on BioMedical Computing. He has been an editor-in-chief of the International Journal of Privacy, Security and Integrity, and International Journal of Social Computing and Cyber-Physical Systems. He has served as a conference general chair, a program chair, a publicity chair, a workshop chair, and a program committee member for 150 international conferences; he has also served as an editor-in-chief, editor, associate editor, guest editor, editorial advisory board member, and editorial board member for 30 journals. He has published 250 articles in peer-reviewed journals and conferences and delivered more than 30 keynote speeches and invited talks. He has been involved in more than 50 projects as a principal investigator (PI) or a Co-PI, which were funded by the National Science Foundation, Department of Defense, National Institute of Health, etc.

Bridging Data and Learning Deep

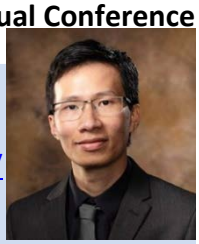
Data has become the central driving force to new discoveries in science, informed governance, insight into society, and economic growth in the 21st century. Abundant data is a direct result of innovations including the Internet, faster computer processors, cheap storage, the proliferation of sensors, etc, and has the potential to increase business productivity and enable scientific discovery. However, while data is abundant and everywhere, people do not have a fundamental understanding of data. Traditional approaches to decision making under uncertainty are not adequate to deal with massive amounts of data, especially when such data is dynamically changing or becomes available over time. These challenges require novel techniques in data analytics, data-driven optimization, systems modeling and data mining. In this talk, a number of recent data science projects will be presented to address various data analytics, mining, modeling and optimization challenges.

Keith Bush, Ph.D.
Associate Professor, Brain Imaging Research Center (BIRC)
Department of Psychiatry
University of Arkansas for Medical Sciences (UAMS)
Arkansas, USA

Dr. Keith Bush received his bachelor degree in Chemical Engineering from the University of Pennsylvania and his doctoral degree in Computer Science from Colorado State University. He also trained in neuroscience applications of machine learning as a postdoctoral fellow under Dr. Joelle Pineau at McGill University (Montreal, Canada) in conjunction with neurophysiologists at the Montreal Neurological Institute. Dr. Bush began his career as a faculty member in the Department of Computer Science at the University of Arkansas at Little Rock (UA LITTLE ROCK) before transitioning to the Brain Imaging Research Center at the University of Arkansas for Medical Sciences (UAMS) in 2016. His research explores machine learning and control theoretic approaches to human neuroimaging, using both real-time fMRI and fMRI-based neurofeedback to understand and exploit emotion processing and regulation. By understanding how the human brain decodes and integrates neurofeedback signals into its processing, he hopes to optimize neuroimaging studies and develop new control theoretic treatments for emotional disorders.

AI for the Brain

The core principles of artificial intelligence (AI) – search, knowledge representation, and learning – combine to form a general framework for experience-driven sequential decision-making, termed Reinforcement Learning (RL), which serves as the foundation of AI advancements in both challenging real-world decision domains (e.g., autonomous vehicles) as well as and algorithms capable of achieving human expert performance from self-play across a broad suite of popular tabletop and computer games. Over the past twenty years, psychologists and cognitive scientists have adopted RL as a theoretical foundation for characterizing decision-making in human behavioral experiments. Today, non-invasive neuroimaging methods allow us to observe the functional neurocircuitry supporting human decision-making and critically test the RL framework. This presentation will provide introductions to both RL and human neuroimaging, survey current progress integrating these fields, and delineate existing challenges that offer fruitful opportunities for novel AI research and technology.



Khoa Luu, Ph.D.
Assistant Professor
Director, Computer Vision and Image Understanding (CVIU) Lab, https://cviu.uark.edu/
Computer Science & Computer Engineering Department
University of Arkansas, Fayetteville, AR

Dr. Khoa Luu is an Assistant Professor and the Director of Computer Vision and Image Understanding Lab, Computer Science and Computer Engineering Department, University of Arkansas in Fayetteville. He was the Research Project Director in Cylab Biometrics Center at Carnegie Mellon University (CMU) where he led the Biometrics and Computer Vision projects and collaborates with research scholars to build stand-alone A.I systems. Before, he was also a research scientist in CMU. He received his Ph.D degree in Computer Science at Concordia University, Montreal, Canada. His Ph.D. thesis was nominated for the Governor General Gold Medal in Canada. He was also a graduate student for the second Ph.D degree in Electrical and Computer Engineering (ECE) at CMU to strengthen his knowledge in signal processing but not completed yet. He has coauthored 80+ papers, some are in top-tier conferences, e.g. CVPR, ICCV, and high-impact journals, e.g. IJCV, TPAMI, TIP, etc. Most papers are about Computer Vision, low-power Deep Learning, Biometrics, 3D Face Modeling, million-scale Face Recognition and detection, automatic scene understanding, temporal deep learning and reinforcement learning. He registered six patents and inventions and received 10+ awards and scholarships. He received two best paper awards in IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems in 2011 and 2012. He was a vice chair of Montreal Chapter IEEE SMCS in Canada from September 2009 to March 2011. Dr. Luu has professional experience in both research and real-world deployed projects in deep learning in IOT devices, million-scale biometrics, long-range vision, real-time face modeling, video understanding and low-power deep learning technologies.

Advanced Approaches in Deep Learning and Computer Vision to 3D Face Emotion Recognition in Food Eating

Human face recognition and 3D face modeling can be classified as one of the key research topics in Biometrics. It is a branch of machine learning and computer vision that works on understanding visual data related to human face characteristics for authentication and identification. Over the last few years there have been many important breakthroughs in the theory and practice of human face recognition and 3D face modeling. Among these practices, deep learning has emerged as field leading the effort towards building the state-of-the-art systems. The practice of deep learning however still requires extensive experience and expertise to produce satisfactory results. Our experience in face processing and building state-of-the-art face detectors, facial landmarking, face recognition and 3D face modeling systems allows us to have a deep understanding of the field of deep learning. We have demonstrated ideas that have resulted in systems that have achieved highly competitive performance in the challenging field of 3D face modeling and real-world biometrics products.

In this talk, we will review the development of such completed architectures in the related applications. We showcase state-of-the-art systems we have developed in applications of face preprocessing, face detection, face landmarking, face recognition, 3D face modeling and face video-analytics. Specifically, we showcase how to deploy these 3D face modeling systems from a single 2D camera for the application of emotion recognition in Food Eating. Finally, we will address the problem at hand – facial emotion recognition and 3D Face Reconstruction in real-time and in low-cost devices.

Kyle P. Quinn, Ph.D.
Assistant Professor
Department of Biomedical Engineering
University of Arkansas - Fayetteville
Arkansas, USA

Dr. Kyle P. Quinn is an assistant professor of biomedical engineering at the University of Arkansas. Dr. Quinn received his B.S. degree in Biomedical Engineering from the University of Wisconsin-Madison, and earned his Ph.D. in Bioengineering from the University of Pennsylvania. He joined the Department of Biomedical Engineering at Tufts University as a postdoctoral fellow, where he was awarded a Ruth L. Kirschstein National Research Service Award and an NIH Pathway to Independence Award (K99). Since joining the faculty in the Department of Biomedical Engineering at the University of Arkansas in 2015, he has received funding through the NIH R00, R21, and R01 mechanisms, DoD STTR and CDMRP projects, and the NSF CAREER award. His overall research interests are in developing and utilizing non-invasive quantitative optical methods to characterize the spatiotemporal patterns of disease progression and tissue repair processes. His lab utilizes label-free multiphoton microscopy to obtain high-resolution 3D images of cell metabolism and extracellular matrix organization using endogenous two-photon excited fluorescence, second harmonic generation, fluorescence lifetime imaging, and coherent anti-Stokes Raman spectroscopy. With these imaging techniques, his group has developed and implemented a variety of advanced image analysis tools to characterize the organization of the extracellular matrix, inter-cellular arrangement and morphology, as well as subcellular organelle distributions within cells. Using these label-free imaging approaches, his research group is currently focusing on establishing non-invasive optical biomarkers of wound healing and cellular aging. He has published 61 peer-reviewed journal articles and book chapters.

Deep learning for automated segmentation of skin wound features

Histopathological staining and imaging of skin tissue sections using stains such as hematoxylin and eosin (H&E) has served as a fundamental gold-standard technique to evaluate disease states and skin wound healing. However, interpretation of histologically processed tissue by even trained pathologists and researchers is time-consuming and subject to inter-observer variability. Neural networks have emerged as powerful image classification tools sensitive to subtle changes in tissue features. Our group has been developing convolutional neural networks capable of classifying and segmenting wound features to assist clinical and laboratory evaluations of skin wounds. Specifically, we have created a deep learning neural network capable of classifying key wound regions in both unstained and H&E stained tissue section images to assist with measuring key outcomes from wound histology. We demonstrate an ability to automatically identify and segment six different skin wound regions including: the epidermis, dermis, granulation, stratum corneum, hair follicles, and connective tissue in less than 10 seconds from entire tissue sections with true positive rates greater than 85%. Using the network outcomes, we were further able to collect automated measurements of dermal thickness, wound size, and percent re-epithelization that were within 5% error. Using the optimized H&E network's feature detectors, a new network was also initialized via transfer learning and trained with autofluorescence images of unstained wound tissue sections. This autofluorescence segmentation network demonstrated similar agreement between predicted pixel classification and ground truth measurements as the H&E-based network, and demonstrated an ability to detect wound features from in vivo multiphoton microscopy images of the wound edge using a mouse model of wound healing. The ability to automatically identify and measure wound features from H&E sections and non-invasive autofluorescence measurements creates new opportunities for rapid, quantitative tissue analysis to assist in wound care.

Leihong Wu, Ph.D.
Visiting Scientist, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Leihong Wu has a solid background and significant experience in diverse areas of data science including bioinformatics, chemoinformatics, transcriptomics, network pharmacology, and next generation sequencing. In the past decade, his research has made substantial contributions in the field of machine learning and artificial intelligence. He has been instrumental in developing novel modeling algorithms and systemic approaches specializing on various types of biological and pharmaceutical data analysis. Dr. Wu has played a major role in multiple projects critical to the FDA's missions. More specifically, he is a technical lead (1) to develop an AI algorithm for food safety concerning filth detection with imaging data which is in collaboration with ORA, (2) to conduct text mining of the FDA drug labelling documents for drug safety by developing and mining the FDALabel database, (3) to develop predictive models for drug-induced liver injury (DILI) with both in vitro assay data and chemical structure data, and (4) to conduct network analysis of genomics data to gain understanding of underlying mechanisms of cancer. The successful AI approaches developed by Dr. Wu have expanded AI capabilities to other types of biological data related to FDA regulated products, including drug labeling documents, biomarkers, genomics and sequencing data. Dr. Wu has published more than 40 papers and book chapters.

FIDE: an AI-based Recognition System to advance ORA regulatory operations at Importing Mail Facilities

Every year hundreds of millions of mails and packages entered the U.S., and about 9% of these packages are estimated to contain some drug products that need to be examined. FDA investigators at nine U.S. International Mail Facilities (IMFs) monitor suspected packages by conducting comprehensive inspection and detection, which is primarily a time-consuming manual process. Unfortunately, due to the sheer volume of packages, FDA was only able to inspect less than 0.2% of the packages presumed to contain drug products with the alarming finding that over 87% of them contain illegal, unapproved, counterfeit and potentially dangerous products. This critical public health challenge urgently needs an innovative and efficient solution.

Artificial Intelligence (AI), due to the recent success of deep learning, has powered many aspects of the modern society from self-driving cars to face recognition. Developing an AI-based product recognition tool will hold the potential to reduce inspection time and efforts of FDA investigators, and to prevent the entry of adulterated, misbranded, or other violative drug products. With that said, we developed an AI-based recognition tool, called FIDE, to assist FDA investigators in their making compliance actions. The tool is designed to rapidly recognize the product and find the most relevant product from the historical importing records, to accelerate current FDA reviewing process at IMFs. FIDE will be a standalone tool but accessible to the current screening system through predefined data exchange interfaces. Such interfaces will pave the way for future integration with the current screening system.

Mahmoud Moradi, Ph.D.
Assistant Professor
Department of Chemistry and Biochemistry, and
Cell and Molecular Biology Program
University of Arkansas, Fayetteville, AR, USA

Dr. Mahmoud Moradi is an assistant professor of chemistry and biochemistry at the University of Arkansas, Fayetteville. He received a PhD in physics in 2011 from North Carolina State University and was a postdoctoral fellow at the Beckman Institute, University of Illinois at Urbana-Champaign from 2011 to 2015. He joined the faculty of chemistry and biochemistry at the University of Arkansas in 2015. Dr. Moradi is an editorial board member for Scientific Reports. Research in Moradi Lab is centered around two inter-related questions: (i) how do proteins function by changing their conformation and undergoing concerted motions? and (ii) how can we simulate these functionally important conformational changes at an atomic level? Moradi Lab employs Molecular Dynamics (MD) based enhanced sampling techniques to tackle both problems. Answering these questions would shed light on the structure-function relationships in proteins, and could improve our understanding of disease at a molecular level. Moradi Lab attempts to narrow the gap between the state-of-the-art computational methodologies and biologically/biomedically relevant applications. Moradi Lab is particularly interested in the study of large-scale conformational changes of proteins such as those involved in membrane transport and signal transduction.

Deep Learning Inferential Methods for Protein Structure Predictions

Despite many advances in pharmaceutical technology, the drug development process cost continues its exponential increase. One of the key reasons for the relatively poor performance of current drug design, discovery, and development approaches is the lack of an accurate molecular level understanding of disease and the way drugs interact with their protein targets. Physics-based computational models can lower the cost of the common trial-and-error approaches in drug discovery process. Structure based drug design has already proven quite successful; however, many limitations prevent this approach from reaching its full potential. The fundamental issue with conventional molecular docking techniques is that the physical model used to describe the protein structure and its interaction with its environment is often too simplistic. The protein structure, for instance, is often considered to be static or only locally flexible. Proteins, however, are known to be highly dynamic structurally and many important protein functions are associated with large conformational changes. The all-atom molecular dynamics (MD) simulation technique provides an alternative approach to conventional structure-based drug design by considering the protein structural dynamics. Unfortunately, many biomolecular processes such as those involved in drug-protein interactions are associated with timescales inaccessible to brute-force all-atom MD, making this technique prohibitively expensive computationally. Various enhanced sampling techniques have been developed over the last three decades to address this timescale gap between all-atom MD (typically nanoseconds) and functionally relevant events (typically milliseconds). These methods increase the chance of sampling various functionally relevant protein conformations that are not predictable using conventional experimental structure determination techniques. Despite much progress in the development of enhanced sampling techniques, these methods are rarely applicable to systems other than toy models and small proteins. This is partly due to the fact that the available enhanced sampling techniques are either too simplistic in their assumptions or too computationally costly to be applicable to realistic proteins. Here, we have employed deep learning inferential methods to improve the performance of enhanced sampling MD techniques to efficiently sample the protein structural space in an unbiased way and predict the structure of various functionally relevant conformations. In this approach, the MD technique is iteratively coupled with a neural network to adaptively enhance the sampling of the conformational space of protein. The methodology has been successfully applied to predict several functionally distinct structures of a P-glycoprotein homolog.

Mariofanna Milanova, Ph.D.
Professor of Computer Science
University of Arkansas at Little Rock
Arkansas, USA

Dr. Mariofanna Milanova is a Professor of the Computer Science Department at the University of Arkansas at Little Rock, USA since 2001. She received a M.Sc. in Expert Systems and Artificial Intelligence and Ph.D. in Telecommunications from the Technical University, Sofia, Bulgaria. Dr. Milanova conducted post-doctoral research in visual perception at the University of Paderborn, Germany. Dr. Milanova has extensive academic experience at various academic and research organizations worldwide.

Dr. Milanova is IEEE Senior Member, Fulbright U.S. Scholar, and NVIDIA Deep Learning Institute University Ambassador. Dr. Milanova's work is supported by NSF, NIH, DARPA, DoD, Homeland Security, NATO, Nokia Bell Lab, NJ, USA and NOKIA, Finland. Prof. Milanova's research areas of interest are: Artificial Intelligence, Machine Learning, Image/Video Processing, Brian-like Computing and Computer Graphics. Dr. Milanova serves as a book editor for two books and as associate editor for various international journals.

She has published more than 120 publications, over 53 journal papers, 35 book chapters, and numerous conference papers and has 2 patents.

Artificial Intelligence for Biomedical Image Analysis

This presentation will briefly cover the main concepts of Artificial Intelligence for biomedical imaging. The first part of the talk will include a discussion on the medical image segmentation challenges and evaluation metrics. Segmentation methods such as the Classic approach, Deep Learning-based approaches and Classifier Ensemble will also be covered. Recent advances in the field of image analysis using X-ray , computer tomography (CT), magnetic resonance imaging (MRI) and Positron emission tomography (PET) images will also be highlighted. The second part of the presentation will cover examples and applications related to the problems of classification and decision making.

Shraddha Thakkar, Ph.D.
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research
US FDA
Jefferson, AR

Dr. Sraddha Thakkar works at FDA's National Center for Toxicological Research. Her research interests are in applying bioinformatics and cheminformatics for study of toxicity and drug development with specific interest in drug-induced liver injury. She has received multiple research and leadership awards regionally and nationally and with FDA. That includes Genentech Innovation in Biotechnology Award from American Association of Pharmaceutical Scientist (AAPS), Margret C. Etter Student lecturer award from American Crystallography Association, and Outstanding Service award from FDA. Dr. Thakkar has adjunct appointments at both University of Arkansas for Medical Sciences and University Arkansas at Little Rock (Assistant Professor). Furthermore, Dr. Thakkar was elected as Board member of the Mid-South Computational Biology and Bioinformatics Society (MCBIOS) in 2014 and served as President for the Society from 2016- 2017. She is also the Chair of Pharmacogenomics Group (2018-19) and serves on the Awards Committee at American Association of Pharmaceutical Scientist (AAPS). For her contribution to AAPS, she is elected to the AAPS board of directors for 2020-22.

AI/ML for predicting drug-induced liver injury

Drug-induced liver injury (DILI) is one of the leading causes of drug failure in clinical trials and drug withdrawal from the market. An alternative approach that can accurately predict DILI at the early stage has the potential to reduce the drug attrition, and it's in great need by regulators as well as developers. In the 21st century, toxicology experienced the paradigm shift, and now it relies heavily on high throughput technologies such as genomics and could be used as a source to predict DILI related signatures. With the advent of AI-based approaches, the deep learning-based algorithms have demonstrated the capability of learning import features from a large number of inputs with hidden signals. This presentation will not only highlight the variety of genomics data to predict liver injury , provide description of steps taken along the way for the development of effective model. The presentation will emphasize how 1) Big dataset 2) ratio of DILI vs nonDILI and 3) Choice of model can contribute to the effective prediction for the complicated endpoint.

Weida Tong, Ph.D.
Director, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served science advisory board for several multi-institutional projects in Europe and USA. He also holds adjunct appointment at several universities. In addition, he is the founder and board chairperson of newly established international MAQC Society. His division at FDA is to develop bioinformatic methodologies and standards to support FDA research and regulation and to advance regulatory science and personalized medicine. The most visible projects from his group are (1) conducting the Microarray and Sequencing Quality Control (MAQC/SEQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (4) *in silico* drug repositioning for the enhanced treatment of rare diseases; and (4) development of various tools such as ArrayTrack™ suite to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with specific interest in estrogen, androgen, and endocrine disruptor. Dr. Tong has published more than 250 papers and book chapters.

FDA efforts on ML for genomic biomarker

Artificial intelligence (AI) has made a significant mark in the past decade and demonstrated its utility in the broad area of predictive medicine. The rapid advancement in AI also poses several challenges and opportunities for regulatory agencies such as FDA: what is the regulatory structure to approve ever evolving nature of AI-based devices and application and how we implement the AI-based framework to improve regulatory process. Reproducibility is a key element to realize the potential of AI in biomedical application and regulatory implementation. The FDA has led a large consortium, called MicroArray and Sequencing Quality Control (MAQC/SEQC), which has interrogated various machine learning approaches in developing gene-expression based biomarkers for both clinical and preclinical applications. Specifically, the questions relating to reproducibility have been extensively investigated such as whether a reproducible result is (1) dataset-dependent, (2) AI methodology dependent, (3) experience-dependent, and (4) technology-dependent. The presentation will conclude with key lessons learned about what is needed to close the gap in reproducibility of AI in predictive medicine.

Xiaowei Xu, Ph.D.
Professor
Department of Information Science
University of Arkansas, Little Rock
Arkansas, USA

Dr. Xiaowei Xu is a professor in the Department of Information Science at the *University of Arkansas, Little Rock* (*UA LITTLE ROCK*), received his Ph.D. degree in Computer Science at the *University of Munich* in 1998. Before his appointment in *UA LITTLE ROCK*, he was a senior research scientist in Siemens, Munich, Germany. His research spans data mining, machine learning, bioinformatics, database management systems and high-performance computing. He has published over 100 papers in peer-reviewed journals and conference proceedings. With over 23,891 citations, he is one of the most cited researchers in his field according to Google Scholar. His groundbreaking work on the density-based clustering algorithm DBSCAN has been widely used in textbooks and software implementations; and has received over 16,453 citations based on *Google scholar* to date. He is a recipient of the prestigious ACM SIGKDD Test of Time award for his contribution to the density-based clustering algorithm DBSCAN. Recently Dr. Xu has been recognized as a Most Influential Scholar in the field of Data Mining for his "outstanding and vibrant contributions to the field of Knowledge Discovery and Data Mining" by AMiner (https://aminer.org/mostinfluentialscholar/datamining), a free online service for academic social network analysis and mining.

Artificial Intelligence (AI) for Text Mining

There is enormous amount of texts that cannot be processed by us. How can we teach a computer to read and decipher the information in the text is what text mining about. This talk will give an overview to the artificial intelligence techniques for text mining. The focus will be on the recent breakthroughs in deep learning, more specifically neural network models and algorithms for text mining, as complex semantic meaning in natural language is hard to be mined using conventional approach. Deep neural language models learning a hierarchical representation proved to be a powerful tool for natural language processing, text mining and information retrieval. This talk will cover deep neural language models for word embedding and learning representations of text for information retrieval and text mining. The topic includes an introduction of language models using self-supervised learning. It is followed by a presentation of recent multi-resolution models that represent documents at multiple resolutions in term of abstract levels. More specifically, we form a mixture of weighted representations across the whole hierarchy of a given deep language model, so that all resolutions of the hierarchical representation are preserved for the downstream task such as question answering. Finally, the application for information retrieval, question answering and other text mining tasks is presented.

Xiuzhen Huang, Ph.D.
Professor of Computer Science, Arkansas State University

Dr. Xiuzhen Huang is Professor of Computer Science at Arkansas State University. Dr. Huang conceived and defined the concept of No-Boundary Thinking (NBT). She is Founding Director of the Arkansas Artificial Intelligence AI-Campus and Founding Director of the Joint Translational Research Lab of Arkansas State University and St. Bernards Medical Center Internal Medicine Residency Program. Her research interests include bioinformatics and biomedical informatics; artificial intelligence, machine learning, deep learning; graph theory and algorithms, parameterized computation and complexity, theory of computation.

Biomedical Imaging Segmentation

The Arkansas Artificial Intelligence AI-Campus is a unique state-wide training program. The program exists outside of our traditional academic environment. This interactive training program allows participants to work closely with experts in the areas of Machine Learning, Artificial Intelligence, and Deep Learning. The participants learn frontier knowledge and technology, and work on hands-on projects. Specifically, one project in medical imaging will be presented.

Zhichao Liu, Ph.D.
Principal Investigator, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Zhichao Liu's background spans the fields of chemistry, biology, and computer science. In the past eight years, he took part in several cutting-edge projects in both industry and academia. Specifically, he developed several high-efficacy, quality-control strategies of processing analytical techniques and applied them to the tobacco industry. Furthermore, Dr. Liu focused on developing the standard pipeline to balance the efficacy and safety in drug repositioning and drug-safety areas. The research aims to provide the standard in silico pipeline for drug repositioning and early drug-safety detection by retrieving, integrating, and organizing the information from chemical, biological, and clinical spaces. This helps the industry seek the optimal route to accelerate the drug-development efficacy from an advanced regulatory-sciences perspective. Recently, Dr. Liu's interest has shifted to apply Artificial Intelligence (AI) to solve regulatory-related questions and promote advanced regulatory sciences. His Research Interests include (1) Applying AI and deep learning for promoting precision medicine; (2) Developing innovative approaches for in silico drug repositioning for rare diseases; (3) Developing flexible and integrative risk-prediction systems for drug-safety evaluation; (4) Developing strategies and frameworks to facilitate AI-based text-mining performance for diverse document types and infrastructures; (5) Designing and developing databases and visualization systems that allow interactive exploration of complex interior relationships embedded in biological-data profiles

Drug Repurposing in Ear of Artificial Intelligence (AI)

Advances in artificial intelligence (AI) are gaining ground for their applicability in drug development. In silico drug repurposing approaches have been showing great promise to provide safer, quicker, and cheaper solution for drug development. However, how to utilize implement advanced AI approaches into computational drug repositioning framework for augmenting successful rate of disease therapies is still an open question. In this presentation, we will elaborate on and exemplify a few deep learning-based drug repurposing strategies including (1) Deciphering immune and mitochondria continuum with augmented representation learning; (2) An autoencoder-based patient stratification for advancing precision medicine-based drug repositioning; (3) A deep learning-based sentiment analysis for drug indication extraction. The presentation will be ended with key lessons learned from real-word applications of deep learnings.

POSTER ABSTRACTS

FDALabel: Amazon Cloud Database for Drug Labeling information

Hong Fang[1], Steven Turner[1], Joe Meehan[1], Stephen Harris[1], Junshuang Yang[1], Guangxu Zhou[1], Taylor Ingle[1], Zhichao Liu[1], Joshua Xu[1], Leihong Wu[1], Darshan Mehta[1] Catherine Li[2], Crystal Allard[2], Shraddha Thakkar[1], Minjun Chen[1], Lilliam Rosario[2], and Weida Tong[1]

[1] NCTR, U.S. FDA, Jefferson, AR 72079, USA; [2]OTS/CDER, U.S. FDA, Silver Spring, Maryland 20993, USA

Abstract:
FDA's Structured Product Labeling (SPL) archive, which stores drug labeling documents submitted by manufacturers, contains labeling information including product indications, dosing recommendations, contraindications, drug interactions, warnings and precautions, adverse reactions, and information for patients to help ensure the safe and effective use of the product. The continual increase in the number of labeling documents and large amount of data contained in these documents necessitates an advanced bioinformatics tool with powerful drug data management and search capabilities. We have developed the FDALabel database as a web-based application, containing over 100,000 drug labeling documents from FDA's SPL archive. FDALabel allows the public to perform customizable (any combination of sections, document types, market categories, and other information), full-text searches of product labeling on a relational oracle database. A new version of FDALabel (*v2.3*), available at Amazon Cloud: https://nctr-crs.fda.gov/fdalabel/, was developed to search human prescription drug and biological product labeling and human over-the-counter (OTC) drug labeling. To demonstrate the FDALabel database's utility, we have selected study cases including a pharmacogenomics study for Precision Medicine and an ADR (Adverse Drug Reaction) study that applied Medical Dictionary for Regulatory Activities (MedDRA) standard terminologies. We identified 224 drugs with 289 drug-biomarker pairs across different therapeutic areas such as oncology (103), psychiatry (33), and infectious diseases (32). We also found severe ADRs were prevalent in MedDRA System Organ Classes such as Nervous system disorders, Psychiatric disorders, and Cardiac disorders. The FDALabel database search tool offers the public, researchers and regulatory reviewers an efficient and user-friendly means of accessing and searching the large amount of information contained in drug labeling. An Amazon Cloud version of FDALabel (*v 2.3*) supports and promote translational medicine and regulatory science by employing advanced computer technologies to deliver end-users a reliable, effective, and efficient search tool.

Statistical Methods to Analyze Uncertain Liquid Biopsy qPCR Data

Wei Vivian Zhuang[1], Luísa Camacho[2], Camila S Silva[2], and Huixiao Hong[1]

1-Division of Bioinformatics and Biostatistics, NCTR, U.S. FDA, Jefferson, AR, U.S.A. 72079. 2-Division of Biochemical Toxicology, NCTR, U.S. FDA, Jefferson, AR, U.S.A. 72079

Abstract:

Introduction: Liquid biopsy tests offer great promise of providing a noninvasive or minimally invasive means to rapidly and cost-effectively detect toxicity or disease when it is most reducible or curable. qPCR (Quantitative Real-time Polymerase Chain Reaction) is often used to measure the levels of RNAs in liquid biopsy. Besides complete data, investigators have observed inevitably incomplete and uncertain qPCR data due to low or completely absent levels of transcripts. **Methods:** Investigators often intervene with incomplete qPCR data in an intrinsic way, such as setting incomplete observations equal to the maximum number of qPCR cycles (MC), applying the complete-observation method by deleting the incomplete observations from analysis (CO), or choosing not to analyze targets with incomplete observations (CNA). The three methods tend to cause biased inference, decrease research reproducibility across replicate experiments, and thus be ineffective and inefficient in helping investigators learn from the qPCR data with incompleteness. To overcome the shortcomings, we propose a nonparametric cycle-to-threshold method (CTOT). CTOT incorporates qPCR-specific features and the time-to-event statistical methodology, and is built around extracting information from all subjects, censored or not. We used simulations and a dataset of rat serum microRNAs to compare the performance of CTOT, MC, CO, and CNA. **Results:** Our simulations show that CTOT may improve the power of detecting differentially expressed biomarkers without generating excess type I errors. In the application, CTOT detected potentially differential expression that would be otherwise overlooked. **Conclusion/Implications**: CTOT helps leverage qPCR technology, increases the power to detect novel biomarkers, and improves research reproducibility.

Efficient Accession Management and Characterization of the USDA-ARS Rice Germplasm Collection through Phenotyping and Genotyping

Huggins, T.D[1]., McClung, A.M.[1], Edwards, J.D.[1], Jia, M.H.[1], Bockelman, H.E.[2], Ali, M.L.[3], and Eizenga, G.C.[1]

[1]USDA-ARS, Dale Bumpers National Rice Research Center, 2890 Hwy 130 E, Stuttgart, AR 72160
[2]USDA-ARS Small Grains and Potato Germplasm Research, Aberdeen, ID 83210
[3]Plant Sciences Dept., Univ. of Missouri-Columbia Delta Center, Portageville, MO 63873

Abstract:
Genebanks are an important source of genetic diversity for food crops world-wide and offer valuable information that can be used by plant breeders to improve agricultural productivity and nutritional quality. Collections of major crop species maintained by genebanks often have over 10,000 accessions, thus there is the potential to generate "big data" that captures a wide range of genotypic and phenotypic diversity across all accessions. There will be abundant opportunities for artificial intelligence (AI) to enable high-throughput phenotyping and mining of the collections for novel genetic diversity. The USDA-ARS National Small Grains Collection (NSGC) currently maintains over 19,000 accessions of cultivated Asian rice (*Oryza sativa* L.), 193 African cultivated rice (*O. glaberrima* Steud.) and 54 *Oryza* wild species. In this study, subsets of the rice collection were genotyped and phenotyped to better characterize the accessions and to evaluate redundancy. We phenotyped a random set of 1,993 Rice-NSGC accessions, for two phenological traits, plant height, 13 morphological traits, six grain quality traits, four production traits, resistance to three diseases, and two stress-related traits and genotyped these accessions with 11 fingerprint markers (FPM), one subspecies marker, and 14 trait specific markers (TSM). The TSM were used to validate phenotypic data for fragrance, pericarp color, blast disease resistance, leaf and hull pubescence, apparent amylose content, starch pasting properties and gelatinization temperature, and plant height. The markers classified accessions by species, subspecies, subpopulation, bran color, aroma, and in some cases by variety. Even with the limited number of markers, the genotyping was adequate for differentiating varieties and predicting phenotypes. Some traits such as plant height, yield, disease detection, and stress detection have been adapted to AI across species. Continued efforts to genotype more accessions with additional trait specific markers will help breeders select valuable germplasm from the NSGC for use in variety development.

Deep sequencing for dissecting the roles of the enzymatic and scaffolding functions of EXO1 *in vivo*

Shanzhi Wang[1], Kyeryoung Lee[2], Richard Chahwan[3], Yongwei Zhang[2], Catherine Tang[2], Sergio Roa[2], Rikke B. Morrish[3], Elisa Pesenti[2], Isabelle Lykens[1], Paula E. Cohen[4], Thomas MacCarthy[5], Winfried Edelmann[2], Matthew D. Scharff[2]

[1]Chemistry Department, University of Arkansas at Little Rock, AR 72204; [2]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461; [3]Department of Biosciences, University of Exeter, Exeter EX2 4QD, United Kingdom; [4]Department of Biomedical Sciences, Cornell University, Ithaca, NY, 14853; [5]Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794

Abstract:
Activation induced cytidine deaminase (AID) generates U:G mismatches in immunoglobulin (Ig) variable-regions (V) genes during somatic hypermutation (SHM) and switch-regions (SR) during class switch recombination (CSR). Exonuclease 1 (EXO1) is essential for both processes. It has been reported that mice expressing EXO1 with the cancer associated E109K mutation (*Exo1*^{EK}) did not have the defects in SHM and CSR seen in *Exo1*^{null} mice, suggesting that the enzymatic activity of EXO1 was not required for antibody diversification. However, subsequent biochemical work revealed that the untagged full length EXO1-E109K protein retained WT exonuclease activity while and EXO1-D173A (*Exo1*^{DA}) active site mutation had negligible enzymatic activity (1, 2). We therefore examined this question by generating new mice expressing the enzymatically nonfunctional EXO1-D173A mutation (Exo1^{DA}) and comparing them to EXO1-knockout (Exo1^{null}) and WT mice. Deep-sequencing of 186.2 heavy chain variable region in B cells from (4-hydroxy-3-nitrophenyl)acetyl (NP) immunized mice showed the mutation spectra of Exo1^{DA} mice was different from both the Exo1^{null} mice and the WT mice. In the meantime, mutation frequency of Exo1^{DA} mice was in-between Exo1^{null} mice and the WT mice. This supports our previous observation in Exo1^{EK} mice that scaffolding functions of EXO1 are critical for ncMMR at the Ig variable regions of B cells, while its exonuclease activity is not partially compensated for by other unknown mechanisms. In contrast, CSR efficiencies of Exo1^{DA} and Exo1^{null} mice were defective compared to WT mice, and residual switching events were slightly enriched in blunt and short SR microhomologies. This could be due to a requirement for EXO1 exonuclease activity in the resections required to create the double stranded DNA breaks. Taken together, the data suggest that EXO1 recruits other factor(s) to promote ncMMR repair in the Ig V regions during SHM, while its enzymatic role is clearly required during CSR.

1. Shao H *et al. Nucl. Acids Res.* 2014, 42:7104-7112.
2. Bregenhorn SJ *et al*. *Nucl. Acids Res.* 2014, 42: 7096-7103.

proteiNorm – A user-friendly tool for normalization and analysis of TMT and label-free protein quantification

Stefan Graw [1,2], Jillian Tang [1], Stephanie D Byrum[1,2,*]

1 Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR. 2 Arkansas Children's Research Institute, Little Rock, AR

Abstract:
The technological advances in mass spectrometry allow us to collect more comprehensive data with higher quality and increasing speed. With the rapidly increasing amount of data generated, the need for streamlining analyses becomes more apparent. Proteomic data is known to be often affected by systemic bias from unknown sources and failing to adequately normalize the data can lead to erroneous conclusions. To allow researchers to easily evaluate and compare different normalization via a user-friendly interface, we have developed “proteiNorm”.
The current implementation of proteiNorm accommodates some preliminary filter on peptide and sample level, followed by an evaluation of several popular normalization methods and identification of missing value pattern. The user then selects an adequate normalization method and one of several imputation methods used for the subsequent comparison of different differential abundance/expression methods and estimation of statistical power. The application of proteiNorm and interpretation of results is demonstrated on a Tandem Mass Tag mass spectrometry example, where the proteome of three different breast cancer cell lines was profiled.
With proteiNorm, we provide a user-friendly tool to identify an adequate normalization method and to select an appropriate method for a differential abundance/expression analysis.
KEYWORDS: Proteomics, Normalization, Differential analysis

* Corresponding author: Stephanie D Byrum

Genomic characterization of plasmids containing antimicrobial resistance and virulence genes of the susceptible recipient strain

Nesreen Aljahdali[A], Kennedi Weston[AB], Bijay K. Khajanchi[A], Yasser M. Sanad[AB], Jing Han[A], and Steven L. Foley[A]

[A] FDA National Center for Toxicological Research, Jefferson, AR 72079, [B] University of Arkansas at Pine Bluff, Pine Bluff, AR 71601

Abstract:

Introduction: Mobile genetic elements, such as plasmids, can potentially increase the ability of bacteria to infect and persist in host cells. Plasmids can contain antimicrobial resistance and virulence factors and can facilitate the horizontal transfer of these genes among bacteria via conjugation. *Salmonella enterica* is known as the source of multiple outbreaks linked to contaminated foods causing illness in animals and humans. Among 2600 different *Salmonella* serovars, *S. enterica* serovar Typhimurium is one of the predominant serotypes that can contribute to significant morbidity and mortality worldwide. The aim of this study was to determine the transferability of plasmids from the donor cell to the recipient cell. **Methods:** Twenty-four donor *S.* Typhimurium isolates originating from different food sources were sub-cultured on blood agar plates. *Salmonella* strains served as the potential donor cells and the *E. coli* J53 strain as the recipient cell for the conjugation experiments. Transconjugants were selected using ampicillin, streptomycin, or tetracycline along with sodium azide. Transconjugants were sub-cultured on blood agar plates and incubated at 35°C for 24 hours. The bacterial genomic DNA was extracted and sequenced using an Illumina MiSeq. Reads were assembled using CLC Genomics Workbench and the sequences annotated using the Pathosystems Resource Integration Center (PATRIC) software. Assembly results were further analyzed using PlasmidFinder and ResFinder to identify predicted plasmids and antimicrobial resistance genes. **Results:** PlasmidFinder analyses confirmed that seventeen (71%) transconjugants contained different plasmids including IncFIA (n=10, 42%), IncFIB (n= 10, 42%), IncFII (n=10, 42%), IncI1 (n=5, 21%), IncA/C2 (n=3, 12.5%), IncColpVC (n=2, 8%), and IncX4 (n=1, 4%). ResFinder results revealed multiple antimicrobial resistance genes included *sul2* (n=10, 42%) of total transconjugants, followed by *tet(B)* (n=8, 33%), bla_{CMY-2} (n=6, 25%), *tet(A)* (n=4, 17%) *sul1* (n=3, 12.5%), bla_{TEM-1B} (n=2, 8%), and *floR* (n=2, 8%). PATRIC results identified several virulence genes including *ssp* (17, 71%), *iut* (9, 38%), *iuc* (8, 33%), and *sit* (7, 29%). These genes likely play a role in the ability of bacteria to survive in food animal environment and cause infections in humans.

Keywords: *Salmonella enterica*, Conjugation, Whole genome sequencing (WGS).

Genomic Prediction and Bayesian Network Analysis of Multiple Root Architecture Traits in Rice

Santosh Sharma, David R. Gealy, Shannon R.M. Pinson, and Jeremy. D. Edwards

Abstract:

Root system architecture (RSA) is a crucial factor in resource acquisition and plant productivity. High-throughput root image analysis systems can expedite the genetic study of RSA. Roots are difficult to phenotype in the field, thus new tools for predicting phenotype from genotype are particularly valuable for plant breeders aiming to improve RSA.

The objectives of this study were (1) to discover quantitative trait loci (QTL) for RSA traits in rice, (2) to evaluate the accuracy of genomic prediction methods for RSA traits, and (3) to model the network of interactions among the RSA traits, RSA QTL, and their interactions with above-ground agronomic traits.

A recombinant inbred line (RIL) population derived from parents with contrasting RSA (PI312777 x Katy) was genotyped with the 7K SNPs (C7AIR) and phenotyped for RSA on seedlings grown on agar plates and agronomic traits on plants in the field. QTL mapping and genomic prediction were used to investigate the RSA traits and their interactions with agronomic traits. The genomic prediction methods evaluated included rrBLUP, GBLUP and BRR. Multi-trait QTL and trait interactions were modeled using the "Bnlearn" Bayesian network (BN) analysis package in R.

Genomic prediction explained a higher proportion of genetic variance (56% for grain yield, 32% root surface area, 58% root hair surface area/total root area) for most quantitative traits compared to multi-QTL models (26% grain yield, 30% root surface area, 34% root hair surface area/total root area). Multi-trait BN analysis was found to improve genomic prediction (3-20%). The BN analysis confirmed QTL identified by single-trait QTL analysis and identified new QTL including one proximal to the pleiotropic gene *NAL1* on chromosome four affecting root length, leaf width and chlorophyll content. This study demonstrates the effectiveness of genomic prediction and multi-trait Bayesian network methods to model RSA traits and their above-ground trait relationships.

Developing imaging methods for automated species level identification of food contaminating beetles

Tanmay Bera[1], Hongjian Ding[2], Leihong Wu[1], Zhichao Liu[1], Amy Barnes[2], Monica Pava-Ripoll[3], Himansu Vyas[2], Weida Tong[1], Joshua Xu[1*]

1 Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079. 2 Food Chemistry Laboratory 1, Arkansas Laboratory, Office of Regulatory Affairs, US Food and Drug Administration, Jefferson, AR 72079. 3 Office for Food Safety, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD 20740.

Abstract:
Food contamination due to insect pests is a perpetual problem. If not managed properly, it can cause serious health risks and/or lose of resources. To remedy this, food samples are routinely analyzed for the presence or remains of food contaminating beetles. Presently, this is done manually by expert entomologies which makes the process time consuming and cumbersome. It has been proposed that automated species identification through image recognition could be an efficient alternative. In our previous studies we demonstrated that automated species identification can be done through elytral pattern recognition for various species of pantry beetles. However, to make this process effective, one would need large repository of high-quality elytra images that clearly reveal the elytra patterns. High clarity of images is essential in distinguishing one species from another especially if they belonged to the same genus. Thus, it is imperative to optimize and standardize imaging technique that allows any operator to collect high-quality images of beetle elytra in a consistent manner. Herein, we present a method that yields enough pattern clarity that one species could easily be distinguished from another, even for the species belonging to the same *family* or *genera* that often have near-identical pattern. This method was tested against beetles with some of the darkest and thickest elytra. It found to be robust enough to be applied to a wide range of food contaminating (or of other type) beetles. The images obtained were subsequently analyzed and statistical investigated to quantify the variations in size and shape distribution of the patterns for each species. We observed a clear variation in elytral pattern distribution, even for the species within the same family or genus. This method held true when tested on images of fragmented elytry isolated from real-time food contamination samples. It was then extended to capture high resolution images for 30 different species to create a high-quality image repository. In summary, our work demonstrates the method of image acquisition that produces quality images that can be used for automated species level identification of food contaminating beetles.

*Corresponding Author: Joshua Xu (Email: Zhihua.Xu@fda.hhs.gov)

Global Profiling of miRNA Expression for Candidates Associated with Ketoconazole-induced Liver Injury

Dongying Li[1], Binsheng Gong[1], Zhichao Liu[1], Bridgett Knox[1], Weida Tong[1], and Baitang Ning[1*]

[1]National Center for Toxicological Research (NCTR), U.S. Food and Drug Administration (FDA), 3900 NCTR Road, Jefferson, AR 72079, USA

Abstract:
Drug induced liver injury is a major reason for drug withdrawal from the market and termination of clinical trials during drug development. Increasing studies have indicated that microRNAs (miRNAs) are critical modules in hepatotoxicity. However, the potential of using miRNAs as robust and reliable biomarkers for hepatotoxicity remains to be further elucidated, especially via uncovering the underlying mechanisms. In this study, we employed a systematic approach to identify mechanistically-meaningful miRNAs associated with hepatotoxicity induced by ketoconazole. We performed miRNA-sequencing with liver tissues from Sprague-Dawley rats treated with ketoconazole at three doses (10, 30 and 100 mg/kg) and four time points (3, 7, 14 and 28 days). We profiled miRNA expression using the miRDeep2 software and identified differentially expressed miRNAs (DEMs) at different treatment conditions using DESeq2, an R/Bioconductor package. The time- and dose-dependence of miRNA differential expression were then examined to explore miRNAs with early and sensitive response to ketoconazole treatment. To understand the mechanism underlying the association of the DEMs with ketoconazole-induced hepatotoxicity, we identified protein coding genes targeted by these DEMs via computation prediction. Moreover, we conducted pathway analysis to identify biological pathways that DEMs and their target genes were enriched in. Finally, candidate miRNAs responsible to ketoconazole-induced hepatotoxicity were proposed.

*Corresponding Author: Baitang Ning (Email: Baitang.Ning@fda.hhs.gov)

Understanding Sex-specific Drug Properties and Adverse Outcome Pathways of Drugs with Potential for Drug Induced Liver Injury (DILI)

Kapil Khadka[1], Ting Li[1], Minjun Chen[1], Weida Tong[1], Dong Wang[1]

1-Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, FDA, Jefferson, AR

Abstract:

Understanding drug properties and adverse outcome pathways at a finer scale of target group could be of critical importance for improving precision public health by detecting safety signals that otherwise could be missed by traditional analyses. This exploratory study leveraged the information from the curated and standardized (i.e., missing value imputed and deduplicated) version of the FAERS database to detect DILI-concern drugs with unusual sex disparity. Likelihood ratio statistic of drug-adverse event combination was calculated for each sex for all the drugs reported for liver toxicity related events, and only the drugs with statistically significant sex disparity and housed as DILI concern drugs in Liver Toxicity Knowledgebase (LTKB) were used for analyses. Significance level was determined via Monte Carlo simulation, and drugs were grouped as male-biased and female-biased by comparing calculated baseline frequency with observed frequency of drug-AE combination. Daily dose and lipophilicity (logP) values of male and female-biased drugs were found statistically significant with considerably higher average values of logP for female-biased and daily dose for male-biased drugs. Also, association between sex disparity and reactive metabolite (RM) formation was detected. Regarding sex specific mechanistic pathway for adverse outcomes, both the drug groups had considerably higher and comparable hits to MIEs - “Inhibition, Bile Salt Export Pump (ABCB11)”, “Activation, Long term AHR receptor driven direct and indirect gene expression changes”, “Activation of CYP2E1”, and “Alkylation, Protein”. We expect that this exploratory study provides opportunities for hypothesis generation for further investigation in the field of precision public health.

*Corresponding Author: Dong Wang (Email: Dong.Wang@fda.hhs.gov)

Microbiome shift and immune responses: differential effects of corn oil on different rodent models

Gokulan K[1], Kumar, A[1], Lahiani MH[1], Sutherland V[2], Cerniglia CE[1] and Khare S[1*]

[1] Division of Microbiology, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Rd, Jefferson, AR, 72079
2 National Institute of Environmental Health Sciences, National Institute of Health, Research Triangle Park, NC. 27709

Abstract:
The gut microbiome has been identified as a key contributor in the metabolism of ingested xenobiotics. However, the gut microbiome itself could be adversely affected by xenobiotics. In biomedical studies with water insoluble xenobiotics, corn oil is one of the most commonly used non-aqueous vehicles for assessing the toxicological end-points in animal models. To test the toxicity of xenobiotics and related health effects, it is crucial to have a diligent selection of (a) an appropriate animal model to translate toxicity assessment in predicting human exposure, and (b) an appropriate non-interfering vehicle. This study evaluated the suitability of corn-oil as a vehicle in adult female Harlan Sprague Dawley rats and adult female CD-1 mice; the rodent models that are often utilized in toxicological studies.

Here, we focused on gastrointestinal toxicity in the host in terms of (a) shift in gut microbiota, and (b) gene expression of permeability related and immune related genes, at the transcription level. The results showed that corn oil (2mg ml^{-1} kg^{-1}) tested against a water control, has no significant effect on the rat gut microbiome, intestinal cytokine/chemokine secretion, and mRNA expression of intestinal permeability, and immune response genes. Whereas, mice treated with corn oil showed significant shifts in the abundance of bacterial community structures at the phyla, genera and species levels in the ileum, as well as revealed significant changes in the mRNA expression of membrane permeability and immune response genes. In conclusion, our study emphasizes appropriate selection of a rodent model, as well as maintaining proper controls when a vehicle (like corn oil) is used to test the toxicity of water insoluble xenobiotics

*Corresponding author: sangeeta.khare@fda.hhs.gov

Bandwidth measurement of GeSn based high speed photodetectors

Manoj K. Shah, Abdulla S. Ali, Chisom Okeke, Mansour Mortazavi and Aboozar Mosleh

Department of chemistry and Physics, University of Arkansas at Pine Bluff, Pine Bluff, AR, USA

Abstract:
Integrated Microwave Photonics (IMWP) incorporates the functions of MWP components in monolithic or hybrid photonic circuits with the aim of meeting future needs. Hence, IMWP offers the promise of reduction of size-weight-and-power and low production cost. The traditional MWP has confronted with 2 photon absorption issue. Thus, it requires a shift of the operating wavelength to 2 µm and beyond in order to be utilized in microwave photonics applications.
Recently, the Intriguing properties of band gap and lattice constant tunability, true direct-band gap, wavelength coverage up to 12 µm and CMOS compatible process of Germanium-tin (GeSn) has drawn much attention in the photonic society. Over a decade, many GeSn based photodetectors have been reported with dramatic improvement in their performance. The responsivity of photodetector as well as Sn incorporation in the materials keeps on increasing, which extends cut-off wavelength to mid-infrared regimes. To date, only few reports are available discussing the bandwidth measurement of GeSn photodetectors. It is highly desirable to fully understand the potential of GeSn photodetectors as well as other high frequency devices operating at 2 µm wavelength and beyond.
In this work, we report n-i-p vertical photodiodes with circular mesa 50 – 100 µm and Sn % of 8% for the detection up to 2 µm, which can be extended to mid and far infrared spectra with higher Sn compositions. The measured -3 dB bandwidth of the devices achieve nearly 1 GHz, however, showed discrepancy with the simulations, resulted due to the leakage. This result indicates GeSn high speed photodetectors have promising perspectives in next-generation infra-red optic communications.

Risk Factors Associated with Severe Drug Induced Liver Injury

Kristin Ashby[1,2], Ayako Suzuki[3], Minjun Chen[2]

1-Department of Information Science, University of Arkansas at Little Rock and University of Arkansas for Medical Sciences, Little Rock, Arkansas. 2-Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas. 3-Division of Gastroenterology and Hepatology, Duke University, Durham, North Carolina

Abstract:
Drug induced liver injury is a rare adverse drug reaction that can at times be very severe, requiring a liver transplant or resulting in patient death. Currently, there are no reliable methods for predicting patient outcome or clinical course; however, certain drug properties and host factors are known to play a role in both patient susceptibility and DILI severity. In this study, we investigate clinical host factors and drug properties associated with DILI severity in N=462 cases collected by the International Drug-Induced Liver Network Consortium iDILIC dataset. Amoxicillin-clavulanate is overrepresented in this dataset, accounting for 31.3% (N=145) of cases. In order to consider the interaction between drug properties and host factors, we analyzed amoxicillin-clavulanate cases separately. Flucloxacillin was excluded from the analysis because it lacked mild severity cases with only 3 of 101 flucloxacillin cases categorized as mild. In the dataset excluding amoxicillin-clavulanate, we found that patients who were over fifty-five and also taking a high dose medication were at a significantly greater risk of developing a more severe injury (OR: 4.04, CI: 1.24-14.42). Immunosuppressant medications were associated with mild DILI cases (OR: 0.25, CI: 0.09-0.64, p=0.006). Age and sex were significant only amongst amoxicillin-clavulanate cases. Further investigation may reveal additional interactions between drug properties and host factors. Understanding risk factors that are specific to certain medications and populations will improve patient care.

PTMViz: A tool for analyzing and visualizing histone post translational modification data

Kevin Chappell, Stefan Graw, Aaron Storey, Eric Peterson, Stephanie Byrum

Abstract:
Histone post translational modifications play an important role in our genome by regulating the structure of chromatin, therefore contributing to the regulation of gene and protein expression within our cells. Given their importance in any genetic system, irregularities in histone PTMs can lead to a variety of different diseases including various forms of cancer. As researchers aim to learn more about histone PTMs and their role in any given system they will inevitably produce large datasets that will require fast detailed analysis and graphical figures for publication purposes. To meet this demand, the PTMViz tool was created to provide quick detailed analysis for histone PTM datasets. This tool will provide a user with information such as, the overall structure of their data, the significant proteins within their samples, and a global view of histone PTM levels within the sample with the inclusion of interactive data tables and graphs to help the user to navigate their data and quickly find the answers they are looking for.

Extracting Patterns in Medical Claims Data for Opioid Overdose

Ryan Sanders[1], Dr. Shengfan Zhang[1], Dr. Art Chaovalitwongse[1], Dr. Brad Martin[2]

[1] Industrial Engineering, University of Arkansas, Fayetteville, Arkansas, USA
[2] Pharmacy Practice, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

Abstract:
The goal of this project is to develop an efficient methodology for extracting features from time-dependent variables in transaction data. Transaction data is collected at varying time intervals making feature extraction more difficult. Unsupervised representational learning techniques are investigated, and the results compared with those from other feature engineering techniques. A successful methodology provides features that improve the accuracy of any machine learning technique. This methodology is then applied to insurance claims data in order to find features to predict whether a patient is at risk of overdosing on opioids. This data covers prescription, inpatient, and outpatient transactions. Features created are input to recurrent neural networks with long short-term memory cells. Hyperparameters are found through Bayesian optimization. Validation data features are reduced using weights from the best model and compared against those found using unsupervised learning techniques in other classifiers.

Genomic, and Virulence Comparisons of Different Bacterial Isolates from BCO Lesions in Broilers

Nnamdi. S. Ekesi, Adnan A. Al-Rubaye, Robert F. Wideman, Douglas D. Rhoads*.

University of Arkansas, Fayetteville, AR, USA

Abstract:
We used embryo lethality assay (ELA) to examine the virulence of bacteria isolates from bacterial chondronecrosis with osteomyelitis (BCO) lesions of lame broilers. BCO is the leading cause of lameness in broiler chickens. Lameness poses serious animal health and welfare issues, as well as, significant economic losses. The full etiology of BCO is not fully understood but our hypothesis is that bacteria cross the gut and respiratory epithelia. Some of these bacteria survive in the blood and colonize the proximal growth plates of the rapidly growing leg bones through weaknesses in the vascular system. We have been comparing *Escherichia coli*, and *Staphylococcus* species, obtained in our research (when we induced lameness using the wire-flooring model, as well as from commercial broiler farms experiencing lameness outbreaks). Differences in ELA, especially among the *E. coli* strains, prompted us to examine phylogenies using whole genome comparisons. *E. coli* isolates 1409 and 1413, from neighboring farms of the same integrator, showed very different ELA results and affiliate with divergent *E. coli* clades. *E. coli* isolate 1527, from a different farm and integrator, had similar ELA results to and a genome very similar to *E. coli 1413*. Isolate *Staphylococcus aureus 1516* represented a common BCO isolate from the third farm and is mildly virulent in the ELA relative to *E. coli 1413, E. coli 1527* and a very pathogenic human *S. aureus* isolate. The genome of *S. aureus* 1516 is most similar to isolates from deep wounds/lesions from chickens in Poland and more distantly to chicken isolates derived from human *S. aureus* in the United Kingdom. ELA allows virulence comparisons of distinct isolates when containment facilities are not available for live bird work.

A Distance Based Multisample Test for High-Dimensional Compositional Data with an Application to the Human Microbiome

Thy A. Dao[1], Qingyang Zhang[2]

Department of Mathematics and Statistics
University of Arkansas, Fayetteville, Arkansas -USA.

Abstract:
Compositional data refer to the data that lie on a simplex, which are common in many scientific domains such as genomics, geology and economics. As the components in a composition must sum to one, traditional statistical tests based on unconstrained data become inappropriate, therefore more sophisticated techniques are needed to analyze this special type of data. In this presentation, we consider a general problem of testing for the compositional difference between K populations. Motivated by microbiome and metagenomics studies, where the data are often over-dispersed and high-dimensional, we formulate a well-posed hypothesis from a Bayesian point of view and suggest a nonparametric test based on similarity graphs for evaluating the statistical significance. The performance of the proposed method is tested by simulated high dimensional data, a real microbiome dataset to study the difference in throat microbiome between smokers and nonsmokers.

Risk Factors of Lung Cancers in the American Poultry Workers

Saly Abouelenein[a], Mohammed F Faramawi [a,b], Eric Johnson [b]

[a] Biomedical Informatics Department, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas, [b] Epidemiology Department, College of Public Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas.

Abstract:
Subjects who work in poultry slaughtering and processing plants have one of the highest human exposures to chemicals and viruses that cause cancer in chickens and turkeys. Certain viruses infect and cause cancer in chickens and turkeys, like the avian leucosis /sarcoma viruses (ALSV), Reticuloendotheliosis viruses (REV) and Marek's disease virus (MDV)
Generally, the viruses cause cancer in primates, and transform human cells in vitro. About 100 years ago, Ellerman & Bang were first isolated these viruses in chickens. For the first time, clear direct evidence of excess occurrence of cancer in well-defined subjects highly exposed to these viruses has been provided by Dr. Eric S. Johnson In three separate retrospective cohort mortality studies of poultry workers

We conducted a large-scale case-cohort mortality study, 339 lung cancer deaths and 457 controls were selected randomly from combined cohorts of 30,411 poultry plant workers and 16,405 non-poultry workers, belonging to United Food & Commercial Workers unions. We explored the role of 410 potential occupational and non-occupational risk factors for lung cancer in these workers. We split risk factors to 49 occupational risk factors and 361 non occupational risk factors. We made the random forest analysis for occupational and non-occupational features. Finally, we included all risk factors for lung cancer together in the random forest algorithm. In all the analyses, the data were randomly split into two datasets: training and test datasets.

The poultry workers have never been assessed before for long-term hazards associated with these exposures. In regards to chemicals, they are exposed the following chemicals: Polyvinyl chloride when they use plastic films to wrap meat; 2) Polyaromatic hydrocarbons when they inhale smoke emitted during smoking of birds ; 3) Heterocyclic amines, when they inhale aerosols emitted during frying of birds; and 4) nitrosamines when they add them to the spices. Additionally, these workers are among the lowest paid in the food industry. Therefore, their personal habits and surrounding environment (non-occupational factors) associated with low socioeconomic standards may contribute in their elevated risk for lung cancer.

The most important variables associated with lung cancer in these workers were smoking, exposure to irradiation, eating processed food, killing and wrapping chickens. This study provides critical evidence for the first time, necessary to demonstrate the extent to which workers in poultry plants who are exposed to cancer-related risk factors. Measures should be implemented to protect them from these harmful workplace exposures.

Genomics model to predict Drug-Induced Liver Injury with deep learning: DeepDILI+

Ting Li[1,2], Ruth Roberts[3,4], Zhichao Liu[1*], Weida Tong[1*], Shraddha Thakkar[1*]

1 Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA. 2 University of Arkansas at Little Rock and University of Arkansas for Medical Sciences Joint Bioinformatics Program, Little Rock, AR, USA. 3 ApconiX Ltd, Alderley Park, Alderley Edge, SK10 4TG, UK. 4 University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Abstract:
Drug-induced liver injury (DILI) is one of the leading causes of drug failure in clinical trials and drug withdrawal from the market. An alternative approach that can accurately predict DILI at the early stage has the potential to reduce the drug attrition, and it's in great need by regulators as well as developers. In the 21st century, toxicology experienced the paradigm shift, and now it relies heavily on high throughput technologies such as genomics and could be used as a source to predict DILI related signatures. With the advent of AI-based approaches, the deep learning-based algorithms have demonstrated the capability of learning import features from a large number of inputs with hidden signals. In this study, we propose a deep neural network (DNN) model to predict DILI signatures based on gene expression data that was captured using human immortalized cell lines. An in-depth analysis was conducted to interpret the DNN based predictions. Also, it was compared with the other state-of-the-art methods, i.e., K-nearest neighbors (KNN), Support Vector Machine (SVM) and Random Forest (RF) to have a comparative understanding of advantage/disadvantage of deep learning-based predictions. The DNN model was trained on 3840 samples and predicted an internal validation set of 1200 samples with an accuracy of 74.3%, the sensitivity of 83.9%, the specificity of 60.3%, and area under the curve (AUC) of 0.798. The optimized DNN model was the most balanced compared with KNN, SVM, and RF in the internal validation set and external validation set. We also observed that the optimized DNN model had a higher accuracy of 87.2% for drugs in the pharmacological group of antineoplastic and immunomodulating agents. We demonstrated the utility of the DNN based predictive models for predicting underlying DILI signatures using cancer cell lines. Deep learning-based models produced a balanced prediction compared to the conventional predictive models for drug safety-related evaluations.

*Author of Correspondence:
Thakkar S, Shraddha.Thakkar@fda.hhs.gov, Liu Z, Zhichao.Liu@fda.hhs.gov and Tong W, Weida.Tong@fda.hhs.gov

Directed Genome Evolution Identifies Deoxyribose Phosphate Aldolase as a Macrophage Survival Factor in Staphylococcus agnetis.

Abdulkarim Shwani, Sura Zaki, Sonali Lenaduwe, Sohita Ojha, Robert F. Wideman, Douglas Rhoads

Cell and Molecular Biology program, University of Arkansas

Abstract

We have employed Directed Genome Evolution (DGE) to identify two copies of the deoxyribose-phosphate aldolase gene as essential factors for survival of *Staphylococcus agnetis* (isolate 908) in a chicken macrophage cell line. A specific amino acid substitution appears to be associated with survival and killing. Macrophage survival may be a key component of the hypervirulence of isolate 908 in inducing bacterial chondronecrosis with osteomyelitis (BCO) in broilers. We first reported the isolation of this staphylococcal species from the bones and blood of lame broilers at the University of Arkansas. We have demonstrated high incidences (>60%) of BCO through administration of *S. agnetis* 908 through aerosols or drinking water. The annotated complete genome of isolate 908 has been published. BCO primarily affects the growth plate of the proximal femur and tibia of fast-growing broilers, but survival in the blood may be essential for transmission to the growth plate. Our phylogenomic analyses of chicken and cattle isolates of *S. agnetis* and *Staphylococcus hyicus* suggest a very close relationship between cattle and chicken isolates. Chicken isolate, 908, is closely related to a cattle isolate, strain 1379. Yet, more than 40 genes and 3 plasmids from strain 908 are absent or poorly conserved in any of the cattle *S. agnetis* isolates. We have found that isolate 1379 is efficiently killed by chicken macrophage but isolate 908 not only survives phagocytosis, this isolate efficiently kills immortalized chicken macrophage within 2 days. We produced and sequenced more than 13 independent DGE transformants of 1379 with 908 DNA. Our analyses demonstrate that multiple elements can be efficiently transferred but the unifying property of all DGE selections is a single amino acid change in either copy of deoxyribose-phosphate aldolase. Our current efforts are aimed at confirmation using purified PCR products, and to understand the relevance of this “stress-response” determinant in survival and killing.

Identification of QTLs Associated with Drought Resistance Traits at Reproductive Stage in K/Z RILs Rice Population

Yheni Dwiningsih[1], Julie Thomas[1], Anuj Kumar[1], Chirag Gupta[1], Charles Ruiz[1], Sara Yingling[1], & Andy Pereira[1]

Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR, USA

Abstract

Rice is the staple food for half of the world population, with USA as the third largest exporter of rice, with an export value of US$ 1.8 billion. Rice also uses 2-3 times the water as other food crops, which totals 30% of the world's freshwater resources world-wide. Stability of rice production is facilitated by economic use of water, which is most essential during flowering and grain formation. In our research we screened adapted US rice cultivars, comprising tropical *japonica* rice genotypes, for drought resistant (DR) and water use efficiency (WUE) traits to search sources for breeding US rice cultivars for a water saving agricultural system. And then identify quantitative trait loci (QTLs) of drought-related trait, such as spikelet per panicle (SP), panicle length (PL), primary panicle branch number (PPB), filled grain per panicle number (FG), hundred grain weight (HGW) and grain yield per plant (GY). A RIL population derived from varieties Kaybonnet (DR) and ZHE733 (sensitive), termed K/Z RILs, was available from the USDA Dale Bumpers National Rice Research Center and chosen for genetic analysis of DR/WUE traits. The RIL population was screened in Fayetteville (AR) by controlled drought stress treatment at reproductive stage, and the effect of stress quantified by measuring the drought-related traits. Furthermore, drought related traits such as SP, PL, PPB, FG, HGW, and GY have a strong correlation with drought conditions. QTL analysis was performed with 28595 Single Nucleotide Polymorphisms (SNPs) markers on 138823 loci mapping data using WinQTL Cartographer 2.5 V. Composite interval mapping (CIM) was used. A total of 15 QTLs were identified for drought-related traits: 1-SP, 3-PL, 1-PPB, 6-FG, 1-HGW and 3-GY. These QTLs will be useful for marker assisted breeding to improve rice productivity and stability under drought stress environments and to explore the relationship between the QTLs and phenotypic traits.

Key words: rice, drought, mapping population, SNP markers, composite interval mapping

Corresponding author: apereira@uark.edu

Reference2Vec: Teach Machine to Read People References

Xinming Li, John R. Talburt, Ting Li

Information Science Department, University of Arkansas at Little Rock, Little Rock, Arkansas, USA
{xxli3, jrtalburt, txli1}@UA Little Rock.edu

Abstract:
Traditional Entity Resolution (ER) depends on intensive human effort. Expecting deep learning can reduce human effort, we propose Reference2Vec, a fundamental functionality that transforms textual references to fixed length vectors (machine language). In other words, this research is to teach machine to read people references.

This research focuses on the challenging unstructured references, which don't follow a uniform attribute schema. While the traditional ER approaches require a uniform attribute schema, the deep learning approach can go beyond the attribute boundary. Specific to ER, to capture both syntactic and semantic similarity, Reference2Vec is designed at character level rather than the general word level. For training data, our synthetic references provide deep learning model with enough information to learn. First, the synthetic references simulate real world people behaviors, such as changing name after marriage, and changing address after moving. Second, the synthetic references also simulate real world data quality problems like missing values and typos. Both supervised and unsupervised learning are designed to teach machine to learn. The supervised training is based on the Siamese structure, within which convolutional neural network (CNN), long short-term memory (LSTM), and the combination of CNN and LSTM are used as feature extractors. We also explore unsupervised auto-encoder.

Our experiments show that the Siamese neural network using CNN to extract N-gram features from references gets the best performance. The dense "character embedding" is better than one-hot vector for encoding reference in CNN. Through controlling quality of training data, we find that, for the same test data, model trained on low quality data performs better than model trained on high quality data; this shows the potential for transfer learning of the Reference2Vec. Based on Reference2Vec teaching machine to read people references, the way of doing ER can be re-designed through machine's eye.

Fall 2019 Cheminformatics OLCC

Hunter L. Tiner[1], Robert E. Belford[1], Sunghwan Kim[2], Jordi Cuadros[3], Ehren Bucholtz[4], Dean Johnston[5], Kristen Fulfer[6], Jason Moore[7] and Delmar Larsen[7]

1. UA LITTLE ROCK, Little Rock, AR, USA, 2. National Institutes of Health, Bethesda, MD, USA; 3. IQS Ramon Llull, Barcelona, ES; 4. Saint Louis College of Pharmacy, St. Louis, MS, USA; 5. Otterbein University, Westerville, OH, USA; 6. Centre College, Danville, KY, USA; 7. UC Davis, Davis, CA, USA

Abstract:
In the Fall of 2019, five campuses participated in the Cheminformatics On-Line Chemistry Course (OLCC), which is an intercollegiate, collaborative teaching project. Throughout this course, students had the opportunity to interact with scientists who develop and maintain the NIH's PubChem, the largest chemical database in the public domain. The course consisted of eight modules on various cheminformatics topics, including the basics of chemical representation on computers, accessing data through databases, predicting Quantitative Structure Property Relationships, Molecular Similarity, Computer Aided Drug Design and Machine-Learning basics. The modules involved python and R programming activities built in Jupyter notebooks that students ran on their own computers or the LibreText Jupyter Hub at UC-Davis. OLCCs have been organized by the American Chemical Society Division of Chemical Education's Committee on Computers in Chemical Education since 1996, with this being the third one on cheminformatics and the first one to be run at the LibreText HyperLibrary and involve a Jupyter Hub. The content is freely available at https://chem.libretexts.org/link?143689.
This poster is being presented by a student who participated in the OLCC and was involved with converting some of the python scripts to R. It will describe what an OLCC is, give an overview of the modules, and describe the use of a Jupyter Hub in a class designed for undergraduate students who had no prior programming experience. This site is intended to be used as a resource for future students who wish to learn cheminformatics and be introduced to some of the basics of machine learning.

Semantic Representations of Multi-Modal Data, NeuroInformatic Processing Pipelines, and Derived Neuroimaging Results in the Arkansas Image Enterprise System (ARIES)

Aaron S. Kemp, MBA[1,2], Jonathan Bona, PhD[2], Tracy S. Nolan[2],
Linda Larson-Prior, PhD[1,2,3], Tuhin Virmani, MD, PhD[3], Lakshmi Pillai, MS[3],
Mathias Brochhausen, PhD[2], Lawrence Tarbox, PhD[2], and Fred Prior, PhD[2]

[1]Psychiatry, University of Arkansas for Medical Sciences (UAMS), Little Rock, AR; [2]Biomedical Informatics, UAMS, Little Rock, AR; [3]Neurology, UAMS, Little Rock, AR

Abstract:

Introduction

Neuroimaging is among the most active research domains for the creation and management of open-access data repositories[1]. Heavy emphasis has been placed on functional magnetic resonance imaging (fMRI) data for both disease specific collections and healthy brain function[2]. The Cancer Imaging Archive (TCIA) has been the National Cancer Institute's principal imaging resource and has encouraged and supported open-science research by acquiring, curating, hosting and managing collections of multi-modal information[3]. The TCIA technology stack is currently being refactored into a more streamlined, easily maintained, containerized package which we have labeled PRISM: Platform for Imaging in Precision Medicine. One of the first applications of PRISM is the establishment of a neuroimaging research data management system at the University of Arkansas for Medical Sciences, which is known as the Arkansas Image Enterprise System (ARIES). As the first instantiation of the PRISM infrastructure, the ARIES project aims to explore the practical utility and usability of the full set of capabilities that this new platform provides. In particular, the integration of semantic representations of multi-modal data elements from a variety of disparate sources (e.g., imaging, behavioral, or cognitive assessments), across image processing stages (e.g., preprocessing steps, neuroanatomical segmentation schemes, analytic pipelines), as well as descriptions of the derived results would ensure greater reproducibility and comparability of scientific findings across large-scale neuroimaging research projects.

User Groups and Pilot Data

Pilot testing of the ARIES instantiation of PRISM is being conducted with three collaborating investigative teams who are using ARIES in a project designed to identify common pathways of neurodegeneration. The dataset for the pilot test includes neuroimaging measures (structural and functional MRI and EEG) as well as endophenotypic data obtained from a variety of assessments designed to measure neuro-motor integration (wearable body sensors, gait-assessment floor mat, digitized gloves, and handwriting/drawing assessments on a digitizing tablet) and neurocognitive functions (performance scores on standardized neuropsychological tests and cognitive activation tasks from functional imaging) in three unique study cohorts diagnosed with Parkinson's disease (PD), Mild Cognitive Impairment (MCI), or Cancer-Related Cognitive Impairment (CRCI).

Semantic Integration Approach

To integrate and manage these data we are building semantic representations using axiomatically-rich ontologies. These will be used to instantiate a knowledge graph that combines the data from these unique study cohorts into a shared semantic representation that explicitly accounts for relations among these data. This amounts to providing queryable relationships across the source data sets. This knowledge graph is stored in a triple store database that supports reasoning over and querying these integrated data. We believe that these unigue capabilities will facilitate the discovery of important new linkages among endophenotypic expressions of disturbed neural functions and discrete neuroanatomical markers of neurodegeneration obtained from the derived neuroimaging results.

Conclusion

Semantic integration of neuroinformatic processes in the ARIES pilot project demonstrates the capabilities of the PRISM infrastructure to effectively represent detailed neuroanatomical segmentation schemata and processing pipelines for image analyses, integrate a diverse set of multi-modal data elements, and provide detailed descriptions of the results obtained across the analytic processing stages and in relation to the supporting endophenotypic data. Such capabilities are essential to ensure greater reproducibility in large-scale neuroimaging research projects.

References

1. Eickhoff S, Nichols TE, Van Horn JD, Turner JA. Sharing the wealth: neuroimaging data repositories. Neuroimage. 2016;124(Pt B):1-8.
2. Hodge MR, Horton W, Brown T, et al. ConnectomeDB—sharing human brain connectivity data. Neuroimage. 2016;124:1102-7.
3. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging. 2013;26(6):1045-57.

Design and Simulation of GeSn based Waveguide for Infrared Photonic at λ = 2μm – 3μm from 1% to 10% of Sn Composition

Abdulla Said Ali, Chisom Okeke, Manoj K. Shah, Mansour Mortazavi.

Department of Chemistry and Physics, University of Arkansas at Pine Bluff
Pine Bluff, Arkansas USA.

Abstract:
Photonics roles will clearly be impacting different areas as high-speed communication for devices and optical communication within computers. The operation of waveguide devices is well researched and understood; their particular performance relies on many parameters, geometry, wavelength and initial field distribution, material data, and electro optic driving conditions. The indirect band-gap nature of bulk Si has been one of the foremost downsides of the material, It has positioned Si away from numerous optical applications for example LEDs and lasers, in favor of the other III–V semiconductors, Si has a larger band-gap of 0.7 eV than that of Ge 0.2 eV, and because of this, the phenomenon of thermal pair generation is smaller in Si than in Ge. Because of the boundaries of Germanium (Ge) and Silicon (Si) as linear and nonlinear waveguide medium to its interaction with other waveguide materials, a design and simulation of photonic waveguide devices using Germanium Tin (GeSn) with 1% of Sn up to 10% of Sn composition on a silicon-on-insulator (SOI) substrate was performed.

A one-dimensional (1D) calculation has been performed to define the slab waveguide modes, as supported by the wafer as well as two-dimensional (2D) calculation. Various wavelengths from 1310nm, 1550nm, 2µm 2.5µm and above, depends on the materials. This has been done numerically and analytically using various advanced simulation software available on the market including Lumerical Mode Solution and Matlab. We reported that for GeSn waveguide with 2% Sn Composition, a waveguide below 495nm at λ = 2μm will be needed to operate in a single mode, for a wider waveguides, another TE-like mode is present, at 910nm a second TM-like mode appears.

Comparison of the Impacts of Broadband and Selected Infrared Wavelengths on Inactivation of Microbes on Rough Rice

Abass A. Oduola, Rebecca Bowie, Shantae A. Wilson, Zeinab Mohammadi Shad, Griffiths G. Atungulu*

Department of Food Science, University of Arkansas Division of Agriculture, Fayetteville, Arkansas, USA.

Abstract:
Proliferation of microorganisms on rough rice during storage is one of the major challenges facing the rice industry. In order to minimize microbial contamination, freshly harvested rice must be dried to a safe moisture content and existing microbes inactivated. Infrared (IR) heat is an effective technique that is capable of simultaneously drying and deactivating microbes on rough rice. This study investigated the effects of broadband and selected infrared (IR) wavelength treatments of rough rice on microbial inactivation. Rough rice was treated at different IR wavelengths and product-to-emitter distances (110, 275, and 440 mm) followed by tempering at 60 °C for four hours. The total mold and aerobic plate counts (APC) on non-treated and treated samples were determined. Significant total mold reductions of 1.14 and 3.11 log CFU/g were obtained after IR heating using broadband and selected wavelengths, respectively ($p<0.05$). The most significant reduction of APC using selected IR wavelength was 1.09 log CFU/g; the broadband IR wavelength had no effect on the mean APC. The IR treatments followed by tempering step resulted in greater reductions of total mold counts and APC (4.03 and 3.50 log CFU/g) in comparison to IR treatments without tempering (3.11 and 1.09 log CFU/g). Overall, bacteria showed more resistance to IR treatments than molds. This study gives an insight into how microbes (mold and bacteria) react to both broadband and selected IR wavelengths and therefore, help rice processing industry to use only the effective wavelengths and intensities to avoid wastage of energy and cost.

Keywords: Rough rice, fungi, bacteria, wavelength, intensity, inactivation

*Corresponding author email: atungulu@uark.edu

Industry Payment to Vascular Neurologists: A Six-year Analysis of the Open Payments Program from 2013 through 2018

Krishna Nalleballe*, M.D[1]; Sen Sheng*, M.D, Ph.D[1]; Chenghui Li, Ph.D[2], Ruchira Mahashabde, B Pharm, M.S[3], Amarnath R. Annapureddy, M.D[4], Kamran Mudassar, M.D[5], Krishna Pothineni, M.D[6], Poornachand Veerapaneni M.D[1], Yohei Harada M.D[1], Abhishek Chilkulwar, M.D[7], Saritha Ranabothu, M.D[8], Aliza Brown, Ph.D[1], Nidhi Kapoor, M.D[1], Sanjeeva Onteddu, M.D[1], Xixi Wang[9]

1-Department of Neurology, Little Rock, AR. 2-Department of Pharmacy Practice, University of Arkansas for Medical Sciences, Little Rock, AR. 3- Department of Pharmaceutical Evaluations and Policy, University of Arkansas for Medical Sciences, Little Rock, AR. 4-Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut. 5-Department of Interventional Radiology, University of Arkansas for Medical Sciences, Little Rock, AR. 6-Department of Cardiology, University of Arkansas for Medical Sciences, Little Rock, AR. 7-Department of Hematology & Oncology, Saint Louis University, St Louis Mo. 8-Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR. 9-Department of Math and Statistics, University of Arkansas, Little Rock, AR

Abstract:

Industry payments to physicians raise concerns regarding conflicts of interest that could impact patient care. We explored non-research and non-ownership payments from industry to vascular neurologists to identify trends in compensation. Using Centers for Medicare and Medicaid Services and American Board of Psychiatry and Neurology data, we explored financial relationships between industry and US vascular neurologists from 2013-2018. We analyzed payment characteristics, including payment categories, payment distribution among physicians, regional trends, and biomedical manufacturers. Furthermore, we analyzed the top 1% (by compensation) of vascular neurologists with detailed payment categories, their position, and their contribution to stroke guidelines. The number of board-certified vascular neurologist increased from 1169 in 2013 to 1746 in 2018. The total payments to vascular neurologist increased from $99,749 in 2013 to $1,032,302 in 2018. During the study period, 16%–17% of vascular neurologists received industry payments. Total payments from industry and mean physician payments increased yearly over this period, with “consulting fee (31.1%)” and “compensation for services other than consulting (30.7%)” being the highest paid categories. The top 10 manufacturers made most of the payments, and the top 10 products changed from drug or biological products to devices. Physicians from south region of the United States received the highest total payment (38.72%), which steadily increased. Payments to top 1% vascular neurologists increased from 64% to 79% over the period as payments became less evenly distributed. Among the top 1%, 42% specialized in neuro intervention, 11% contributed to AHA/ASA guidelines and around 75% were key leaders in the field. In conclusion, A small proportion of US vascular neurologists consistently received the majority of industry payments, the value of which grew over the study period. Only 11% of the top 1% receiving industry payments have authored AHA/ASA guidelines, but ~75% appear to be key leaders in the field. Whether this influences clinical practice and behavior requires further investigation.

Isolation and Molecular Characterization of Shiga-Toxin producing *Escherichia coli* from Food and Clinical samples

Khulud Alotaibi and Ashraf Khan

Division of Microbiology, National Center for Toxicological Research, USFDA, Jefferson, AR 72079

Abstract:

Forty-six Shiga toxin producing *Escherichia coli* strains, isolated from clinical and food samples were analyzed for antibiotic resistance, virulence genes, plasmids and plasmid replicon types. To assess the genetic diversity, *XbaI* restriction enzyme pulsed-field gel electrophoresis (PFGE) fingerprinting and plasmid profiles were performed. Forty clinical and two from food isolates were resistant to tetracycline and eight from clinical and two food were resistant to ampicillin. Six isolates from clinical resistance and one from food sources were resistance to trimethoprim-sulfamethoxazole and kanamycin. All isolates from food and clinical samples were sensitive to nalidixic acid and gentamicin except one from clinical was resistant to gentamicin. PFGE typing of forty-six isolates by *XbaI* resulted in 15 to 20 bands and were grouped to more than ten clusters, each with similarity from 70% to 30%. Most isolates were positive for one or more of the fifteen virulence genes, which are mostly found in shiga toxin producing *E. coli* (*Eae*, *HlyA*, *Stx1*, *Stx2*, *Stx2f*, *Stx2c*, *Stx2e*, *Stx2*, *NelB*, *PagC*, *Sen*, *ToxB*, *Irp*, *Efa*, *Efa1*). All isolates carried a typical more than 20 kb plasmid, and most of the isolates had large and small plasmids. Twenty-nine clinical isolates were positive by PCR for X, twenty clinical isolates for FIB plasmid and seven isolates from clinical were positive for Y plasmid replicon type. Three clinical isolates positive for I1 and HI1 genes. All food isolates were negative for genes of plasmid typing replicon except two were positive for B/O and W. These results indicate that shiga toxin producing *E. coli* have a diverse clonal population among the clinical and food samples and are resistant to several antimicrobials. Since, these shiga toxin producing *E. coli* isolates have several virulent and antibiotic resistant genes which can be on mobile genetic elements which can factor into and contribute to the potential transmission to other microorganisms. Surveillance and characterization of STEC can provide useful information about the trend of STEC infection and will help to identify STEC serotypes that are highly pathogenic to humans that may emerge as a threat to public health.

Improving the Phylogeny of *Arceuthobium* (Dwarf Mistletoe)

Kate Sanders[1], Jacob Idec[2], Adam Schneider[2]

[1] Department of Mathematics and Computer Science, Hendrix College, Conway, Arkansas
[2] Department of Biology, Hendrix College, Conway, Arkansas

Abstract:
Arceuthobium, also known as dwarf mistletoe, is a widespread genus of parasitic angiosperms that primarily affect pine and cypress trees. Since *Arceuthobium* does not rely solely on photosynthesis, rapid gene loss occurs in the chloroplast and nucleus. This leads to extreme morphological reduction, so a traditional approach to phylogeny construction is challenging. Most older taxonomies rely almost exclusively on geography for species identification. For more accurate classification, the use of molecular data in constructing a phylogeny is needed. We improve previous attempts at building a molecular phylogeny of *Arceuthobium* by creating a properly rooted maximum-likelihood phylogeny using a larger number of taxa and genes. We performed genome skimming with reference guided assembly on the nrDNA cistron and several plastid regions for 1-3 samples per species. This phylogeny will enhance our understanding of *Arceuthobium*'s evolution and support future taxonomic refinement of this group.

Evaluation of lightweight alignment tools for RNA-seq quantification in plants

Anjala Jiji [1], Dr. Asela Wijeratne[2]

Department of Biological Sciences, Arkansas State University, State University, AR 72467

Abstract:

RNA sequencing (RNAseq) is a valuable tool that can be used to monitor global gene expression. RNAseq involves aligning millions of short DNA sequences to a reference DNA sequence and counting the number of reads aligned to given loci to quantify the gene expression. Traditionally, the alignment step is carried out using short-read aligners, which is time-consuming. However, novel alignment-free methods, such as Kallisto and Salmon, reduce the analysis time significantly compared to standard tools like BWA.

Both Kallisto and Salmon have been tested using datasets from humans. However, it is unclear how they perform with plant datasets. A systematic evaluation of these tools with plant datasets is critical, given plants have complex genomes with many redundant DNA sequences, which could affect the performance of these tools. Our goal is to evaluate the performance of Kallisto and Salmon using Arabidopsis and soybean (soybean is estimated to have 70% of its genes as duplicates) datasets and compare the results obtained from BWA.

We will use Rsem-simulate-reads to simulate RNAseq data for Arabidopsis and soybean. The simulated RNAseq data will be aligned to respective reference sequences using BWA, Kallisto, and Salmon. DESeq2 will be used to find differences in gene expression levels. Results from the BWA analysis will be considered a control/standard and will be used to evaluate the performance of Kallisto and Salmon. Similarly, we will use published data for Arabidopsis and soybean to perform the same analysis to evaluate the two tools using real datasets.

RNAseq is central to many plant researches and having a fast way to analyze these massive datasets will speed up the analyses, and the results obtained from this study will enable us to achieve this goal. Also, using these data, we will create a standardized analytical pipeline that will be beneficial for many researchers.

Differences in learning and gene expression in brains of male and female *Bicyclus anynana*

Gabrielle A. Agcaoili[1], David Ernst[1], Erica Westerman[1]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
One way to understand the variation in the behavior of animals is by looking at the genes involved. I am particularly interested in behavioral differences between the sexes. How might these differences be manifested in the brain? This study worked to answer this question by using male and female butterflies of the species *Bicyclus anynana*, examining what goes on in the brains of males and females as they learned from a social exposure. We focused on how sex plays a role in behavior and learning; it has been seen that males and females respond to the same social experience in similar yet different ways. *B. anynana* males and females learn from the same experience but exhibit sex biases in the traits they look for as well as in what each are good at learning. We explored whether sex specific biases in learning are due to sex specific variation in gene expression in perception or higher processing. Males and females were given the same social experience, and the behaviors they exhibited during this training period were recorded. Each treatment - naive versus learning and male versus female - consisted of 10 individuals for a total sample size of 40. At the end of the training period, their heads were flash frozen for later dissection and RNA extraction for eyes and brains. A total of 40 eyes and 40 brains were collected. The RNA was sequenced to look for differentially expressed genes between the sexes. Butterflies, like many species, including humans, are social animals. So, the conclusions drawn from this experiment could be applied to better understanding differences in behavior and genetics of many animals, possibly including humans. This can help in understanding more about human biology. This is especially important in today's world as individualized medicine becomes more and more prevalent.

Using Procalcitonin Levels: Benefits to Antimicrobial Stewardship and Aspiration Diagnosis

Jessica Gilmore[1], Di' Granger[1], Samantha Robinson[2]

1-Department of Health, Human Performance and Recreation, University of Arkansas, Fayetteville, AR USA. 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Procalcitonin (PCT) is a peptide precursor of the hormone calcitonin, often produced in response to bacterial infection. PCT levels increase within 3-6 hours after onset of a bacterial infection. This makes PCT levels beneficial to antimicrobial stewardship programs (ASPs), as PCT levels can assist with antibiotic initiation and/or discontinuation. ASPs are necessary to improving the use of antibiotics and reducing antibiotic resistance in medical centers. Results of a retrospective, quasi-experimental cohort study show that physician compliance with PCT-guided ASP recommendations led to a significant reduction in length of antimicrobial therapy (LOT), demonstrating the beneficial aspects of using PCT levels in antimicrobial stewardship.

Additionally, the significant increase of PCT levels in the blood in response to bacterial infection, suggests that PCT levels could assist healthcare practitioners in the often difficult to distinguish diagnosis of aspiration pneumonitis and aspiration pneumonia. High levels of morbidity and mortality are associated with both pneumonitis and pneumonia after aspiration, making it clinically relevant to distinguish between the two illnesses. The difficulty in diagnosis often results in inappropriate treatment and unnecessary use of antibiotics in medical centers. Results of a second retrospective, quasi-experimental cohort study demonstrate that PCT levels are significantly lower for patients with confirmed aspiration pneumonitis. Given that antibiotic therapy is not recommended for aspiration pneumonitis, this preliminary finding suggests that PCT levels could be used to effectively diagnose aspiration illnesses and guide therapeutic recommendations.

Identification of Individual and Regional Features Impacting Childhood Nutrition in Tanzania

Benjamin Ingle[1], Samantha Robinson[2]

1-Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA; 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Despite recent declines in childhood mortality, nearly 60% of children in Tanzania have some form of anemia. In the Lake Zone of Tanzania, anemia is actually the second leading cause of morbidity in children under the age of five (37%). Moreover, over 1/3 of children in the country are stunted (i.e., these children are too short for their age). While the nutritional status of children in Tanzania has improved during the last 25 years, reports continue to indicate chronic undernutrition. As such, it is apparent that there is still much work to be done to improve childhood nutrition in Tanzania.

Data for this exploratory study consists of survey responses from the most recently released Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS). Utilizing Classification and Regression Tree (CART) methods as well as Random Forests (RF), this study identified certain maternal factors and living characteristics that most influence nutritional outcomes among children. This exploratory analysis (as well as the visual appeal of the tree-based methodological approach) can motivate, inform, and facilitate the creation of policy that improves childhood nutrition across all of Tanzania.

Exploring the Relationship between Maternal Health Characteristics and Preterm Birth using Classification and Regression Tree (CART) Methods

Stetson Ledbetter[1], Emily Bell[1], Samantha Robinson[2]

1-Department of Health, Human Performance and Recreation, University of Arkansas, Fayetteville, AR USA; 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
With modern medical procedures, research interest in preterm births would seem obsolete and unnecessary. While preterm births are only a fraction of total births in the United States (US), the preterm birthrate has actually risen to more than 10%. In fact, the US is one of the 10 countries with the greatest number of preterm births in the world.

The symptoms and complications resulting from preterm birth can be partially alleviated. For instance, the use of magnesium sulfate injections, the ability to plan for Cesarean births, and the use of certain holistic methods have proven effective. Delaying birth is also possible through the use of steroids yet can cause several problems itself. While there is the potential to reduce risk, the preterm birthrate continues to climb and the number of potentially preventable preterm births continues to grow.

Using natality data from the National Vital Statistics System (NVSS), the purpose of this study was to predict preterm birth status and gestational age using maternal health characteristics. Classification and Regression Trees (CART) methods were applied and, while these methods displayed high rates of overall accuracy, the sensitivity of the predictions was quite low. Maternal health characteristics seem to contribute to preterm birth but, due to imbalance in the natality data, prediction of preterm birth status and gestational age was limited. Results highlight the importance of greater data collection and further study to more accurately predict and prevent preterm births.

Forecasting the Number of Illnesses Related to Salmonella Outbreaks

Abigeal Omolewu[1], Samantha Robinson[2]

1-Department of Poultry Science, University of Arkansas, Fayetteville, AR USA. 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Salmonella, a rod-shaped gram-positive bacterium that belongs to the family Enterobacteriaceae is one of the most common types of foodborne pathogens that result in illness. According to the Centers for Disease Control and Prevention (CDC), Salmonella is estimated to cause about 1.35 million infections, 26,500 hospitalizations, and 420 deaths in the United States every year. Children under the age of five, older adults and individuals with weakened immune systems are more susceptible to Salmonella infection, which can then lead to death. Although Salmonella can be found mostly in poultry and poultry product, it can also be found in other food such as vegetables, port, nut butter etc.

Outbreaks linked to Salmonella were consistently under reported for many years. However, with the establishment of the National Salmonella Surveillance System in 1962, there has been improvement in the report of outbreaks and the resultant numbers of illnesses. National disease surveillance systems help to collect data on laboratory-confirmed Salmonella outbreaks using state and territorial laboratories. CDC's National Salmonella Reference Laboratory confirms and/or further characterizes unusual and untypable serotypes. The serotype-specific surveillance for Salmonella in the United State was established to determine endemic patterns of Salmonella infection, detect outbreaks, monitor trends in disease transmission, and evaluate control efforts.

Using data from the National Outbreak Reporting System (NORS), this study implements various time series modeling techniques to investigate past trends in the number of illnesses resulting from future Salmonella outbreaks. Given the strong seasonality observed in the illness trend data, a seasonal method of exponential smoothing is then utilized to forecast the number of future illnesses related to Salmonella. This study, which explores Salmonella illness trend data between 1998 and 2017 and provides future prediction of outbreak illness numbers, helps guide further control efforts.

Regional Differences in Factors Influencing Arkansas Teenage Pregnancy Rates

Andrew Palmer[1,2], Samantha Robinson[3]

1-Department of Sociology, University of Arkansas, Fayetteville, AR USA. 2-Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA. 3-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA.

Abstract:
Within the field of public health, there is growing interest in the elucidation of factors that influence the incidence of reproductive health phenomena. Teenage pregnancy rates in the United States have been declining in recent decades, though the rates remain significantly higher than many other developed nations.

Previous research has focused on global linear regression models that place the entire nation in a regression without weighing potential regional social or economic differences. These global models assume that the relationships between teen pregnancy rates and various predictor variables remain constant throughout the entire nation, which can obscure regional effects and relationships that exist on a more disaggregate level.

Geographically Weighted Regression (GWR) is a local spatial regression technique that allows for spatially-varying parameter estimates. This advancement in regression modeling makes it possible to retain regional effects in model building. Shoff and Yang (2012) used GWR to examine regional differences in teen birth rate determinants between rural and metropolitan areas. The present study applies a similar methodology (i.e., GWR) to illustrate the differential importance of teen birth rate determinants at the county level across Arkansas.

Paramount to the understanding of teenage pregnancy as a social phenomenon is the examination of social determinants and predictors of teen pregnancy in areas where teen pregnancy is the most prevalent, namely, Arkansas. Further, an in depth study of Arkansas on a county level yields novel results, unobservable on the national scale and of greater importance to policy in Arkansas. This study adds to a growing body of research into determinants of teenage pregnancy and may serve to better inform region-specific public policy.

Clustering Public Perception of Vaccination on Social Media

Allison Peters[1], Sharifa Williams-Lewis[2], Samantha Robinson[3]

1-Department of Rehabilitation, Human Resources and Communication Disorders, University of Arkansas, Fayetteville, AR USA. 2-Department of Biomedical Engineering, University of Arkansas, Fayetteville, AR USA. 3-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Vaccines are one of the greatest achievements in modern medicine and within the scientific field. With the growing movement of the anti-vax campaign in the recent years, this has attracted extensive attention throughout social media and has created a debate in mainstream media. The debate has brought to light controversy and the rapid spread of misinformation. Public concerns regarding vaccine safety and effectiveness have altered public perception on vaccination.

Using data extracted from Twitter, this study examines current public perception of vaccination by exploring the sentiment of social media posts. Sentiment analysis is used to visualize the overall attitude of the public regarding vaccination. Topic Clustering, a form of Latent Dirichlet Allocation (LDA), is then used to form clusters of related words/terms within the Twitter data. Four clusters emerge that describe the differing opinions and attitudes of the public. Specifically, the four emergent clusters relate to vaccine risks, vaccine benefits, administrative requirements, and public outcry/condemnation over the reemergence of several 'old' diseases. By understanding public perception of vaccination on social media, this study can inform public officials about how best to address both sides of the debate and maintain the health of our communities.

Analyzing the Effect of Food Environment on Diabetes Prevalence in Arkansas Counties across Space and Time

Mackenzie Reed[1], Samantha Robinson[2]

1-Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA. 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Type 2 diabetes mellitus (T2DM) is a chronic disorder in which the body is resistant to the effects of glucose on blood sugar levels. An estimated 30.3 million Americans have diabetes, with 1.5 million new cases diagnosed each year. A blend of both genetic and non-genetic risk factors influence an individual's chances of developing T2DM. Non-genetic risk factors, such as lifestyle and dietary intake, play a significant role in the development of this disorder. In the United States, Arkansas ranks fourth highest for the prevalence of diabetes.

The purpose of this study was to examine the relationship between food environment and diabetes prevalence between Arkansas counties during a five year time-span (2015-2019). Geographically Weighted Regression (GWR), a local spatial modeling technique, was implemented utilizing a boxcar kernel weighting scheme with a fixed 60 kilometer bandwidth for each individual year between 2015 and 2019. Results indicated that the relationship between food environment and diabetes prevalence in Arkansas has changed not only across time but across space.

Future work, incorporating further explanatory variables such as ethnic makeup of each county, adult smoking rate, access to exercise, and rate of excessive drinking, will add to the current results and help Arkansas policy makers confront diabetes prevalence at the county level in a targeted, systematic way.

An Exploration of Stress: Leveraging Online Data from Crowdsourcing Platforms

James Roddy[1], Samantha Robinson[1]

1-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Despite continued media coverage and public interest in those epidemics that dominate current health policy initiatives (e.g., opioid abuse, obesity, mental illness, etc.), stress is an often overlooked health risk factor. Stress, having an adverse effect on overall health, has been linked to substance abuse, sleep deprivation, obesity, and depression. Consequently, it is imperative to explore trends in the stress of both adults and children.

Current work utilizes a crowdsourced online sample recruited from Amazon Mechanical Turk (MTurk). As recommended to increase sample representation, survey data was collected from MTurk via TurkPrime with micro-batching and additional panel features enabled. Participants completed a demographic questionnaire and a 10-item Perceived Stress Scale (PSS) for themselves and one of their children.

Differences in item functionality for the 10 PSS items as well as differences in overall stress level, as measured by the average response on the full set of PSS items, were explored.

Rasch trees, a differential item functioning (DIF) method based upon model-based recursive partitioning, provided evidence that individual items on the PSS functioned differently based upon a self-reported rating of physical health.

Analysis of Variance (ANOVA) methods revealed that neither geographic location nor race had a significant effect on overall stress level. However, parent education level, employment status, and characteristics of social media usage did impact overall stress level.

These preliminary findings reveal features that contribute differentially to stress levels and, in leveraging survey data from a popular crowdsourcing platform, suggest that policy makers could utilize such platforms to generate cross-sectional snapshots of health risk efficiently and at low cost.

Prediction of Human Immunodeficiency Virus Infection Rates using Naïve Bayes

Olivia James[1], Kayla Simon[2], Samantha Robinson[3]

1-Department of Health, Human Performance and Recreation, University of Arkansas, Fayetteville, AR USA. 2-Department of Health, Human Performance and Recreation, University of Arkansas, Fayetteville, AR USA. 3-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
The current number of estimated cases of Human Immunodeficiency Virus (HIV) in the United States is 1.2 million. HIV is a virus that attacks one's immune system and eventually develops into Acquired Immunodeficiency Syndrome (AIDS). At this point in time, there is no cure for HIV/AIDS. Since its emergence in the 1980s, HIV has been a prevalent public health concern in the United States.

Data for this study consisted of merged state-level demographic data, 2017 HIV surveillance data from the Centers for Disease Control and Prevention (CDC), sexual risk data from the 2017 Youth Risk Behavior Surveillance System (YRBSS), and sexual education data from the Guttmacher Institute. The purpose of the current study was twofold: (1) create a classification rule to accurately predict high HIV infection rates at the state-level and (2) identify factors that contribute to HIV infection rates.

A Naïve Bayes classifier was shown to be significantly more accurate than the No Information Rate (NIR) in predicting states with high HIV infection rates, with 87% overall accuracy. Factors that contributed the most to HIV infection rates included the reported sexual activity in a state as well as various sexual education requirements (or lack thereof).

Findings indicate that sexual education requirements, which policy makers can adjust, are predictive of HIV infection rates. Further research would allow us to use these and future findings to target communities in need of public health interventions and to help fight the spread of HIV.

Twitter Sentiment at the Hospital and Patient Level as a Measure of Experience

Ellie Vicha[1], Samantha Robinson[2]

1-Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA. 2-Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR USA

Abstract:
Childhood cancer is the leading cause of death among children between the ages of birth to 19 years. One out of every eight children diagnosed with cancer will not survive and, with nearly 16,000 new diagnoses each year of children having an average age of six years old, a pediatric oncology unit is one of the most bittersweet places to be. While there is still work to be done to improve survival rates among children diagnosed with cancer, there is also a need to understand the patient experience.

Interest in the analysis of patient-focused measures, including patient-reported experience measures (PREMs), has grown in recent years as attempts are made to understand and assess latent (i.e., unobserved) traits such as quality of life and mental health. While most PREMs involve lengthy survey instruments and questionnaires, posts on social media platforms might serve as more authentic measures of patient reported experience while admitted in a pediatric facility.

Using data extracted from Twitter, sentiment analysis was used to compare the attitudes, perceptions, and overall impressions of patients (or their parents) with St. Jude experiences to those with more general childhood cancer experiences. Results indicated that patients reflected on their care with more relative negativity when speaking of experiences other than St. Jude. While there were some notable differences between the two comparison groups, the sentiment of both was still mostly positive. Pediatric oncology differs greatly from adult oncological care and, thus, this sentiment analysis of patient social media posts is likely to differentiate even more between medical providers when assessing adult patient experience. This study suggests that physicians and anyone involved in the care of patients can use this type of authentic, real-time feedback to improve care not only for pediatric oncology patients, but for all patients seeking care in any specialty.

Evaluating Liver Function After Treatment with an AAV-delivered Antibody for the Treatment of Methamphetamine & MDMA Abuse

James Cameron Tilley[1], Chris T. Bolden[2], Kennede McLeroy-Charles[2], Charles E. Hay[2], Eric Peterson[2]

1-Department of Biological Sciences, University of Arkansas, Fayetteville, AR USA. 2-Department of Pharmacology and Toxicology, University of Arkansas for Medical Sciences, Little Rock, AR USA

Abstract:
Methamphetamine (METH) and 3,4-Methylenedioxymethamphetamine (MDMA) substance abuse have greatly impacted society. To date there are no medications proven successful at preventing overdose and dependence on METH or MDMA. Anti-METH monoclonal antibodies can quell METH accumulation within the brain. MDMA has a similar molecular structure that allows a similar result. These antibodies do have restrictions, one major constraint is the multiple injections needed to allow recovery from addiction to become a possibility. An adeno-associated viral (AAV)-delivered DNA sequence for a single-chain variable fragment could offer long-term, continuous expression of anti-METH antibody fragments. For these studies, we injected mice via tail vein with one of two AAV8 scFv constructs and measured long-term expression of the antibody fragments. The scFvs decreased concentrations of METH and MDMA in the brain. These results suggest that AAV-delivered scFv could be a promising therapy to treat METH and MDMA abuse and addiction. Body temperature was measured throughout the experiments to monitor hyperthermia. Hyperthermia is caused by many stimulating drugs such as the ones administered in this study. Individuals receiving AAV antibody treatment saw shielding from hyperthermia, while those without treatment did see an increase in temperature. Data was also collected post sacrifice of the mice. This data included blood and tissue samples to test ALT and Bilirubin levels. The purpose being to see any possible damage the treatment may have had on the liver.

Electrochemically doped Halide Perovskites for Simple Structure Electro-Optical Devices

Huddoy Walters[1], Jintian Jiang[1] and Qinglong Jiang[1]

1-Department of Chemistry and Physics, University of Arkansas, Pine Bluff, 71603

Abstract:
In less than 10 years, organic-inorganic halide perovskite materials have been used in solid state solar cells (over 23% efficiency). The use and application of halide perovskites are found in areas such as light emission, photoluminescence, solar cells, sensors and more. Halide perovskite materials can be synthesized by wet chemistry. Doping can be achieved by an electrochemical method. Halide perovskite semiconductors $MPbX_3$ (M= Cs^+, $CH_3NH_3^+$ etc.; X=Br^-, Cl^-, I^-, SCN^- etc.) have found specific applications in solar cells, LED and sensors. This project it involves the synthesis and electrochemical doping of halide perovskite materials and nano materials for solar cell and light emission. We explored electrochemical doping of cesium lead tribromide ($CsPbBr_3$).

Ref:
1. Qinglong Jiang, Xiaoqiao Zeng, Ning Wang,* Zhanhu Guo,* Zhili Xiao, Jun Lu,* ACS Energy Lett., 2018, 3, 264–269.
2. Qinglong Jiang, Mingming Chen, Junqiang Li, Qiangfeng Xiao, Xiaoqiao Zeng, Tiglet Besara, Jun Lu, Yan Xin, Bicai Pan, Changchun Wang, Theo Siegrist, Zhibin Yu,*ACS Nano, 2017, 11, 1073–1079.
3. Ying Guo, Tao Liu, Ning Wang,* Qiang Luo, Hong Lin, Jianbao, Li, Qinglong Jiang,* Lili Wu, Zhanhu Guo,*Nano Energy, 2017, 38, 193-200.
4. Qinglong Jiang, Dominic Rebollar, Jue Gong, Elettra Piacentino, Chong Zheng, Tao Xu,*Angew. Chem. Int. Edit. 2015, 54, 7617-7620, VIP.
5. Junqiang Li, Xin Shan, Sri Ganesh R. Bade, Thomas Geske, Qinglong Jiang, Xin Yang, Zhibin Yu,* J. Phys. Chem. Lett. 2016, 7, 4059–4066,
6. Qinglong Jiang, Xia Sheng, Xinjian Feng, Tao Xu,* Chem. Commun. 2014, 50, 14720-14723, inside cover page.
7. Qinglong Jiang, Xia Sheng, Bing Shi, Xinjian Feng, Tao Xu,* J. Phy. Chem. C, 2014, 118, 25878-25883.

Structural Analysis of the Ge_xSn_{1-x} Alloy Thin Film for Optoelectronics Applications

Samantha Gibson[1], Kayleigh Johnson[1], Fatma Merve Yurtsever[1], Aboozar (Salman) Mosleh (P.I.)[1], Seyed Amir Ghetmiri[*1], Mansour Mortazavi[1]

1-Chemistry & Physics department, University of Arkansas at Pine Bluff, Pine Bluff, AR, USA

Abstract:
Si is the material that is the basis of all electronic devices and electronic industry. Next generation industry requires optoelectronic materials to be incorporated in electronics to help the industry overcome reaching the physical limits of size reduction in electronic devices. Silicon is widely used for fabrication of photodetectors as well, however, the light generation is not successful because silicon has an indirect bandgap. In this research, we are investigating other group IV alloys that are having direct bandgap and better optical properties such as Germanium Tin (GeSn). These materials are studied because they could be produced in a cheaper method to match consumer and supplier demands. Raman spectroscopy is being used to analyze the incorporation of Sn in the Ge lattice to form GeSn crystal. Different peaks in the Raman spectroscopy are studied to show uniform incorporation of Sn in order to determine if the materials are suitable for incorporation into optoelectronic devices.

The authors acknowledge the Air Force Office of Scientific Research (AFOSR) (grant Nos. FA9550-14-1-0205 and FA9550-18-1- 0361), and National Science Foundation Established Program to Stimulate Competitive Research (NSF EPSCoR) (grant No. OIA-1457888) and the Arkansas EPSCoR Program, ASSET III.

Precision Oncology Initiatives in Arkansas

Donald J. Johann, Jr.[1], Mathew Steliga[1], Jason Muesse[1], Katy Marino[1], Ik Jae Shin[1], Sergio Pina Oviedo[1], Sanjay Maraboyina[1], Nikhil Meena[1], Mary Yang[2], Karl Walker[3], Andy Pereira[4], Erich Peterson[1], Jason Liem[1], Owen Stephens[1], Joshua Xu[5], Weida Tong[5]

[1]University of Arkansas for Medical Sciences, Little Rock, AR, [2]University of Arkansas at Little Rock, Little Rock, AR, [3]University of Arkansas at Pine Bluff, Pine Bluff, AR, [4]University of Arkansas, Fayetteville, AR, [5]National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR

Background:

The 21st Century Cures Act calls for the development of more sensitive diagnostic tests for cancer. Its intent is to support research with the potential to transform the scientific field, that has inherently higher risk, and that seeks to address major challenges related to cancer. Blood-based testing for cell free DNA (cfDNA) for oncology applications and particularly lung cancer, are examples of this type of research, and are under rapid scientific development and regulatory evaluation.

Liquid biopsy methodologies allow for plasma genotyping of solid tumors via cfDNA, and are an application of deep next generation sequencing (NGS). These assays are rapidly entering the clinical domain of research-based monitoring in translational oncology, especially for thoracic malignancies, where they offer a much safer alternative to traditional tissue-based biopsies. Potential applications for these cfDNA assays include: i) initial diagnosis, ii) response to therapy and follow-up, iii) tumor evolution, and iv) minimal residual disease evaluation.

Results:

A number of national and local endeavors involve AR-BIC members. Via the Sequencing Quality Control Phase 2 (SEQC2) project, standard analysis protocols and quality control metrics for fit-for-purpose use of NGS data to enhance regulatory science research and precision medicine are in progress. The project consists of three specific aims: (1) to develop quality metrics for reproducible NGS results from both whole genome sequencing (WGS) and targeted gene sequencing (TGS), (2) to benchmark bioinformatics methods for WGS and TGS towards the development of standard data analysis protocols, and (3) to assess the joint effects of key parameters affecting NGS results and interpretation for clinical application.

The UAMS IRB approved an advanced lung cancer clinical trial and patient enrollment is in progress. An assortment of liquid biopsy methodologies are in development. A variety of novel and complex bioinformatic approaches, designed to address the needs of the clinical trial, are developed and continue to be enhanced. Solid tumor material and cfDNA from routine blood draws are being subjected to molecular profiling with NGS and ddPCR with results being compared, contrasted, and integrated.

As a BloodPAC member, standards and best practices are being developed for liquid biopsy assays. First, minimal technical data elements have been specified and approved by FDA (CDRH) regarding preanalytical steps or everything related to a sample before any assay is run. Second, a generic analytical variable protocol is in progress to aid in the development of liquid biopsy assays following best practices. The aim is to provide a detailed starting point for assay developers.

Conclusions:

In order to improve outcomes for cancer patients, there is an urgent need for advanced translational research involving: i) innovative assays, and ii) molecular profiling with cutting-edge bioinformatics, and iii) advanced clinical trials. This is all happening in Arkansas.