UF Research AI Catalyst Fund Awardees 2020

PI LAST NAME	PI FIRST NAME	TITLE	COLLEGE
Azad	Hassan	OR-DRD-AI2020: Application of Machine Learning in the Prediction and Modeling	DCP

Co-PI: Ian Flood

Environmental noise is a critical pollutant directly associated with health and work-related problems such as hearing loss, children's cognition problems, and reduced work productivity. Estimation methods have been used to calculate and model environmental noise since around 40 years ago. Over the past few years and with the help of high-performance and parallel (GPU) computing, physics-based models have been developed for environmental noise problems in larger scales. These numerical modeling techniques are computationally expensive and time consuming. The primary goal of this research is to use machine learning to replace physics-based solutions to the problem of environmental noise modeling with its AI-generated counterpart. It is anticipated that the outcomes of this research have long-term impacts on 1) helping urban and disaster planners to evaluate their design scenarios and city policy makers to adjust their plans with respect to the harmful effects of environmental noise using the prediction noise models 2) proposing integrated and optimized solutions to the idea of smart and connected communities and civic infrastructures 3) creating a healthier and sustained built environment, and consequently 4) elevating the overall human life experience.

Barmpoutis	Angelos	AI-driven Movement Classification and Analysis across Clinical and Cultural Application Areas	COTA

Co-PIs: Jill Sonke, Osubi Craig, James Oliverio, Markus Santoso, Timothy Difato

An interdisciplinary team from the Digital Worlds Institute, the Center for Arts in Medicine, and the Center for the Arts, Migration, and Entrepreneurship proposes in this project to automate Laban analysis of human movement using artificial intelligence. Human movement has been studied in multiple disciplines, including health sciences (biomechanics, kinesiology, neurology, sports medicine) and the Arts (theater and dance, cultural studies) as well as their intersection (arts in medicine, dance therapy), resulting in a large but disparate assortment of multi-modal datasets, including video, skeletal motion capture, manual annotations, and clinical metadata. Traditional data collection processes often include Laban movement analysis, a standardized form of human movement annotation that parameterizes the observed motion changes in a pre-defined 4-dimensional feature space (effort, space, shape, body timing/phrasing). Such analysis requires lengthy manual input from professionals who annotate the recorded data through a time-consuming "watch and pause" process, which is also prone to human errors. In this project, we propose to use AI to fully automate the annotation process involved in Laban analysis by using deep learning algorithms on existing human motion datasets of video and skeletal sequences. The trained model will then be tested in Laban-annotating existing video and skeletal sequences and validated by arts in medicine practitioners and experts in Laban analysis. This AI-driven project will have a significant impact as it will enable automated classification and understanding of human motion across a spectrum of movement-centered disciplines, including clinical and telehealth settings, orthopedic centers, choreographic practice, and cross-cultural movement analysis.

Chen	Huan	Fairness in Information Access through Culturally Competent AI Systems	CJC

Co-PIs: Sylvia Chan-Olmsted, My Thai

Organizations and public agencies often disseminate information uniformly at scale without any cultural considerations, leading to unfairness in information access for certain social groups due to their inherent, unique cultural backgrounds. By integrating machine learning and social science theories to address bias and inequality with both technical solutions and realistic assumptions of social behavior, our proposed project is aiming at 1) identifying and validating cultural features that influence social groups information access needs and receptivity; and, 2) building a culturally aware machine leaning (ML) system for information dissemination purposes. The proposed project will facilitate certain social groups' information access in specific contexts, resulting in better information receptivity because of cultural resonance. More importantly, it will help to build a large-scale fairer information access system via culturally competent AI mediated model, advancing the research fronts of fairness in ML where we not only exclude the bias features but also promote the cultural sensitivity features.

Fang	Ruogu	VCA-DNN: Neuroscience-Inspired Artificial Intelligence for Visual Emotion Recognition	HWCOE

Co-PI: Mingzhou Ding
From the profound grief of losing a loved one to the thrilling joy of welcoming a new life into the world, human emotion permeates everyday life and influences our psychological well-being, as well as our ability to adapt and survive in an ever-changing environment. Scientifically, human emotions are dynamic, multidimensional responses to challenges and opportunities, which emerge from network interactions in the brain. Disruptions of these network interactions underlie emotional dysregulation in many mental disorders including anxiety and depression. Recognizing the limitations of empirical approaches to the study of human emotion, in this project, we seek to create an AI-based model system that is informed and validated by known biological findings and that can be used to complement empirical studies to better understand how the human brain processes emotional information. In the long term, the model will be further enriched and expanded so that it becomes a platform for testing a wider range of normal brain functions, as well as a platform for testing how various pathologies affect these functions in mental disorders.

Glenn	Alina	Parasitic Nematode Identification with Deep Learning	HWCOE

Co-PI: Peter DiGennaro
Nematodes are the most abundant and ubiquitous animals on Earth. Of crucial import are nematodes that infect animals/humans and plants. Worldwide, nematode infections affect over 50% of all humans, causing severe illness and hundreds of thousands of deaths (Chan 1997; Horton 2003). Plant-parasitic nematodes (PPN) are one of the most important pests of all major cultivated crops, responsible for over 125 billion dollars in yield losses annually (Chitwood, 2003). Identifying nematode species and abundances allows for informed management recommendations and is a platform for the detection of new and emerging pests that can direct quarantine and sanitation procedures. Developing effective management strategies begins with the critical step of identifying the pests present, and their destructive potential to human health and the economy. Accurate and rapid nematode identification is paramount to informing critical management decisions in human/animal health and agriculture. In this work, we will leverage the large number of samples sent to the UF Nematode Assay Lab to annotate microscopic images of nematodes important to agriculture. Using these samples, a deep algorithm will be developed to automate the differentiation between species and allow for interaction with expert to continually improve the approach over time.

Huffaker	Ray	Symbiosis of machine learning, nonlinear time series analysis, and novel supercomputing to reconstruct soil-biome nonlinear dynamics from field and remote-sensing large data	IFAS

Co-PIs: Rafael Munoz-Carpena, Kati Migliaccio
Many important components of biophysical systems are inherently non-linear and often appear to behave erratically. As a result, they are notoriously difficult to characterize and predict. One such system is the soil, which mediates Earth's water, mass and energy exchanges, and provides essential life services like natural waste treatment, food production, and biodiversity. Soil is living matter that constantly changes in response to complex abiotic and biotic interactions. Converting new satellite big-data products into knowledge of real-world soil dynamics remains a challenge. AI offers a promising computation approach for understanding complex soil dynamics that have remained undetected by past empirical approaches. A recent proof-of-concept demonstrates that new types of AI (Echo State Networks or "reservoir computing" -RC) can learn complex nonlinear dynamics from clean data generated by closed 'toy' models with known covariates, causal interactions, and dynamics. However, in modeling open real-world biophysical systems like the soil, we do not know all covariates involved or whether/how they interact. We do not directly observe dynamics, and must

infer them from noisy data. We will address these challenges by preprocessing data inputted into RC with nonlinear time series analysis (NLTS) to filter noise and test for stationarity, detect causal interactions, and reconstruct state-space dynamics from denoised data. This increases the dynamic efficiency of RC since reservoirs learn complex dynamics of stationary real-world dynamic systems from denoised data taken from causally-interactive covariates. We will employ the new UF-NVIDIA-GPU infrastructure to automate a large parallel grid search of RC hyper-parameters, and screen for architectures learning dynamics reconstructed with NLTS. We will deliver a fully-automated RC-NLTS graphical user interface (GUI) software that will serve as reliable soil 'digital-twin' to analyze remote-sensing data provided by previous UF research trials, and the long-term wireless sensor network SoilNet at the Wüstebach test site, Jülich Research Center (Germany). Operationalizing AI RC for real biophysical data has broader impacts in addressing pressing environmental challenges involving nonlinear biophysical processes and creating AI-based early-warning systems for quick prediction of extreme environmental events (flooding, drought, landslides, fires).

Liu	Hongcheng	A New Stochastic Gradient Algorithmic Paradigm for Training Massive AI Models in Network-Wide Traffic Anomaly Warning	HWCOE

Co-PI: Lili Du
While the mainstream AIs are increasingly larger and more powerful, their scalability and accessibility are frequently hindered by the current limitations in the AI's training algorithms. The state-of-the-art optimization theory stipulates that the amount of computational effort taken by these training algorithms would be sensitive to the growth of the size of the AI model. As it is nowadays common to see AI models with thousands of millions and even billions of fitting parameters, training such large-scale models stretches the existing optimization paradigms. The above issue is particularly looming in some recent AI applications to traffic management. To enable timely and proactive traffic control for mitigating traffic jams, early predicting traffic anomalies prior to a traffic-influencing public event (TIE), e.g., sports and ceremonies, has been made possible through AIs. However, these AI models tend to be cumbersome to train, especially for the ones targeted on serving a large traffic network. On this challenge, this research is aimed at the derivation and verification of a novel stochastic gradient descent (N-SGD) algorithm that exploits the problem structures within a traffic model. Our conjecture is that the N-SGP's computational complexity will be provably insensitive to the growth of the problem dimensionality, improving upon the traditional AI training schemes significantly.

Li	Chenglong	A Machine Learning Approach to Drug Hit Optimization	COP

Co-PI: Adrian Roitberg
It is estimated that the process of bringing a single new drug to market can take up to 10 years, and cost billion dollars. The discovery process starts with the identification of a hit compound, a new molecule with promising activity against some disease target, usually by screening a large chemical database with experimental and computational techniques. This is then followed by the hit optimization phase, in which researchers attempt modifications of the initial structure to maximize the activity and minimize possible side effects. At this stage, besides activity, many properties need to be optimized at the same time such as absorption, distribution, and toxicity, making it a very complex and difficult task. It normally takes multiple cycles of design, synthesis, and experimental testing, resulting in the most expensive and laborious stage in drug discovery. Still, many drug candidates still fail approval due to unanticipated effects, and the whole process needs to be restarted. We are developing a new system that will harness the power of Artificial Intelligence to speed up this phase. Using AI, the computers will learn to suggest new molecules while optimizing many variables at the same time, generating new drug suggestions which already incorporate the desired traits. This system will greatly reduce the number of optimization cycles needed to reach a new candidate with greater chances for regulatory approval. It should help make the drug discovery process more efficient, significantly reducing the time and cost for a new drug to reach the market.

<table>
<tr><td>Lu</td><td>Qing</td><td>A Kernel Neural Network for High-dimensional Genomic Risk Prediction</td><td>PHHP</td></tr>
<tr><td colspan="4">Co-PIs: Yan Gong, Xiaoxi Shen
Risk prediction capitalizing on emerging human genome findings holds great promise for new prediction and prevention strategies. While the large amounts of genomic data generated from high-throughput technologies offer us a unique opportunity to study the role of a deep catalog of variants in risk prediction, the high-dimensionality of genomic data and massive data bring tremendous challenges. Advanced tools are in great need to address the analytical and computational challenges from ongoing genomic risk prediction. The goals of the application are to develop a kernel-based neural network (KNN) for high-dimensional risk prediction on large-scale genome data and to use it in the collaborative research of studying a coronary artery disease (CAD) prediction model. While neural networks (NN) have been well developed in many areas (e.g., computer vision), there is almost no NN-based method specially developed for high-dimensional genomic risk prediction. The proposed research will fulfill this gap by developing one of the first NN-based tools that is capable of handling millions of genetic variables, is computationally efficient for large-scale datasets, and is able to consider complex genotype-phenotype relationships. The application of KNN to a large-scale UK Biobank (UKB) dataset will also lead to a novel CAD risk prediction model that could be further investigated for clinical use.</td></tr>
<tr><td>Nino</td><td>Juan Claudio</td><td>AI-Enabled Imaging Biomarker Identification for Early Detection and Treatment of Alzheimer's Disease</td><td>HWCOE</td></tr>
<tr><td colspan="4">Co-PI: Marcelo Febo
In this project we are interested in using mathematical algorithms run on the advanced computers of HiperGator 3.0 to identify biomarkers or early signals of Alzheimer's Disease (AD). We will do so by extracting information from nuclear magnetic resonance imaging (MRI) and then trying to track and understand the evolution of the connection between the neurons in the brain (neural network) as the disease progresses. This will allow us to first identify key connections or areas of the brain that are most prominently affected as a result of the disease and therefore be able to identify or formulate therapy or treatment of AD. In addition, using artificial intelligence analysis techniques such as deep learning for image classification, we anticipate being able to provide early detection and propensity for AD from a MRI scan. In combination with established protocols for treating neurodegenerative diseases, we expect to develop key building blocks towards a comprehensive disease progression model for AD diagnosis, prognosis, and treatment</td></tr>
<tr><td>Odegaard</td><td>Brian</td><td>Combining Deep Neural Networks and Large-Scale Brain Data to Predict Human Cognition and Behavior</td><td>CLAS</td></tr>
<tr><td colspan="4">Co-PIs: Natalie Ebner, Yunmei Chen, Karim Oweiss
In the last decade, Deep Neural Network models (DNNs) have facilitated remarkable advances in brain decoding, revealing how distinct patterns of neural activity correspond to different thoughts, sensations, and behaviors. From brain activity alone, these models can predict what images a person is seeing or imagining in each moment, current states of pain, specific auditory sensations, and many other mental states. When trained on sufficiently large data, they can not only be used to decode current mental states, but also predict specific phenotypes (e.g., average weekly alcohol intake, fluid intelligence, etc.). In this investigation, we will combine DNNs with a "big data" approach to decode cognitive and behavioral phenotypes from existing brain data samples collected at UF. Specifically, by leveraging the computing power of the HiPerGator supercomputer, we will build predictive models from large, openly accessible MRI/fMRI datasets and then apply these algorithms to predict cognitive and behavioral phenotypes from specific populations of interest (e.g., aging individuals). This work will (i) facilitate development of (and make publicly available) a new research algorithm for deep learning; (ii) provide insights into how brain structure and function are linked to different phenotypes; and (iii) identify brain regions of interest for future targeted interventions, such as using neurofeedback training to enhance cognition and behavior.</td></tr>
</table>

Ozrazgat Baslanti	Tezcan	The Artificial Intelligence learns optimal treatment strategies for hypotension in surgery.	MED-GNV

Co-PIs: Tyler Loftus, Crystal Johnson-Mann

In the United States, where the average American can expect to undergo seven surgical operations during a lifetime, each year 1.5 million patients develop a medical complication. Intraoperative hypotension is a well-recognized modifiable risk factor for postoperative complications. The pathophysiology of postoperative complications is influenced by patients' preexisting health, the type of surgery, the management of intraoperative events and especially those events relating to hypotension. Current practices in the administration of intravenous fluids and vasopressors during surgery as a treatment of hypotension are suboptimal. Patients and physicians make essential decisions on which diagnostic and therapeutic interventions should be performed or deferred under time constraints and uncertainty regarding patients' diagnoses and predicted response to treatment which may lead to cognitive and judgment errors. Reinforcement learning have the potential to play complementary roles in delivering high-value surgical care through sound judgment and optimal decision-making. Reinforcement learning is a subfield of machine learning that identifies a sequence of actions to increase the probability of achieving a predetermined goal. The algorithm mimics human trial-and-error learning process to calculate optimum recommendation policies. We propose use of highly granular electronic health record from adult hospitalized patients undergoing surgery to determine optimal treatments for hypotension. We aim to develop and validate a deep reinforcement learning model that would provide individualized and clinically interpretable treatment decisions that could help balance blood pressure during surgery and decrease postoperative complication risks in patients undergoing surgery. Precise data-driven artificial intelligence augmented identification of optimal treatment strategies in the highly dynamic environment could assist clinicians in timely decisions to guide treatment and improve patient outcomes.

Saxena	Shreya	Artificial neural networks meet biological neural networks: designing personalized stimulation for the data-driven control of neural dynamics	HWCOE

Co-PIs: Marcelo Febo, Yong-Kyu Yoon

Personalized neurostimulation using data-driven models has enormous potential to restore neural activity towards health. However, the inference of individualized high-dimensional models and the design of stimulation strategies from data remains challenging due to their under-constrained nature. To address these issues, PI Dr. Shreya Saxena in the Department of Electrical and Computer Engineering will leverage the wealth of data that multi-subject experiments provide, as well as the computational resources newly available at UF to develop novel AI algorithms. This work will be performed in collaboration with Co-PI Dr. Marcelo Febo in the Departments of Psychiatry and Neuroscience, who is an expert in functional Magnetic Resonance Imaging (fMRI) to examine neural activity related to memory and cognition. Promising stimulation strategies will be validated *in-vivo* using an fMRI-compatible neural probe in collaboration with Co-PI Dr. Yong-Kyu Yoon in the Department of Electrical and Computer Engineering, who is an expert in the fabrication of biomedical devices. Together, the team will develop and validate new AI methods for performing data-driven control for high-dimensional neural data: a novel approach that can have an enormous impact on neurostimulation applications for diseased systems.

Stucky	Brian	Using AI to uncover decades of global ecological change	FLMNH

Co-PIs: Robert Guralnick, Pamela Soltis, Douglas Soltis, José Fortes, Matt Gitzendanner

Plant phenology -- the seasonal timing of when plants develop leaves, flowers and fruit -- is highly sensitive to environmental change. Plant phenological data are therefore essential for understanding and forecasting ecosystem responses to climate and land use changes. We will use a suite of artificial intelligence techniques to extract and mobilize new phenological information from the millions of historical plant specimen images currently available at UF through the iDigBio specimen digitization project (https://www.idigbio.org/). Data from this project will provide vital historical baselines for forecasting long-term ecological and biodiversity trends, with impacts on both conservation and agriculture. This work will also establish a foundation for building new research infrastructure for generating, integrating, and delivering plant phenological data.

Subhash	Ghatu	AI-assisted accelerated discovery of novel materials for ballistic applications	HWCOE

Co-PIs: Richard Hennig, Salil Bavdekar
Exploring and understanding the relationship between the material structure at the atomic scale and its influence on the material properties at different stages of processing and performance is a key issue in material science research. The "big data" generated by previous experimental and computational studies has spawned new avenues for this research through material informatics, which can accelerate the discovery and development of novel materials. With the support of UF Catalyst Fund, we will conduct a pilot study to lay foundation for identifying new material chemistries and structures, specific to ballistic applications, by leveraging AI-based techniques. In the 'exploitation' phase of the study, we aim to develop chemical design rules in a smaller, well-understood, design space of the Zr-Cr-Si-B-C-O system by targeting new stoichiometries and structures with promising ballistic properties. Genetic algorithms and machine learning models will be employed to learn the experimentally measured hardness of the predicted structures from atomic parameters and answer the fundamental question of how chemical additions to the structure affect its hardness and stability. These approaches will then be extended to the 'exploration' phase, where a larger design space consisting of hard metals (W, Cr, Ti, Zr, Hf, etc.) and light elements (B, C, N, O, Al, Si) will be considered. In this phase, statistical learning and deep learning methods will be used to rapidly identify pathways to navigate and uncover hidden relationships between material structure, properties, and performance utilizing both experimental and computational approaches. This pilot study will kickstart a long-term AI-driven research that facilitates accelerated discovery of new materials with the ultimate goal of creating an AI-based material design strategy to tailor material microstructure for targeted applications in numerous fields (nuclear, refractory, energy storage, etc.).

Tepe	Emre	Spatio-Temporal Modeling of Land-Use Changes Using Big Data	DCP

Co-PI: Abolfazl Safikhani
Understanding human activities forming present and past land use patterns is essential to predict future land developments. Extractions of complex relationships involved in land developments are challenging. In this research project, computationally feasible methodological approaches will be developed using Artificial Intelligence and Machine Learning methods in order to track past and present land use patterns in Florida. The developed modeling approaches can be used to simulate future impacts of anticipated changes in land developments. Also, these methods can be easily applied to a wide spectrum of research areas such as: hedonic pricing, traffic forecasting, urban energy consumption, ecological systems, disease spread, trade relations, and business network.

Triplett	Eric	Deep learning prediction of autoimmune disease from early childhood gut microbiome composition.	IFAS

Co-PIs: Bryan Kolaczkowski, Desmond Schatz, Mark Atkinson
The proportion of the US population afflicted with autoimmune diseases, such as type 1 diabetes and celiac disease, continues to increase each year. The annual cost of managing just two of these diseases, type 1 diabetes and celiac disease, in this country is about $34 billion. Complications that can arise from these diseases drive this cost much higher. A simple means for early life prediction of autoimmune diseases would allow physicians and parents time to plot of course for prevention. Our team recently discovered gut bacteria in children at one year of age that are associated with autoimmune diseases that occur up to 16 years later. In this project, we will employ UF's new AI computing power artificial to our current datasets to obtain an accurate a prediction of future autoimmune disease in one-year old children. In our early efforts, gut microbiome data alone can predict future autoimmune disease in the study group of 1800 children with 74% accuracy. This work will improve that

prediction accuracy substantially by: 1) performing a wider array of analyses on the new UF AI computing infrastructure; 2) adding more data to the model including diet, probiotic use, prebiotic use, and infectious episodes during the first year of life; and 3) increasing the datasets through other analyses. Discovering those parameters associated with future autoimmune diseases are expected to suggest interventions that may prevent disease.

Xing	Wanli	Fair AI Responding to Online Education "FAIR_EDU"	Education

Co-PIs: Bahar Basim
Online learning raises methodological difficulties for instructors as they work to identify academically at-risk students remotely and provide in-time interventions. The emerging fields of artificial intelligence in education (AIED) and learning analytics (LA) provide a promising solution as researchers rely on machine learning (ML) techniques to analyze big data to automatically infer students' learning status. These constructed ML models predict students at risk of dropout or failure in advance for an online course to support effective intervention design. However, limited attention has been given to the fairness of prediction with ML in educational settings, which can produce and reinforce inequalities across groups. Using a popular online learning platform as a testbed, the goal of this study is to build ML models by leveraging the GPU computing infrastructure that can (1) fairly predict students' performance without being biased with demographic factors; (2) examine the robustness and transferability of the fair prediction model to other online learning platforms. Results from this project will contribute to the scientific understanding of the effects of applying ML for education bias reduction.

Weisberg	Steven	What mind matters? Machine learning approaches to linking structural variation in the brain to individual differences in spatial behavior	CLAS

Co-PIs: Alina Zare
A foundational challenge in understanding the human brain is how macroscopic neural structure can be associated with cognitive function. Exploiting advances in AI and machine-learning in this area, we propose a data-driven approach to close this gap. Spatial navigation is a widely-studied problem in this area, in part because spatial navigation varies drastically across people, despite being an essential task everyone must solve. A prominent hypothesis about the neural instantiation of spatial navigation is that the cognitive map is constructed by the hippocampus, a structure in the medial temporal lobe of the human brain. But evidence on whether hippocampal volume relates to variability in navigation behavior is mixed. Expert navigators and patients with lesions to or atrophy in the temporal lobes support the hypothesis, but two recent experiments with large sample sizes of healthy young adults do not. Here, we will apply a data-driven machine learning approach interrogating those datasets to 1) determine which regions of the brain covary with navigation ability and which structural properties of those regions that predict variability; and 2) create a pipeline for use with other structural MRI datasets for improved exploratory data analysis in neuroimaging. In the short-term, this research will provide evidence in favor (or against) the structure-function hypothesis - a critical question in human cognitive neuroscience. In the longer term, this research will promote the development of machine learning strategies that can be applied toward general cognitive traits, measured in different ways, across different populations. Such an approach could unlock exciting new avenues for disease diagnosis and prediction, new targets for clinical intervention, and new theories about brain-behavior associations.

Zhao	Xilei	Real-Time Management of Micromobility Services for Smart Cities	HWCOE

Co-PIs: Xu Sun, Yu Yang
In 2019, micromobility, i.e., the small, lightweight vehicles such as dockless e-scooters, emerged on the urban streets of dozens of cities almost overnight. These devices are flexible, convenient, fun to use, and especially attractive for serving short-distance trips. In the U.S., half of all personal trips are no longer than three miles, and more than 70 percent of these trips are made by private vehicle. If these short-distance driving trips were made by micromobility instead, streets will be much less crowded and greenhouse gas emissions will be

significantly reduced. To promote the modal switching from personal vehicles to micromobility, we propose to design a highly efficient micromobility system that operates optimally in real-time and offers the best service to urban travelers by leveraging state-of-the-art AI technologies. Under this scope, we have two specific objectives. First, we aim to forecast spatio-temporal micromobility demand using deep learning algorithms. Second, we will integrate the demand-forecasting model with the rebalancing optimization model to provide real-time reliable near-optimal solutions to managing a large fleet of micromobiltiy devices. This research leverages robust AI for improving our understanding and modeling of micromobility systems for smart cities and leads to a possible leap in advancing our knowledge of real-time management of micromobility services.