The National Academies of SCIENCES ENGINEERING MEDICINE

Government-University-Industry Research Roundtable

AIFS
AI INSTITUTE FOR FOOD SYSTEMS

Ilias Tagkopoulos, Ph.D.

Director, AI Institute for Next Generation Food Systems
Professor, Computer Science and Genome Center,
University of California, Davis

US Food System: History

Industrialization
&
Globalization
(19th Century)

Agrochemicals, Nutrients &
Monocultures
(20th Century)

Precision Agriculture
(21st Century)

The Food 3.0 (R)Evolution

Enablers
- Sensors And Devices
- AI, Data And Analytics
- Connected Global Markets
- Consumer Demand For Personalization
- Global Food Safety And Sustainability Concerns

Barriers
- Gap Between Tech And Ag Sector
- Data, Software And Hardware Infrastructure
- Innovation Gap
- Access To Markets
- Education And Outreach

US Food System: Fact Sheet

WORKFORCE

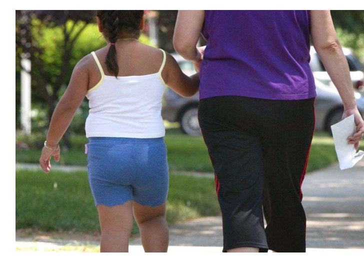
- Farmers account for **1%** of the US population, with roughly 1/3 of them over the age of 55.
- About **one third to half** of the hired agricultural labor force **lacks authorization** to work in the US (48%, 2014-2016).
- **60%** of production ($) comes from **5%** of large-scale family farms.

ENVIRONMENT

- Total **cropland has decreased by 15% over the past 20 years** (460ma – 392ma between 1992-2012). Soil erosion and Groundwater depletion are also an issue.
- **Pesticide and herbicide** use **increased 10-20% over 5 years** (899 pounds, 2012). Most corn, cotton, and soybean varieties (>90%) are GMOs.
- Agriculture is responsible for **9.3% of total U.S. greenhouse gas emissions** (2018). Livestock and soil management are key factors.
- About a **third of the food supply is lost** (EPA, 2010), accounting for 15% of municipal solid waste, 2% of energy use.

HEALTH

- **40%** of U.S. adults, **18%** of 12-19 y.o. are obese (BMI>30).
- **One in eight** families in America are hungry (12.3%). World-wide, 820m people go hungry and 680m are obese (FAO, 2018, $2T annually).



AIFS:
The AI Institute for Next Generation Food Systems

AIFS MISSION: Develop and leverage transformative AI for the ethical production and distribution of **safe, sustainable, nutritious food** with fewer resources.

AIFS Partnerships

Partner Organizations

- Research I Universities
- Land-grant, Agricultural Schools
- Strong in AI and Computational Science
- Direct connection to Cooperative Extension offices
- USDA research programs

Team Make-up

- Diversity across career stages, gender, ethnicity
- History of collaboration and multidisciplinary approach
- Shared vision

Individual Research Strengths

- Breadth of the food system
- AI and computational science
- Education, workforce development, knowledge transfer

UC DAVIS UNIVERSITY OF CALIFORNIA

CORNELL UNIVERSITY

Berkeley UNIVERSITY OF CALIFORNIA

ILLINOIS

USDA ars Agricultural Research Service

UNIVERSITY OF CALIFORNIA Agriculture and Natural Resources

$20M Funding by USDA for 5 years – one of the 7 AI Institutes funded by NSF/USDA

AIFS Approach

A holistic, integrated view of the food system…

Molecular Breeding

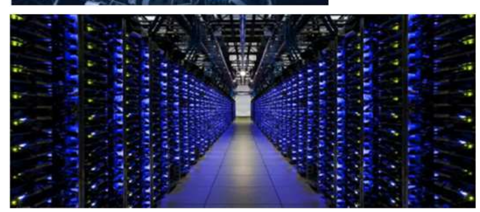
Agricultural Production

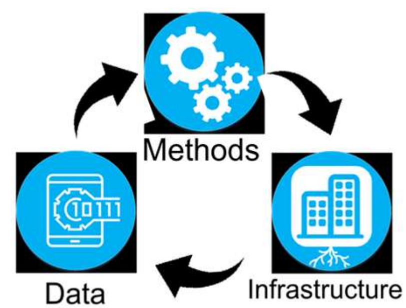
Food Processing & Distribution

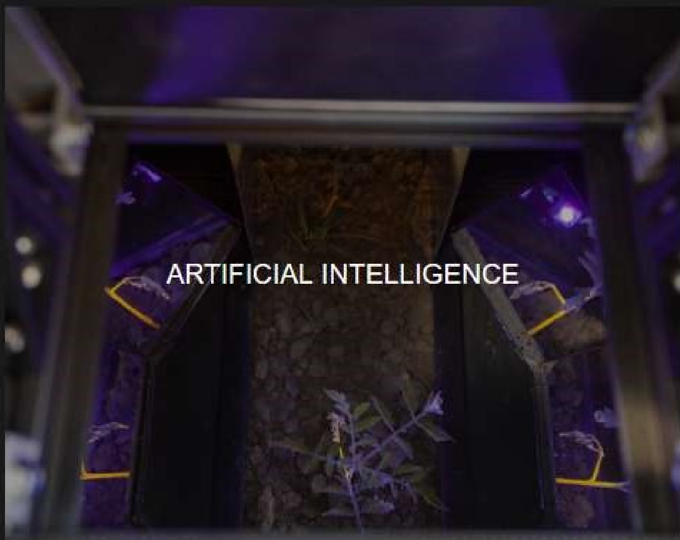
Nutrition

…and the people behind it…

…with AI and technology as an enabler

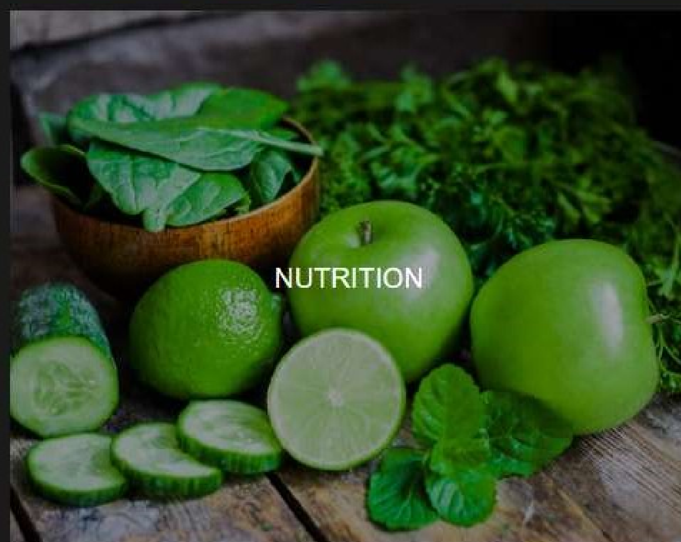





ARTIFICIAL INTELLIGENCE

MOLECULAR BREEDING

AGRICULTURAL PRODUCTION

FOOD PROCESSING & DISTRIBUTION

NUTRITION

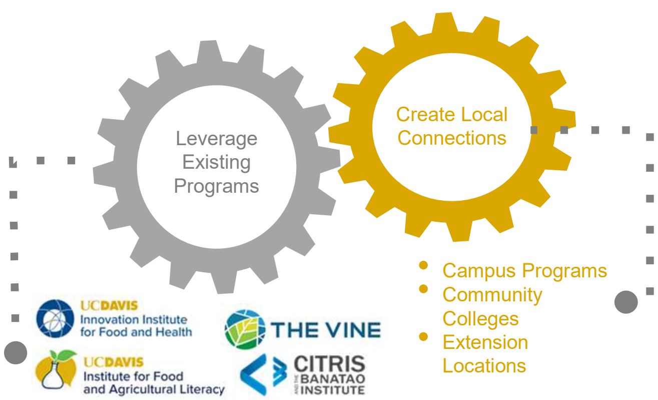
ETHICS & SOCIOECONOMICS

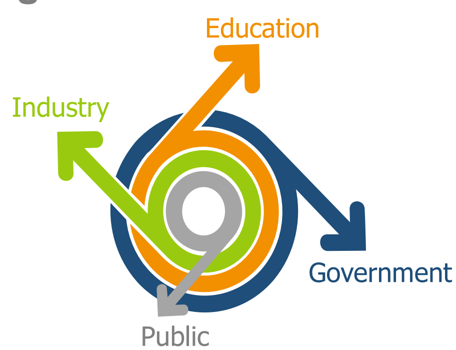
AIFS Participants

AIFS AI INSTITUTE FOR FOOD SYSTEMS

Artificial Intelligence (AI)	Molecular Breeding	Agricultural Production	Food Processing	Nutrition	Socioeconomics & Ethics

Engagement





Coordination

UC DAVIS UNIVERSITY OF CALIFORNIA

UNIVERSITY OF CALIFORNIA Agriculture and Natural Resources

CORNELL UNIVERSITY

Berkeley UNIVERSITY OF CALIFORNIA

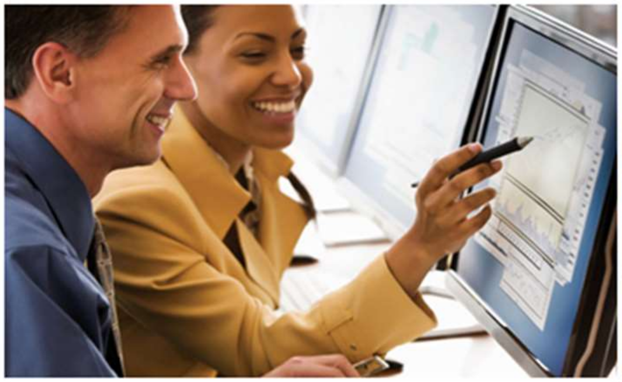
USDA ARS Agricultural Research Service

ILLINOIS

AIFS AI INSTITUTE FOR FOOD SYSTEMS

AIFS Engagement

Internships & Fellowships



Field Days

Teacher Fellowships

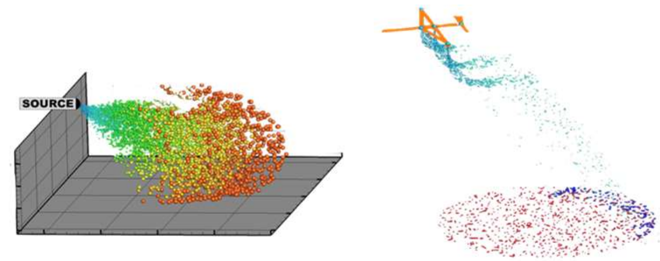
Curricula Enrichment

Industry+Public Engagement

Conferences & Workshops

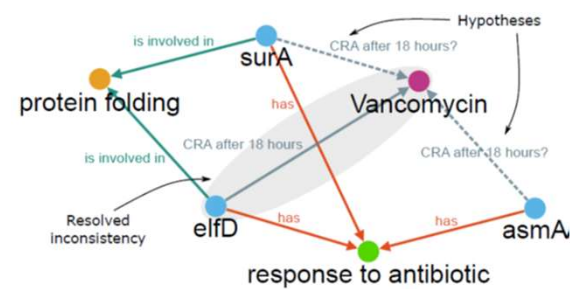
17 FUNDED PROJECTS for 2021

Digital Twin and Machine-Learning for Optimized Sensor-driven Precision Agriculture Utilizing Autonomous Ground and Aerial Delivery



Digital Twin and Machine-Learning for Optimized Pathogen Contact-tracing, Sanitation and Decontamination

Ultrasonic Cleaning of Produce

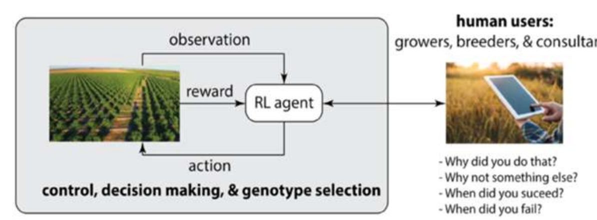


Predicting Health Effects From Food Composition via Large-scale Information Extraction

AI-Enabled Water and Nitrogen Stress Sensing and Forecasting for Sustainable Agricultural Production

Sequential Design for Accelerating Multi-trait Breeding

Prediction of glycan content from real-world food diaries

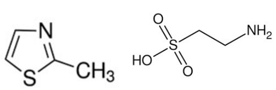


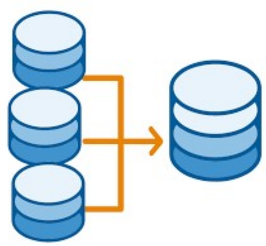
Performance Based Engineering for Resilient Food Industry

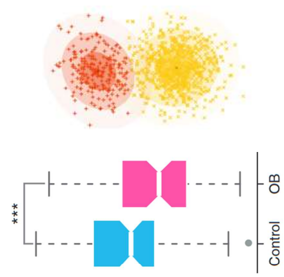
Case study: Nutrition AI

Analytical data collection	Preprocessing & integration	Hypothesis testing and discovery	Predictive modeling and simulation	Capturing and generating Knowledge	Translating to practice

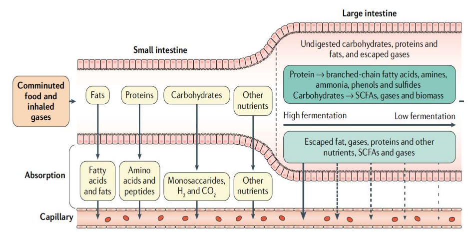
Vitamins
Minerals
Omics
Lipidomics
Glycomics
Metabolomics
Transcriptomics
Microbiome
Performance
Clinical

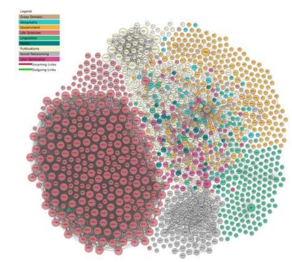


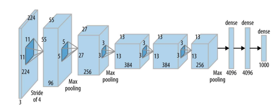




$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m} J(W,b;x^{(i)},y^{(i)})\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2$$
$$= \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}(x^{(i)}) - y^{(i)}\right\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2$$







Global and Integrated

Standardized and Open to all

Accelerating discovery

Challenges and Opportunities

AI Ethics

Datasets

Standards and Benchmarks

Mentality

Economics

Secure, fair collaboration

Education, awareness and initiatives (Rome Call)

Funding for AI-focused dataset generation

Consortia and working groups

Inclusion and trusted parties

DARPA-type long-term vision

Awareness, training, and ensuring compliance

UCDAVIS
UNIVERSITY OF CALIFORNIA

ILLINOIS

UNIVERSITY OF CALIFORNIA
Agriculture and Natural Resources

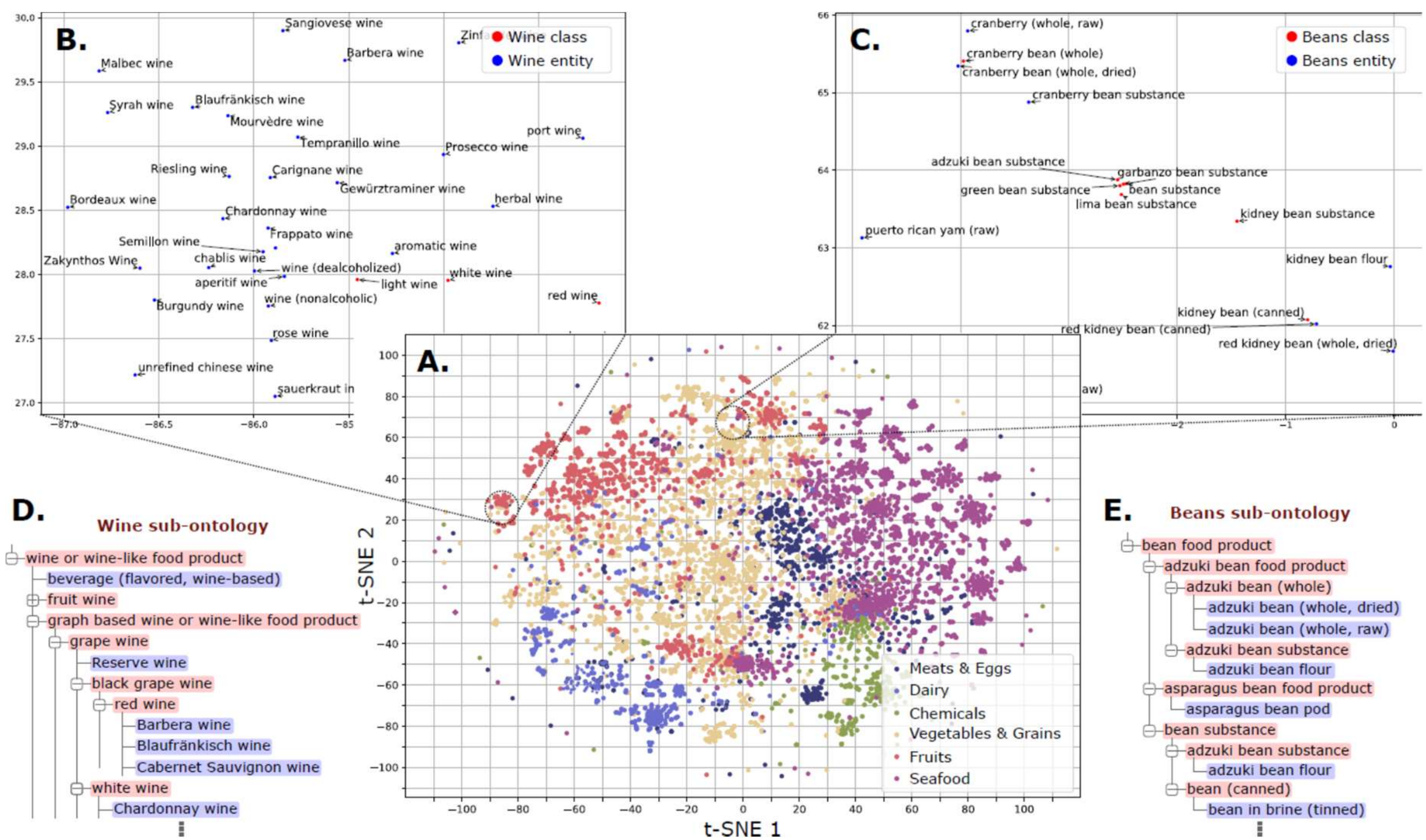
AIFS
AI INSTITUTE FOR FOOD SYSTEMS

CORNELL UNIVERSITY

Berkeley
UNIVERSITY OF CALIFORNIA

USDA ARS Agricultural Research Service

Case study: Nutrition AI



Digital Twin and Machine-Learning for Optimized Sensor-driven Precision Agriculture Utilizing Autonomous Ground and Aerial Delivery

SPECIFIC AIMS:

Model and simulate precise optimal crop-dependent sensor placement and delivery via ground-based and/or aerial-based systems. This also includes the placement of other agricultural products such a fertilizer, insecticides, etc.

Specifically, we will develop clean and easy-to-use digital tools that are repeatedly in demand by food system researchers such as:

- Data collection
- Compression and assimilation from the environment
- Rapid ODE and PDE solvers
- Neural nets and genetic algorithms with embedded detailed models of fundamental physics (fluid flow, heat transfer, stress analysis, electromagnetism, optics, robotics, etc.)

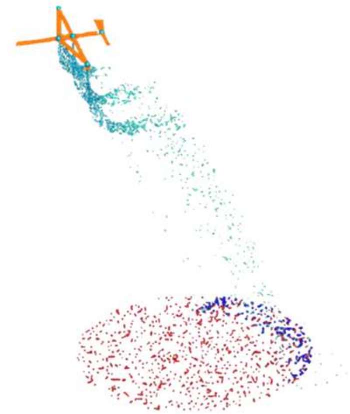
APPROACH:

This work develops a digital twin and machine-learning optimization framework for model problems combining:

A. Fluid-dynamics of released objects ranging from powders to encapsulated packets and sensors
B. Ground-based vehicle or aircraft (unmanned or manned) dynamics
C. Energy-efficient path planning of multiple ground or air vehicles
D. Lidar-based and hyperspectral data extraction for agricultural mapping for planning and assessment.

The framework is designed to enable digital twin type technologies to run at much faster rates, in order to enable machine-learning to optimize the planning.

The overall guiding motivation is to provide useful tools to enable rapid path planning for autonomous vehicle operators in real-time and to train operators in large surface area food systems.



Digital Twin Example: Aerial simulation frames illustrating a digital twin of an airdrop of materials onto a target possessing strong updrafts. (Zohdi, https://msol.berkeley.edu/)

A. Arias　M. Earles　I. Kisekka　E. Ligon　X. Liu

S. Mäkiharju　K. Mosalam　M. Mueller　I. Tagkopoulos　T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Digital Twin and Machine-Learning for Optimized Pathogen Contact-tracing, Sanitation and Decontamination

SPECIFIC AIMS:

Develop a Digital Twin and Machine Learning algorithms for addressing key food safety challenges associated with introduction and spread of pathogens in food facilities and enable guided decontamination for food contact surfaces to reduce risk of cross-contamination.

APPROACH:

Fundamental Approach #1) Computational framework for real time computation using digital twin models: The overall goal is to develop a computational framework for digital twin models that can be run in real time with the physical system. The examples of these frameworks include parallel compute functionality in agent-based models to characterize spread of pathogens in a food facility based on interaction among people, food, equipment surfaces and processing operations.

Fundamental Approach #2) Integration of validated digital twin model for AI enabled prediction of risks of spread of pathogens and decontamination effectiveness: This project will provide foundation for integrating multifactorial output from digital twin models to train machine learning models.

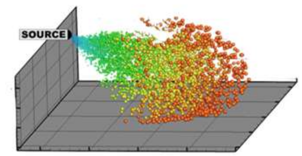
PROJECT PART I: We focus on two food safety scenarios:

A. The spread of viral pathogens (Covid-19) in a simulated work environment based on particle dynamics under diverse air flow and ventilation conditions
B. The spread of Listeria in a processing facility based on interactions among people, equipment and food products.

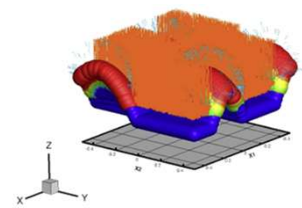
For each of these scenarios, the team will develop digital twin models to predict the biophysical dynamics of pathogens, validate these digital models using experimental data sets using surrogates under simulating industry conditions or field data and use validated models to train machine learning based predictions for the spread of pathogens in food facilities.

PROJECT PART II: We plan to evaluate guided decontamination of the food contact surfaces using light based antimicrobial technologies including UV light. The proposed research plan will build on the digital twin models to predict the spread of pathogens as well as experimental approaches to predict the localization of biofilms on food contact surfaces. The outputs of the models and the experimental data will be used for developing machine learning models to guide deposition of light energy and predict inactivation of viruses and biofilms.

Success in developing this framework will enable industrial deployment of these models and provide rapid response associated with characterizing the risk of spreading of pathogens and sanitation and decontamination.



Cough simulation. (Zohdi, https://msol.berkeley.edu/)



UV-c pulse applied to a surface. (Zohdi, https://msol.berkeley.edu/)

R. Abergel A. Arias R. Ivanek X. Liu S. Mäkiharju N. Nitin

C. Simmons I. Tagkopoulos L. Wang M. Wiedmann T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Digital Twin and Machine-Learning for Optimized Robotic Production of Complex Multiphase Foods

SPECIFIC AIMS:

The objectives are to develop Digital Twin and Machine Learning algorithms for optimized components associated with robotic production and handling of complex multiphase foods.

APPROACH:

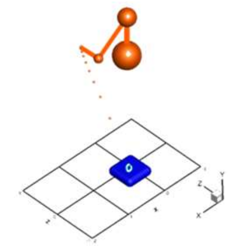
PROJECT PART I: This entails expertise of faculty in Optimization, Machine-learning, High-performance computing, Robotics, Fluid Mechanics, Micromechanics.

Specifically, we will develop an MLA to ascertain the appropriate robotic motion needed to create/assemble food products which would be difficult or impossible by guesswork. These models will then be used for designing high-performance feedback control systems; especially their use for predictive control and receding horizon estimation.

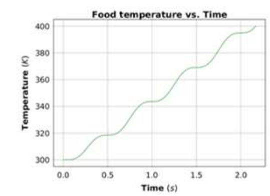
Accurate control of robotic motion requires accurate estimation of the interior state of the robotic manipulators in the face of poor sensor data, including noise, biases, etc. Additionally, optimal and safe decision making requires the prediction of how uncertainty will propagate in response to control actions, and what control actions will minimize cost while achieving satisfactory performance. We will combine, specifically, model-based approaches to uncertainty quantification (such as the Kalman-filter or particle filter) using large datasets. The trade-offs between the approaches, their robustness to process changes, and their computational costs will be investigated.

PROJECT PART II: Across many modern industries, as technologies have matured, the use of more complex processes involving multiphase materials has increased. In the food industry, multiphase fluids are now relatively wide-spread, in particular, because of the desire to have faster throughput for large-scale food production. In many cases involving transport, such materials consist of a fluidized binder material with embedded particles. As one increases the volume fraction of particles, a corresponding increase in effective overall viscosity occurs. Often, during the process, the material must be heated, for example, to ensure food safety, induce pasteurization, sterilization, etc.

For real-time control, this requires rapidly computable models to guide thermal processing, for example by applied electrical induction. In the present analysis, models are developed for the required heating field (typically electrically induced) and pressure gradient needed in a delivery system (pipe-flow) to heat a multiphase material to a target temperature and to transport the material with a prescribed flow.

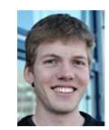


Example of precise robotic deposition of heated complex media. (Zohdi, https://msol.berkeley.edu/)



Thermal behavior of multiphase foods. (Kim, Zohdi, Singh, https://msol.berkeley.edu/)





F. Borrelli K. Mosalam M. Mueller



N. Nitin I. Tagkopoulos T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Ultrasonic Cleaning of Produce

SPECIFIC AIMS:
The overall goal of this study is to develop and optimize an AI-assisted ultrasonic fresh produce sanitation system to significantly improve the efficacy of current fresh produce sanitation operations. The specific objectives include:

A. Develop a sub-pilot scale continuous-flow ultrasonic fresh produce washing testbed,
B. Quantify cavitation activity on produce surfaces using high-speed imagining and X-ray techniques,
C. Develop electrical sensors measuring chlorine concentration and pH level based on graphene field-effect
D. transistor (FET) sensors, and
E. Perform machine learning (ML)-based response surface modeling of fresh produce washing process.

APPROACH:
The proposed study will tap into the power of AI to develop strategies to optimize an ultrasonic fresh produce sanitation process. It will for the first time enable an intelligent produce washing process to ensure maximized sanitation efficacy for different produce types.

A. Continuous-flow ultrasonic fresh produce washing testbed for data generation (Feng, UIUC)
B. Determination of cavitation activity on produce surface (Mäkiharju, UC Berkeley)
C. Determination of graphene field-effect transistor (FET) chlorine concentration sensors (Nam, UIUC)
D. Development of develop an ML-based response surface model (Shao, UIUC)

AI guided ultrasonic cleaning of produce will be examined. The initial experiments will focus on enhanced cleaning effectiveness due to shear induced on the produce surface by bubble enhanced ultrasonic cleaning. Range of amplitudes considered is such that we span cases from oscillating non-condensing gas bubbles to cavitation near an artificial leaf surface. The artificial plant leaf surface is developed in Feng lab (UIUC) in order to mimic the surface topology, hydrophobicity, and surface chemistry of real lettuce leaf, but allowing for repeatable well-controlled studies. After this model case is examined, enhancing our physical understanding of the relevant flow, time-averaged cavitation within complex media (including real produce) will be examined. We will utilize X-ray computed tomography with photon counting energy resolving detector arrays.

Deep Neural Network with dense layers and multiple outputs will be trained to take the data yielded by up to 128 energy-level computed tomography reconstructions and classify the voxel material composition. We'll use hyperbolic tangent, ReLU or sigmoid activation function, and tentatively idea is to both code ourselves and use Tensor Flow (in Colab) – which in the simple classification case should yield comparable results if our code works correctly. Training of the network will be accomplished by a hybrid dataset consisting of experimental phantom CT scan data and two types of simulated data.

Digital twin of enhanced cleaning produce process informed by laboratory experiment data.


H. Feng


S. Mäkiharju


S. Nam


N. Nitin


C. Shao

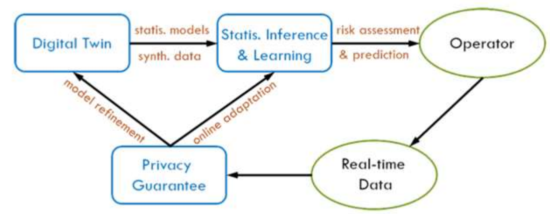
The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Statistical Learning for Food Safety Assessment under Privacy and Resource Constraints

SPECIFIC AIMS:

We aim to develop data analytics for food safety assessment, specifically,

1. to build digital-twin models of processing facilities, utilizing pre-existing industry datasets and developing the ability of integrating diverse streams of data for improved model parameterization;
2. to develop statistical learning and inference algorithms for rapid and reliable pathogen detection and risk-averse prediction of effective corrective measures, by integrating synthesized data from the digital-twin models with the facility-owned data;
3. to address practical constraints, in particular, privacy constraints to incentivize data sharing and improve human trust, as well as resource constraints in terms of computation and memory complexity.



The figure above illustrates the integrative approach that consists of three components.

APPROACH:

Digital-twin models: We aim to develop agent-based modeling tools to facilitate decision-making in complex systems and to minimize the risk of making wrong decisions. We will augment the agent-based models on *Listeria* dynamics in food processing facilities to allow integration of diverse industry datasets and real-time data streams.

Statistical learning for rare event detection: Food safety events are typically rare. The challenges in detecting such rare events lie in:

1. the massive search space and the high cost and risk in assessing potential corrective measures;
2. the need for high detection accuracy due to the costly and potentially catastrophic consequences associated with false positives and negatives;
3. the time sensitivity of the problem due to the urgency for taking recourse measures.

Our technical approach is rooted in active hypothesis testing that sequentially and adaptively determines where to search based on the current estimate on where the rare events may reside.

Privacy guarantee: Data privacy is a primary concern for the food industry where leakage of information about detection of a pathogen can have significant financial and societal impacts. Differential privacy is a noise addition mechanism to ensure privacy of entities that contribute data. This comes at the price of a loss in utility. A solution to this challenge is the notion of context- aware privacy that provides stronger privacy guarantees for events of concern. Performing active hypothesis testing on privatized data samples is also of interest.


J. Acharya


R. Ivanek


N. Nitin


M. Wiedmann


Q. Zhao


T. Zohdi

The projects are jointly conducted with partners through the AIFS Network:
https://aifs.ucdavis.edu/

Predicting Health Effects From Food Composition via Large-scale Information Extraction

SPECIFIC AIMS:
The goal of this project is to (a) create a software package for the automated curation of nutrition data from published papers, (b) apply the software package to a corpus of hundreds of food and nutrition papers, and organize the resulting information in a Knowledge Base (KB), (c) use the novel KB to augment published datasets linking food intake to health outcomes to create a predictor of health effects from chemical composition of food.

APPROACH:
Aim 1. Create a software package for the automated curation of nutrition data: The project will use Python, Natural Language Processing toolkits such as spaCy, and tabular and chart data extraction software such as Tabula and WebPlotDigitizer to extract nutrition composition information from free text, charts, and tables from scientific literature documents. Documents will be obtained from a suitable scientific search engine such as Web of Science.

Aim 2. Create a knowledgebase of food and chemical compound information: The project will apply the automated curation package from Aim 1 to extract quantified nutrition composition information from scientific literature for the top foods in a ranked list, sorted by their contribution to dietary intake in datasets used in Aim 3; good coverage of popular foods in the study cohort is required to demonstrate the additional predictive power we expect from the literature-extracted features. The corpus for data extraction will be derived from paper texts and supplemental information from nutrition studies using resources such as Web of Science.

Aim 3. Obtain datasets for use in a predictor of health effects from food chemical composition: The project will identify multiple datasets linking food intake data to health outcomes. Dataset size and accuracy of food intake observations are relevant considerations for dataset selection. Large datasets (such as EPIC and TwinsUK) offer the opportunity to train more expressive machine learning models, but noise is a problem in nutrition studies with self-reported food intake. Supervised feeding studies provide more accurate reporting at the cost of smaller cohort sizes. To address this tradeoff, datasets will be identified from at least one large cohort study and at least one supervised feeding study. For each dataset, predictors using baseline food composition information and the extracted information from Aim 2 will be compared to assess the predictive value of the extracted information.

D. Lemay

I. Tagkopoulos

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

AI-Enabled Water and Nitrogen Stress Sensing and Forecasting for Sustainable Agricultural Production

SPECIFIC AIMS:
This project aims to develop a data-efficient, AI-enabled framework for water (W) and nitrogen (N) stress sensing and prediction. To do this, we will pursue the following sub-objectives: (Obj. 1) Design, build, and deploy W and N status sensors that will accelerate the rate of ground-truth data collection, (Obj. 2) Develop a synthetic data generation pipeline that couples process-based biophysical modeling to deep learning for prediction of W and N stress, and (Obj. 3) Create a deep learning framework that fuses multiple data streams to directly predict plant W and N stress at multiple scales from individual tree to orchard (Obj. 1 and 2). We will pursue these objectives for almond, which is the most widely irrigated and economically valuable crop in California.

APPROACH:
[Obj. 1] To dramatically accelerate the rate of ground-truth W and N data collection, we will design, build, and deploy W and N stress sensors. Specifically, we will instrument existing commercial farms with in-situ, proximal, and remote sensing systems for measuring W and N status in soil and plants at multiple scales. Currently, Kisekka, Pourreza, and Bailey are working with commercial partners to deploy a wide range of sensors.

[Obj. 2] To substantially increase data-efficiency and model generalizability, we will develop a synthetic data generation pipeline that couples biophysical modeling and deep learning for W and N stress prediction. A 3D crop model called Helios will generate millions of 2D synthetic sensor imagery of almond cultivars across combinations of environmental and management scenarios, with trees of estimated W and N stress values. We will then use this synthetic data to pre-train models to estimate W and N stress in actual production conditions.

[Obj. 3] To overcome current models' inability to predict W and N stress with high accuracy at individual tree scale, we will create a deep learning framework that fuses multiple ground-based and aerial data streams to directly predict W and N. We will build on current efforts by Kisekka, Jin and Pourreza to predict W and N stress in almonds. Novel deep learning architectures will be developed to fuse multiple data.


A. Arias


B. Bailey


C. Diepenbrock


M. Earles


Y. Jin

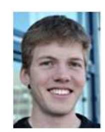

I. Kisekka


Z. Kong


X. Liu


S. Mäkiharju


K. Mosalam


M. Mueller


A. Pourreza


A. Smith


S. Vougioukas


T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Data-efficiency in the Food Systems

SPECIFIC AIMS:
Data efficiency plays a vital role in using AI-based solutions in next generation food systems. In this project, we will identify key challenges in food systems using domain-specific examples and develop technical approaches to address data efficiency in food systems. In this project, we highlight and abstract the challenges in food systems with the goal of developing generalizable solutions to food systems.

APPROACH:
To address the common data-efficiency challenge, we consider the following two-pronged approaches.

Improving models through active learning and efficient exploration: In the application problems, we face large-scale learning problems where obtaining labeled data is costly and time consuming. This is a challenging issue for not only real data but also simulation data. In molecular breeding, it takes multiple growing seasons and significant cost to obtain a new sample. In simulation models, typically, the better the model, the higher the computational complexity. For example, it takes five minutes to generate one good image created by a 3D-model generated in ag production. We will pursue an active learning approach where the decision maker actively chooses which data points to use or which experiments to carry out for the purpose of gathering the most relevant information for the learning task at hand. The focus is on exploiting latent structures in the learning space determined by the specific underlying applications in food systems. Furthermore, when the state-action space is large, we will explore efficient approximation and exploration techniques. Various techniques have been developed in the literature, using different mechanisms to measure state-action uncertainties. Building on our existing work using GAN-based models, we will investigate efficient exploration strategies, incorporating domain knowledge.

Bridging the gap between simulation models and real environments: Simulation plays a key role in studying complex real systems in a safe and cost effective manner. Many projects have existing simulation tools or models built using existing data. Several proposed projects depend on building digital-twin models. While simulation tools are powerful, there typically exists a gap between simulation models and the real world. How to bridge the gap between the two is a key research question that is relevant to all application clusters. We plan to explore two types of approaches. 1) domain adaptation, where one adapts a model trained in a source domain (simulation) to a new target domain (a real environment). 2) Domain randomization, where discrepancies between the source and target domains are modeled as variability in the source domain. We plan to study the integration of model-based and model-free sequential decision processes. We plan to investigate the idea of meta learning in this scenario, also incorporating invariant features, identified both through training and through utilizing domain knowledge. We also plan to investigate GAN (generative adversarial network) based approaches, where we augment training data using GAN.


M. Earles


D. Lemay


X. Liu


N. Nitin


D. Runcie


I. Tagkopoulos


Z. Yu


Q. Zhao


T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

An Ethics Plan for AIFS

SPECIFIC AIMS:
We propose to complete two projects that will contribute to socially trustworthy AI for agricultural applications. One project will result in a set of recommendations that AIFS research cores can adopt to help assure the trustworthiness of their research. The other will create an ethics curriculum for AIFS researchers, graduate students, and post-doctoral fellows.

APPROACH:
To complete both projects, we will conduct interviews with the researchers in each AIFS cluster. To develop the recommendations, we will ask interviewees a series of questions expressed in lay terms:

- What is it that AI developers and producers are asking people to trust them with, to accomplish what purpose?
- What are the accountability, safety, and precautionary methods and practices that are in place to assure that AI developers/producers can be trusted to do the things they claim to want to do with the data they are soliciting?
- How effective are those methods and practices thought to be?
- How vigilant are AI developers and producers in attempting to identify new concerns and issues that can arise from their AI applications and who is responsible for guarding against them? To whom are they accountable and what are the mechanisms used to hold them accountable?

To develop the ethics curriculum, we will interview researchers about the following three themes: Bright Lines, Big Picture, and Deep Questions. This thematic approach has been successfully piloted by Mark Yarborough and Larry Hunter for ethics instruction for both the computational biosciences program at the University of Colorado Denver and the UC Davis Clinical and Translational Science Center. Interviewees will also be asked to respond to the following topics:

- What are the **bright lines** that cannot be crossed in research?
- What is the **big picture** in which science is situated today?
- What are the **deep questions** posed by AI in the food system?

We will record and transcribe the interviews, use NVivo or comparable data analysis software to help analyze and synthesize the responses, and then produce our recommendations for helping to assure the trustworthiness of AIFS research, as well as the ethics curriculum. We will work with the socioeconomics and ethics cluster personnel and the leads of the other clusters to refine the interview questions for both projects. Engagement from all components of AIFS is crucial for the success of this project and the institute overall.



L. Flemming E. Ligon P. Ronald


A. Smith M. Yarborough

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Sequential Design for Accelerating Multi-trait Breeding

SPECIFIC AIMS:

Breeding programs need to improve multiple characteristics of their crops simultaneously, including yield, disease resistance, and quality. However, when traits are genetically correlated, selection on one trait leads to indirect selection on other traits. Classic quantitative genetics theory provides a solution, defining theoretically optimum weights of two (or more) traits for simultaneously improving both. However, actual breeding programs differ from the simplifying assumptions underpinning these theories in several critical ways:

- Genetic parameters are not directly observable and are often estimated with considerable uncertainty
- Some traits can be measured more precisely, cheaper, or at different growth stages than other traits
- Responses to selection are proportional to the available genetic variation in the breeding population
- Correlations between traits can dramatically hinder or improve genetic gain.

A holistic, simulation-based approach to breeding program optimization could address all of these challenges, and provide precise and explainable recommendations to breeders that take into account the specifics of each individual program.

APPROACH:

Aim 1: We will develop a simulation tool to rapidly simulate outcomes of breeding decision frameworks over multiple generations and calibrate the simulations around the current state of the UC Davis strawberry and pepper breeding programs. This will be used to evaluate the probability of creating specific genotypes under different decision frameworks.

Aim 2: The simulations of Aim 1 will quantify the probability of discovering specific genotypes. But predicting the quality of these genotypes requires modeling the genotype-phenotype relationship, accounting for uncertainty. Most genotype-phenotype models in breeding programs produce point-estimates of genetic quality. This is sufficient for single-trait breeding objectives, but the uncertainty of genetic correlations among traits is much higher; therefore accurately quantifying the uncertainty in genetic relationships among traits may change breeding decisions. We will re-analyze phenotypic data from recent breeding trials using multiple Bayesian models implemented in MegaLMM.

Aim 3: The outputs of Aim 1 and Aim 2 will be used to optimize breeding program designs for both programs, using the principles of Sequential Design of Experiments, including which lines to cross (using which selection criteria), which to phenotype, and how to allocate limited resources between these objectives.

A. Van Deynze K. Knapp Z. Kong X. Liu D. Runcie

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Optimization of Enhanced Agricultural Product Ultrasonic Multiphase Cleaning System with AI

SPECIFIC AIMS:

The overall goal of this study is to develop and optimize an AI-assisted ultrasonic fresh produce sanitation system to significantly improve the efficacy of current fresh produce sanitation operations. The specific objectives in year 1 of the project include:

- Develop a sub-pilot scale continuous-flow ultrasonic fresh produce washing testbed
- Quantify cavitation activity on produce surfaces using high-speed imagining and X-ray techniques,
- Develop electrical sensors measuring chlorine concentration and pH level based on graphene field-effect transistor (FET) sensors, and
- Perform machine learning (ML)-based response surface modeling of fresh produce washing process.

APPROACH:

Task 1: Development of sub-pilot scale continuous-flow ultrasonic fresh produce washing testbed for data generation. Since a significant amount of data will be needed for AI and ML purposes, and since produce washing with the pilot scale produce washer requires large amount of water and produce, we will design and fabricate a sub- pilot scale produce washer with ultrasound capacity for data generation and sensor evaluation purposes. This produce washing testbed will be built with transparent material so the flow pattern, mixing of the produce during washing will be visualized. Sensors developed in Task 3 will be installed in the testbed to collect data about changes in sanitizer concentration in different location of the washer as a function of produce type, sanitizer to produce ratio, washing time, ultrasound power density, and cavitation activity (determined in Task 2).

Task 2: Determination of cavitation activity on produce surface. The laboratory experiments start by replacing the lettuce by an artificial plant leaf surface developed in Feng lab that mimics the surface topology, hydrophobicity, and surface chemistry of real lettuce leaf, but allowing for repeatable well-controlled studies. Cavitation on the surfaces will be characterized by readily available high-speed imaging up to 820,000 fps.

Task 3: Determination of graphene field-effect transistor (FET) chlorine concentration sensors. Phenyl-capped aniline tetramer (PCAT) film will be deposited on graphene FET surface to enable specificity toward chlorine. Oxidation of PCAT by chlorine induces p-doping of graphene FET and allows real time electrical detection of chlorine level. The p-doped PCAT-graphene can be electrochemically un-doped by polarizing it cathodically to allow re-use of our sensors. Our chlorine sensor will be simpler in construction while allowing higher sensitivity compared to conventional electrochemical sensors.

Task 4: Development of develop an ML-based response surface model. We will develop an ML-based response surface model for leafy vegetable washing processes.





H. Feng S. Mäkiharju S. Nam N. Nitin C. Shao

The projects are jointly conducted with partners through the AIFS Network:
https://aifs.ucdavis.edu/

Prediction of glycan content from real-world food diaries

SPECIFIC AIMS:

- Test off-the-shelf computer vision algorithms on food images for single core foods
- Generate benchmark data set of real-world mixed meal photos paired with food records
- Engage with SCIO to pilot test whether NIR can distinguish a pair of visually similar foods
- Develop glycan library for core foods
- Build library of images for single core foods being analyzed for glycan content
- Determine optimal set of mixed meals to analyze for glycan content

APPROACH:

We will curate a dataset of 'core foods' corresponding to simple (i.e. non-recipe) foods evaluated by the Lebrilla lab for glycan content (n > 200). For each core food, 100 images will be downloaded using Google image search. We will manually select images from the Recipe1M dataset, which contains images from online recipes, to obtain a subset of images of mixed foods and images with multiple foods per image. The mixed food images will include foods with the core foods as ingredients (e.g. core food = 'oatmeal', mixed food = 'oatmeal cookies') in addition to other randomly selected mixed food images. Images will be annotated with ground truth bounding boxes via Mechanical Turk.

The core and mixed food image dataset will be used to develop a model for identifying core and mixed foods. We will use transfer learning to leverage existing pre-trained models. First, we will try faster-RCNN or YOLO models to classify whether food in an image is a core or mixed food, and each food will be cropped to the identified bounding area. Cropping to the area of the food will enable the input images to contain multiple food or plates (i.e. meals) instead of only accommodating close ups of a single food. Next, we will try the Inceptionv3, ResNet-50, and VGG16 models to classify the specific labels of the core foods (e.g. apple, oatmeal, brown rice). Since the model will be developed using images obtained from the internet, the final ensemble will be evaluated using the real-world images collected from human subjects.

To complement visual recognition, we will test near infrared (NIR) technology from SCIO tech. We will test using visually similar foods (e.g. mashed cauliflower and mashed potatoes) to determine if the foods can be distinguished.

Because we cannot determine the glycan profile of an infinite number of mixed meals, we will use an optimization algorithm, OPEX to identify which mixed meals to analyze to create the data set needed to build a mixed meal glycan predictor.


E. Bonnel


M. Earles


B. German


C. Lebrilla


D. Lemay


E. Ligon


X. Liu


C. Martin


D. Mills


J. Seigel


J. Smilowitz


C. Stephenson

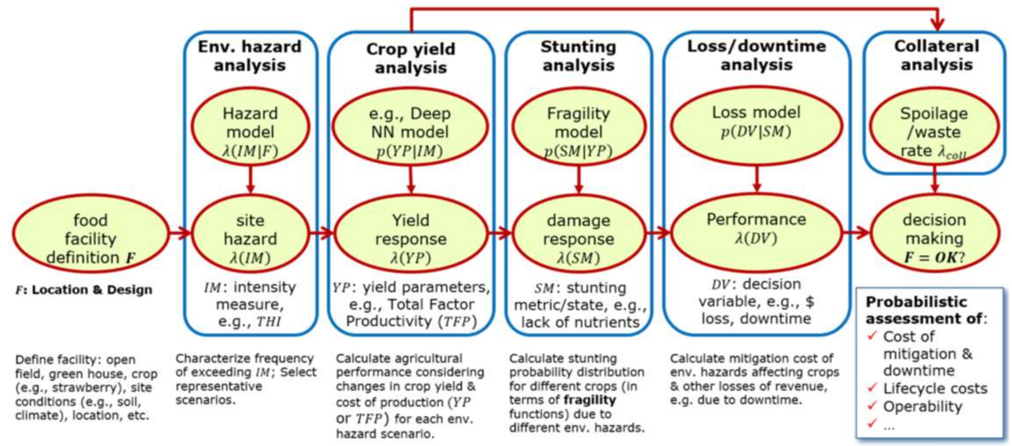

I. Tagkopoulos

The projects are jointly conducted with partners through the AIFS Network:
https://aifs.ucdavis.edu/

Performance Based Engineering for Resilient Food Industry

SPECIFIC AIMS:

Develop a food system management framework by combining Performance Based Engineering (PBE) and Artificial Intelligence (AI), to increase efficiency, safety & resilience of agricultural production by improving food yield & quality, controlling crop diseases, decreasing resource consumption & waste, and increasing traceability. Traditionally, agricultural management is based on empirical judgement resulting from experiments & experience, which is no longer adequate because of increased complexity & added uncertainty of food systems due to the increasing demands of the world population. Thus, uncertainty quantification in food systems has been recently adopted. However, a robust framework is still lacking and this proposal aims at filling this gap.



Overview of the proposed PBE methodology for food systems.

APPROACH:

Crop yield prediction is of great importance to global food production. Policy makers rely on accurate predictions to make timely import and export decisions for national food security. During the last decade, scientists and engineers have made significant headway in developing and deploying tools and devices that deliver massive, yet too often raw, data streams to food system stakeholders at unprecedented spatiotemporal resolution. Concurrently, AI algorithms repeatedly break benchmarks in Computer Vision applications, Natural Language Processing (NLP), and automation. Despite its significant benefits, the complex food systems face several challenges in the application and adoption of AI. Frameworks developed in other domains can be integrated with AI approaches to tackle these challenges. PBE is one of these frameworks with a robust formulation for decision making under inherent uncertainty due to the complex nature of food systems. PBE originated in the natural hazards field, particularly earthquake engineering, and is now adopted by wind & fire engineering.

K. Mosalam

I. Tagkopoulos

T. Zohdi

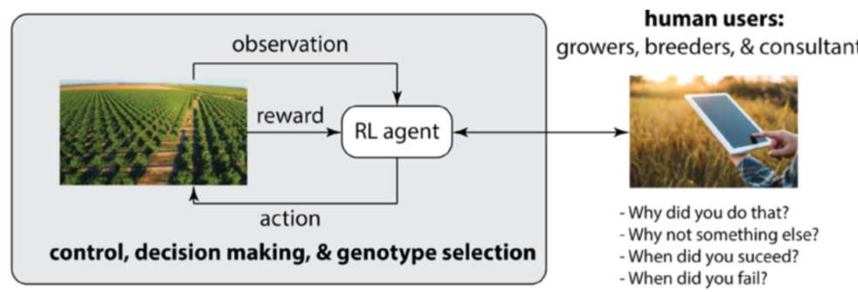
The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Explainable Reinforcement Learning for Next Generation Food Systems

SPECIFIC AIMS:

Trust is an essential prerequisite of adopting AI-based solutions in next generation food systems. Explainability has been identified as one key component in increasing both users' trust and their acceptance of AI-based solutions. Furthermore, explainability justifies an AI's decisions and enables them to be fair and ethical. Reinforcement learning (RL) will play a key role in solving control and decision-making problems in agricultural production and food processing/distribution as well as genotype selection problems in molecular breeding.

The long-term objective of this project is to advance theoretical and algorithmic foundations of explainable RL (XRL) and create a (or multiple) proof-of-concept RL system(s) pertaining to the aforementioned problems. In Year 1, we will adopt a human-centric AI methodology and work closely with our collaborators to identify our stakeholders' needs and develop a preliminary XRL framework in accordance with their needs.



Schematic representation of our XRL framework

APPROACH:

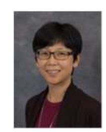
This project will build upon PI Liu's and PI He's expertise on RL and PI Kong's expertise on explainable machines. Our team will adopt a human-centric strategy to develop our XRL framework. From social science, humans prefer causal explanations that are contrastive. Our proposed XRL framework can access a simulator/model of the food system under control and a list of external and internal variables relevant to the system. A domain expert will provide these variables. Our proposed XRL framework consists of two interleaving phases. First, a structural causal model (SCM) that describes the relationships between the variables (listed by the expert) will be learned from the RL agent's interaction with the simulator. Second, the RL agent will learn a control policy, which both maximizes the expected reward and respects the learned SCM. Our XRL framework will cycle between these two phases until a consistent SCM and control policy have been obtained. Finally, our XRL agent can answer human inquiries based on both the learned SCM and control policy. One crucial feature that distinguishes our framework with existing XRL methods is that it can generate counterfactual explanations.

M. Earles

N. He

I. Kisekka

Z. Kong



X. Liu

N. Nitin

D. Runcie

I. Tagkopoulos

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

AI-Enabled Yield Sensing and Forecasting for Agricultural Production

SPECIFIC AIMS:

This project aims to develop a crop-generalizable, data-efficient, AI-enabled framework for agricultural yield sensing and forecasting. Specifically, we will:

1. Design, build, and deploy yield monitoring sensors to accelerate ground-truth yield data collection
2. Develop a synthetic data generation pipeline that couples a 3D crop model to a deep learning framework for yield prediction
3. Create a deep learning framework that fuses multi-modal groundbased and aerial datastreams to directly predict yield from large-scale ground-truth data that incorporates human decisions

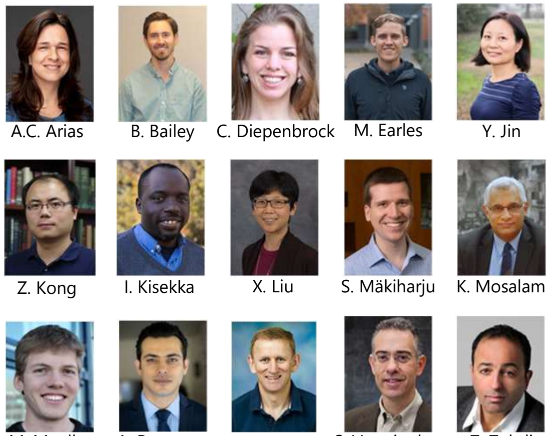
We will pursue these objectives for the three most economically valuable crops in California: strawberry, grape, and almond.

APPROACH:

For Objective 1) in Y1 are the generation of groundtruth yield data covering 50,000 strawberry, 100,000 grapevine, and 25,000 almond plants at farms where we will also collect or obtain a suite of ground-based and aerial sensor data as well the timing and quantity of input use (see Obj. 3).

For Objective 2) in Y1 are generation of over 100,000 synthetic training data images for each strawberry, grape, and almond, pre-trained models for each validated against real-world ground-truth data from Obj. 1.

For Objective 3) in Y1 are the curation of training datasets of ground-based and aerial sensor data with corresponding ground-truth yield values, and an accompanying end-to-end deep learning training and evaluation pipeline.



A.C. Arias B. Bailey C. Diepenbrock M. Earles Y. Jin

Z. Kong I. Kisekka X. Liu S. Mäkiharju K. Mosalam

M. Mueller A. Pourreza A. Smith S. Vougioukas T. Zohdi

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

AI prediction of fruit bruising, serum leakage, and spoilage during transportation of processing tomatoes to processing facilities

SPECIFIC AIMS:

1. Develop a database of processing tomato quality at the point of delivery to processing facilities and associated meta data for harvesting and hauling;
2. Develop digital model to predict mold growth based on quality parameters and agricultural meta-data;
3. Develop an AI model of tomato damage, spoilage, and serum leakage based on said meta data;
4. Validate model to determine prediction accuracy

APPROACH:

An accurate AI regression model will be constructed to predict tomato mold severity and likelihood of load rejection at processing facilities based on 10 years of data available from the Processing Tomato Advisory Board (PTAB) and the digital model of mold growth on the fruit based on physiological and environmental factors. The data include the location of harvest, date and time of harvest, tomato variety, tomato maturity, tomato color, mold level, damage frequency, moisture content, pH, number of loads, and number of rejected loads spanning millions of loads of tomatoes inspected at processing facilities. Based on this database, a digital model of mold growth and potential risk of spoilage will be modeled. An accurate AI regression model will be structured to predict tomato spoilage and load rejection based on the given fruit properties and sourcing metadata. We will start with classical machine learning models such as linear regression, logistic regression, Support Vector Machine (SVM), and convolutional neural networks. We will also evaluate ensemble models, e.g., AdaBoost and XGBoost, especially if the accuracies of classical models are not satisfactory. Before model training, data engineering will be used to manage noise in data and transform the clean data into a format that is suitable for a machine learning. We will compare the performance of models and determine the best model that results in the highest accuracy.

Leveraging domain knowledge and active learning techniques, we will study how to improve the PTAB model by collecting additional data and incorporating serum leakage as an additional response. Specifically, force data for individual tomato loads will be obtained by placing load cells and accelerometers within gondolas used for hauling tomatoes to processing facilities as well as Vis/NIR spectroscopy data to represent physiological parameters in tomatoes. Additional meta data will be collected regarding the routes traveled from the field to the processing facility as well as the duration and temperatures for transportation. These data will be combined with corresponding PTAB data for the loads and measurements of serum volume in the gondolas. Overall, these experimental data sets will be supplemented with a digital model of mold growth based on the plant physiological and environmental conditions. The resultant data set will be used to retrain the initial multimodal AI model in order to determine if the additional predictors improve the accuracy of predicting tomato damage and spoilage. Model validation will be achieved using


X. Liu


N. Nitin


C. Simmons


E. Spang


L. Wang

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/

Sensing and modeling of leaf biochemical and physiological traits, including early vigor

SPECIFIC AIMS:

To use AI approaches to quantify, model, and predict relationships between remote sensing observations and in-situ measured leaf quality, vigor, and yield traits. Specific aims: A. Non-destructive and simultaneous sensing of leaf biochemical and physiological traits at field + population scales, to be selected upon in the breeding process. B. Design of mechanistic and AI-enabled models of crop growth and development (in leafy greens) and broadscale crop structural properties (e.g. canopy dimensions, in legumes), with a focus on simulation and prediction of yield and quality – moving towards functional-structural models.

APPROACH:

We will examine opportunities in four crops in Year 1 and focus (based on the success metrics below) on two or three crops in subsequent years.

Spinach: Conduct hyperspectral imaging of spinach breeding accessions, and extract physiological and biochemical traits (the latter to be validated via colorimetric methods) using machine learning techniques that isolate specific wavelengths sensitive to functional and structural traits. Success metric: moderate accuracies compared to groundtruth for rate and extent of above-ground biomass accumulation (with a focus on early vigor and yield); nitrogen use efficiency; greenness (and discoloration); and/or certain carotenoid traits.

Lettuce: analyze and collect hyperspectral data from previous and ongoing field experiments in lettuce using wide diversity and recombinant inbred line (RIL) mapping populations for which genotyping is already available, alongside groundtruth agronomic and biochemical data. Success metric: identification of genomic regions controlling traits of interest and comparative data from remote sensing and groundtruth field/lab data; likely to focus on leaf color/phytonutrients and vigor, validated from wet biochemistry and proximal imaging.

Alfalfa: extract physiological (photosynthetic, early vigor, canopy volume) and nutritional traits (crude protein, neutral and acid detergent fiber; to be validated against benchtop near-infrared spectroscopy [NIRS]). Success metric: high cross-validation accuracies in trait extraction; moderate correlations between remote sensing and NIRS data.

Common bean: examine photochemical efficiency and non-photochemical quenching under drought vs. well watered conditions in 300 genotypes of a common bean multi-parent advanced-generation intercross (MAGIC) population (Berny Mier y Teran et al. unpublished), and 20 tepary bean genotypes. This data collection is part of an existing NIFA grant. Success metric: moderate heritabilities of remote sensing collected traits.


B. Bailey


C. Brummer


C. Diepenbrock


M. Earles


T. Magney


D. Runcie


G. Taylor


L. Wang

The projects are jointly conducted with partners through the AIFS Network: https://aifs.ucdavis.edu/