ID	Science keywords	AI/ML/data science keywords	Preferred skills	ARS unit and location	Research summary	Potential AI/ML/data science component(s)
1	apiculture, honey bees, pollination, stock evaluation, behavior, ecology	remote sensing, audio, temperature, activity, SQL	knowledge of applied statistics, experience working with SQL files	Honey Bee Breeding, Genetics, and Physiology Unit	We have two honey bee activity projects that we are using to test remote sensing technologies (passive data collection techniques). One complete dataset is in-hand (pollination experiment) and one is in the process of being collected (overwintering experiment), which allows an incoming student to play with the complete dataset and to potentially apply their learning with the other dataset. The initial (completed dataset) experiment was a honey bee stock	Discovering if different honey bee stocks are more or less efficient at foraging compared to others is important to commercial beekeepers as well as almond producers. However, we have not had the chance to develop pipelines to analyze this data yet due to understaffing and time constraints on the staff we have. AI/ML/data science expertise is what we need to complete this project, and it also gives students a chance to learn about a new biological system (pollination
2	Crop improvement, host resistance, genomics	Hidden markov model, data mining	Python or Perl programming experience, Linux working environment and shell programming, interest or exposure to bioinformatics	Crop Production and Pest Control	The project involves genome assembly and annotation of Phytophthora sansomeana (an emerging soybean pathogen) using short and long sequence reads. Following this step, we will use data mining to identify effector genes in the genome that are likely responsible for pathogenicity. We also plan to compare this genome to the genomes of other Phytophthora species to identify genomic signatures (particularly, their repertoires of effector genes)	Genome assembly and annotation require experience in script language programming and familiarity with Linux working environment and its shell language. Machine learning/data mining skills are required to build/improve hidden markov models to identify effector genes.
3	Disease vectors, Animal Health, Genomics, Genetics, Annotation	computational genomics, pattern matching, unsupervised learning	bash command line, nextflow or scientific workflow systems, some bioinformatics in genetics and/or genomics	Arthropod-borne Animal Diseases Research Unit, Manhattan, Kansas	Currently, we are sequencing several vector insects to gain insights into the genetic basis of environmental stress and insecticide responses. The mentee will assemble several HiFi/HiC genomes utilizing existing Machine Learning pipelines, create a baseline gene annotation, and will further annotate for Cis-regulatory modules using specialized existing machine learning pipelines (see: SCRMshaw_HD). CRM annotation will provide insights into gene promoters, a	Incorporating AI/ML and data science expertise into this genomics project can significantly enhance the efficiency, accuracy, and depth of the research. These technologies automate time-consuming tasks, uncover intricate genomic patterns, and provide predictive insights that contribute to a more comprehensive understanding of insect disease vectors and their responses to environmental and insecticidal stressors. Several steps in the proposed analysis
4	plant genomics, plant breeding, crop wild relatives	genome assembly, genome annotation	Unix/bash, R, and/or Python programming knowledge	Kearneysville, WV	The graduate student summer intern will be joining a collaborative project to develop a super-pan-genome for the Malus (apple) genus. This project includes the assembly and annotation of over 25 wild crop relatives of apple genomes. Specifically, the graduate student will be given the opportunity to learn and perform genome assembly and annotation of one Malus species. They will be taught principles of genome assembly, gain experience in annotation procedures,	Genome annotation relies on the use of AI/ML expertise. The current methods for gene annotation require the training of ab initio gene prediction programs that can search a genome sequence for signatures that are associated with genes. Having expertise in AI/ML for this process can improve the prediction accuracy and precision of the annotation programs. Many of the current ab initio gene prediction programs are flexible, allowing for modification or
5	statistics and machine learning, environmental drivers, pathogen prevalence and persistence, poultry farms, pre-harvest, broiler chicken	Bayesian MCMC modeling, logistic regression, random forest, XGBoost, supervised learning	knowledge of applied statistics, R programming, machine learning experience	Egg and Poultry Production Safety Research Unit, Athens, GA	Infections caused by Listeria and Campylobacter through poultry consumption pose significant challenges to public health and the economy. The risk of these infections may escalate with the expansion of the poultry management system. This project is primarily focused on developing predictive models to identify variables and factors that could be linked to the reduction of Listeria and Campylobacter pathogens. The mentee involved in this project will gain hands-on	The datasets to be collected from multiple pastured poultry farms are excellent examples for statistical and machine learning analyses. The size and quality of the data, along with the scale of variables, are likely to circumvent technical difficulties that might otherwise discourage interns.
6	landscape ecology, climate refuge, insect pest forecasting	causal inference, machine learning, deep learning, GIS/remote sensing	knowledge of applied statistics, machine learning and GIS/remote sensing, experience with Python/R computing environment and commonly	Pest Management Research Unit, Sidney, Montana	Mormon cricket outbreaks typically originate on rangeland where infestations can result in substantial forage losses and cause major damage when they migrate into crops. Currently, a forecast model for Mormon cricket outbreaks does not exist. Recent research has demonstrated that Mormon cricket half-lives in the egg stage exceed seven years at high elevations, and thus identification of multi-generational egg beds is imperative to	Deep learning methods employing deep neural networks have proven to outperform more conventional machine learning and statistical techniques like random forest and generalized linear models particularly when dealing with complex spatiotemporal dynamics and non-linear relationship with environmental and climatic variables. The developed deep learning methods will elucidate the spatiotemporal patterns of Mormon cricket populations and provide accurate
7	hail damage assessment, high-resolution imagery, yield deduction	computer vision, predictive modeling	Python, R, Image processing	Water Management and Systems Research Unit, Fort Collins, CO 80526	This research project aims to develop an integrated framework for assessing hail damage in crop fields, leveraging advanced techniques in computer vision, statistical modeling, and data integration. The goal is to provide farmers and agronomists with a reliable tool that can rapidly and accurately evaluate the impact of hail on maize crops and enable informed decision-making for crop management and yield optimization. The intern will contribute to the development of	Besides the benefits of image processing, automated feature extraction, and data fusion techniques, AI/ML expertise will enhance the accuracy and precision of hail damage detection and provide more reliable insights for farmers and agronomists. Machine learning models will provide a predictive capability that will allow stakeholders to anticipate potential crop damage more effectively. Data science experts can ensure that the developed models and tools are
8	biological nitrogen fixation, photosynthesis, crop improvement	genomics/transcriptomics, gene network inference	coding in one of Python/R/C/C++/etc., machine learning	Global Change and Photosynthesis, Urbana IL	Our lab studies how nitrogen assimilation and photosynthesis are coordinated in crops and model plants to achieve optimal growth. We are particularly interested in how the relationship will be affected by climate change. We integrate detailed physiological measurements with genomic datasets to determine the genes and metabolic processes that respond to changes in nitrogen dose and CO2 concentration. Our goal is to identify the transcriptional regulators that	Using the gene expression data we have generated from multiple plant species, the student intern will perform a robust comparison of several machine learning network inference tools. After optimizing and running each tool, the accuracy will be compared based on their ability to predict validated interactions and target gene expression. Several newer algorithms can also incorporate "priors" based on known regulatory interactions which they claim can improve
9	phosphorus, water quality, nutrient management, soil science, spatial ecology	ensemble learning, causal inference, supervised and unsupervised learning, spatial statistics	knowledge of applied statistics, experience with geospatial data, programming experience (Python, R, etc.)	Water Quality and Ecology Research Unit, Oxford, MS	We propose to adapt a geospatial and machine learning approach developed by Dolph et al (2023) (or similar ensemble learning methods) to predict high resolution soil phosphorus (P) concentrations within the Lower Mississippi River Basin. We will utilize an in-house database of soil properties in the Mississippi portion of the Lower Mississippi River Basin, augmented with data compiled from literature and USGS databases to develop and test (an anticipated) random	Data science and ML expertise are needed to integrate the geospatial datasets with the soil chemistry dataset to analyze for predictive relationships.
10	root phenotyping, root architecture, climate resilience, tree physiology	computer vision, structure analysis, image processing	familiarity with image data, programming experience in Python or C++, familiar with use of deep learning libraries	USDA-ARS-AFRS, Kearneysville, WV	Commercial tree fruit production in the U.S. relies heavily on rootstocks. Rootstocks chosen for the current planting conditions by fruit growers are likely vulnerable to future climate scenarios, for example, more frequent and intensive drought. As root architectural traits such as depth, angle, and branching order are indicative of drought stress tolerance and mechanisms, a better understanding of belowground phenotypes is critical to develop climate resilient rootstocks. To	Root architectural traits and their associations with stress tolerance and crop productivity in woody perennials remains largely unexplored. The new tools developed in this project will greatly improve the throughput of root phenotyping. The knowledge of apple roots generated from the proposed work will also allow characterizing the effects of rootstocks on whole-tree physiology related to stress resilience, accelerating the development of climate-smart varieties
11	Bacteriophages, next-generation sequencing	deep learning, natural language processing	Python programming experience, NPL model-related experience etc	Produce Safety and Microbiology Research Unit, Albany CA	In our research project, we will develop a natural language processing-based bioinformatic AI tool to classify bacteriophage lifecycles based on their sequences. First, the interns can help with data collection from public databases to set up the comprehensive training dataset and test dataset. Further, they can assist in model training with different key parameters.	Bacteriophages are viruses of bacteria and play an important role in shaping the ecology and evolution of microbial communities. Based on their lifecycles, bacteriophages can be classified as virulent phages and temperate phages. Virulent phages, the natural predators of bacteria, have become promising antimicrobial agents for antibiotic-resistant bacterial pathogens. Temperate phages are favorable to several applications, such as phage display, genetic
12	sustainable agriculture, crop production, precision agriculture	UAV imaging, machine learning, deep learning, big data	knowledge of statistics, machine/deep learning, and remote sensing, and skills of GIS, Python and graphic user interface (GUI) programming	Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS	We have field experiments related to studies of cover crops, nutrient treatments, and irrigation for the responses of gas emission and microbiology to optimize row crop productivity to close the yield gap. For these experiments, in addition to traditional agronomic methods, innovative remote sensing approaches are being developed and applied. In particular, unmanned aerial vehicles (UAVs) offered a flexible, high-performance platform for crop growth monitoring and	Machine learning provides a model-free paradigm for high-performance crop growth process modeling for biological parameter estimation and yield prediction. Traditionally, linear regression was used to fit the data. However, the data we collected exist complex nonlinear relationships. For known nonlinearity linear regression still can be used by transformed input variables, such as log, exponential, polynomial etc., which may be the rare case in practice.
13	hydrologic modeling, conservation assessment, climate adaptation	random forest, deep machine learning, model optimization	Python programming, hydrologic modeling, water resources management	Grassland Soil & Water Research Laboratory, Temple, Texas	The Soil & Water Assessment Tool (SWAT) is a watershed modeling system used internationally to estimate impacts of land management practices on water quality in complex watersheds. Among other applications, it has been used extensively within the USDA Conservation Effects Assessment Program (CEAP) for evaluating the effectiveness of conservation practices to improve water quality in U.S. watersheds. There are also many possible SWAT applications	The benefit of AI, ML, and data science for SWAT modeling lies primarily in efforts to calibrate the model for applications across large watersheds. Model calibration efforts now require great computational expense, which can push the practical limits of existing high-performance computers. Modern AI and ML technologies can be trained to learn how current model optimization techniques arrive at optimal modeling solutions. This could help reduce
14	Natural resource modeling, spatial connectivity, spatial dependency, soil erosion	Recurrent Artificial Neural Networks	Machine Learning, Python, GIS concepts	National Sedimentation Laboratory, Watershed Physical Processes RU, Oxford, MS	Management of natural resources in fields and watersheds often requires modeling of complex physical processes to guide policy and decision making, remediation design and performance, and monitoring. In natural systems, such processes present complex spatial and temporal dependencies. For example, generation of runoff in a field depends on the timing, intensity, and amount of storm precipitation, and on time-varying hydraulic properties of the soil. The	Although natural resources management relies heavily on computational modeling of spatially distributed physical processes, machine learning has been shown to efficiently replace expensive computations, allowing for more flexible predicting systems that fully consider the natural world complexities. The capability of recognizing contributions from spatially connected neighbors or sequences of neighbors is applicable to a large variety of spatial problems.
15	Unmanned Aerial Systems (UAS), Sentinel-2, precision crop management	Artificial Neural Networks (ANNs) / Deep Neural Networks (DNNs) for Computer Vision, Time Series Forecasting, geospatial data analysis	AI programming experience, e.g., Python, C++, JavaScript, PyTorch, TensorFlow, geospatial data analysis	University Park, PA	Although use of UAS' may be an effective strategy for targeted management practices, costs increase with spectral range, they are time consuming and impractical to cover large acreage over time, and are impacted by a range of environmental conditions such as wind, precipitation, and cloud cover. Satellites provide lower resolution imagery, but it is comprehensive over space and time, and generally includes greater range of spectral bands. Our objective is to try to	The proposed strategy includes using a super resolution CNN (SRCNN) AI model.
16	climate change, microbiome, ruminant livestock, methane, GHG, soil, fertilizer	bioinformatics, computational biology, random-forest ML	graduate students with multidisciplinary interests in Microbiology, Livestock Agriculture, Environment, Veterinary Medicine,	Livestock Nutrient Management Research Unit, Bushland, Texas	The intern will be directly involved in an (1) ongoing meta-analysis for investigating the rumen microbiome's relationship to current methane emissions mitigation strategies. Additionally, the intern could serve as a key contributor to (2) bioinformatic analyses to microbial profiles of soils are primed with cattle manure as an organic fertilizer, its relationship to emissions, and potentially (3) investigating microbiome differences in feedlot surfaces when water is applied	The projects are heavy in computational biology. Therefore, expertise in these tools and techniques are essential for the projects to be successful. Experience in next-generation sequencing and the accompanying computational pipelines for analysis is required.
17	insecticide, physiology, insect behavior, mating	computer vision, causal inference, behavior chain estimation	Python programming, image detection and automation	Center for Grain and Animal Health Research, Manhattan, KS	The graduate student will contribute to a project on quantifying the mating behavior of insects that have been exposed to sublethal doses of insecticides. The goal is to develop an automated systems to track male-female interactions and determine how mating patterns are disrupted by insecticide exposure and resulting female fecundity. The student will have a large set of previously recorded video and still images available and will need to determine meta-data	Many hours can be spent scoring and describing behavioral data by hand, which significantly delays dissemination of information critical for understanding insecticide efficacy and resistance management. By using AI/ML/data science expertise, we can cut time spent hand scoring videos and speed up the analysis of insect behavior and sublethal consequences of insecticide applications. This information is critical for the current and future development of insecticide
18	crop improvement, phenomics, quantitative genetics	computer vision, object detection, segmentation	Linux shell scripting experience, Python programming experience	Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND	Measuring disease response is an important component of plant breeding and genetics research. My lab is working to improve disease measurements of a highly volatile foliar disease called oat crown rust, which is caused by the fungal pathogen, Puccinia coronata f. sp. avenae. Infection of crown rust on susceptible oat varieties elicits strong visual and physiological cues that are important for characterizing response to infection and rate of disease progression on	Expertise in model construction and pipeline development will be necessary for model flexibility and the accuracy of final segmentation maps that serve as foundation (input) for calculation of disease metrics.
19	agriculture, effector biology, plant-microbe interactions, genomic	bioinformatics, data analytics, computational biology, omics integration	knowledge of applied statistics, familiarity with bioinformatics tools, transcriptomics, pathway analysis, Phyton or R, interest in plant-pathogen interactions	Temperate Tree Fruit & Vegetable Research, Wapato, Washington	The overarching goal is to enhance our understanding of the molecular interactions between Phytoplasma pruni, the leafhopper vector, and their plant host. Effectors, key proteins secreted by pathogens to manipulate host defenses, play a pivotal role in the establishment of pathogenicity. This project aims to identify and characterize effectors associated with Phytoplasma pruni and the leafhopper vector, shedding light on their functions and the	Incorporating AI/ML/data science expertise into the project will not only expedite the identification and characterization of effectors but also enhance the depth and precision of our understanding of the complex interactions within the patosystem. Efficient Data Analysis: AI and machine learning algorithms can streamline the analysis of large-scale omics data, facilitating the identification of potential effectors in a more efficient and systematic manner.
20	precision aquaculture, crop protection, remote sensing, image processing	computer vision, unsupervised learning, self-supervised learning	image processing, Python programming, machine learning, GIS	Sugarbeet and Potato Research Unit, Fargo, ND	The student will work on AI-based plant classifications for real-time weed identification and site-specific spray application. Weeds and crop canopy appear similar in color but can be discriminated by spectral signature in beyond-visible bands, morphological features, and geospatial (inter- and intra-row) locations. HR weed images will be used to train learning-based models to extract those features for plant classification. The student will contribute to the CRIS	HR weeds are a significant threat to crop production with their fast growth of 2 inch per day and populating a half million seeds per plant that would destroy sugarbeet fields if not managed. Machine learning (ML)-based classifications can make outstanding improvements for weed identification. High-throughput image analytics using AI algorithm is a key to meet early detection and treatment for timely weed management, reducing yield loss and environmental
21	Crop improvement, Biology, Agriculture, Genetics, Breeding	computer vision, active machine learning, convolution neural network, unsupervised learning, deep learning	Python or R programming experience, knowledge of applied statistics or agriculture or biology	Plant Science Research Unit, St Paul, MN	Maturity-based harvesting time for alfalfa is critical to have a balanced high digestibility and biomass yield so that farmers can harvest maximum digestible biomass yield for maximum profit. Currently, maturity is measured by estimating the percentage of flowers, and an estimated 25% of flowering by human eyes is used to decide the harvesting time. We have developed a deep-learning model to count the number of flowers from images taken by mobile	The project will benefit from AI/ML/data science expertise by applying computer vision to analyze the images, deep learning for object detection to find the flowers, and differentiating flower pixels from background pixels, thus improving breeding efficiency by applying artificial intelligence algorithms to modern agriculture.
22	plant pathology, weed science, microbial community ecology, invasion biology	supervised and unsupervised machine learning	Programming experience (any language), basic knowledge of applied statistics, introductory level coursework in microbiology and/or genetics	Fort Detrick, MD	Non-native invasive weeds cause over nine billion dollars in damage to agricultural productivity and natural resources annually. Microbes can be applied as biological control agents to effectively manage these invasive weeds, but biological control agent development is limited by the rate at which candidate microbes are discovered using conventional methods. The student intern would contribute to research aimed at improving our ability to detect candidate	The project requires computationally intense analyses, large scale data curation, and the use of machine learning approaches. Specifically, the project will benefit from the application of established classification tools for the prediction of gene function and genome content, as well as the adaptation of recently developed machine learning algorithms to categorize microbial lifestyles.
23	drought stress, crop physiology, remote sensing	supervised learning, unsupervised learning	experience with Python programming, specifically Sci-kit learn, Pytorch and/or Tensorflow	Adaptive Cropping Systems Laboratory, Beltsville Agricultural Research Center, Beltsville Maryland	The inhibition of photosynthetic activity and metabolic changes induced by severe drought stress exposure at flowering can be remotely detected via canopy reflectance. These spectra are sensitive enough to differentiate responses to drought that are mediated by genotype, environment and irrigation management in crops, and contain stress-associated signatures which are directly influenced by changes in leaf pigment and metabolite content that occur at	This project will use AI/ML methods to identify signature patterns in the reflectance spectrum that are strongly associated with physiological crop stress responses including leaf water potential (an indicator for drought stress impact on the plant), photosynthetic pigment content and drought-impacted metabolites in cotton and soybean plants that were well-watered or drought-stressed during the critical reproductive stage. The student's role in the project will be
24	crop development, food security, data findability, data accessibility	Data sharing. Data discoverability, findability, and accessibility	Experience with Unix programming, website programming (PHP), application programming interface (API) development (JavaScript),	Crop Improvement and Genetics Research Unit, Albany, CA	Climate change is threating global food availability by straining global food networks by increasing the risk of lowering crop yield due to changing temperatures, as well as abiotic (e.g., flood, drought) and biotic (pest, diseases) factors. There is a dire and critical need to understand genotypic-phenotypic relationships, such as what genes contribute to what crop traits, so that improved crops can be developed to withstand to challenges posed by climate change.	This project will be a superb opportunity for a student who would like to work alongside with programmers and PhD-level scientist at an established ARS biological database for small grains, GrainGenes. The student will be exposed to various data management concepts such as data storage, discoverability, access, and retrieval through a hands-on learning will implement a robust and a secure API that interface a distributed network of servers.
25	foodborne pathogens, assay development, long-read sequencing	bioinformatics, analysis workflow, pipeline development	knowledge of bioinformatics, Galaxy, Python, and command line	ERRC CIFP Wyndmoor, PA	My lab is investigating the use of long-read sequencing for foodborne pathogen detection. The goal is to develop a method that reduces the time and labor needed for foodborne pathogen detection from current methods. Long-read sequencing could be highly advantageous for foodborne pathogen testing because it provides information on the whole genome of a pathogen. Therefore, species, serotype, virulence genes, and antibiotic resistance profile can all	We are optimizing methods to process the meat samples and extract high quality DNA for sequencing, but another important aspect of this method will be the bioinformatics workflow that will be needed to analyze the sequencing data. A basic pipeline has been established, but it would be valuable for us to have an intern that could dedicate time to refining and automating the pipeline. The goal is for a user to input the DNA sequencing file and then receive a report
26	nutrition, ultra-processed foods, nutrients, food categorization, NOVA	multi-class random forest classifier, machine learning model, machine learning prediction	knowledge of Python 3.6.10, MATLAB 2022a, and R, ability to generate PCA plots, familiarity with the jupyter ecosystem of products	NP 107 Healthy Body Weight Research Unit Grand Forks, ND	Earlier this year, a machine learning algorithm developed to predict the degree of processing for foods using the NOVA scale (not an acronym) was published in Nature Communications. This publication included links to the source code for the FoodProX model developed by the investigators. My lab is currently working on several projects related to food processing level, and we are interested in comparing the food processing categorizations of this machine	Currently, the GFHRNC does not have any personnel currently with expertise on ML. A graduate student with a background in these skills would be crucial to our ability to implement the FoodProX model with our current food processing categorization research.
27	Soil moisture, Cover crop, Nutrient management, Digital Soil mapping	Support vector regression (SVR), artificial- and deep neural network	knowledge of soil physics, R/python programming experience	Soil, Water, and Air Resources Research Unit, NLAE, Ames, IA	Fragipan soils are characterized by dense subsurface layers that severely restrict water flow and root growth. Located in southern Illinois, two on-farm studies have been conducted to determine whether the adoption of annual ryegrass (Lolium multiflorum) as a cover crop can improve drainage during the wet portions of the year and reduced crop water stress during the summer for fragipan soils under corn-soybean production. Profile (60 cm) soil moisture and	The graduate intern will contribute toward the calibration and validation of machine learning models, linear regression, support vector regression and recurrent neural network, in predicting soil water movement and soil hydraulic properties like, soil water retention curve. Main three activities of the graduate student will be (i) compare various machine learning model to simulate soil water retention curve of fragipan soils, (ii) training and testing of sensor data to
28	emerging diseases, zoonotic, zoonoses, biothreat, virus, pathology, ticks, Crimean Congo hemorrhagic fever, livestock	agent-based modeling, unsupervised learning, climate, geospacial, ecological	knowledge of applied statistics, Python programming experience, experience with AI/ML	ZEDRU Manhattan KS	I have three projects that would benefit from a graduate student. The first is the development of models of the risk of distribution of Crimean Congo Hemorrhagic fever virus. There is significant concern that the virus may be expanding its geographic range and may find new hosts. There is a need to build models using available data sets. Of particular interest is if the primary ticks if accidentally introduced to the US would be able to establish populations. This	

29	Potato, Breeding, Quantitative genetics, Image analysis	Machine vision, Classification, Deep learning	Python programming skills, Knowledge relating to classification/deep learning	Prosser, WA	The student will build models to classify features of potato tubers (sprouts/eyes, defects, anatomy) and/or processed products (fry quality evaluation) from thousands digital images of potato tubers and fry planks collected by our breeding program this winter (11/2023 – 4/2024).	This will provide cross-discipline training opportunities for the student and the PI (Feldman). The AI/ML/data science models/solutions will be applied directly in the context of our potato breeding program. Results and utility will be communicated to ARS and state university cooperators as well as stakeholders at regional and national meetings.
30	Forage grasses, Rangeland restoration ecology, Unmanned Aerial Systems UAV, High throughput phenotyping HTP, Ecological Services	Deep learning, Neural networks, Supervised learning	R and/or Python programming experience, multispectral UAV imagery	Forage and Range Research Lab - Logan Utah	We are actively developing a HTP project for both turf and forage grasses using unmanned aerial systems - "drones" to estimate several plant traits from high-resolution RGB and multispectral imagery. For many of our main dual-purpose (forage and grain) grasses, we would like to model seed yield (pounds of seed per unit area). Accurate estimations of seed yield modeled from drone imagery can help plant breeders to accelerate their plant material	We envision a process in which deep learning models can be developed and validated to automate the estimation of seed yield directly from drone imagery. Seed yield from grasses is a particularly challenging trait to model with time series of drone imagery. It is even more challenging to detect temporal signals in late spring or early summer imagery that can be used to model future (end of season) seed yield. In addition to being able to detect these signals
31	life cycle analysis, cover crop, climate change, watershed	semi-supervised sequence learning	knowledge of applied statistics and machine learning algorithms	Hydrology and Remote Sensing Lab, Beltsville, MD	This project aims to quantify environmental impacts of agricultural systems with winter cover crops (WCC). The impacts of WCC on reducing agricultural releases are not fully understood. The performances of WCC are often spatially heterogeneous and climate dependent. The graduate student intern will integrate spatially explicit life cycle assessment, machine learning and SWAT modeling approaches to quantify environmental impacts of the WCC systems under	The graduate student intern will build, train and deploy semi-supervised sequence learning models to predict life cycle environmental impacts of the winter cover crops systems under future climate. First, the intern will build and compare semi-supervised sequence learning models like vanilla recurrent neural network, long short-term memory and gated recurrent units to detect all hidden relationships between the features of the in-house datasets. Then, the
32	plant physiology, plant microbiome, plant drought tolerance	DNA sequence analysis, microbiome data analysis, phylogenetics, Clustering, statistical testing of alpha and beta diversity, linear models	knowledge of applied statistics and R programming	Crop Disease, Pests and Genetics, Parlier, CA	In sub-objective 3E of our in-house appropriated research project 2034-22000-015-000D, normal and drought-treated grapevine and citrus root and rhizosphere/bulk soil samples will be collected for DNA extractions. Bacterial ribosomal 16S RNA sequences will be amplified for sequencing. Raw sequences will be demultiplexed and processed through QIIME2 for the identification of amplicon sequence variants (ASVs). Bacterial taxonomy will be assigned using	From the 16S sequencing data, we need to extract useful information such as microbial species constitution, taxonomy, diversity, and changes in response to drought treatment. Analysis of these microbiome data needs the expertise in AI/ML/data science.
33	crop improvement, computational biology	machine learning	knowledge of applied statistics, Python programming experience, and machine learning	Southern Horticultural lab, Poplarville, MS	Step 1: This aspect of the project focuses on using image processing techniques to analyze the color spectrum of blueberries and muscadine grapes. The aim is to classify the fruits based on their color variations, which may indicate different stages of ripeness or quality. This classification can be achieved by capturing high-resolution images of the fruits and applying digital image processing algorithms to extract color features. Step 2: The second part of the	AI/ML techniques can significantly improve the accuracy of color classification and anthocyanins prediction by learning from large datasets of fruit images and corresponding fruit quality data. Automation through AI/ML can process large volumes of image data rapidly, which is beneficial for large-scale agricultural operations. Machine learning models can uncover patterns and relationships in the data that may not be immediately apparent, offering deeper insights into
34	big data, database, data curation, data integration, breeding information management system (BIMS), breeding data management	computer vision, causal inference, agent-based modeling, unsupervised learning, machine learning pipeline, etc	knowledge of applied statistics, Python or comparable language programming experience, use of AI tools for data extraction and manipulation, etc	National Clonal Germplasm Repository, Corvallis, Oregon	The intern would be working with the results of a National Agricultural Library (NAL) - National Plant Germplasm System (NPGS) Rayyan search to extract publications with phenotypic and genotypic data associated with small fruits (strawberry, raspberry, blackberry, blueberry, cranberry) and pear. The data from these publications would then be extracted, converted to specific templates to upload to and make available through public databases (GRIN-Global,	The Rayyan search results cover all crops included in the NPGS and will need to be filtered for the crops of interest to this project. Once the publications on the crops of interest are identified, AI/programming tools would be needed to extract the data and format it for upload into the databases. These methods will be used across the NPGS system for the other crops we preserve. The results will be used to measure impact of genetic resources on various research
35	food composition, nutrition	text mining, natural language processing, data mining, data modeling	programming experience, knowledge of applied statistics	Beltsville Human Nutrition Research Center, Beltsville, Maryland	Our research projects all relate to the data we collect and disseminate in USDA's FoodData Central (https://fdc.nal.usda.gov/), integrated data system that provides expanded nutrient profile data and links to related agricultural and experimental research. We have two data projects that would greatly benefit from support of a graduate student intern as follows: 1. Research Literature Review. This project involves using AI for text mining and natural language	The first project will require expertise in text mining and natural language processing of peer-reviewed literature. An AI scientist would be able to identify the best AI/ML methods for mining and synthesizing information that best accomplishes the goals of the project. The Historical Review of Analytical Food Composition will require data science experience and knowledge in data mining and data modeling. The database includes a diverse collection of connected but not
36	Plant disease management	computer vision, bioinformatic workflow	knowledge of applied statistics, Python programming experience, machine-learning	San Joaquin Valley Agricultural Sciences Center, Parlier, California	Grape viral diseases cause major economic losses worldwide. The most common viruses in grape-grow regions are grapevine fanleaf virus, the grapevine leafroll viruses, red blotch viruses, and the grapevine vitiviruses. High-throughput sequencing (HTS) has been widely used for plant viruses detection and discovery. Commercial software packages are usually costly and lack full annotation capability. In this project, we are seeking to establish standard bioinformatic	Currently available software packages are costly and lack full annotation capability. With the contribution of a graduate student carrying AI/ML/data science expertise, we can reduce the data analysis cost and customize the bioinformatic pipelines for virus detection and discovery in grapevines.
37	aquaculture, oyster	image analysis, image recognition	knowledge of image analysis/recognition AI/ML techniques	Orono, Maine	The primary task for this project is to develop an algorithm that automatically measures width and length of oysters based on a semi-standardized group photograph. The secondary task, if time permits, is to develop an algorithm that identifies an individual oyster based on a photograph of that individual at an earlier time period (smaller size). We have collected images of oysters at several time points in the grow-out cycle for manual image analysis (measuring	This project involve image analysis and recognition that is prohibitively slow, labor-intensive, tedious, or impossible to complete manually. AI/ML/data science expertise would help in developing tools that efficiently accomplish this process for both researchers and industry.
38	Landscape Ecology, Crop Rotations, Remote Sensing	Clustering, Regionalization, Feature Extraction, Unsupervised Learning	Knowledge of applied statistics, Python programming experience, R (optional)	Mandan, ND	The National Agricultural Statistics Service and Economic Research Service recently released the Crop Sequence Boundaries geospatial tool that maps the agricultural fields and historical crop sequence since 2007. This project would use this new tool and the SCINet geospatial common data library to conduct a clustering and regionalization analysis to understand how crop sequence decisions cluster together in the Northern Great Plains. To answer, for example, if	This project involves answering a simple question with complex multidimensional data. Dealing with such complex and variety of data types and identifying the most influential variables to develop efficient models cannot be handled by traditional methods, hence these advanced AI/ML methods. An individual with AI/ML data science expertise will be able to condense this information so that it can be used by producers to better understand the rotational options in
39	livestock disease ecology, continental disease spread, climate, drought, early warning	causal inference, data assimilation, random walks, Bayesian statistics	programming (e.g., R, Python), version control (GitHub)	ABADRU Manhattan, KS	As part of an interdisciplinary group of ecologists, virologists, entomologists, and informaticians at the USDA ARS, we explore the spread of arthropod-borne livestock diseases towards better management practices. Vesicular stomatitis is one such disease that quasi-periodically spreads northwards from endemic regions in south central Mexico to cause large outbreaks in horses and cattle in the United States. Our group links this spread to environmental conditions that	Vesicular stomatitis is a complex disease system, featuring multiple likely insect vector species that have different ecological constraints; quasi-periodic outbreak frequency; and suspected biases in mandatory and centralized reports. AI/ML/ data science play a crucial role filling spatial and temporal observation gaps and creatively generating hypotheses of facilitation and constraints on disease spread, especially when systematically updated with new disease and
40	crop production & protection, sterile insect technique, navel orangeworm	computer vision, deep learning, neural networks	programming experience, knowledge of applied statistics, commitment to use automated imaging software & hardware	Commodity Protection and Quality, Parlier, California	Amyelois transitella, commonly known as the navel orangeworm (NOW), is the most damaging agricultural insect pest for almonds, pistachios, and walnuts in the U.S. The systems approach by ARS scientists in Parlier also involves sterile insect technique (SIT) investigations with collaborators to solve the problem with NOW in tree nut production in California. This research supports the collaborative NOW Program between California Department of Food & Agriculture	This project will bring precision agriculture to the SIT component of the NOW Program by tearing down barriers to the flow of field data for ARS researchers and collaborators. Maximal throughput using a robust AI algorithm will enhance the accuracy of decision-making by CDFA and APHIS-PPQ. NOW SIT operations could be optimized to benefit the tree nut industry in the U.S.
41	precision nutrition, disparities, digital interventions	causal inference, agent-based modeling, data visualization	knowledge of applied statistics, Python, SAS programming experience, familiarity with advanced statistical methods and data visualization techniques	Children's Nutrition Research Center, Houston, TX	Depending on the intern's interest and skill-set, Dr. Thompson has two available projects with large datasets: a videogame study that promoted healthy nutrition to elementary-aged children (n=400); and an online obesity prevention intervention designed for 8-10 year old Black girls (n=324).	For both projects, AI/ML will be used to identify digital phenotypes – i.e., characteristics associated with greatest change. This information will help identify for whom an intervention of this type is most likely to be effective. An intern would help run the analyses, interpret the results, and assist with manuscript development – i.e., they would be an active participant in the team meetings.
42	Handwriting recognition, Image-to-Text-Conversion, Soil Erosion, Conservation	Image processing, environmental modeling, process-based modeling	Knowledge of applied statistics, Python and R programming experience	National Soil Erosion Research Laboratory (NSERL), West Lafayette, IN	The student will apply and modify AI/ML algorithms to extract machine-actionable format from imagery of already scanned record logs of about 50 agricultural field stations records in 29 States. These records from the 1940s and 50s were the foundation of the famous Universal Soil Loss Equation (USLE), the widely used, first quantitative erosion prediction technology. The goal of this project is to develop, train, and validate algorithms with an existing partial digital data	AI/ML techniques can aid the extraction of digital records from imagery of scanned handwritten records. Commercially available software and external services are either too costly or cannot be trusted to be processed without agricultural engineering expertise outside ARS. AI/ML can determine the best algorithms of image-processing and information extraction for quality assurance purposes by comparing the data with validated historical records of crop yields, soils
43	crop improvement, viticulture, metabolomics, genomic, phenotyping	causal inference, unsupervised learning, canonical discriminant analysis	knowledge of applied statistics, Python programming experience, image analysis	Crop Disease, Pests and Genetics, Parlier, CA	The student will work on a combination of high-throughput phenotyping by imagining analyses to evaluate effects on plant health of diseased and water-stressed citrus and grapevines. The automated approaches include obtaining measurements of discoloration of leaves (yellowing or browning), plant size, and plant architecture (i.e., leaf and branch spacing). The student may also develop automated programs to measure differences in fungal cultures to allow quicker	AI/ML/data science benefits this research by allowing faster, more precise measurements to occur. Currently, measurements of plant health are often on a subjective scale (such as 0-5), which is dependent on the human observer. Developing a AI/ML pipeline to make measurements will lead to more objective, consistent and precise results.
44	lipidomics, bovine milk nutrition, dairy forage improvement, dairy big data	statistical analysis by AI, machine learning for big data statistical analysis, R programming for analysis of bovine milk, big data lipidomics of milk	R programming (required), RStudio, interest in AI, ML	Beltsville Human Nutrition Research Center, Beltsville, MD	We have all of the R code written for the data analysis of the first sequence of 88 runs and are in the process of writing those results for publication. We are now ready to expand the data analysis to the full set of all 88-run sequences that we have run by LC-MS as part of the Dairy Grand Challenge. So far, we have 17 sets of 88 runs, each set representing four replicates of two sets of six samples from different times in the feeding study plus five replicates of six	All of the code is written in R, which is free, and open-source, bringing "data science to the masses". The gap in knowledge is the need for more free, widely available applications of R to real-world examples of lipidomic data analysis, with all data freely available online with all code free and downloadable, which is a primary goal of this project.
45	Dairy cultures, Streptococcus thermophilus, antimicrobial peptides, gene expression	supervised learning, causal inference, whole-genome sequencing	Python programming experience, bioinformatics	Dairy and Functional Foods	Our laboratory works on the production of antimicrobial peptides, called bacteriocins by Streptococcus thermophilus, a bacterium used in the production of yogurt. Bacteriocin production in S. thermophilus has been shown to occur naturally within some strains, but most strains are not able to produce these antimicrobials even though they appear to possess the necessary genes. Our laboratory has performed whole-genome sequencing on 15 S.	We believe machine learning could be an essential tool for identifying genetic elements which positively or negatively regulate bacteriocin production within S. thermophilus strains. Within our collection we have strains that: 1) naturally produce, 2) can be induced to produce, or 3) cannot produce bacteriocins, even though they appear to possess the necessary gene clusters. We believe that computational analysis of the sequenced genomes could assist in predicting the
46	Mass Spectroscopy, Biodiesel	Agent-based modeling	The intern should be proficient in a programming language such as Python and be able to employ AI as well as Machine Learning within the framework of that	SBCP, ERRC, Wyndmoor, PA	The parent project for this proposal involves the conversion of lipids to biodiesel. A major challenge faced by the biodiesel industry is the presence of hard-to-remove sulfur-bearing compounds. The sulfur in fuel forms particulates that are a primary contributor to air pollution and the cause of harmful corrosion in the engine. Gas Chromatography-Mass Spectrometry (GC-MS) has been used to identifying the sulfur-bearing molecules in biodiesel. Unfortunately, the	Unfortunately, the GC-MS library is generic and non-specific to chemical pathways. Therefore, computational, and intelligent systems may advance efforts to identify these compounds and assist in the development of chemical or physical means of removal.
47	dairy, rheology, protein formulations, edible films, 3D printing, transport processes	machine learning, multi-parameter regression, optimization	programming experience in Python, and R or MATLAB	PA/Wyndmoor/ERRC/DFFRU	The research project concerns the formulation of edible films and 3D printable structures primarily consisting of dairy proteins, especially casein. Dairy-based edible films have applications as coatings added directly to foods to extend shelf life by acting as a secondary package, thereby preventing waste, and to improve quality and nutrition, or to serve as stand-alone packaging to maintain freshness. They have additional benefits over conventional plastic-	The above project would benefit from AI expertise with the development of a machine learning algorithm to test formulations for film making and 3D printing. Input parameters would potentially include protein type (sodium caseinate, micellar casein, casein and whey protein isolate mix), protein concentration, cross-linking type, fat content, plasticizer content, etc. Model outputs would likely be parameters including yield stress to characterize flowability and
48	crop improvement	Object detection AI models (e.g. YOLO, Faster R-CNN, Single Shot Detector (SSD))	Python, computer vision	FNPRU, Beltsville, MD	Detection and counting systems of tillers are important to enhance yield efficiency. Deep learning-based models that can effectively evaluate genetic variations in tillers of interspecific bontgrass mapping population will be developed. Most of the time, accuracy of image detection is a critical factor; however, when temporal progression of trait development (e.g. tillering, stress symptoms, etc) is being evaluated, image processing speed is as important as detection	Such artificial intelligence (AI)-based high-throughput phenotyping platforms can significantly increase the effectiveness of germplasm evaluation. For example, time to count tillers of a few thousands of genotypes would take a month while this platform (once developed) can reduce 1/10 of the total processing time. It takes time to develop a platform; however, once it is developed, it will allow us to effectively and precisely evaluate genetic performance of yield
49	rangeland ecology, ecological restoration	neural networks, time-series analysis, spatial predictions, remote sensing	knowledge of spatial statistics, R or python programming experience, familiarity with or willingness to learn about rangeland ecosystems	Jornada Experimental Range - Range Management Research Unit, Las Cruces, NM	We are seeking a graduate student intern to contribute to a research project focused on evaluating the viability of the Rangeland Analysis Platform (RAP) vegetation models to evaluate restoration treatment effectiveness in southern New Mexico. This project will be an integral part of the RestoreNM project, which aims to enhance our understanding of ecological restoration practices in arid systems so that we can most efficiently implement future rangeland	The project described would greatly benefit from data science expertise given the substantial size and complexity of the datasets, particularly the fractional cover products. An intern with applied data science skills could optimize tools like the Rangeland Analysis Platform for processing and interpreting this data, not only for the current project but also for future applications in assessing treatment effectiveness. This expertise would streamline data analysis,
50	bees, pollinators, genomics, conservation, beenome100	machine learning, unsupervised learning, supervised learning	Experience with Python, Linux, Bash scripting, genomics and artificial intelligence	Pollinating Insect - Biology, Management, Systematics Research (PIBMSR), Logan, UT	My research unit is a hub for USDA's beenome100 project, which seeks to collect, sequence, assemble, annotate, and analyze 100+ genomes of important U.S. bee species, relying heavily on SCINet resources to do the computational work. The project has acquired data for over 90 bee genomes and is starting to analyze the data in a rigorous, efficient, and reproducible way. My unit has a postdoc who is specifically assigned to working on beenome100, Dr. Rena	Artificial intelligence methods are just starting to be applied to the analysis of genomic data. Areas of interest include genome annotation, gene family curation, automated genome scaffold curation, and protein conformation prediction. Having a student learn what AI tools could be incorporated into and/or developed for beenome100 would speed up productivity and create new knowledge.
51	plant disease, bacterial pathogen, Xylella, fastidious prokaryotes	computer vision, DNA sequence analyses, classical machine learning statistics, dimensional reduction, supervised and unsupervised learning	knowledge of applied statistics, DNA sequence handling and manipulation, programming experience, basic bacteriology and bioinformatics	Crop Diseases, Pests, and Genetics Research Unit, 9611 South Riverbend Avenue, Parlier, CA 93648	The graduate student is expected to utilize machine learning (ML) tools using ARS computational resources to analyze available genome sequences of Xylella and/or Liberibacter available in GenBank and in our lab collections to identify/describe pathogen population diversity. One immediate goal is to resolve current taxonomy issues that facilitate pathogen detection and disease diagnostics. The first approach is to use classical ML tools such as	Analyses of large sequence data requires helps from computational expertise. Our lab is in the frontline of biology research and working on using ML approach to further advance Xylella and Liberibacter research. This is stated in our current CRIS plan (2034-22000-015-000D) Sub-objective 1B: Characterize metagenomes of Xylella spp. and "Ca. Liberibacter spp." infected samples using machine learning (ML) focusing on improvement of taxonomic identification.
52	swine, feeding behavior, precision livestock farming, illness detection, animal welfare	convolutional neural network, time series prediction, support vector regression	R programming experience, experience with time series data, knowledge of convolutional neural networks	Genetics and Animal Breeding Research Unit, U.S. Meat Animal Research Center, Clay Center, NE	The U.S. Meat Animal Research Center (USMARC) currently possesses feeding behavior data for more than 12,000 pigs and continues to add data for ~1,900 pigs every year. We have analyzed various aspects of feeding behavior in 4 published and 2 soon to be published studies. We have developed a modeling approach that allows us to predict a pig's feeding behavior in real-time. We have used this model to identify anomalous feeding behaviors related to the onset of	Development of the ML approach for anomaly detection will require assimilating and evaluating large data sets for subtle nuances reflecting critical changes in pig behavior associated with specific stressors.
53	. insect detection, insect acoustic trapping, insect pest monitoring, mating disruption	unsupervised learning, convolutional neural networks, acoustic signal/noise discrimination	Python programming experience is helpful, general knowledge of SAS or R is helpful	Insect Behavior and Biocontrol Research Unit, Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, FL	Ongoing projects that involve 1) the use of convolutional neural networks to optimize disruption of vibrational mating communication in Asian citrus psyllid, and 2) acoustic detection of hidden infestations of palm weevil in palm trees in the presence of background noise would benefit from research that a graduate student could perform. One former high school student participated in the Asian citrus psyllid project and has won awards at the International	There are many options for applying convolutional neural networks to insect acoustic signal identification and discrimination questions. For example, a student paper in preparation in my laboratory described comparisons among feature extraction procedures such as Mel Filterbanks, LEarnable Audio Front End (LEAF), and a custom-written front end. Once feature extraction is optimized, Generative Adversarial Networks with different hyperparameters can
54	Tribal Lands, Space-based water management, Precision decisions, Food sovereignty	Neural network analysis, Supervised learning, Remote sensing	Python programming experience, MATLAB, Remote sensing postprocessing. What is your approach to mentoring and why would you be a strong	Fayetteville, AR	Current irrigation systems result in over half of irrigation and nutrients being lost to the environment. Digital soil mapping-based precision irrigation management zones, nutrient and micro drip systems, soil and weather sensors, and satellite-based crop stress methods are all estimated to improve 'crop per drop' by ~30% when carried out in isolation, however, no research has been done to combine water and nutrient saving methods for precision decisions.	Irrigation and water management is a complex decision made by humans with limited data, particularly under a changing climate. PIs have access to proprietary 10 m resolution, repeated every 4 hrs thermal, hyperspectral, and multispectral data (only scientists in USDA with access). The use of these proprietary data in this project will further advance water-carbon-energy efficiency gains in water-deficit systems. Tribal reservations in the US have limited information
55	nutrition, food science	natural language processing (NLP), large language models, GPTs	Python programming, shell scripting, git, API usage, NLP basics	Immunity and Disease Prevention Unit, Davis, CA	Last summer, the graduate student's (Alex's) task was to create generalizable "food mapper" software to map unlinked databases in nutrition and food science. Alex developed software for both database mapping and for ingredientization (determining the ingredients in a mixed dish, e.g. pizza = flour, cheese, tomatoes, etc.): Mapper: https://github.com/alexraposo2000/String-Matcher-USDA Ingredientizer + thesaurus: https://github.com/alexraposo2000/Ingredientizer-	Although both postdoctoral scientist co-mentors have applied some natural language processing in their projects, neither are trained as a computer scientist (both are nutrition scientists).
56	Chronic Wasting Disease (CWD)	machine learning, AI, big data, computer vision	Python and/or R programming experience, familiarity with YOLO, knowledge of applied statistics	Produce Safety and Microbiology Research Unit; Albany, California	Over the summer a graduate student developed a method to identify the location of the free ranging captive Tule elk in Tomales Point (Point Reyes National Seashore, Northern California). I would like to extend this approach to areas of the country that contain free ranging elk that have a high incidence of chronic wasting disease (CWD). In addition, I would use this analysis to measure an animal's physical characteristics (size and orientation) and propinquity to other	Wild cervids are susceptible to a transmissible and invariably fatal prion disease called Chronic wasting disease (CWD). CWD-infected animals tend to drift away from their herd mates, males are disproportionately infected with CWD at a higher rate, and CWD infected females produce fewer fawns. In principle, AI and machine learning approaches may be used to identify wild cervids and measure female/fawn ratios, animal propinquity, gender propinquity, density, etc.

57	climate smart agriculture, hydrologic modeling	code modules, data extraction, conversion of data to model readable formats, high performance computing	ability to code modules in programming languages such as Python, R or MATLAB, basic understanding of high-performance computing and Big	Sustainable Water Management Research Unit, Stoneville, MS	The candidate will aid in developing a module to project the outcomes from existing hydrological models like Agricultural Policy Environmental Extender (APEX) into the future using different climate scenarios from global circulation models (GCMs) recommended by the Intergovernmental Panel on Climate Change. To do this the candidate will a) collect the climate data from various sources, b) develop or upgrade module to convert the data to the model	An intern with AI/ML/data science expertise would greatly benefit this project by expediting a lot of our data mining and data conversion time. By shortening the link between obtaining the climate data from GCMs and producing outputs from different climate scenarios in a calibrated model, we will be able to more efficiently and effectively produce visualization tools for climate scenario analyses through statistical inferences that aid stakeholders in decision-making in a
58	crop genetics and genomics, crop improvement, polyploids, genetic introgression, plant-pathogen interaction	variant calling, SNP-machine learning, deep learning	knowledge of bioinformatics, experience in programming languages such as Python, R, and Perl	Dawson, Georgia	The research project to which a graduate student intern could contribute, is part of the peanut genetics and germplasm development program at the National Peanut Research Laboratory. The program has generated genome and transcriptome data for a variety of peanut species with different ploidy levels and genome types. One of the objectives of this project is to identify and annotate genetic variants, primarily single nucleotide polymorphisms and short insertion-	The development of variant calling pipelines, together with machine learning-based variant refinement, will aid in the accurate identification of true SNPs in peanuts with different levels of genome complexity. The incorporation of ML algorithms will also increase the precision of SNP discovery in wild diploid relatives of peanut, particularly those with genomes that are relatively distant from currently available references. High-confidence SNPs/INDELs could serve as
59	diapause, dormancy, overwintering, stress, transcriptomics, gene expression	artificial neural networks, decision trees, support vector machines	R and/or Python programming experience, applied statistics, machine learning	Insect Genetics and Biochemistry Research Unit, Fargo, ND	Dozens of transcriptome studies have borne insight into the mechanisms that facilitate insect dormancy; a life stage that is critical for the proper management of a variety of economically important species. However, tools for identifying conserved patterns of gene expression among these large transcriptomic datasets are not well developed, making valuable comparative analyses of these publicly available data not feasible. Therefore, we propose the development	The Insect Genetics and Biochemistry Research Unit has an extensive background in stress physiology and using transcriptomics to investigate the mechanisms underlying complex phenotypes. However, our Unit lacks experience applying machine learning to investigate these mechanisms at a broader scale. Additionally, these machine learning techniques are applicable to a variety of other 'big-data' projects currently underway and planned for our Research Unit
60	soil health, soil spectroscopic sensing, soil spectrometer	robotic AI, machine learning, vision Transformers, long-short-term memory networks, PDE networks	knowledge of statistics, programming experience	Genetics and Sustainable Agriculture Research Unit	Soil health is the key to ensuring the sustainability of agriculture and climate resilience. A preeminent challenge to improve soil health is the lack of accurate and widespread data and models for robust assessment. We propose Machine Learning (ML) techniques to develop, calibrate and validate models for measuring soil health parameters using robotic spectroscopic sensors. The graduate student intern will contribute to ML algorithm design, data analysis, and	Big data, artificial intelligence, and machine learning are powerful tools that rely on high quality data input, particularly when working with complex, interconnected datasets in agriculture. Multisource high-dimensional spectroscopic images have high computational requirements and require high throughput algorithms to process effectively. This research will utilize advance state-of-the-art deep neural networks (including Vision Transformers, Long-Short-Term Memory
61	biological invasions, ecology, and arthropod systematics	data analysis, linear regression, statistical power, machine learning, data visualization, Python programming	Python, data visualization, dashboards, familiarity with applied statistics What is your approach to mentoring and why would you be a strong mentor	Systematic Entomology Laboratory, Washington, DC & Beltsville, MD	Biological invasions cost billions in loss of ag-important crop commodities, and this project aims to evaluate the predictive power of existing pest interception data to anticipate high-risk arthropod biological invasions. We aspire to generate a predictive machine learning model that leverages >1.6M port- and border-intercepted arthropod records (e.g., commodity, year, season, country of origin, and taxonomy) to identify patterns within the interception data to	