T32: AI + Precision Nutrition Research

Trainee Handbook

Advanced Training in Artificial Intelligence for Precision Nutrition Science Research (AIPrN) Institutional Research Training Programs

NIH T32 Grant

PROGRAM TITLE:

Increasing and Diversifying Future AI-Precision Nutrition Research Workforce to Promote Nutrition Health Equity among Underserved Populations

This document is subject to annual revision.

Last edited date: *August 20, 2024*

I. Purpose

The objective of this training program is to address the shortage and lack of diversity of individuals trained in using computational approaches and precision nutrition to address nutritional health disparities. One goal is to provide predoctoral and postdoctoral students with rigorous multidisciplinary training in computational data skills, including Artifiical Intelligence (AI) and machine learning (ML), along with providing precision nutrition knowledge to address health disparities questions pertaining to diet quality, obesity, type 2 diabetes, cardiovascular disease, and cancer. Students will also be mentored by faculty with various areas of expertise in data science, biostatistics, nutritional sciences and research methods, in order to foster a cross-disciplinary training experience.

II. Description

This program is designed to increase the number of underrepresented predoctoral students or postdoctoral fellows at Prairie View A&M University (PVAMU). Underrepresented students include various characteristics such as:

1. Racial and ethnic groups that are demonstrated to be underrepresented in health-related sciences by the National Science Foundation
2. Individuals with disabilities, defined as a physical or mental ailment that substantially limits one or more life activities
    a. Individuals from disadvantaged backgrounds, defined as
    b. Individuals who come from a family with an annual income below established low-income thresholds, which are published by the U.S. Bureau of the Census based on family size.
    c. Individuals who come from an educational environment such as that found in certain rural or inner-city environments that has demonstrably and directly inhibited the individual from obtaining the knowledge, skills, and abilities necessary to develop and participate in a research career

Each year over the project period, four trainees will participate in the program. This will include 3 predoctoral students from the Eectrical and Computer Engineering department with an AI/ML background, along with 1 postdoctoral fellow with precision nutrition or nutritional sciences background stationed at the College of Agriculture, Food and Natural Resources. Predoctoral trainees will rotate at the Baylor College of Medicine (BCM) USDA/ARS Children's Nutrition Research Center (CNRC) to gain hands-on research experience in nutrition-related projects. Each trainee is expected to have spent at least one year in their respective educational program prior to admission to the T32 program, and to remain in the T32 training program for a period of 2 years.

III. **Trainee Role and Expectations**

Curriculum

Trainees are expected to complete both a didactic training curriculum and a research training curriculum.

The didactic training curriculum consists of various courses offered by Eectrical and Computer Engineering Department and Department of Agriculture, Nutrition and Human Ecology at PVAMU, and the Graduate School of Biomedical Sciences at BCM. See **Appendix I** for a list of the courses offered to predoctoral students, and Appendix II for post-doctoral students. The list includes suggested core and elective courses. Course offerings may vary based on training needs.

The goal of the research training curriculum is to conduct interdisciplinary research under the mentorship of one or more Program Faculty mentors. Trainees will be assigned to the Program Faculty co-mentors to engage in various sample research projects, laboratory rotations, and presentations at conferences.

Responsible Conduct of Research (RCR)

Objective: foster a culture of integrity in precision nutrition and health disparities science.

Format: Responsible conduct of research (RCR) education will be provided through a variety of training formats with attention to in-person (face-to-face) instruction, online, attendance of seminars and workshops on research ethics, real-time discussions with their peers during weekly 'trainee group' meetings, and as informal conversations during mentoring sessions and lab group meetings about ethical conduct. Trainees will be asked to include RCR practices in their written summary or progress report of their research experience. At least two RCR focused discussions per year will be included in journal clubs which will meet monthly.

Faculty Participation: the MPIs (Drs Kim, Antwi, Musaad) will be responsible for ensuring the trainees obtain this training, either in-person at their respective institutions (PVAMU, BCM) or online. Other key faculty in the training program from both institutions will be asked to include dimensions of RCR in their interactions with the trainees. In addition to virtual training, a series of in-person trainings (e.g. faculty mentor, compliance office, etc.) will be organized to cover RCR related topics. These seminars will total up to four hours per month of in-person training, and will cover the following topics:

- Mentor/mentee responsibilities and relationships,
- Collaborative research including collaborations with industry
- Peer review

- Data acquisition and laboratory tools, management, sharing and ownership
- Responsible authorship and publication
- Other safety courses, such as laboratory safety

Duration and Frequency of Instruction: Trainees are expected to devote at least 4 hours per month on RCR instruction on at least one of the 5-year training duration except for time spent reviewing/refreshing Collaborative Institutional Training Initiative (CITI) training modules.

Monitoring of participation:

Trainees will be required to provide documentation on the contents and dates of their participation in RCR activities (e.g., form, seminar requirement, certificate) will be included within a dedicated section in the Progress report (twice per year).

Faculty will be required to provide documentation on their inclusion of RCR discussion and other related activities as part of their engagement with the trainees. RCR documentation will be included within a dedicated section in the respective reports or in a separate standard form.

Theme	Topic	Instruction format/responsible person(s)/area	Institution
Scientific research	Ethical considerations in health disparities research	Face-to-face (Dr Antwi, Dr O'Connor). Discussion include measurement methods and cultural sensitivity in recruitment and retention of minorities, low-income families	PVAMU, BCM
	CITI training modules	Online. Biomedical Research - Basic/Refresher Biomedical Refresher 101 Biomedical Refresher 200 Biomedical Refresher 201 Social & Behavioral Research Refresher 101 Social & Behavioral Research Refresher 201 Health Information Privacy and Security (HIPS) Basics of Health Privacy Health Privacy Issues for Clinicians Health Privacy Issues for Researchers Basics of Information Security, Part 1 Basics of Information Security, Part 2 Research Privacy & Security Training Good Clinical Practice module	Responsible Conduct of Research (RCR) | CITI Program (https://about.citiprogram.org/series/responsible-conduct-of-research-rcr/)
Data use	Big Data Ethics	Face-to-face (Dr Musaad, Dr Kim). Discussions about privacy, security, data validation, transparency, responsible use of predictive analytics, use in healthcare.	PVAMU, BCM
	Bias in biomedical and precision nutrition research and relevance to heath disparities	Face-to-face (Dr Musaad, Dr Kim, Dr O'Connor). Discussion topics include data integration, racial/ethnic bias, implicit racial bias, equity of methods, unbalanced representation of individuals	PVAMU, BCM
	NIH Data management and Sharing Policy	Online: Webinar I: Understanding the New NIH Data Management and Sharing (DMS) Policy**,** Webinar II: Diving Deeper into the new NIH Data Management and Sharing Policy	NIH
Research environment	Equity and inclusion	Online: Unconscious Bias Course. Topics include addressing personal unconscious bias, microaggressions, a solutions toolkit, develop self-awareness, bias and disparities in medicine and healthcare.	AIM-AHEAD Courses: Unconscious Bias
	Equity and inclusion	Face-to-face (Dr Musaad, Dr O'Connor, Dr Antwi). Discussion on sexual, racial, ethnic, disability and other forms of discriminatory harassment.	PVAMU, BCM
Publications	Responsible authorship	Face-to-face (Dr O'Connor, Dr Xu) and in-lab seminars. Discussion topics include publication bias, acknowledgement of	BCM

Theme	Topic	Instruction format/responsible person(s)/area	Institution
		courses, interpretation of findings, case studies (e.g. retracted manuscript) or examples.	
	Responsibilities as a reviewer	Face-to-face (Dr O'Connor, Dr Musaad). Discussions on conducting objective review based on scientific content and quality of research, unbiased and justifiable review, consider ethical validity of the study.	BCM

Reproducibility and responsible workflow

Trainees will be instructed on the importance of rigor, reproducibility and responsible workflow in the data science, statistics, and nutrition field and how to identify and avoid bias in nutrition precision research. Given the grave implications of lack of rigor on study conclusions, it is imperative to clarify and demonstrate how elements of rigor and reproducibility are engrained in every step of the research process. The instruction in methods for enhancing research reproducibility will be integrated into the overall training program and in accordance with the stages of the trainee development cycle. This strategy will advance their skills and confidence in computation, data management, documentation, and interpretation, skills that are transferable to research jobs and elsewhere.

Format: The training experience will be provided via a variety of methods including in-person (face-to-face) discussions with peers and mentors, attendance of seminars and workshops, as informal conversations during mentoring sessions, lab group meetings about ethical conduct, and as part of coursework or educational curricula. Moreover, several of the didactic AI/ML courses at PVAMU such as *Machine Learning for Engineering Applications* and *Statistical Learning for BigData* already address statistical rigor and reproducibility of data analysis and this will be reinforced.

Faculty Participation: the MPIs (Drs Kim, Antwi, Musaad) will be responsible for ensuring the trainees receive the required instruction and that supporting faculty mentors explicitly include elements of enhancing rigor and reproducibility in the trainees' research projects.

Duration and Frequency of Instruction: Trainees are requested to devote at least 2 hours per month per year of training for receiving instruction in methods for identifying and enhancing rigor and reproducibility in precision nutrition research, AI/machine learning and health disparities research.

Monitoring of participation:

- Trainees will be required to record the content and dates of their participation in instruction on methods for enhancing rigor and reproducibility and document this information in the Progress reports.
- Faculty will be required to provide documentation or information about their plan to include methods for enhancing rigor and reproducibility, releavnt discussion topics, and other related activities as part of their mentoring.

A list of suggested topics by component is provided below. It is not exhaustive and other topics may be considered based on resarch objectives ore program goals.

Component: Designing powerful studies

- Sample size calculation and power assumptions
- Developing the analysis plan a priori
- Defining the experimental unit
- Controlling sources of variation
- Assessment of dietary intake
- Bias in collecting nutrient intake data
- Issues with correlated measurement errors
- Avoiding sampling bias
- Integrating social determinants of health by design

Component: Ensuring rigorous data analysis

- Influence of sampling bias on predictive models
- Detecting and evaluating batch effects in gene expression studiesStatistical and inferential challenges when using clustered data in nutrition research
- Transparency in data analysis strategy
- Estimates of effect size
- Methods for addressing Data heterogeneity and missingness

Component: Evaluating your results

- Prediction evaluation
- Validation methods in big data
- Statistical conclusions validity
- Transparency in reporting and presentation of results
- Generalizability of findings

Component: Special topics

- The power of good data documentation and successful collaboration in ensuring reproducible research

- The American Statistical Association's Statement on p-Values: Context, Process and Purpose
- Advantages and pitfalls of Machine Learning methods to model non-linear relationships and confounding in nutritional data

Mentorship and Professional Development

After choosing a lab group and with concordance of the proposed co-mentor, the trainee will need to develop an Individual Development Plan (IDP) with their co-mentor. This will include measurable career goals with congruent research/learning opportunities that pertain to the training program and trainees' career level. The IDP will also include training objectives and plans for presentations and conferences, participation, applied methods and technical experience, grant application method, and plan for responsible conduct of research.

Trainees will receive extensive one-on-one mentoring on attending conferences and presenting data, career development training, professional communication and networking. This mentoring and training will be conducted in conjunction with the PVAMU Graduate School and Career Services. Trainees will take a career assessment questionnaire using the Focus 2 Career & Education Planning Systems to determine their career interests, values, skills and career preparedness. In addition, mock interviews and a resume development workshop will instruct trainees on how to prepare their curriculum vitae, resumes, and professional online profiles to highlight the impact of their accomplishments. Finally, training will be provided in other topics to increase awareness and understanding of how to develop fruitful mentoring experiences, make racial equity more explicit in their projects, education about marginalized groups and biases, and importance of access to healthy lifestyles and sustainable dietary practices for the prevention of chronic diseases.

Rotations and lab project

Prior to starting the research project, trainees will be given the option to rotate for 2-3 weeks across different CNRC lab groups. This experience will expose the trainees to different lab environments (e.g., 'dry' vs 'wet' lab setting, plant vs. Mouse, etc.), develop leadership and enquiry skills, and explore project possibilities with potential mentors that fit their background, interests, and career objectives.

All research projects will be developed in collaboration with PVAMU to ensure consistency with their PVAMU training experience and career trajectory. PVAMU and CNRC faculty will help narrow down the research questions and help identify available data sets that they could explore or use to address the research questions. Projects can utilize publicly available data through sources like the USDA or NHANES, or the co-mentor's own data.

Once the research project is developed and assigned (mostly year 2), each trainee is expected to

1. Spend at least one in person day per week in CNRC
2. Interact with other members of the training grant via weekly trainee group meeting and monthly journal clubs to discuss their training experience, lessons learnt, and research topics in AI/ML advances in nutrition, obesity, and health disparities research.
3. Submit a written report of activities consistent with program guidelines upon completion of their experience, an annual progress report, and give a presentation summarizing their experience.

IV. Mentor Roles and Expectations

There are several roles and responsibilities of mentors. Mentors may develop and teach courses or training materials in a vairety of subjects including nutrition, genetics, imaging, microbiome, physiology, behavioral factors, cultural and social influences, and environmental exposures. They must contribute to trainees professional development in areas like leadership, career development, and grant writing and manuscript development. Finally, they must engage in research mentorship through laboratory rotations for sample research project activities.

There are several selection criteria to become a faculty mentor. These include: history of successful research and mentorship, commitment to training, research focus that fits within the area of study of the program, existing funding, collaboration with program faculty, engagement with the program, and access to big data.

Current PVAMU faculty mentors include Lijun Qian, Xishuang Dong, Andrea McDonald, Gloria Regisford, Victoria Mgbemena, Javad Barouei.

Current BCM USDA-ARS CNRC mentors include Teresia O' Connor, Yong Xu, Deborah Thompson, Jennette Moreno, Kathleen Motil, Jayna Dave, Alli Antar, Alexis Caroline Wood, Erica Soltero, and Robert Waterland.

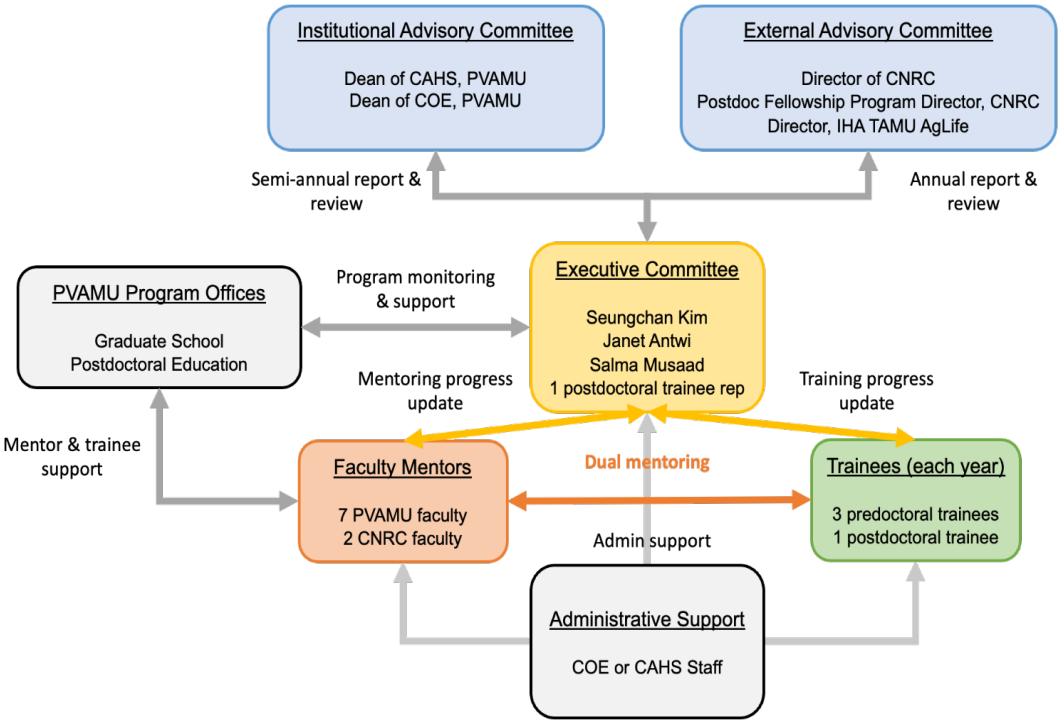
V. Timeline

Activity	Year 1	Year 2	Year 3	Year 4	Year 5
Conduct coursework	X	X			
Course progress and completion	X	X			
Conduct rotations		X	X		
Identify mentor		X	X		

Participate in online messaging system (Slack)		X	X	X	X
Develop IDP		X			
AIPrN annual workshop		X	X	X	X
Progress report (2/year)	X	X	X	X	X

VI. Research Training Program Organizational Structure

The training program is a cross-disciplinary, an inter-departmental, inter-college and inter-institutional program and will have three PD/PIs. The three persons to serve as multiple program directors are Dr. Seungchan Kim, Dr. Janet Antwi, and Dr Salma Musaad. The three program directors are strongly committed to the success of the trainees and will effectively coordinate with each other to ensure adequate mentoring and management that the program will require. They will handle the administration of the grant, scientific leadership, monitor trainees progress, supervise and manage recruitment and appointment of trainees, and monitor the progress and overall project accomplishments as well as responsible for all evaluations and progress reports. **Figure 1** presents the adminsitrative structure of the training program.



The training program will be monitored and evaluated by two Advisory Committees (Institutional Advisory Committee and External Advisory Committee) and the Office of Research and Graduate Studies at PVAMU. The Advisory Committees will be closely engaged in review of the program materials and methods, activity selections, evaluation of the quality of work and the decisions on

outcomes in the project curriculum. The Advisory Committees will also evaluate the overall effectiveness of the training program relative to other comparable programs, with specific programmatic guidance and review related to trainee selection and progress, effectiveness of the didactic curriculum and research enrichment activities, and program assessments to ensure that overall program requirements and goals are met. The criteria assessing the career development and progression of trainees include, but not limited to, diversity in gender identity and underrepresented groups, publications, proposals submitted and external fundings secured, degree completion, and post-training position. Evaluations will be conducted twice a year to allow for implementation of feedback for the following academic year. Feedback will be received from the trainees, faculty mentors, and Advisory Committees.

VII. Forms

Trainee feedback about the Mentor

Mentee progress report

Individual Development Plan

Project Description

Sustainability Assessment (TBD)

VIII. Abbreviations

To be added

IX. Appedix

Predoctoral courses

Year	Course Name	Course Type
Pre-T32	ELEG 6360 Modern Artificial Intelligence	Elective or Core
Pre-T32	ELEG 6316 Statistical Learning for Big Data	Core*
Pre-T32	ELEG 6318 Deep Learning	Elective
Pre-T32	GNEG 5315 Data Ethics	Core
1	NUTR 5304 Fundamentals of Integrated AI-Precision Nutrition	Core
1	NUTR 5310 Nutrition Assessment	Core
1	NUTR 5312 Social Cultural Influences on Nutrition	Elective
1	ELEG 4377 Machine Learning for Engineering Applications	Core*
1	ELEG 6360 Modern Artificial Intelligence	Elective or Core

1	ELEG 6380 Introduction to Bioinformatics	Core
1	ELEG 6316 Statistical Learning for Big Data	Core*
1	ELEG 6318 Deep Learning	Elective
1	GNEG 5315 Data Ethics	Core
2	NUTR 5311 Nutrition in Public Health	Elective
2	NUTR 5323 Nutrition and Metabolism II	Elective
2	NUTR 5324 Research Seminars in AI-Precision Nutrition	Core
2	NUTR 5326 The Human Microbiome: Comprehensive Methodologies & Data Analysis	Elective
2	NUTR 5328 Advanced Research Methods, Statistics and Data Analysis in Nutrition	Core**
2	NUTR 5329 Analytical methods for health equity research in nutrition and health	Core**
2	GS-GG-6204 Method and Logic in Genetics & Genomics	Elective
2	GS-IY-6205 Microbiome Methodology & Data Analysis	Elective

Appendix II Post Doctoral Courses

Year	Course Name	Course Type
1	NUTR 5304 Fundamentals of Integrated AI-Precision Nutrition	Core
1	ELEG 6360 Modern Artificial Intelligence	Core
1	ELEG 6316 Statistical Learning for Big Data	Core
1	ELEG 6380 Introduction to Bioinformatics	Elective
1	GNEG 5315 Data Ethics	Core
1	NUTR 5302 Nutrition Informatics	Elective
1	NUTR 5327 Diet and the Gut Microbiome	Elective
2	NUTR 5324 Research Seminars in AI-Precision	Core
2	ELEG 6318 Deep Learning	Elective
2	NUTR 5328 Advanced Research Methods, Statistics and Data Analysis in Nutrition	Core
2	NUTR 5329 Analytical methods for health equity research in nutrition and health	Core
2	GS-GG-6204 Method and Logic in Genetics & Genomics	Elective
2	GS-IY-6205 Microbiome Methodology & Data Analysis	Elective