Fangzhou Li

Davis, CA 95616

✉ fzli@ucdavis.edu | 🔗 in/fangzhou-li-8a9359155 | 👤 Personal Website

SUMMARY

Fifth-year Ph.D. candidate in Computer Science developing in machine learning (ML) solutions to interdisciplinary problems. Specialized in large language models and knowledge graph construction. Proficient in fostering cross-disciplinary collaboration, working alongside experts from diverse fields.

EDUCATION

Ph.D. Candidate in Computer Science, *University of California, Davis*   *Davis, CA*
- GPA: 3.9/4.0   *September 2020 - Present*
- Expected Graduation: July 2026

B.S. in Computer Science and Applied Mathematics, *University of California, Davis*   *Davis, CA*
- Award: Luther D. and Marie M. Davis Scholarship   *September 2015 - July 2019*
- Honor: Phi Kappa Phi, Dean's Honors List

MAIN RESEARCH PROJECTS

FoodAtlas - *Website*, *GitHub*   *University of California, Davis*
First Author - *Computers in Biology and Medicine* | **Keywords** - Deep Learning, PyTorch   *Ongoing*
- Developed a semi-automated pipeline that extracted domain-specific knowledge from scientific literature using the BioBERT large language model (LLM), achieving an F1 score of 0.83.
- Incorporated active learning workflow into the pipeline, accelerating annotation efficiency by 38%.
- Constructed the largest evidence-based food knowledge graph with 285K triplets extracted from 155k scientific literature.
- Collaborated with web developers to coordinate the project's delivery of a web service, contributing to user study and outreach events.
- Currently prompting OpenAI GPT-4 and fine-tuning Meta Llama 3.1 to extend the pipeline to extract more complicated data.

Personalizing Protein Digestion   *University of California, Davis*
First Author - Under review at *NPJ Science of Food* | **Keywords** - Multi-omics Integration, Statistical Analysis   *Ongoing*
- Created a data analysis and visualization tool for omics data, which assisted the collaborating biochemists in identifying sample outliers, reducing 50% of the manual work.
- Developed a data normalization specific to the biological data, significantly reducing noise for the downstream analyses.
- Integrated and analyzed the biological data from human and food samples, providing statistical evidence for novel findings.
- Documented approaches and results in a manuscript, instrumental in securing funding for the subsequent study phase.

A2H (Animal-to-Human Trial Translation) - *GitHub*   *University of California, Davis*
First Author - *Frontiers in Artificial Intelligence* | **Keywords** - Scikit-learn, Explainable AI   *2023*
- Developed a machine learning model to predict preclinical-to-clinical translational outcomes of drugs for *Clostridium difficile* infection, a common gut disease, improving the efficiency of drug development by 58%.
- Conducted a feature importance analysis based on variance, Gini impurity, and Shapley values, discovering experimental variables influential for drug translation success.
- Performed a feature selction using recursive feature elimination, reducing 90% features while achieving the similar F1 score.

Glioblastoma Classification - *GitHub*   *University of California, Davis*
Co-Author - *Biomedicine* | **Keywords** - Scikit-learn, Explainable AI   *2023*
- Developed a random forest classifier to categorize glioblastoma patients using metabolites in patients' plasma samples, achieving an accuracy score of 0.81.
- Collaborated with neuro-oncologists and biochemists to develop the data processing strategy, reducing noise for the downstream analyses.
- Discovered 10 out of 340 plasma metabolites critical for classifying patients treatment stages, paving the way for the future diagnostic and prognostic tool for glioblastoma.

OTHER RESEARCH PROJECTS

MSAP (Model Selection and Analysis Pipeline)   *University of California, Davis*
Project Lead | **Keywords** - Machine Learning Pipeline, Reproducibility, Parallel Computation   *Ongoing*
- Developed a machine learning pipeline which streamlined the workflow encompassing data processing (outlier removal, missing value imputation, feature transformation), model development (model selection, data augmentation, hyperparameter tuning), and performance evaluation (metrics, visualization).
- Deployed MSAP successfully to research projects with reproducible results, contributing to three manuscripts.
- Currently extending the pipeline to allow time series forecasting and GPU functionality.

Techno-Economic Analysis for Lab-Grown Meat - *GitHub*   *University of California, Davis*
Co-Author - *Foods* | **Keywords** - Sensitivity Analysis   *2020*
- Collaborated with food scientists and implemented a mathematical model with 67 variables and 47 equations to estimate the total production cost for lab-grown meat.
- Performed a comprehensive sensitivity analysis using 6 algorithms, identifying 9 variables most influential to the production expense for lab-grown meat, corroborated by other experts in the field.

PUBLICATIONS

- **Li, F.**, Barboza, M., Weng, C.-Y., Cheang, S.E., Abrieux, A., Strandberg, E., Schaal, K., Bacalzo, N., Santos, E., Caster, R., Vairavamoorthy, V., German, B., Ahmed, S., de la Parra, J., Baar, K., Tagkopoulos, I., Lebrilla, C.B. and Siegel, J.B. (2024) "Molecular Basis of Protein Digestion of Animal and Plant Food Through Integrated Proteomics and Human Serum Metabolomics", under review at *NPJ Science of Food*.
- **Li, F.**, Youn, J., Millsop, C. and Tagkopoulos, I. (2024) "Predicting Clinical Trial Success for Clostridium difficile Infections Based on Preclinical Data", *Frontiers in Artificial Intelligence*. *https://doi.org/10.3389/frai.2024.1487335*
- **Li, F.**, Yoo, A., Youn, J., Guan, J., Guyer, A.E., Hostinar, C.E. and Tagkopoulos, I. (2024) "Prediction of Adolescent Depression from Prenatal and Childhood Data from ALSPAC Using Machine Learning", *Scientific Reports*. *https://doi.org/10.1038/s41598-024-72158-9*
- **Li, F.**, Youn, J., Simmons, G., Kim, S. and Tagkopoulos, I. (2024) "FoodAtlas: Automated knowledge extraction of food and chemicals from literature", *Computers in Biology and Medicine*. *https://doi.org/10.1016/j.compbiomed.2024.109072*
- **Li, F.**, Youn, J. and Tagkopoulos, I. (2023) "Semi-Automated Construction of Food Composition Knowledge Base", in *2nd AAAI Workshop on AI for Agriculture and Food Systems*. *https://doi.org/10.48550/arXiv.2301.11322*
- Aboud, O., Liu, Y., Dahabiyeh, L., Abuaisheh, A., **Li, F.**, Aboubechara, J.P., Riess, J., Bloch, O., Hodeify, R., Tagkopoulos, I. and Fiehn, O. (2023) "Profile Characterization of Biogenic Amines in Glioblastoma Patients Undergoing Standard-of-Care Treatment", *Biomedicines*. *https://doi.org/10.3390/biomedicines11082261*
- Risner, D., **Li, F.**, Fell, J.S., Pace, S.A., Siegel, J.B., Tagkopoulos, I. and Spang, E.S. (2020) "Preliminary Techno-Economic Assessment of Animal Cell-Based Meat", *Foods*. *https://doi.org/10.3390/foods10010003*

WORK EXPERIENCE

Department of Computer Science, University of California, Davis   *Davis, CA*
Teaching Assistant   *April 2021 - July 2022*
- Led weekly discussions and office hours for 100+ students in ECS170 - Intro to AI in Springs.
- Designed test suites and graded assignments regarding A* search and perceptron.
- Gave an introductory lecture for deep learning for ECS170 in Spring 2022.

Genome Center, University of California, Davis   *Davis, CA*
Junior Specialist   *August 2019 - August 2020*
- Collaborated with environmental scientists to create a toxicology database. Developed a full-stack web application using MEAN framework, along with REST API and command-line interface for programmatic users.
- Developed a pipeline for data retrieval and imputation for a multi-omics platform using Python. Optimized the pipeline using multiprocessing, achieving more than 10x speedup.

AWARDS

- UC Davis Graduate Group in Computer Science PhD Fellowship Award ($20,000)   *March 2024*

OTHER SKILLS

- Linguistics: Mandarin (Native), Japanese (Native)
- Hobbies: Combat sports (Mixed martial arts, Brazilian jiu-jitsu)