

Sunan Gao

sgao57@jh.edu || +1 (667)335-6100 || Baltimore, Maryland,

EDUCATION

Johns Hopkins University, Bloomberg School of Public Health 09/2023-Now

Major: MHS in Epidemiology (GPA: 4.0/4.0)

- **Courses (in progress):** Methods in Biostatistics (I-II), EPI Methods (I-II), Probability and Statistical Inference (II), Statistical Programming Paradigms and Workflow, Scalable Computational Bioinformatics, Data Structure.

Renmin University, School of Statistics 09/2022-07/2023

Major: MS in Epidemiology and Biostatistics (Transfer, GPA: 3.90/4.00)

- **Courses:** Advanced Mathematics Statistics, Statistical Computing, Epidemiology, Biostatistics, Causal Inference, Generalized Linear Model, Survival Analysis, Advanced Model Selection

Zhejiang University, School of Biosystems Engineering and Food Science 09/2018-06/2022

Major: BS in Food Science and Engineering (GPA: 3.96/4.00) **Minor:** Psychology (GPA: 4.00/4.00)

- **Courses:** Statistical Machine Learning, Probability Theory, Data Mining, Multivariable Calculus, Linear Algebra, Bioinformatics, Nutrition, Molecular Biology, Biochemistry, Psychometrics, Developmental psychology
- **Awards:** 2019-2022 National Scholarship (1/50, 3 times), Outstanding Undergraduate Thesis (1/150), Provincial Innovation Competition Gold Medal (top 1%, twice), Chu Kochen Scholarship Nomination (18/7000, Top 0.2%)

PUBLICATIONS

- **Sunan, G.**, Heming D., Shaobo W., ... Yu, W. (2023). Effects of accelerated biological aging on depressive symptoms in a causal reasoning framework. *J AFFECTIVE DISORDERS*, 339, 732–741.
- Qing, Y.[†], **Sunan, G.**[†], Junfen, L., ... Ming, C. (2022). A machine learning-based data mining in medical examination data: a biological features-based biological age prediction model. *BMC BIOINFORMATICS*, 23(1), 411.
- **Sunan, G.**, Mingyi, Y., Zisheng, L., ... Li, L. (2022). Soy protein/chitosan-based microsphere as stable biocompatible vehicles of oleanolic acid: an emerging alternative enabling the quality maintenance of minimally processed produce. *FOOD HYDROCOLLOID*, 124(Part B), 107325.
- Heming D., **Sunan, G.** *CRAN R Package* (2023). uotm: Uncertainty of Time Series Model Selection.
- Xuzhi, W., Yiju, Z., **Sunan, G.**, ... Yu, Z. (2023). Machine learning prediction of exposure to acrylamide based on modelling of association between dietary exposure and internal biomarkers. *FOOD CHEM TOXICOL*, 170: 113498.
- A brain-targeted lipid nanocapsule drug-carrying system and its preparation method and application. (Application number/Patent numbe: 2021102356580). *National Invention Patent*.
- Heming D., **Sunan, G.**, Tongxu, L., ... Yu, W. (2023). Association between negative mental health, medical burden, and economic development: evidence from an emerging composite anxiety-depression index. (*Under Review*)
- **Sunan, G.**, Yimei T., Xuzhi W., ... Yu, Z. (2023). Urinary metabolic analysis under 3-Chloro-1,2-propanediol di-ester and Eicosapentaenoic acid co-exposure: based on non-targeted metabolomics and molecular docking. (*Under Review*)
- Yimei T.[†], **Sunan, G.**[†], Xuzhi W., ... Yu, Z. (2023). Internal and external exposure associations of 3-monochloropropane-1,2-diol and glycidol using machine learning. (*Under Review*)
- **Sunan, G.**, Tongxu L., Heming D., ... Yu, W. (2023). Serum metabolites mediate causal relationships between intestinal flora and non-cognitive and cognitive features: . (*Working Paper*)

PROJECT EXPERIENCE

Serum metabolites mediate causal relationships between intestinal flora and cognitive features: Based on multivariable mendelian randomization (MR) and model averaging 04/2022-08/2023

Research Assistant (Biostatistics & Epidemiology Lab, Branch of Aging)

- Performed two-sample MR (6 methods) around intestinal flora (n=211), serum metabolites (n=483) and cognitive characteristics features. Sensitivity and outlier tests were carried out simultaneously.
- Calculated the indirect and direct mediation effect by MVMR and Delta methods; Evaluated causal strength of metabolites through Bayesian model averaging (BMA); Drew metabolic profiles from significant mediators.

Effects of Accelerated Biology Aging on Depressive Symptom (DS) under Causal Frame 09/2022-04/2023

Team Leader (Biostatistics Lab, Branch of Aging)

- Adopted mixed-effects logistic regression to obtain correlation results; Used DAG and backdoor criteria to identify confounding factors; Controlled covariates by fast large-scale almost matching exactly approach (FLAME);
- Recognized subgroups by generalized mixed-effects model tree and compared association stability under stratification.

- New Aging Indicators based on Population Characteristics and Ensemble Learning** 02/2022-09/2022
Team Leader (Bioinformatics Lab, Branch of Medicine)
- Compared interpolation effects of AutoEncoder, KNN, MICE, etc. in EHR data (70000+ entries with 100+ features);
 - Performed Pearson correlation and variable selection methods (Lasso, Elastic Net) to identify aging-related features;
 - Developed STK-BA based on features and Stacking with GAM, GBDT, CNN, etc. to solve the influence of model overfitting on the stability of association result; Adopted Poisson and logistic model to indicate effects.
- Biomarker-Based Predictive Models for Typical Food Contaminants** 12/2021-08/2022
Research Assistant (Epidemiology & Food Lab)
- Measured concentrations of vivo biomarkers of acrylamide and 3-MCPD from urine samples using UHPLC-MS/MS;
 - Built and compared models by MLR, SVM, Random Forest, XGBoost, LightGBM and CatBoost algorithms with internal exposure data to predict dietary exposure to acrylamide and assessed their performance using R and Python.
- Metabolic Profile Analysis of Contaminant Exposure and Dietary Interventions Based on Non-Target Metabolomes and Computational Biology (Molecular Docking)** 10/2021-06/2022
Undergraduate Thesis (Epidemiology & Food Lab)
- Compared the difference between single and co-exposure through stepwise criterion; Discovered the differential metabolites by OPLS-DA and volcano plots; Conducted enrichment analysis to explore affected metabolic pathways;
 - Draw metabolic fingerprints to analyze exposure differences; identified metabolic enzymes that linked dysregulated metabolites based on KEGG and PDB databases; Explored the docking mechanism by CB-Dock 2 and AutoDock Vina.
- Construction and Application of Biocompatible Carrier for Plant Wax** 10/2020-08/2021
Team Leader (Preservation & Nutrition Lab)
- Developed a new food-grade emulsion (SCM-OA); Characterization methods include confocal microscope, Cryo-SEM rheological behavior, Encapsulation efficiency (HPLC), Antibacterial and antioxidant properties, quality traits, etc.;
 - Used SPSS and Prism for data analysis to obtain optimal formula; Visualized the results by Origin and Photoshop;
- Brain Targeted Drug Delivery Based on Computer Aided Virtual Screening** 02/2019-08/2020
Research Assistant (Nutrition & Human Health Lab)
- Compared the applicability of molecular docking models: Ledock (Linux), AutoDock (Python), etc.; Screened 40000+ molecules by using affinity, clustering and similarity indicators; Visualized the 2D and 3D interactions on Pymol;
 - Constructed modified Nano-carriers, and assisted to conduct in vivo imaging and cell uptake experiments

WORKING EXPERIENCE

- Investment Analyst Intern (Healthcare)** 06/2023-08/2023
K2VC (K2 Venture Partners), Beijing
- Deeply involved in 10+ company/industry studies, covering high-value consumables, in vitro diagnostics, medical imaging, etc.; Presented 4 reports to Founding Partner and prompted 1 collaboration; Analyzed 6 business plans.
- Industry Analyst Intern (Innovative Drug Group, Top 3 in 2022)** 04/2023-08/2023
Soochow Securities Research Institute, Shanghai
- Skilled use of WIND, medical cube, etc. Analyzed business conditions through fundamental and valuation models; Participated in writing 8 weekly reports, 2 Industry/Company in-depth report
 - Using Python, crawler and ChatGPT to optimize automatic information extraction, judgment and text integration
- Physical and Chemical Analysis Intern** 09/2021-03/2022
CDC in Zhejiang Province, Hangzhou
- Carried out long-term exposure experiments (OGTT, ITT, etc.); Introduced multi-omics analysis and identified glucolipid metabolism disorders related markers by dose-response relationship from 24000+ potential variables;
 - Responsible for visualization and model verification; My analysis process and results were praised by mentors
- Entrepreneurial Team Co-founders** 05/2020-12/2021
AnGene Care Co., Ltd, Hangzhou
- Promoted portable medical testing devices; Responsible for incubation, business plan and enterprise cooperation;
 - Participated in national and provincial start-up competitions three times (2 Gold and 1 Silver Gold), took charge of road shows, and contacted angel investors; Our teams were recommended by 6 academic and business mentors

SKILLS

- **Software:** R, Python, Stata, SQL, Databricks, SAS, SPSS, Prism, Linux, LaTeX, Origin, Photoshop, InDesign,
- **Laboratory:** Fruit Preservation, Microbiology, Cellular, Mouse and Human Population Experiments' basic skills; CLSM, Turbiscan, Rheometer, Refractometer, GC-MS, HPLC, Virtual Screening, Homologous Modeling, etc.
- **Language:** English, Mandarin (Native)