Taehyo Kim

RESEARCH FOCUS **High-Dimensional Statistics, Interpretable ML/DL, Multimodal Representation Learning**

CONTACT taehyo97@gmail.com New York, NY, USA

EDUCATION **New York University**, New York, NY
Ph.D. in Biostatistics (Committee: Hai Shu, Yang Feng, Wen Zhou) Expected Spring 2027
M.S. in Computer Science 2022

University of Toronto, Toronto, ON
B.A.Sc in Computer Engineering, Biomedical Engineering minor 2020

PROFESSIONAL **FDA-OCE-ASA Oncology Educational Fellow** Oct 2025 - present
EXPERIENCE U.S. Food and Drug Administration (FDA) / American Statistical Association (ASA) Remote
- Deepened expertise in oncology drug development, emphasizing regulatory policy, clinical trial design, and survival analysis.

Machine Learning Researcher Sept 2020 - present
Hai Shu Lab, NYU School of Global Public Health New York, NY
- **Thesis:** Advancing Statistical Machine Learning for Alzheimer's Disease Research: Methods for Multiple Testing, Multi-modal Feature Fusion, and Survival Modeling
- Designed computationally efficient, spatial false discovery rate control methods for high dimensional neuroimaging, adapting unsupervised W-Net and Bayesian graphical modeling.
- Formulated an L0 penalized sparse canonical correlation analysis method for interpretable multimodal feature fusion with application to imaging-omics data.
- Proposed LLM-RAG guided survival modeling with knowledge-augmented penalization for bias robust feature selection.

Deep Learning Research Assistant Sept 2024 - present
Biofeedback Intervention Technology for Speech Lab, NYU Steinhardt New York, NY
- Trained transformer-based self-supervised representation learning models (ViViT, BYOL, Video-MAE) on video linguistic data, achieving 0.745 MCC, 87.2% ACC on rhotic classification.

Statistical Fellow Summer 2024
Biostatistical Collaboration and Consultation Core, NYU GPH New York, NY
- Prepared statistical analysis plans for clients and conducted statistical analyses (survey weighted multivariate regression) to support manuscript development.

Deep Learning Research Assistant Summer 2020
Multimedia Laboratory, University of Toronto Toronto, ON
- Fine-tuned CNNs (ResNet, U-Net) to classify histological tissue types and co-authored a large-scale survey on computational pathology analyzing over 800 papers.

Neural Coding Research Intern Summer 2019
N.1 Institute for Health, National University of Singapore Kent Ridge, Singapore
- Wrote MATLAB functions to automate cleaning and quality checks for a 76TB neural dataset.

SKILLS **Programming Languages:** Python, R, C/C++/C#, SQL, Java, Shell, MATLAB, SAS
Frameworks: PyTorch, TensorFlow, JAX, OpenCV, Spark, Hadoop, Numba, Git, Multiprocessing
Machine Learning: Transformers, Masked Autoencoders, VideoMAE, CNNs, Video Vision Transformers, U-Nets, Diffusion Models, Graphical Modeling, Large Language Models

PUBLICATIONS	Kim, T., Shu, H., Jia, Q., de Leon, M. J. DeepFDR: A Deep Learning-based False Discovery Rate Control Method for Neuroimaging Data. (2024). *AISTATS, 238: 946–954. Runner-up Winner of JSM Student paper competition* [Paper] [Code]

Hosseini, M. S., Bejnordi, B. E., Trinh, V. Q., Chan, L., Hasan, D., Li, X., Yang, S., Kim, T. et al. Computational Pathology: A Survey Review and the Way Forward. (2024). *Journal of Pathology Informatics, 15, 100357.* [Paper]

Kim, T., Jia, Q., de Leon, M. J., Shu, H. A False Discovery Rate Control Method Using a Fully Connected Hidden Markov Random Field for Neuroimaging Data. (2025). *Major Revision at Medical Image Analysis.* [Paper] [Code]

Tang, T., Chen, Y., Kim, T., Shu, H. UKAN-EP: Enhancing U-KAN with Efficient Attention and Pyramid Aggregation for 3D Multi-Modal MRI Brain Tumor Segmentation. (2025). *Accepted to BMC Medical Imaging.* [Paper] [Code]

Lai, A., Kim, T., Dahlen, A., Lomas, T. A Global Understanding of Work Enjoyment and Human Wellbeing. (2025). *Major Revision at Social Indicators Research.*

MANUSCRIPTS IN PREPARATION	Kim, T., Shu, H. L0-IPLS: An L0 Penalized Sparse Canonical Correlation Analysis Method with Application to High-Dimensional Imaging-Omics Data.

Chen, Y.*, Kim, T.* et al. Causal Determinants of Blood Pressure Control among US Adults with Hypertension: A Data-Driven Causal Graphical Learning, NHANES 2013 to 2023.

Eads, A., Benway, N., Kim, T. et al. Identifying clinically relevant articulatory patterns within perceptually inaccurate rhotic productions using self supervised learning on lingual ultrasound data.

Chen, Y., Kim, T., Shu, H., Feng Y. Transfer-guided Conditional Score-based Diffusion Network for Replenishment Sampling Imputation.

Kim, H., Cardoso, D. d. M., Kayahara, G. M., Kim, T. et al. Pre-treatment pain phenotypes and their association with disease progression and post-treatment pain in head and neck cancer.

MEMBERSHIP AND AFFILIATIONS	Member, American Statistical Association	2022 - present
	Member, Eastern North American Region of the International Biometric Society	2022 - present
	Member, Korean International Statistical Society	2024 - present
	Journal Reviewer, The American Statistician	Aug 2025

AWARDS	Student Paper Award, ASA Statistics in Imaging	2024
	DataFest Finalist, ENAR	2024
	Student Travel Award, American Statistical Association	2024
	PhD Fellowship Award, New York University	2022
	Certificate for Artificial Intelligence Engineering, University of Toronto	2020
	International Summer Research Award, University of Toronto	2019
	Gold Award, The Duke of Edinburgh's Award	2015

TALKS AND PRESENTATIONS	"A Global Understanding of Work Enjoyment and Human Wellbeing" Poster, Joint Statistical Meetings, Nashville, TN	August 2025

"Causal Determinants of Blood Pressure Control among US Adults with Hypertension: A Data Driven Causal Graphical Learning, NHANES 2013 to 2023"

2

Poster, ENAR Spring Meeting, New Orleans, LA   March 2025

"DeepFDR: A Deep Learning-based False Discovery Rate Control Method for Neuroimaging Data"
Oral, Joint Statistical Meetings, Portland, OR   August 2024

"DeepFDR: A Deep Learning-based False Discovery Rate Control Method for Neuroimaging Data"
Poster, International Conference in Artificial Intelligence and Statistics, Valencia, Spain   May 2024

"Enhancing AI-based Speech Therapy through Acoustic to Articulatory Mapping"
Poster, AI Research Symposium: Bridging AI Innovation and Societal Impact
New York, NY   April 2024

"Machine Learning-driven Risk Factor Identification on Post-2013 Blood Pressure Control Decline
in Hypertensive Populations"
Oral, ENAR Spring Meeting, Baltimore, MD   March 2024

TEACHING
EXPERIENCE

Teaching Assistant, New York University, New York, NY

Survey Design, Analysis, and Reporting (GPH-GU 2387)	Fall 2025
Applied Bayesian Analysis in Public Health (GPH-GU 2272/3372)	Fall 2024
Applied Survival Analysis (GPH-GU 2368/3368)	Spring 2024
Statistical Inference (GPH-GU 3225)	Fall 2023

Graduate Student Mentor, New York University, New York, NY

Pathways into Quantitative Aging Research Summer Program	Summer 2024
Pathways into Quantitative Aging Research Summer Program	Summer 2022

3