

YOUNG-GEUN KIM

Younggeun.Kim@nyspi.columbia.edu ◇ Personal Website ◇ Google Scholar

PROFESSIONAL APPOINTMENTS

Adjunct Associate Research Scientist

Jul. 2021 - Present

Department of Biostatistics, Columbia University

Mentor: Ying Liu, Ph.D.

Postdoctoral Researcher

Jul. 2021 - Present

Department of Psychiatry, Columbia University

Mental Health Data Science, New York State Psychiatric Institute

Mentor: Ying Liu, Ph.D.

Postdoctoral Researcher

Mar. 2021 - Jun. 2021

Department of Statistics, Seoul National University

Mentor: Myunghee Cho Paik, Ph.D.

EDUCATION

Seoul National University

Mar. 2015 - Feb. 2021

Ph.D. in Statistics

Graduated with the Best Dissertation Award

Advisor: Myunghee Cho Paik, Ph.D.

Dissertation: Statistical distance of conditional distributions and its applications

Seoul National University

Mar. 2010 - Feb. 2015

Triple Major

Graduated with Honors (Cum Laude)

B.S. in Industrial Engineering

B.S. in Statistics

B.S. in Mathematical Sciences

HONORS & AWARDS

Outstanding Reviewer Award

Jul. 2022

Thirty-ninth International Conference on Machine Learning

Best Dissertation Award

Feb. 2021

College of Natural Sciences, Seoul National University

Seoul National University Innovation Program Scholarship

Mar. 2017 - Feb. 2018

Seoul National University

* Awarded to the Ph.D. student with the highest GPA in the department.

Student Paper Competition 1st Prize

June 2017

Korean Statistical Society

RESEARCH INTERESTS

My research interests revolve around developing innovative data science tools and promoting their dissemination on biomedical data. Research topics include, but are not limited to:

- Deep generative models for multi-modal biomedical data (e.g., neuroimaging and multi-omics)
- Deep learning for identifying biomarkers associated with mental illness
- Reinforcement learning-based health care

PUBLICATIONS & PREPRINTS

*: First author; ‡: Corresponding author

Journal

- **Kim, Y.-G.***, Lee, K., and Paik, M.C.‡ (2022). Conditional Wasserstein generator. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. [\[Paper\]](#) [\[Code\]](#)
- **Top 1** Applied Mathematics journal (H-index: 397; upper 0.2%)
- **Kim, Y.-G.***, Kwon, Y., and Paik, M.C.‡ (2019). Valid oversampling schemes to handle imbalance. *Pattern Recognition Letters*, 125 (1): 661-667. [\[Paper\]](#) [\[Code\]](#)
- **Top 13** AI journal (H-index: 170; upper 4.6%)

Peer-reviewed Conference

- **Kim, Y.-G.***, Liu, Y.‡, and Wei, X. (2023). Covariate-informed representation learning to prevent posterior collapse of iVAE. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2023)*. [\[Paper\]](#) [\[Code\]](#)
- **Top 6** AI conference (H5-index: 85)
- Kim, M.*, **Kim, Y.-G.**, Kim, D., Kim, Y., and Paik, M.C.‡ (2021). Kernel-convoluted deep neural networks with data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2021)*. [\[Paper\]](#) [\[Code\]](#)
- **Top 4** AI conference (H5-index: 180)
- **Kim, Y.-G.***, Kwon, Y., Chang, H., and Paik, M.C.‡ (2020). Lipschitz continuous autoencoders in application to anomaly detection. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*. [\[Paper\]](#) [\[Code\]](#)
- **Top 6** AI conference (H5-index: 85)

Patents

- Paik, M.C.‡, **Kim, Y.-G.**, and Lee, K., Method and apparatus for conditional data generation using conditional Wasserstein generator. Republic of Korea Patent. [\[Info\]](#)
- Paik, M.C.‡, **Kim, Y.-G.**, and Chang, H., Learning method and learning device for high-dimension unsupervised anomaly detection using kernalized Wasserstein autoencoder to lessen too many computations of Christophel function, and testing method and testing device using the same. Republic of Korea Patent. [\[Info\]](#)

Preprints

- Kim, S.*, **Kim, Y.-G.**, and Wang, Y.‡ (2023). Temporal generative models for learning heterogeneous group dynamics of ecological momentary data (under *Revision Invited* at Biometrics). [\[BioRxiv\]](#)
- **Kim, Y.-G.***, ..., and Paik, M.C.‡ (2023). Wasserstein geodesic generator for conditional distributions (under review at Journal of Machine Learning Research). [\[ArXiv\]](#)[\[Code\]](#)
- **Kim, Y.-G.***, Ravid, O.*, ..., and Zhu, X.‡ (2023). Explaining deep learning-based representations of resting state functional connectivity data: focusing on interpreting nonlinear patterns in autism spectrum disorder (under review at Hippocampus). [\[BioRxiv\]](#) [\[Code\]](#)

GRANT

I submitted the following grant proposal as the **PI**.

- **Development of reinforcement learning-based tools for evaluating contingency management intervention in** *Apr. 2024 - Mar. 2029*
(if accepted)

substance use disorder treatments

Under review at NIH/NIDA K99/R00: Pathway to Independence Award

Total Grant Amount: \$1,113,066.

I participated the following researches as a **research scientist**.

- **A data science framework for empirically evaluating and deriving reproducible and transferrable RDoC constructs in youth (R01)** *Jul. 2021 - Present*
Funded by NIH/NIMH
- **Deep learning with incomplete and sequential data: Application to biomedical data** *Mar. 2020 - Jun. 2021*
Funded by National Research Foundation of Korea
- **Development of low-yield trackers via causal inference** *May 2019 - Nov. 2019*
Funded by SK Telecom
- **Statistical approaches to deep learning: New methods for convolutional neural networks in application to medical imaging data** *Mar. 2017 - Feb. 2020*
Funded by National Research Foundation of Korea
- **Deep Learning for the CT based Acute Cerebral Infarction Classification and Lesion Segmentation** *July 2016 - May 2019*
Collaborated with Seoul National University Bundang Hospital
Funded by National Research Foundation of Korea
- **New Robust Methods for Missing or Censored Covariates** *Mar. 2016 - Nov. 2016*
Funded by National Research Foundation of Korea

SELECTED TALKS

Invited Talks

- (Scheduled) **Kim, Y.-G.**, Luo, S. X., Brandt, L., Cheung, K., Nunes, E. V., Roll, J., and Liu, Y. (2024). Optimizing Contingency Management Interventions in Substance Use Disorder Treatment with Reinforcement Learning. *The Joint Statistical Meetings (JSM), Portland, OR.*
- **Kim, Y.-G.**, Liu, Y., and Wei, X. (2023). Covariate-informed representation learning to prevent posterior collapse of iVAE. *The Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2023), Palau de Congressos, Valencia, Spain.*[†]
- Liu, Y, **Kim, Y.-G.**, and Wei, X (2023). Covariate informed identifiable variational autoencoder to learn representations from brain imaging measures. *Eastern North American Region (ENAR), Nashville, TN.*
- Kim, M., **Kim, Y.-G.**, Kim, D., Kim, Y., and Paik, M.C. (2021). Kernel-convoluted deep neural networks with data augmentation. *The 35th AAAI Conference on Artificial Intelligence (AAAI-21), Virtual conference due to COVID-19.*
- **Kim, Y.-G.**, Kwon, Y., Chang, H., and Paik, M.C. (2020). Lipschitz continuous autoencoders in application to anomaly detection. *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Virtual conference due to COVID-19.*
- **Kim, Y.-G.**, Kwon, Y., Chang, H., and Paik, M.C. (2019). Lipschitz continuous autoencoders in application to anomaly detection. *IMS-China International Conference on Statistics and Probability, Dalian, China.*

Contributed Talks

- **Kim, Y.-G.**, Liu, Y., Brandt, L., Cheung, K., Nunes, E. V., Roll, J., and Luo, S. X. (2023). Optimizing contingency management in substance use disorder treatment using off-policy policy evaluation. *Eastern North American Region (ENAR) 2023 Spring meeting*.
- **Kim, Y.-G.**, Kwon, Y., and Paik, M.C. (2017). Handling imbalance in medical imaging data using convolutional neural network. *Spring Korea Statistical Conference 2017, Seoul, Republic of Korea*.

† indicates a poster presentation.

TEACHING EXPERIENCE

Student Instructor at Seoul National University

- Statistics Lab. (Fall 2015)
 - Freshman course to introduce R programming.
 - Served as a lecturer for 17 students, with full responsibility, including providing the whole 13 lectures, writing exam problems, and giving final grades.

Teaching Assistant at Seoul National University

- Mathematical Statistics 1 (Spring 2016, Summer 2016, Spring 2017, Summer 2017)
 - Major core course to focus on conditional probability, stochastic independence, and the distributions of random variables.
 - Held office hours and graded homework and exams.
- Mathematical Statistics 2 (Fall 2016, Winter 2016, Fall 2017)
 - Major core course to provide a deeper understanding of limit distributions, statistical estimation, and statistical inferences.
 - Held office hours and graded homework and exams.
- Statistics (Spring 2015, Spring 2020)
 - Freshman course to introduce Statistics.
 - Held office hours and graded homework and exams.
- Deep Learning: A Statistical Perspective (Spring 2018, Fall 2018, Fall 2019, Fall 2020)
 - Graduate-level course on deep learning.
 - Gave lectures about deep learning programming languages and deep learning-based object detection algorithms in English.
- Seminar in Recent Development of Applied Statistics (Fall 2017)
 - Graduate-level course on missing data analysis.
 - Gave a lecture about the application of expectation-maximization algorithm in incomplete data in English.

Guest Lecturer at Seoul National University

- Deep Learning: A Statistical Perspective (Fall 2021) at Seoul National University
 - Graduate-level course on deep learning.
 - Gave the lecture “Conditional Image Synthesis and Its Applications” in English.

Co-mentoring A Graduate Student at Columbia University

- Soohyun Kim, Department of Biostatistics Mar. 2022 - Present
 - Had regular weekly meetings with the student and Dr. Yuanjia Wang.

OTHER PROFESSIONAL ACTIVITIES

Conference Organizer

- (Accepted) Invited Session at JSM 2024 (role: **Organizer & Speaker**); Title: Reliable and Cost-effective Mental Health Care with Reinforcement Learning
- Invited Session at ENAR 2023 (role: **Chair**); Title: Advanced Methods for Analyzing Large-Scale Neuroimaging Data from Nationwide Consortia for Mental Health Research [\[Info\]](#)
- Oral Presentation Session at ICML 2022 (role: **Chair**); Title: Theory [\[Info\]](#)

Reviewer

- JAMA Psychiatry
- Expert Systems with Applications
- Pattern Recognition Letters
- International Journal of Computer Assisted Radiology and Surgery
- International Conference on Machine Learning
- International Conference on Artificial Intelligence and Statistics