Yejin Kan

Ph.D. Candidate • Computer Science and Artificial Intelligence
Dongguk University | Seoul, South Korea
kanyj@dgu.ac.kr | Google Scholar | ORCID | DataLab

Research Interests

Machine learning for biological data, especially representation learning and multi-label learning. I develop scalable protein representations informed by functional annotations for function prediction.

Education

Ph.D. in Computer Science and Artificial Intelligence Mar. 2023 – Present
Dongguk University, Seoul, South Korea
Advisor: Prof. Gangman Yi | Expected graduation: Aug. 2027

M.S. in Multimedia Engineering Mar. 2021 – Feb. 2023
Dongguk University, Seoul, South Korea
Advisor: Prof. Gangman Yi

B.S. in Multimedia Engineering Mar. 2017 – Feb. 2021
Dongguk University, Seoul, South Korea

Research Experience

Graduate Researcher, DataLab, Dongguk University Mar. 2021 – Present

- Develop machine learning methods for protein representation learning and function prediction, with an emphasis on multi-label learning.
- Investigate how functional annotations and relationships among protein functions inform representation learning from sequences and structures.
- Evaluate learned representations across diverse protein annotation and classification benchmarks.

Publications

Journal Articles

1. **Yejin Kan** and Gangman Yi. "ProtDML: Label-aware representation learning for broad-spectrum protein function prediction." *Briefings in Bioinformatics*, 27(3):bbag293, 2026. doi:10.1093/bib/bbag293.
2. Seungmin Lee, **Yejin Kan**, Dongyeon Kim, Minseop Kim, and Gangman Yi. "A Method for Consensus Anchor Graph Learning for Incomplete Multi-View Clustering." *IEEE Access*, 14:18155–18168, 2026. doi:10.1109/ACCESS.2026.3659917.
3. **Yejin Kan**, Dongyeon Kim, Jinkyung Yang, and Gangman Yi. "GapSense: Similarity Estimation-Based Gap Filler with TGS-Reads for Genome Assemblies." *Interdisciplinary Sciences: Computational Life Sciences*, 18(2):501–514, 2026. doi:10.1007/s12539-025-00770-y.
4. Mingeun Ji, **Yejin Kan**, Dongyeon Kim, Seungmin Lee, and Gangman Yi. "DeepPI: Alignment-Free Analysis of Flexible Length Proteins Based on Deep Learning and Image Generator." *Interdisciplinary Sciences: Computational Life Sciences*, 16(3), 2024. doi:10.1007/s12539-024-00618-x.

5. Adil Abdu Bushra, Dongyeon Kim, **Yejin Kan**, and Gangman Yi. "AutoSCAN: Automatic Detection of DBSCAN Parameters and Efficient Clustering of Data in Overlapping Density Regions." *PeerJ Computer Science*, 10:e1921, 2024. doi:10.7717/peerj-cs.1921.
6. Dongyeon Kim, **Yejin Kan**, Seungmin Lee, and Gangman Yi. "Parameter-Free Basis Allocation for Efficient Multiple Metric Learning." *Machine Learning: Science and Technology*, 4(4):045049, 2023. doi:10.1088/2632-2153/ad113b.
7. Dongyeon Kim, **Yejin Kan**, YooJin Aum, Wanhee Lee, and Gangman Yi. "Hotspots-Based Patrol Route Optimization Algorithm for Smart Policing." *Heliyon*, 9(10):e20931, 2023. doi:10.1016/j.heliyon.2023.e20931.
8. Mingeun Ji, **Yejin Kan**, Dongyeon Kim, Jaehee Jung, and Gangman Yi. "cPlot: Contig-Plotting Visualization for the Analysis of Short-Read Nucleotide Sequence Alignments." *International Journal of Molecular Sciences*, 23(19):11484, 2022. doi:10.3390/ijms231911484.

Conference Papers

1. Dongyeon Kim, **Yejin Kan**, and Gangman Yi. "T-MDML: Triplet-Based Multiple Distance Metric Learning for Multi-Instance Multi-Label Classification with Label Correlation." *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 39(17):17850–17858, 2025. doi:10.1609/aaai.v39i17.33963.
2. **Yejin Kan**, Dongyeon Kim, and Gangman Yi. "A Novel Gap-Filling Method Based on Hybrid Read Information Analysis." *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3827–3829, 2022. doi:10.1109/BIBM55620.2022.9994889.
3. Dongyeon Kim, **Yejin Kan**, and Gangman Yi. "Novel Patrol Route Optimization Method Based on Big Data Analysis." *IEEE International Conference on Big Data (IEEE BigData)*, pp. 6699–6701, 2022. doi:10.1109/BigData55660.2022.10020159.

Manuscripts Under Review

1. **Yejin Kan** and Gangman Yi. "ProtBiC: Bipartite Context-Supervised Protein Representation Learning for Function Prediction." In revision, 2026.
2. Dongyeon Kim, **Yejin Kan**, Minseop Kim, and Gangman Yi. "Crime-Risk-Based Patrol Route Optimization for Crime Prevention in Urban Road Networks." In revision, 2026.
3. **Yejin Kan** and Gangman Yi. "ProtCon: Hierarchy-Aware Multi-Modal Contrastive Representation Learning for Protein Function Prediction." Under review, 2026.

Conference Presentations

1. Ilkhomjon Sadriddinov, **Yejin Kan**, Dongyeon Kim, Jinkyung Yang, and Gangman Yi. "Novel Algorithm for Finding Maximum k-Balanced Biclique Based on Formal Concept Analysis." *15th International Conference on Frontier Computing (FC)*, 2025. Oral presentation.
2. Jinkyung Yang, **Yejin Kan**, and Gangman Yi. "CNGNN-DTI: Integration of Context and Neighbor Information Based on Graph Neural Network for Drug-Target Interaction Prediction." *29th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, 2025. Poster.
3. **Yejin Kan**, Jinkyung Yang, and Gangman Yi. "Integration of Multi-View Features of Protein Sequences for Improved Protein Function Prediction." *23rd European Conference on Computational Biology (ECCB)*, 2024. Poster.

Conference Papers (Korea)

1. Minseop Kim, **Yejin Kan**, Dongyeon Kim, Ilkhomjon Sadriddinov, and Gangman Yi. "HS Code Classification Using Graph RAG with an HS Explanatory Notes-Based Knowledge Graph." *Proceedings of the Korea Computer Congress (KCC)*, 2026. DBpia. (in Korean)
2. Yerin Lee, Dongyeon Kim, **Yejin Kan**, Seungmin Lee, and Gangman Yi. "Public Data-Based Hotspot Detection and Patrol Route Optimization with Spatiotemporal Data." *Proceedings of the Korea Computer Congress (KCC)*, pp. 414–416, 2024. DBpia. (in Korean)
3. **Yejin Kan**, Dongyeon Kim, Mingeun Ji, Seungmin Lee, and Gangman Yi. "BLAST-Based Gap Analysis for Gap Filling Algorithm." *Proceedings of the Korea Computer Congress (KCC)*, pp. 1625–1627, 2022. DBpia. (in Korean)

Research Grants

Principal Investigator — WISET STEM Research Team Project Apr. 2022 – Oct. 2022

Project: Development of a Hybrid Read-Based Gap-Filling Algorithm

Funded by the Ministry of Science and ICT (MSIT), South Korea.
Administered by the Korea Foundation for Women in Science, Engineering and Technology (WISET).
Grant amount: KRW 8,000,000.

Teaching Experience

Teaching Assistant, Dongguk University 2021–2022; 2024–2026

- **Spring 2026:** Algorithms.
- **Fall 2025:** Algorithms; Database Design.
- **Spring 2025:** Data Structures; Database Systems.
- **Fall 2024:** Database Design.
- **Spring 2024:** Introduction to Programming; Database Systems; Data Structures.
- **Spring 2022:** Database Systems; Data Structures.
- **Fall 2021:** Introduction to Programming; Computer Networks.
- **Spring 2021:** Data Structures; Operating Systems; Database Systems.

Honors and Scholarships

- **Dongguk Global Fellowship**, Dongguk University (2023–2024; four semesters).
- **Dongguk University Alumni Scholarship**, Dongguk University (Spring 2023).
- **B&M (Bachelor's–Master's) Scholarship**, Dongguk University (Fall 2021; Spring 2022).
- **Dongguk Future Leaders Scholarship**, Dongguk University (Spring 2021).

Patent

Dongyeon Kim, Gangman Yi, **Yejin Kan**, Yujin Eom, and Wanhee Lee.
"Hotspots-based patrol route optimization algorithm for smart policing."
Republic of Korea Patent No. 10-2751950. Granted Jan. 3, 2025; filed Dec. 14, 2023.

Last updated: October 2026