

# Sangseon Lee (이상선)

Institute of Computer Technology  
Seoul National University  
Seoul, Korea  
Phone. +82-10-2928-3740  
E-mail. sangseon486@snu.ac.kr

## EDUCATION

---

### Seoul National University, Korea

- Ph.D., Computer Science and Engineering Mar. 2014 – Feb. 2020  
(Advisor: Professor Sun Kim)
- B. S., Computer Science and Engineering (*Summa Cum Laude*) Mar. 2010 – Aug. 2013

## EXPERIENCE

---

### Seoul National University, Korea

- Postdoctoral research fellow,  
SNU Institute of Computer Technology Sep. 2021 – Present  
(Advisor: Professor Sun Kim)
- Postdoctoral research fellow,  
SNU BK21 FOUR Intelligence Computing Dec. 2020 – Aug. 2021  
(Advisor: Professor Sun Kim)
- Postdoctoral research fellow, SNU Bioinformatics Institute Mar. 2020 – Nov. 2020  
(Advisor: Professor Sun Kim)

## RESEARCH INTERESTS

---

**Artificial Intelligence (AI)**, Machine learning, and data mining techniques

**Integration of BT and IT (Bioinformatics)** for biomedical big data, especially graph & network analysis

**Cheminformatics and AI** in Drug discovery

My research area focuses on developing deep learning/machine learning models for analyzing graph-structured data and multi-omics data in the biomedical and chemical informatics fields. Focuses on modeling interactions between biological elements of living organisms as form of a graph, or on modeling molecular graphs incorporating subgraphs and 3D information. Applications include interpretable deep learning model for bioinformatics, computational biology with multi-omics analysis, subpath mining in the biological network, molecular property prediction, and drug response prediction.

## RESEARCH EXPERIENCE

---

### Graduate research in Bio & Health Informatics Lab, Seoul National University, Korea

#### < Deep Learning models for interpreting biological data with prior knowledge >

- Cancer subtype classification and modeling by pathway attention and propagation
  - Designing a pathway-based explainable deep learning model by graph convolutional network and attention mechanism
  - Predicting cancer subtypes using gene expression data and pathway information
  - Paper accepted by *Bioinformatics* journal (SCIE, IF=5.610)

- Learning Cell-Type-Specific Gene Regulation Mechanisms by Multi-Attention Based Deep Learning with Regulatory Latent Space
  - Designing multi-modal deep learning model for learning gene regulation mechanism
  - Elucidating epigenetic gene regulation mechanism by various attention layers
  - Paper accepted by *Frontiers in Genetics* journal (SCIE, IF=3.260)
- Multi-layered Knowledge Graph Neural Network Reveals Pathway-level Agreement of Three Breast Cancer Multi-gene Assays
  - Designing an explainable deep learning model by integrating intra- and inter- pathway level attention
  - Identification of shared regulatory mechanisms of three breast cancer multi-gene assays
  - Under review in SCIE journal

#### < AI with graph structured data in biomedical domains >

- Sparse Structure Learning via Graph Neural Networks for Inductive Document Classification
  - A novel GNN-based sparse structure learning model for inductive document classification
  - Employing structure learning to sparsely select edges between words by considering dynamic contextual dependencies
  - Paper accepted by *AAAI 2022*
- Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers
  - A semantic multi-layer guilt-by-association approach that leverages the principle of guilt-by-association - "similar genes share similar functions" at the drug-gene-disease level.
  - Designing a semantic information-guided random walk to generate embeddings of drugs and disease in a unified embedding space
  - Paper accepted by *Nature Communications journal* (SCIE, IF=16.6)

#### < Cheminformatics and AI in Drug discovery >

- Multi-Task Informed Learnable Prototypes on Few Shot learning for Molecular Property Prediction
  - Leveraging shared knowledge across multiple molecular properties in few-shot learning
  - Incorporating a stochastic attention mechanisms to reflect information from multiple assays
  - Under review in International Conference
- Dual Representation Learning for Predicting Drug-side Effect Frequency using Protein Target Information
  - Utilizing heterogenous features of drugs and side effects to get improved latent representations
  - Compensating for the drugs without clear target proteins using the Adaboost method
  - Paper accepted by *IEEE Journal of Biomedical and Health Informatics* journal (SCIE, IF=7.7)
- Improved drug response prediction by drug target data integration via network-based profiling
  - A novel model-agnostic framework to enhance performances of existing deep learning models for drug response prediction
  - Network-based estimation of gene perturbation by drug treatment
  - Paper accepted by *Briefings in Bioinformatics* journal (SCI, IF=9.5)

## PUBLICATIONS

---

### In review (\*: equal contributors)

- [1] Lee S, Piao Y, Lee D, Kim S. Multi-Task Informed Learnable Prototypes on Few Shot learning for Molecular Property Prediction
- [2] Lee S\*, Park J\*, Piao Y, Lee D, Lee D, Kim S. Multi-layered Knowledge Graph Neural Network Reveals Pathway-level Agreement of Three Breast Cancer Multi-gene Assays
- [3] Cho C, Lee S\*, Piao Y, Bang D, Kim S. ChemAP: Chemical Structure-Based Deep Learning Model via Knowledge Distillation for Predicting Drug Approval before Clinical Trial Phase
- [4] Sung I, Lee S\*, Bang D, Yi J, Kim S. MDTR: A Knowledge-Guided Interpretable Representation for Quantifying Liver Toxicity at Transcriptomic Level
- [5] Shin Y, Lee S\*, Pak M, Cho C, Kim S. DrugPT-Net: Drug Perturbation Guided Visible Neural Network for Drug Response Prediction at Transcriptomic Level
- [6] Lu Y, Lee S\*, Park S, Yi J, Cho C, Lim S, Kang S, Kim S. An ensemble strategy to improve generalization power of deep learning models for DTI prediction
- [7] Piao Y, Lee S, Lu Y, Kim S. Improving Out-of-Distribution Generalization in Graphs via Hierarchical Semantic Environments
- [8] Lu Y, Piao Y, Lee S, Kim S. Context-Aware Hierarchical Fusion for Drug Relational Learning

### International Journals / Conferences (\*: equal contributors)

- [1] Park S\*, Lee S\*, Pak M, Kim S. Dual Representation Learning for Predicting Drug-side Effect Frequency using Protein Target Information. *IEEE Journal of Biomedical and Health Informatics* (SCIE, IF=7.7). Jan. 5, 2024.
- [2] Yi J, Lee S, Lim S, Cho C, Piao Y, Yeo M, Kim D, Kim S, Lee S. Exploring chemical space for lead identification by propagating on chemical similarity network. *Computational and Structural Biotechnology Journal* (SCIE, IF=6.0). Aug. 25, 2023; 21:4187-4195.
- [3] Gu J\*, Bang D\*, Yi J\*, Lee S, Kim DK, Kim S. A model-agnostic framework to enhance knowledge graph-based drug combination prediction with drug-drug interaction data and supervised contrastive learning. *Briefings in Bioinformatics* (SCIE, IF=9.5). Aug. 7, 2023; 24(5):bbad285.
- [4] Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nature Communications* (SCIE, IF=16.6). Jun. 15, 2023; 14(1):3570.
- [5] Pak M\*, Lee S\*, Sung I, Koo B, Kim S. Improved drug response prediction by drug target data integration via network-based profiling. *Briefings in Bioinformatics* (SCIE, IF=9.5). Feb. 8, 2023; 24(2):bbad034.
- [6] Lee S, Lee D, Piao Y, Kim S. SPGP: Structure Prototype Guided Graph Pooling. *NeurIPS 2022 Workshop New Frontiers in Graph Learning (GLFrontiers)*. Dec. 2, 2022.
- [7] Koo B, Lee D, Lee S, Sung I, Kim S, Lee S. Risk Stratification for Breast Cancer Patient by Simultaneous Learning of Molecular Subtype and Survival Outcome Using Genetic Algorithm-Based Gene Set Selection. *Cancers* (SCIE, IF=6.573). [From The 10th International Conference on Intelligent Biology and Medicine (ICIBM 2022)]. Aug. 25, 2022; 14(17):4120.
- [8] Lim S\*, Lee S\*, Piao Y, Choi MG, Bang D, Gu J, Kim S. On Modeling and Utilizing Chemical Compound Information with Deep Learning Technologies: A Task-oriented Approach. *Computational and*

**Structural Biotechnology Journal** (SCIE, IF=6.155). Aug. 4, 2022; 20:4288-4304.

[9] Moon JH, **Lee S**, Hur B, Kim S. MLDEG: A network-based ensemble model to identify differentially expressed genes consistently. **IEEE/ACM Transactions on Computational Biology and Bioinformatics** (SCIE, IF=3.702). Jul.-Aug., 2022; 19(4).

[10] Piao Y, **Lee S**, Lee D, Kim S. Sparse Structure Learning via Graph Neural Networks for Inductive Document Classification. **AAAI 2022**. Feb. 22, 2022.

[11] Sung I\*, **Lee S\***, Pak M, Shin Y, Kim S. AutoCoV: Learning the Spatial and Temporal Dynamics of COVID-19 Spread Patterns from Embedding Space by k-mer Based Deep Learning. **BMC Medical Genomics** (SCIE, IF=3.622). [From The 10th International Conference on Intelligent Biology and Medicine (ICIBM 2022)]. Apr., 2022; 23(3):149.

[12] Kim J, Lim S, **Lee S**, Cho C, Kim S. Embedding of FDA Approved Drugs in Chemical Space Using Cascade Autoencoder with Metric Learning. **IEEE BigComp 2022. AI-BioHealth 2022 workshop**. Jan. 17, 2022.

[13] Kim I, **Lee S**, Kim Y, Namkoong H, Kim S. Probabilistic model for pathway-guided gene set selection. **IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 12th International Workshop on Biomedical and Health Informatics (BHI 2021)**. Dec. 9, 2011.

[14] Kim MS, **Lee S**, Lim S, Lee DY, Kim S. Subnetwork Representation Learning for Discovering Network Biomarkers in Predicting Lymph Node Metastasis in Early Oral Cancer. **Scientific Reports** (SCIE, IF=4.380). Dec., 2011; 11:23992.

[15] Lee T, **Lee S**, Kang M, Sun Kim. Deep Hierarchical Embedding for Simultaneous Modeling of GPCR Proteins in a Unified Metric Space. **Scientific Reports** (SCIE, IF=4.380). May., 2021; 11:9543.

[16] **Lee S**, Lee T, Noh YK, Kim S. Ranked k-spectrum kernel for comparative and evolutionary comparison of exons, introns, and CpG islands. **IEEE/ACM Transactions on Computational Biology and Bioinformatics** (SCIE, IF=3.710). May.-Jun., 2021; 18(30).

[17] Jeong D, Lim S, **Lee S**, Oh M, Cho C, Seong H, Jung W, Kim S. Construction of Condition-Specific Gene Regulatory Network using Kernel Canonical Correlation Analysis. **Frontiers in Genetics** (SCIE, IF=4.599). Mar., 2021; 12:652623.

[18] Park YJ, **Lee S**, Lim S, Hahn N, Ji Y, Huh JY, Alfadda AA, Kim S, Kim JB. DNMT1 maintains metabolic fitness of adipocytes through acting as an epigenetic safeguard of mitochondrial dynamics. **Proceedings of the National Academy of Sciences (PNAS)** (SCIE, IF=11.205). Mar. 16, 2021; 118(11):e 2021073118.

[19] Oh M, Park S, **Lee S**, Lee D, Lim S, Jeong D, Jo K, Jung I, Kim S. DRIM: A web-based system for investigating drug response at the molecular level by condition-specific multi-omics data integration. **Frontiers in Genetics** (SCIE, IF=3.260). Nov., 2020; 11:564792.

[20] Kang M\*, **Lee S\***, Lee D, Kim S. Learning Cell-Type-Specific Gene Regulation Mechanisms by Multi-Attention Based Deep Learning with Regulatory Latent Space. **Frontiers in Genetics** (SCIE, IF=3.260). Sep., 2020; 11:869.

[21] **Lee S**, Lim S, Lee T, Sung I, Kim S. Cancer subtype classification and modeling by pathway attention and propagation. **Bioinformatics** (SCIE, IF=5.610). Jun., 2020; 36(12):3818-3824.

[22] Lim S, **Lee S**, Jung I, Rhee S, Kim S. Comprehensive and critical evaluation of individualized pathway activity measurement tools on pan-cancer data. **Briefings in Bioinformatics** (SCIE, IF=8.990). Jan., 2020; 21(1):36-46.

[23] Hur B, Kang D, **Lee S**, Moon JH, Lee G, Kim S. Venn-diaNet: Venn diagram based network propagation analysis framework for computing multiple biological experiments. **BMC Bioinformatics** (SCIE, IF=2.511). [From Proceeding of 30th International Conference on Genome Informatics (GIW)]. Dec. 27, 2019;

20(23):1-12.

[24] Kang D, Ahn H, **Lee S**, Lee CJ, Hur J, Jung W, Kim S. StressGenePred: a twin prediction model architecture for classifying the stress types of samples and discovering stress-related genes in Arabidopsis. **BMC Genomics** (SCIE, IF=3.501). [From Proceeding of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)]. Dec. 20, 2019; 20(11):949.

[25] Kim MS, **Lee S**, Lim S, Kim S. SpliceHetero: An information theoretic approach for measuring spliceomic intratumor heterogeneity from bulk tumor RNA-seq. **PLoS ONE** (SCIE, IF=2.776). Oct. 23, 2019; 14(10).

[26] Lee DH, **Lee S**, Kim S. PRISM: Methylation Pattern-based, Reference-free Inference of Subclonal Makeup. **Bioinformatics** (SCIE, IF=4.531). [From Proceeding of 27th conference on Intelligent Systems for Molecular Biology (ISMB)]. Jul. 5, 2019; 35(14):i520-i529.

[27] Lee CJ, Kang D, **Lee S**, Lee S, Kang J, Kim S. In silico experiment system for testing hypothesis on gene functions using three condition specific biological networks. **Methods** (SCIE, IF=3.998). May. 25, 2018; 145:10-15.

[28] **Lee S**, Park Y, Kim S. MIDAS: Mining differentially activated subpaths of KEGG pathways from multi-class RNA-seq data. **Methods** (SCIE, IF=3.812). Jul. 15, 2017; 124:13-24.

[29] Moon JH, Lim S, Jo K, **Lee S**, Seo S, Kim S. PINTnet: construction of condition-specific pathway interaction network by computing shortest paths on weighted PPI. **BMC Systems Biology** (SCIE, IF=2.303). [From Proceeding of 15th Asia Pacific Bioinformatics Conference (APBC)]. Mar. 14, 2017; 11(2):15.

[30] **Lee S**, Moon JH, Park Y, Kim S. Flow maximization analysis of cell cycle pathway activation status in breast cancer subtypes. **IEEE International Conference on Big Data and Smart Computing (BigComp) 2017**. Feb. 13-16, 2017. Jeju, Korea.

[31] Chae H, **Lee S**, Nephew KP, Kim S. Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. **BMC Systems Biology** (SCIE, IF=2.208). [From Proceeding of 27th International Conference on Genome Informatics (GIW)]. Dec 23, 2016; 10(4):116.

[32] Chae H, **Lee S**, Seo S, Jung D, Chang H, Nephew KP, Kim S. BioVLAB-mCpG-SNP-EXPRESS: A system for multi-level and multi-perspective analysis and exploration of DNA methylation, sequence variation (SNPs), and gene expression from multi-omics data. **Methods** (SCIE, IF=3.503), Jul. 28, 2016; 111:64-71.

[33] Jeong HM, **Lee S**, Chae H, Kim R, Kwon MJ, Oh E, Choi YL, Kim S, Shin YK. Efficiency of methylated DNA immunoprecipitation bisulphite sequencing for whole-genome DNA methylation analysis. **Epigenomics** (SCIE, IF=4.044), Jun. 8, 2016; 8(8):1061-1077.

## Book Chapters

[1] Pak M, Jeong D, Moon JH, Ann H, Hur B, **Lee S**, Kim S. Network Propagation for the Analysis of Multi-Omics Data. *Recent Advances in Biological Network Analysis* (pp. 185-217). 2021

## Domestic Journal (\*: equal contributors)

[1] 박민우, 이상선, 성인영, 신윤열, 정인욱, 김선. 밀도기반 군집화와 딥러닝 모델을 이용한 COVID-19 바이러스 전장 유전체 임베딩 전략. *Journal of KIISE* [From 2021 한국컴퓨터종합학술대회] 제49권 제4호, pp.261-270. 2022.

[2] 박은화\*, 이상선\*, 김선. 신약개발과 맞춤의학을 위한 그래프 학습 기술. *한국정보과학회지*

제40권 제3호, pp.30-40. 2022.

[3] 박은화, 이상선, 김선. 헬스케어 예측을 위한 전자 건강 기록 기반 그래프 컨볼루션 모델. 한국정보과학회 2020 한국컴퓨터종합학술대회 논문집, Vol.2020, pp.569-571. Jul., 2020.

[4] 강동원, 안홍렬, 이상선, 정우석, 김선. 시계열 유전자 발현 데이터의 스트레스 특이 반응 유전자 검출 및 스트레스 예측 알고리즘 개발. 한국정보과학회 2018 한국컴퓨터종합학술대회 논문집, Vol.2018, pp.1063-1065. Jun., 2018.

## GRANTS

---

**산학과제.** (주)아이겐드릭, Jul., 2023 – Jun., 2024.

**창의·도전연구.** 학문균형발전지원사업, 한국연구재단, Jun., 2023 – May., 2026.

**산학과제.** (주)아이겐드릭, Oct., 2022 – Mar., 2023.

**산학과제.** (주)아이겐드릭, Mar., 2022 – Feb., 2023.

**박사후국내연수.** 학문후속세대지원사업, 한국연구재단, Sep., 2021 – Aug., 2023.

**Google Education/Course Grant (Co-instructor).** Google, Oct., 2020.

## AWARDS

---

**BIOINFO 2022 우수포스터상.** 한국생명정보학회. Oct. 21, 2022.

**The AI Korea 2022 포스터 세션 대상.** 한국정보과학회 인공지능소사이어티. Aug. 18, 2022.

**목암 인공지능 심포지움 Best Poster 우수상.** 목암생명과학연구소. Jun. 29, 2022.

**KCC2021 우수논문상.** 한국정보과학회. Jun. 25, 2021.

**스탠다임 우수논문상.** 한국생명정보학회. Aug. 20, 2020.

**Best Paper Award (Bronze).** 한국정보과학회 인공지능소사이어티 The AI KOREA 2019. Jul. 26, 2019.

**Best Paper Award on Bioinformatics.** IEEE International Conference on Big Data and Smart Computing (BigComp) 2017. Feb. 16, 2017.

**Domestic Scholarship.** 관정 이종환 교육재단. 2014-2015.

## TEACHING EXPERIENCE

---

**Co-lecturer:** Drug target prediction and drug repurposing with graph learning, 2024 Feb, KSBi-BIML 2024 생물정보학 & 머신러닝 워크샵

**Guest lecturer:** AI-BIO 연구인력 양성과정, 2023 Fall, Artificial Intelligence Institute, Seoul National

University, Korea

**Co-lecturer:** Drug target prediction and drug repurposing with graph learning, 2023 Feb, KSBi-BIML 2023  
생물정보학 & 머신러닝 워크샵

**Guest lecturer:** AI-BIO 연구인력 양성과정, 2022 Fall, Artificial Intelligence Institute, Seoul National University, Korea

**Co-lecturer:** Graph Learning for Personalized Medicine and Drug Discovery, 2022 Aug, The AI KOREA, 2022

**Guest lecturer:** Machine Learning in Bioinformatics, 2022 Spring, Seoul National University, Korea

**Guest lecturer:** Machine Learning in Bioinformatics, 2021 Spring, Seoul National University, Korea

**Guest lecturer:** 생물정보학을 위한 IT기초, 2020 Spring, Seoul National University, Korea

**Guest lecturer:** 휴먼바이오공학기초, 2020 Spring, Ewha Womans University, Korea

**Guest lecturer:** 기계학습의 이해, 2019 Oct, Big Data Academy, Seoul National University, Korea

**Guest lecturer:** 휴먼바이오공학기초, 2019 Spring, Ewha Womans University, Korea

**Guest lecturer:** 기계학습의 이해, 2018 Jul, Big Data Academy, Seoul National University, Korea

**Guest lecturer:** Deep Learning for Bioinformatics, 2018 Feb, KSBi-BIML 2018 (생물정보학&머신러닝/  
인공지능 워크샵), Korea