

# Seojin Bang

PhD candidate in Computational Biology  
School of Computer Science, Carnegie Mellon University  
✉ seojinb@cs.cmu.edu | 🏠 seojinb.com

## RESEARCH INTEREST

---

My research interest is in the interpretability of black-box machine learning models. I develop interpretable machine learning approaches and investigate the relationship between interpretability and robustness in machine learning algorithms. I also develop multimodal/multiview machine learning approaches for combining different data modalities. I have been applying the approaches to help to solve healthcare and biomedical problems.

## EDUCATION

---

Aug 2020 (expected)	<b>The School of Computer Science, Carnegie Mellon University</b> PhD candidate in Computational Biology	Advisor: Wei Wu
2013	<b>Seoul National University, Korea</b> MS in Statistics	Advisor: Taesung Park
2010	<b>Sungkyunkwan University, Korea</b> BS in Mathematical Education · BE in Statistics (double major)	

## PUBLICATIONS/PREPRINTS

---

- [1] **Explaining a black-box using deep variational information bottleneck approach.**  
*preprint arXiv:1902.06918*, 2020.  
Bang, Seojin and Xie, Pengtao and Lee, Heewook and Wu, Wei and Xing, Eric.
- [2] **Robust multiple kernel k-means clustering using min-max optimization.**  
*preprint arXiv:1803.02458*, 2020.  
Bang, Seojin and Yu, Yaoliang and Wu, Wei.
- [3] **Detecting attackable sentences in arguments.**  
*preprint*, 2020.  
Jo, Yohan and Bang, Seojin.
- [4] **Kw-han: Knowledge infused hierarchical attention network for enhancing interpretability.**  
*preprint*, 2020.  
Jang, Hyeju and Bang, Seojin.
- [5] **Phased-lstm based predictive model for ehers with complex missing patterns.**  
*preprint*, <https://www.cs.cmu.edu/~epxing/Class/10708-17/project-reports/project8.pdf>, 2019.  
Bang, Seojin and Yang, Yang and Wang, Yuchuan.
- [6] **A mixture model to detect edges in sparse co-expression graphs.**  
*Statistics in Medicine, under review, preprint arXiv:1804.01185*, 2019.  
Bar, Haim and Bang, Seojin.
- [7] **Dropout prediction over weeks in moocs via interpretable multi-layer representation learning.**  
*AAAI AI4EDU*, 2020.  
Jeon, Byungsoo and Park, Namyong and Bang, Seojin.

- [8] **Multiview cluster analysis identifies variable corticosteroid response phenotypes in severe asthma.**  
*American Journal of Respiratory and Critical Care Medicine*, (IF 16.49) **Highlighted Articles**, 2019.  
 Wu\*, Wei and **Bang\*** (co-first), **Seojin** and Bleecker, Eugene and Castro, Mario and Denlinger, Loren and Erzurum, Serpil and Fahy, John and Fitzpatrick, Anne and Gaston, Ben and Hastie, Annette and Israel, Elliot and Jarjour, Nizar and Kerr, Sheena and Levy, Bruce Meyers, Deborah and Moore, Wendy and Peters, Michael and Phipatanakul, Wanda and Sorkness, Ronald and Wenzel, Sally.
- [9] **Phosphoproteomic analysis of the amygdala response to adolescent glucocorticoid exposure reveals g-protein coupled receptor kinase 2 as a target for reducing motivation for alcohol.**  
*Proteomes*, 6(4), 2018.  
 Bertholomey, Megan L. and Stone, Kathryn and Lam, TuKiet T. and **Bang, Seojin** and Wu, Wei and Nairn, Angus C. and Taylor, Jane R. and Torregrossa, Mary M.
- [10] **Naïve bayes ensemble: A new approach to classifying unlabeled multi-class asthma subjects.**  
 In *IEEE International Conference on Bioinformatics and Biomedicine*, 2016.  
**Bang, Seojin** and Wu, Wei.
- [11] **Joint selection of snps for improving prediction in genome-wide association studies.**  
 In *IEEE International Conference on Bioinformatics and Biomedicine. Workshops*, 2012.  
**Bang, Seojin** and Kim, Yong-Gang and Park, Taesung.
- [12] **Ethnic variability in the allelic distribution of pharmacogenes between korean and other populations.**  
*Pharmacogenetics and genomics*, 22(12), 2012.  
 Kim, In-Wha and Im Kim, Kyung and Chang, Hyeu-jin and Yeon, Bora and **Bang, Seojin** and Park, Taesung and Kwon, Ji-sun and Kim, Sangsoo and Oh, Jung Mi.

## RESEARCH EXPERIENCE

### Interpretability and Robustness in Machine Learning

Jul – Aug 2019

*Visiting Graduate Student advised by Dr. Adrian Weller*

*ML Group at University of Cambridge, UK*

- Investigated the relationship between interpretability and robustness of machine learning and deep learning models.
- Developing a knowledge guided interpretable ML model using the posterior regularization. (in progress)

### Interpretable Machine Learning

May – Dec 2018

*Research Intern advised by Dr. Pengtao Xie*

*AI/ML Solution Team at Petuum, Pittsburgh PA*

- Developed a system-agnostic interpretable ML approach using neural networks using information bottleneck principle.
- Developed an approach to build a improve a out-of-sample prediction of epitope-TCR binding using an interpretable model.
- Developed a pytorch based software: **VIBI** ([github.com/SeojinBang/VIBI](https://github.com/SeojinBang/VIBI)) and **TCR** ([github.com/SeojinBang/TCR](https://github.com/SeojinBang/TCR))
- A paper submitted to *ICLR*

### Multimodal/Multiview Machine Learning for Asthma Subtype Identification

2015 – 2019

*Research Assistant advised by Dr. Wei Wu*

*Carnegie Mellon University, Pittsburgh PA*

- Developed a multiple kernel k-means clustering approach that is robust against adversarial features.
- Identified asthma subtypes showing differential responses to corticosteroid by combining multiview clinical data.
- Developed an R-package: **MKKC** ([github.com/SeojinBang/MKKC](https://github.com/SeojinBang/MKKC))
- Three papers published in *AJRCCM*, *IEEE BIBM* and *Proteomes*, and a paper submitted to *AAAI*.

### Other Collaborative Works with CMU Colleagues

2017 – 2019

*Graduate Student*

*Carnegie Mellon University, Pittsburgh PA*

- Developed an ontology-based neural network model for patient need detection. A paper in progress.
- Developed a phased-LSTM based predictive model for EHRs with complex missing patterns. Released a preprint.

## Statistical Modeling of Sparse Gene Networks

2013 – 2015

Research Assistant advised by Dr. Haim Bar

University of Connecticut, Storrs, CT

- Developed a statistical mixture model for better estimating sparse gene network.
- Helped to develop an R-package: **edgefinder**
- A paper under review in *Biostatistics*.

## Statistical Approaches for Biomedical Problems

2012 – 2013

Research Assistant and Scientist advised by Dr. Taesung Park

Seoul National University, Korea

- Developed a joint feature selection method using the elastic-net regularization to high-dimensional data.
- Deployed a time-dependent survival model to identify subtypes of intraductal papillary mucinous neoplasm.
- Two papers published in *IEEE BIBM* and *Pharmacogenetics and genomics*.

## PROFESSIONAL SERVICE

---

2019	<b>Reviewer</b> , <i>NeurIPS ML4H Workshop, IEEE Access, IEEE TNNLS</i>
2018	<b>Reviewer</b> , <i>ACM BCB</i>
2018	<b>Admission Committee</b> , <i>Computational Biology Department, Carnegie Mellon University</i>
2012	<b>Program Committee</b> , <i>International Symposium on Statistical Genetics, Korea</i>
2012	<b>Program Committee</b> , <i>Microarray Analysis Workshop: Statistical Analysis using R language, Korea</i>
2011	<b>Program Committee</b> , <i>The Spring Conference of the Korean Statistical Society, Korea</i>

## HONORS AND AWARDS

---

2018	<b>The Center for Machine Learning and Health Fellowships in Digital Health</b> full tuition and stipend for 12 months and research-related expenses (total \$75,200)	
2013	The Korean Statistical Society Paper Awards (3st Place)	
2012	The Korean Statistical Society Poster Awards (1st Place)	
2006 – 2010	<b>National Science and Engineering Undergraduate Scholarship</b> full tuition for 8 semesters an additional \$500 grant for a high GPA an additional \$500 grant for a high GPA	2006 – 2010 2009 2008

## SOFTWARE

---

<b>Python</b>	<b>VIBI</b> : pytorch implementation of VIBI <a href="https://github.com/SeojinBang/VIBI">https://github.com/SeojinBang/VIBI</a>
<b>Python</b>	<b>TCR</b> : TCR to epitope binding prediction approach using VIBI <a href="https://github.com/SeojinBang/TCR">https://github.com/SeojinBang/TCR</a>
<b>R-package</b>	<b>MKKC</b> : multiple kernel $k$ -means clustering on a multi-view data <a href="https://github.com/SeojinBang/MKKC">https://github.com/SeojinBang/MKKC</a>
<b>LaTeX template</b>	<b>TidyCV</b> : simple and tidy LaTeX template for your curriculum vitae <a href="https://github.com/SeojinBang/TidyCV">https://github.com/SeojinBang/TidyCV</a>

## TECHNICAL STRENGTHS

---

<b>Computer Languages</b>	Python, R, C/C++, MATLAB, Bash, HTML
<b>Library &amp; Others</b>	Pytorch, Tensorflow, Caffe, Keras, Git, LaTeX

## TEACHING EXPERIENCE

---

Carnegie Mellon University

Teaching Assistant

- 2018 Quantitative Cell and Molecular Biology Lab
- 2017 Computational Methods for Proteogenomics and Metabolomics

**University of Connecticut**

Teaching Assistant

- 2014 Mathematical Statistics
- 2014 Introduction to Mathematical Statistics
- 2013 Elementary Concepts of Statistics
- 2013 Introduction to Statistics I and II
- 2013 Statistical Methods

**Seoul National University, Korea**

Teaching Assistant

- 2012 Statistics Laboratory
- 2012 Regression and Analysis and Laboratory
- 2011 Statistics

**Bongyoung Girls' Middle School, Korea**

Student Teacher

- 2009 Middle School Mathematics

**Sungkyunkwan University, Korea**

Teacher

- 2006 Alternative Elementary/Middle School Mathematics