SANG WON BAEK

Phone: (+82)-10-8805-3873 **Email**: Baeksw98@gmail.com **GitHub**: https://github.com/Baeksw98 **Homepage**: https://baeksw98.github.io/

Research Interests & Objectives

Interests: Real-World Data Analytics, Predictive Modeling, Explainable AI, Interpretable ML, Multimodal Learning, Data Mining, Data Visualization

Objectives:

- Deliver interpretable and reliable AI tools to clinicians to provide a holistic view of disease progression, thereby facilitating clinical workflows and enhancing patient care.
- Develop a precision healthcare system using advanced AI tools to generate detailed patient profiles from multimodal physiological data, optimizing the use of comprehensive information for personalized treatment

Research Experience

AI Research Team, Sionic AI **Seoul, Korea**
AI Research Scientist 0*7/2024-Current*

- Leading a multimodal LLM research project to optimize AI reasoning and content detection in various file formats, aiming to develop state-of-the-art models for on-premise deployment

Department of Anesthesiology and Pain Medicine, Seoul National University Hospital **Seoul, Korea**
Medical Data Scientist 0*1/2024-06/2024*

- Developed a reinforcement learning-based intervention system for mechanical ventilators under the supervision of Dr. Hyunkyu Yoon
- Preprocessed **~70M** intraoperative biosignal data points extracted from the surgical data of **40K+** patients for model training and validation

Medical AI Research Center, Samsung Medical Center **Seoul, Korea**
Medical Data Scientist 0*8/2022-12/2023*

- Conducted a government funded (**$9M**), multicenter (**19** hospitals) study to develop and validate a ML-based Robust and Interpretable Early Triaging System (RIETS) for predicting COVID-19 severity progression as the **first author** under the supervision of Dr. Kyunga Kim
  - ✓ Established RIETS through comprehensively evaluating all possible combinations of candidate feature subsets and modeling algorithms
  - ✓ Utilized a tree-based unsupervised learning technique (i.e. DDRTree) to cluster and characterize patients based on the features in RIETS
  - ✓ Applied Shapley Additive exPlanations (SHAP) to interpret the marginal contributions of individual features in RIETS
- Conducted a retrospective observational study on lymphatic metastases patterns in esophageal squamous cell carcinoma (ESCC) as the **first author** under the supervision of Dr. Seongyong Park and Dr. Kyunga Kim
  - ✓ Performed network analyses to elucidate the complex lymphatic metastases patterns in ESCC patients
  - ✓ Implemented association rule mining to quantify the important associations among metastasized lymph nodes

Department of Laboratory Medicine, Gyeongsang National University Changwon Hospital **Changwon, Korea**
Lead Researcher 0*5/2021-12/2021*

- Conducted a retrospective cross-sectional study to develop a multivariable logistic regression-based mortality prediction system for severe patients with or without bacteremic sepsis as the **first author** under the supervision of Dr. Seungjun Lee
  - ✓ Collected clinical and laboratory biomarkers from **300+** severe patients by thoroughly investigating electronic health records
  - ✓ Utilized Kaplan-Meier curve to visualize the patient survival probabilities within a 30-day hospitalized period

Publications

1. **Baek, S.,** Jeong, Y. J., Kim, Y. H., Kim, J. Y., Kim, J. H., Kim, E. Y., Lim, J. K., Kim, J., Kim, Z., Chung, M. J†, Kim, K.† (2024). Development and Validation of a Robust and Interpretable Early Triaging Support System for Patients Hospitalized With COVID-19: Predictive Algorithm Modeling and Interpretation Study. *Journal of Medical Internet Research (JMIR)*, *26*, e52134. DOI: 10.2196/52134 **(IF: 5.8; Q1)**

2. **Baek, S.**, Kim, K. †, Park, S.Y. †, Jeon, Y. J., Lee, J. H., Cho, J. H., Kim, H. K., Choi, Y. S., Zo, J. I., Shim, Y. M. (2024). Application of Network Analysis and Association Rule Mining for visualizing the Lymph Node Metastasis Patterns in Esophageal Squamous Cell Carcinoma. *Under 2nd round review at Scientific Reports*

3. **Baek, S.**, Lee, S.J. (2023). Clinical Characteristics and Laboratory Biomarkers in ICU-admitted Septic Patients with and without Bacteremia: A Predictive Analysis. *medRxiv:* 2023.11.16.23298625

AI & Data Science Projects

SwinGPT: An LMM for Image Captioning and Object Detection tasks ***03/2024-04/2024***

- Fine-tuned a high performing large multimodal model to integrate text and vision inputs for image captioning and object detection tasks
- Accelerated model training using multi-node and multi-GPU setups with advanced optimization techniques (DDP, QLoRA, mixed-precision)

Fingertips Position Estimation of a Robot Hand ***10/2022-12/2022***

- Secured **1st place** for the school-wide Kaggle competition among 110 participants for most precisely estimating the fingertip positions of a robot hand by training RGBD images on 2D convolutional neural network (CNN) based model

Home Credit Default Risk Analysis ***01/2022-05/2022***

- Evaluated the feasibility of deploying an automated decision system (ADS) in a real-world environment from both ethical and legal perspectives through utilizing explainable AI tools (i.e. SHAP, LIME)

Abstractive Summarization for Long Input Text Question & Answering ***01/2022-05/2022***

- Implemented a large language model capable of abstractively summarizing lengthy texts and answering multiple choice questions

Education

New York University **New York, NY**
Bachelor of Arts in Data Science, Minor in Mathematics *09/2020-12/2022*

- GPA: 3.8 / 4.0 (Cum Laude)
- Relevant Coursework: Intro to ML, Data Management and Analysis, Causal Inference, Probability and Statistics, Linear Algebra

Babson College **Babson Park, MA**
Major in Business Analytics and Entrepreneurship (Transferred to NYU after Sophomore year) *09/2016 - 05/2018*

Teaching Experience

Institute of Convergence Medicine with Innovative Technology, Seoul National University Hospital **Seoul, Korea**
AI Research Seminar Speaker *01/2024-05/2024*
- Expanded modeling options in approaching healthcare AI research for **30+** researchers by introducing **10+** state-of-the-art transformer-based text and vision models

Kim Study Online Platform **Seoul, Korea**
AI/ML Programming Lecturer *07/2023-Current*
- Designed and implemented a customized curriculum (introductory and advanced programming techniques, explainable AI, data mining and analysis techniques, and data visualization methods) for **20+** students, primarily college graduates and MS/PhD students
- Guided **4** students into research roles in AI/ML fields by mentoring them with real-world case studies and hands-on projects

Research Institute for Future Medicine, Samsung Medical Center **Seoul, Korea**
Healthcare AI Instructor *04/2023-11/2023*
- Introduced **5+** prevalently used healthcare AI tools to medical doctors and researchers by conducting monthly AI seminars
- Contributed to the increase in thesis completion rate by thoroughly explaining the TRIPOD guideline, which details approaches for conducting appropriate statistical analysis of clinical data
- Facilitated the initiation of research collaboration between the hospital and the AI division of Samsung Electronics by providing feedbacks on their ongoing healthcare AI projects

Leadership Experience

Republic of Korea Air Force (ROKAF), 5th Air Mobility Wing **Pusan, Korea**
Squad Leader, Driver Instructor *08/2018-06/2020*
- Served as the Wing Commander's personal driver to fulfill mandatory military duty
- Created standard operating procedures to reduce operational conflicts between offices and units as a squad leader
- Instructed **70+** soldiers for their preparation of active duty operations

Tie-Off Organization – *Entrepreneurship Study Group* **Babson Park, MA**
Team Leader *09/2017-05/2018*
- Performed benchmark analyses to test the feasibility of a theoretical business model within the prospective US market
- Organized prototype fairs at five nearby colleges (Wellesley College, Olin College of Engineering, Brandeis University, Boston College, Babson College) to sell products and raised **~$8K** in revenue
- Generated a predictive model to calculate break-even time and quantity to maximize the operational efficiency based on collected data

FreshFeet – *College Start-up* **Babson Park, MA**
Co-Founder, Chief Technology Officer *09/2016-05/2017*
- Led the development and implementation of the business model using data analytics to identify market trends and customer preferences
- Initiated three on-campus prototype fairs to sell functional wearables and raised **~$3K** in revenue
- Coordinated panel sessions with industry experts to refine the business revenue model
- Leveraged data-driven insights to optimize financial projections and growth strategies

Academic Services

- Statistical Peer Reviewer for original articles and research letters, ***JAMA Network Open*** *01/2024-Current*

Conference Presentations

1. Oral Spotlight & **Poster Presentation**, "Early Triaging Support System for Hospitalized COVID-19 Patients: a Machine-Learning based Severity Prediction Model using Nationwide Multi-Center Real World Data", *American Society for Microbiology* (ASM) 2023, Houston, USA, June 2023

2. Poster Presentation, "Clinical Characteristics and Laboratory Biomarkers in ICU-admitted Septic Patients with and without Bacteremia: A Predictive Analysis", *American Society for Microbiology* (ASM) 2022, Washington, D.C., USA, June 2022

3. Oral Presentation, "Clinical Characteristics and Laboratory Biomarkers in ICU-admitted Septic Patients with and without Bacteremia: A Predictive Analysis", *European Congress of Clinical Microbiology and Infectious Diseases* (ECCMID) 2022, Lisbon, Portugal, April 2022

4. Poster Presentation, "Clinical performance evaluation of 'Boditech Quick COVID-19 Ag' test that can detect SARS-CoV-2 specific antigen in saliva from COVID-19 suspected patients", *Laboratory Medicine Congress & Exhibition* (LMCE) 2021, Online (Virtual Conference), September 2021

Patents

1. Baek, S., Kim, K. "Apparatus and method for predicting patient prognosis using machine learning model". Korea Patent Pending: No. 10-2023-0129662. 26 Sep. 2023

2. Baek, S., Kim, K., Park, S.Y. "Visualization method for lymph node metastases in esophageal cancer and apparatus". Korea Patent Pending: No. 10-2024-0002349. 05 Jan. 2024

Scholarship & Awards

- **Founders Day award (Honors Scholar)** *New York University*, April 2023
- **Student Government Assembly Conference Attendance Grant ($2K)** *New York University*, June 2022
- **International Student Scholarship ($4K)** *New York University*, April 2022
- **Dean's List for Academic Year** *New York University*, May 2021

Technical Skills

Languages: Python, R, SQL, Java, HTML, LaTeX, MongoDB, Tableau (in order of proficiency)
Frameworks/Libraries: Transformers, PyTorch, PyTorch Lightning, TensorFlow, Keras, Nltk, Peft, Lime, Shap, Dynamo, D3rlpy