

JAEMIN SONG

1100 Hensel Drive Apt College Station, TX 77840

+1 979-344-6080 | jaemin.song@tamu.edu |

github.com/jaemin-song876 | linkedin.com/in/jaemin-song876

EDUCATION

Texas A&M University

M.S. in industrial engineering, Data Science

Dec 26 (Expected)

- Coursework: Engineering Data Analysis, Databases and Computational Tools Used in Big Data

Hongik University, Seoul, Korea

B.S. in Industrial and Data Engineering

Mar 17 – Feb 24

- Coursework: Data Programming and Practices, Application of Data Analysis, Data Science, Spatial Big Data Analysis, Design of Human Sensibility Interface, Fundamentals of Internet of Things, Introduction to Data mining, Machine Learning and Smart City, Database, System Analysis
- Academic Scholarship for 4 semesters (Fall 2017, Fall 2018, Spring 2019, Fall 2019)

WORK EXPERIENCE

ELORILAB, Seoul, Korea

Sep 24 – Dec 24

Internship

- Collected and preprocessed hyperspectral imaging datasets to enable contaminant classification in specific food products
- Calibrated machine learning models for industrial deployment, improving classification accuracy and operational reliability
- Built and deployed a Retrieval-Augmented Generation (RAG) pipeline to enable natural-language querying of spectral and metadata, integrated with the agentic system based on ReACT
- Utilized LLMs into data exploration workflows by developing a chat-based dashboard app using LangChain and FastAPI, automating insight extraction from calibration and test reports and boosting data retrieval efficiency
- Developed an agentic system to orchestrate multi-step food quality analysis tasks, coordinating data retrieval, preprocessing, and result interpretation; deployed food analysis MCP tools on cloud
- Implemented GraphRAG techniques to link heterogeneous datasets (spectral signatures, contaminant profiles, lab notes) into a structured knowledge graph, enabling context-aware model recommendations

RESEARCH EXPERIENCE

Ultrasound Research Laboratory at Texas A&M University

Apr 25 – present

Graduate Research Assistant (PI: Dr. Kenneth Hoyt)

“Improving Tumor Segmentation Accuracy in Breast Cancer Ultrasound Images: A Deep Learning-Based Approach”

- Designed and optimized a U-Net deep learning model enhanced with attention gates for precise segmentation of breast tumors in medical imaging.
- Trained the model on data collected from two different medical centers using a binary cross-entropy loss function and evaluated its performance with the Dice score metric.

Hongik University Artificial Intelligence Application (AIA) Lab, Seoul, Korea

Feb 22 – Aug 24

Undergraduate Research Assistant (PI: Dr. Yong Oh Lee)

“Improving Laryngeal Cancer Detection: Comparing Octave-band Filtering & MFCC in Voice Data Machine Learning”

- Applied Octave Frequency Cepstrum (OFC) using an Octave Band Filter in place of the previously utilized Mel-Frequency Cepstral Coefficients (MFCCs).

“AI-based Diagnosis of Laryngeal Cancer using Voice Data for Enhanced Diagnosis and Treatment”

- Digitized analog voice frequencies by applying signal processing techniques such as Fourier transform, short-term Fourier transform, and discrete cosine transform to preprocess voice data.
- Calculated Mel-frequency Cepstral Coefficients (MFCCs) from vowel sound voice samples and converted the feature-engineered data into images for a proposed computer vision model, developing a novel feature engineering method of human voice data collected from laryngeal cancer patients by.

- Trained ResNet50 encoder via various ML classification models such as Extreme Gradient Boosting (XGBoost), Support Vector Machine, Artificial Neural Network, and Residual Neural Network for hypothesis testing and ablation studies.
- Extracted acoustic features such as jitter, shimmer, and HNR from voice data using Praat to overcome MFCC limitations.

“Predicting Chemical Properties using BiLSTM with BPE tokenization method using SMILES Representation”

- Developed a data-driven model using molecular data to predict compound properties for pharmaceutical companies to accelerate drug discovery and development.
- Tokenized Simplified Molecular Input Line Entry System (SMILES) data by reinventing a traditional data compression algorithm, “Byte Pair Encoding” (BPE) for Natural Language Processing (NLP) tokenization method.
- Improved the BPE algorithm for tokenizing SMILES data to use LSTM architecture and successfully tokenized data.
- Trained a Bidirectional LSTM (BiLSTM) model as a baseline of the proposed sequence classification model to classify toxicity with tokenized molecular data (tox21) using my tokenization pipeline.

JOURNAL PUBLICATIONS

J. Song, H. Kim, and Y.O. Lee. “Laryngeal Disease Classification Using Voice Data: Octave-Band vs. Mel-Frequency Filters.” Heliyon, 2024.

H.B. Kim, J. Song, S.H. Park, and Y.O. Lee. “Classification of Laryngeal Diseases including Laryngeal Cancer, Benign Mucosal Disease, and Vocal Cord Paralysis by Artificial Intelligence using Voice Analysis.” *Scientific Reports*, 2023.

PRESENTATIONS

Conference

J.M. Song, et al. “Enhancing Vocal-Based Laryngeal Cancer Screening with Additional Patient Information and Voice Signal Embedding.” Big Data Analytics for Health and Medicine (BDA4HM 2023), Workshop at 2023 IEEE International Conference on Big Data (IEEE Big Data 2023), 2023.

Oral Presentation

“Prediction and Analysis Utilization Methods of Seoul’s Popular Places,” Korea Data Agency (KDATA), Seoul, Korea, 2022.

HONORS & AWARDS

2nd place, Bio-health Data Competition, National Research Foundation of Korea 2023

- Developed a predictive model using Convolutional Neural Network (CNN) techniques on dental X-ray images to assess the risk of alveolar nerve damage post wisdom tooth extraction.

1st place, Bio-health Academic Competition, National Research Foundation of Korea 2023

- Recognized for the project “Voice-based Deep Learning Model for Screening Laryngeal Diseases,” demonstrating an innovative application of deep learning in medical diagnostics.

3rd place, Data Science Competition, Korean Data Agency affiliated by Government of Korea 2023

- Developed a machine learning model utilizing SVM, LightGBM, XGBoost, Decision Tree, and KNN to analyze the traits of existing popular places in Seoul city, forecasting future popular tourist locations.

TECHNICAL SKILLS

Programming Language: Python, R, C, C++,

Frameworks and Packages: TensorFlow, PyTorch, Scikit-Learn, NumPy, Librosa

Data Engineering: SQL, AutoML

Gen AI: LLMs, LangChain, LangGraph, Semantic Kernel, MCP,

AI tools: Cursor, ChatGPT, Gemini CLI, Claude Desktop