**Hyunjun Choi**

27121 Valley Oak PL, Valencia, CA 91381 • choirene@alumni.usc.edu • 818-838-5373

## Education

**University of Southern California** — Los Angeles, CA
Master of Science, Major in Computer Science. — December 2022
Master of Science, Major in Data Informatics. — December 2018

**Inha University** — Incheon, South Korea
Bachelor of Science, Major in Computer Science. — August 2016

## Work Experience

**Cedars-Sinai Medical Center** — Los Angeles, CA
**Centre for AI Research and Education, Department of Computational Biomedicine**
**Research Data Scientist** — May 2024 – Current 2024
**Research Associate Data Scientist** — October 2023 – May 2024
**Associate Programmer Analyst** — June 2022 – October 2023

- Conducted research focused on classifying patients with **Postural Orthostatic Tachycardia Syndrome (POTS)** by designing and optimizing advanced deep learning models, including **Transformer** architectures, **Temporal Convolutional Networks (TCN)**, and **Long Short-Term Memory (LSTM)** in **PyTorch**, using physiological time-series data derived from electrocardiogram (ECG) and chest-worn accelerometers. Improved latent feature representations of physiological signals through Self-Supervised Learning (SSL)-based Contrastive Learning frameworks, achieving strong performance on an independent holdout test set with Balanced Accuracy of 86.77%, Sensitivity of 93.55%, Specificity of 80.00%, Precision of 96.67%, and F1 Score of 95.08%, thereby demonstrating clinical validity and effectiveness of the proposed methodology.
- Conducted research to improve model performance by utilizing synthetic data generation techniques (Gaussian Copula, **Conditional GAN (CTGAN)**, **Bayesian Network (BN)**, **TVAE** , **RTVAE**, **Denoising Diffusion Probabilistic Models (DDPM)**) using **TensorFlow, PyTorch, and Python** to address data scarcity issues in developing machine learning models for QTc interval monitoring in Cardiac Surgery Intensive Care Unit (CSICU), achieving 92% MCC in collaboration with a Cardiology fellow.
- Designed, implemented, and deployed an AI-powered web application for Alzheimer's drug repurposing research, leveraging **Graph of Thought (GoT)**, **Retrieval-Augmented Generation (RAG)**, and **LLMs** to derive insights from a graph-based knowledge base (Memgraph) summarizing Alzheimer's Disease drugs, genes, pathways, and other biological entities, enabling their effective utilization in research
- Fine-tuned **open-source large language models** using an **Alzheimer's dataset** and compared this with an approach combining **Retrieval-Augmented Generation (RAG)** and **chain of thought**, achieving **80.7% accuracy**, while also co-leading the evaluation of these models and contributing to a **ReactJS and Python-based web application** featuring a leaderboard for **performance metrics**.
- Implemented and tested **RAG-based approaches** with **graph of thought methodology**, combining and evaluating performance with both **fine-tuned large language models** on **Alzheimer's knowledge graph data** and **non-fine-tuned models**, achieving **80.3% accuracy** using **Python and Weaviate**.
- Conducted **data cleaning, exploratory analysis, preprocessing, feature engineering, and machine learning** on a spinal surgery dataset in **Python**, achieving balanced accuracies of 0.75 for binary discharge disposition (home or non-home), 0.73 for binary hospital stay length (greater than 7 days), and 0.64 for binary 90-day readmission.

## Publications

- **Hyunjun Choi**, Nicholas Matsumoto, Xi Li, Debbie Teodorescu, Anxhela Kote, Min-Jing Yang, Xiao Liu, Miguel E. Hernandez, Jason H. Moore, Graciela Gonzalez Hernandez, Peng-Sheng Chen. Deep Learning-based Classification of Patients with Postural Orthostatic Tachycardia Syndrome using Wearable ECG and Accelerometer Data. Pacific Symposium on Biocomputing (PSB), 2026. (Accepted for Proceedings).
- **Hyunjun Choi**, Debbie Lin Teodorescu, Trevor Mears, Gizem Bilgili, Xi Li, Jui-Hsuan Chang, Nicholas Matsumoto, Miguel E. Hernandez, Zhiping Paul Wang, Bernice Coleman, Jason H. Moore. Enhanced QTc Interval Monitoring in the CSICU: Evaluating the Impact of Synthetic Data and Machine Learning Techniques. Pacific Symposium on Biocomputing (PSB), 2025. Poster presentation.
- Matsumoto N., **Choi H.**, Moran J., Hernandez M.E., Venkatesan M., Li X., Chang J.H., Wang P., Moore J.H. "ESCARGOT: An AI Agent Leveraging Large Language Models, Dynamic Graph of Thoughts, and Biomedical Knowledge Graphs for Enhanced Reasoning."Bioinformatics, 2024.
- Walker C., Ghosh A., Freda P., Shahrestani S., Boyke A., Orlenko A., **Choi H.**, Matsumoto N., Obafemi-Ajayi T., Moore J. "Pre-Operative Anemia is an Unsuspecting Driver of Machine Learning Prediction of Adverse Outcomes after Lumbar Spinal Fusion."Journal of Neurology, 2024.
- Matsumoto N, Moran J, **Choi H**, et al. KRAGEN: a knowledge Graph-Enhanced RAG framework for biomedical problem solving using large language models. Bioinformatics. Published online June 3, 2024. doi:10.1093/bioinformatics/btae353.
- Bilgili G, Teodorescu DL, Mears T, Moore JH, Orlenko A, Hernandez ME, **Choi H**, Matsumoto N, Skrebtsova I, Kazerouni S, Nurok M, Coleman B. Watching for (QT)rouble: Live, Automated vs Spot, Manual QTc Monitoring in Cardiac Surgical Critical Care. Poster presented at: American College of Cardiology 2024; April 6-8, 2024; Atlanta, GA. Poster Number: 1421-198. Session: Special Topics: Critical Care: Cardiology.
- **Hyunjun Choi**, Nicholas Matsumoto, Jay Moran, Miguel E. Hernandez, and Jason H. Moore. "Aliro: An Automated Machine Learning Tool Leveraging Large Language Models."Accepted for publication in Bioinformatics 2023.
- Ribeiro, Pedro, Anil Saini, Jay Moran, Nicholas Matsumoto, **Hyunjun Choi**, Miguel Hernandez, and Jason H. Moore. "TPOT2: A New Graph-Based Implementation of the Tree-Based Pipeline Optimization Tool for Automated Machine Learning."In Genetic Programming Theory and Practice XX, pp. 1-17. Singapore: Springer Nature Singapore, 2024.
- Alena Orlenko, Philip J. Freda, Attri Ghosh, **Hyunjun Choi**, Nick Matsumoto, Tiffani J. Bright, Corey Walker, Tayo Obafemi-Ajayi, and Jason H. Moore. "Cluster Analysis Reveals Socioeconomic Disparities Among Elective Spine Surgery Patients."Accepted for publication in the Pacific Symposium on Biocomputing 2024.
- Matsumoto Nicholas, Anil Kumar Saini, Pedro Ribeiro, **Hyunjun Choi**, Alena Orlenko, Leo-Pekka Lyytikäinen, Jari O. Laurikka, Terho Lehtimäki, Sandra Batista, and Jason H. Moore. "Faster Convergence with Lexicase Selection in Tree-Based Automated Machine Learning."Genetic Programming: 26th European Conference, EuroGP 2023, Held as Part of EvoStar 2023, Brno, Czech Republic, April 12–14, 2023, Proceedings. Cham: Springer Nature Switzerland, 2023.