June (Junhyeok) Lee

born-june04 | junh-lee | Portfolio | june0604@uw.edu | +1.206.512.7412

Summary

I develop clinically aligned representation learning systems for biomedical AI. I move beyond black-box optimization by embedding physiological and structural priors into models to improve robustness under real-world distribution shift.

Education

2025 – Present **M.S. in Bioengineering**, University of Washington GPA: 3.9/4.0
Focus: Multi-scale biomechanics and physics-informed deep learning
Expected Graduation: June 2026

2018 – 2024 **B.S. in Biomedical Engineering**, Hankuk University of Foreign Studies GPA: 3.9/4.0
Thesis: *Deep Clustering of Single-cell RNA-seq with Combined Features*

Work Experience

ML Engineer — Monitor Corporation Jul 2023 – Jun 2025

- **Probe-Conditional Ultrasound Denoising (Production DDPM)**:
  - Reduced false positives by **89% across 50K+ clinical scans** by deploying a probe-conditional DDPM with contrastive learning to suppress spectral artifacts across three GE ultrasound probe variants
  - Eliminated per-probe retraining by unifying domain adaptation into a single conditional architecture, explicitly conditioning on probe-specific prior distributions to address hardware-induced domain shift
- **Semi-Supervised 3D Segmentation Pipeline**:
  - Achieved **0.72 Dice** with sub-second inference for tumor segmentation by building a Fast Fourier Convolution (FFC) 3D U-Net with multi-scale frequency filtering
  - Reduced annotation cost by **21,600 hours ($540K)** by designing a MedSAM-inspired semi-supervised framework with reliability scoring, scaling **300 to 21,900 labeled scans** (1:73 ratio) while maintaining fully supervised parity
- **Model Compression for Edge Deployment**:
  - Achieved **2.5× faster inference** preserving **98% accuracy** by applying bfloat16 quantization and structured pruning to compress 3D segmentation models by 50%, enabling real-time clinical deployment

Research Experience

Vector Decomposition from Frozen Segmentation Encoders — Asan Hospital Jan 2025 – Present

- Discovered latent biomarker axes in frozen segmentation encoders recoverable with **zero trainable parameters**, achieving directional stability (5-fold cosθ=0.90/0.96) via centroid-difference projections
- Demonstrated zero-shot generalization to held-out patients (**CVS AUC 0.803, PRL AUC 0.716**) without any task-specific training
- Validated physical grounding with imaging gradients (SWI gradient $p < 0.001$; PRL rim $\rho = -0.49$, $p < 0.001$); submitted to *MICCAI 2026* (2nd author)

WiseFind: Structure-as-Context Multimodal RAG — Harborview Medical Center (HMC) Sep 2025 – Present

- Identified intra-domain visual homogeneity as a fundamental retrieval barrier (silhouette <0.06 post-fine-tuning), motivating text- and structure-based retrieval over visual embeddings
- Achieved **P@1 0.848, MRR 0.872, R@5 0.701—3.3×** above best visual baseline on 200 expert neurosurgical queries by proposing Shotgun Retrieval via THC Tree with three complementary channels (T2I, CI, XRef)
- Raised hard-query P@1 by **+76%** for figures >10 pages from text anchor by fine-tuning ColBERT (PubMedBERT backbone, 30K steps) on 5,544 domain-specific pairs with cross-reference following
- Submitted to *MICCAI 2026* (1st author)

WiseMind: RAG-Powered Neurosurgical AI Assistant — HMC Feb 2026 – Present

- Reduced hallucination in medical LLM responses by designing a **multi-step chain-of-thought pipeline** (Draft → Review → Refine) with retrieval-grounded inline **[N] citations** linking each claim to exact source passages for clinician verification
- Improved clinical recall by engineering **ontology-driven query expansion** (1,730 medical terms) and domain-scoped retrieval with automatic fallback, covering **27,845 indexed passages** across textbook and guideline corpora
- Achieved source-faithful multi-document synthesis by prompting **MedGemma 27B** to present conflicting evidence side-by-side rather than silently resolving disagreements between clinical guidelines
- Building an **RLHF data pipeline** from clinician thumbs-up/down feedback and RAG-provenance-annotated conversations stored in MongoDB, targeting reward model training for domain-aligned LLM fine-tuning; deployed at HMC with 7 release iterations (v1.0–v1.4)

WiseSpine: Patient-Specific Vertebral Fracture Simulation — HMC Sep 2025 – Present

- Reproduced **full A0–A4 AO classification** from first principles by building a patient-specific voxel finite-element engine (234K hexahedral elements, $Ku{=}F$) with HU-derived anisotropic material properties directly from clinical CT
- Achieved **clinically consistent force–fracture thresholds** (A0 at 1 kN → A4 at 8 kN; BMD 0.3 → 77.5% yield) by implementing transversely isotropic elasticity, incremental loading, and progressive continuum damage mechanics with cortical shell blending
- Enabled **per-scenario simulation in <20 min** on a single GPU by accelerating sparse assembly and CuPy-based conjugate gradient solve over 784K DOF, making patient-specific fracture risk assessment clinically feasible
- Formulated segmentation robustness as a **physics-constrained counterfactual augmentation** framework with causal DAG analysis (force, BMD, flexion → AO type), generating balanced A0–A4 training data from normal CT scans; preparing *NeurIPS 2026*

Culture-Positive TB Detection and Physiologically Grounded AI Modeling — Ubicomp Lab Sep 2025 – Present

- Reduced TB diagnostic latency from weeks to real-time, achieving **AUC 0.9519** and **92.7% sensitivity** by developing an acoustic detection system on 7,031 cough segments
- Identified infection-linked spectral signatures by extracting **57 physiological biomarkers**, revealing demographic and bacterial-load heterogeneity (female-specific slope $d = 0.593$; age-dependent reversals)
- Demonstrated interpretability via saliency concentration (25.9% in 500–3000 Hz) and biomarker correlation ($|r|$ up to 0.42); in preparation for *Nature Communications*

Tissue-Aware Drug Response Prediction — Bioinformatics Lab Jan 2022 – Dec 2025

- Achieved $R^2 = 0.947$ (Pearson $r = 0.977$) with **3× MSE reduction** over best baseline on 114,215 drug–cell line pairs (181 drugs, 682 cell lines, 13 tissues) by modeling tissue identity as a structural prior via ROKU gene weighting from GTEx
- Demonstrated tissue weighting accounts for **89% of total R^2 gain** ($\Delta R^2 = +0.148$), confirming tissue context as the dominant modeling factor over architectural complexity
- Validated biological grounding via input saliency: known target recovery (BRAF rank #2/35,830 for Dabrafenib, $p < 10^{-4}$), mechanism-based drug clustering ($p = 3.1 \times 10^{-4}$), and literature co-occurrence correlation ($r = 0.41$, $p = 0.009$)
- Manuscript under review at *Bioinformatics*

MRI Bias Field Correction with Generative AI — MedicalVision Lab Aug 2022 – Apr 2023

- Achieved **9.75× faster inference** than standard DDPM while preserving diagnostic fidelity by developing a diffusion-based MRI bias field correction with Hampel mixture noise modeling; published at **MIDL 2023**

Hackathon

Google MedGemma 2026 Healthcare Hackathon (GitHub, Demo) Jan 2026 – Present

- Achieved end-to-end multi-modal clinical diagnosis in ∼5s on consumer hardware by architecting a **three-agent pipeline** (Audio/HeAR, CXR/MedSigLIP, CT/MedSigLIP) with confidence-weighted late fusion mirroring the multi-specialist case conference model
- Produced interpretable *biological stories* linking findings to diagnoses by designing a **3-level hierarchical reasoning framework** (pathophysiology → pattern recognition → disease-specific biomarkers) with explicit quantitative thresholds
- Formalized **disagreement as a clinical signal**: divergent agents raise a *NEEDS CONFIRMATION* flag with modality-specific weights (CT ×1.2, CXR ×1.0, Audio ×0.8), surfacing co-morbidity hypotheses rather than silently overriding minority agents
- Designed for **modular deployment in resource-limited settings** — audio-only smartphone screening (HeAR) gracefully scaling to CXR then CT; FHIR-compatible JSON outputs; submitted to *MedGemma Impact Challenge 2026* (Kaggle)

Seoul National University Hospital (SNUH) Skin Burn Classification Hackathon — 3rd Place 2022

- Improved burn depth classification without increasing model complexity by introducing a physiologically motivated attenuation-inspired RGB reweighting scheme (0.21, 0.72, 0.07), emphasizing hemoglobin-sensitive green-channel information
- Awarded **3rd Place** among national-level medical AI teams

Publications

Lee, Junhyeok, Alex Ching*, Cynthia Dong*, David Horne, Thomas Hawn, Shwetak Patel. (2026). Rapid Identification of Culture-Positive Tuberculosis from Cough Sounds Using Deep Learning. *Nature Communications* (Manuscript in preparation).

Lee, Junhyeok, Aruna Souri*, Mary Nguyen*, Benjamin Han*, Christopher W. Lewis, Diana Wiseman. (2026). WiseFind: Why Structure Matters for Traceable Retrieval in Neurosurgery. *Medical Image Computing and Computer Assisted Intervention (MICCAI)* (under review).

Jeong, Eunseon, **Lee, Junhyeok**, Yangshin Choi, Yoonho Nam. (2026). Extracting CVS and PRL Biomarker Axes from Frozen Lesion Segmentation Encoders. *Medical Image Computing and Computer Assisted Intervention (MICCAI)* (under review).

Lee, Junhyeok, Jiwon Seo, Younhee Ko. (2026). Tissue-Specific Gene Weighting Enables Interpretable Drug Response Prediction. *Bioinformatics* (under review).

Lee, Junhyeok*, J. Kang*, Y. Nam, and T. Lee. (Apr. 2023). Bias Field Correction in MRI with Hampel Noise Denoising Diffusion Probabilistic Model. *Proceedings of Medical Imaging with Deep Learning (MIDL)*.

Teaching Experience

Teaching Assistant — BIOE 437/537 Computational Systems Biology Fall 2025
University of Washington, Prof. Herbert M. Sauro & Prof. Joseph L. Hellerstein

– Assisting enzyme kinetic simulation and optimization coursework; leading office hours and grading
– Contributing to novel research integrating simulation-generated data for AI training

Technical Skills

Programming Languages	Python, R, Matlab, SQL, C/C++, Flask, SvelteKit, Neo4j, MongoDB, ChromaDB
ML/AI Frameworks	PyTorch, Transformers, DDPM, GenAI, AWS SageMaker, LLM, RAG, ColBERT, FAISS, Ollama
Systems/Infrastructure	PyCuda, TensorRT, Distributed Training, Docker, SLURM, HPC Cluster, Apptainer, nginx, Gunicorn
Project Management	Atlassian Jira, Confluence, Github
Research Methodologies	Physical Simulation, Bayesian Optimization, Statistical Validation, Clinical Trials

Awards & Honors

2022	**Director's Medal**, Korean Information Society Development Institute	ACK
2022	**3rd Place**, Skin Burn Classification Hackathon	Seoul National University Hospital
2021	**Army Commendation Medal (ARCOM)**	U.S. Army