Baiying Lu

+1 410-934-8293 | baiying.lu.gr@dartmouth.edu | LinkedIn | Google Scholar |
HomePage | Hanvoer, NH, USA

Overview

Ph.D. candidate in Computer Science specializing in applied machine learning, time-series modeling, and digital health. Experienced in developing large-scale ML pipelines, wearable data analytics, and mobile healthcare monitoring systems. I am actively seeking **Software Engineer, Machine Learning Engineer, or Research Intern** roles for **Summer 2026**.

Education

Dartmouth College — Hanover, NH
Ph.D. in Computer Science — *2021 – Present (Expected Dec 2026)*

Duke University — Durham, NC
M.S. in Biomedical Engineering; ***GPA: 3.97/4.00*** — *2019 – 2021*

Shandong University — Shandong, China
B.E. in Biomedical Engineering; ***GPA: 91.13/100*** — *2015 – 2019*

Internship Experience

Samsung Research America – Digital Health Team
Research and Development Intern — *Mountain View, CA Jun.–Sep. 2025*

- Designed and implemented a scalable end-to-end data pipeline to ingest, preprocess, and manage 1M+ real-world smartwatch PPG segments, enabling large-scale model training and evaluation.
- Built and optimized deep learning models (ResNet, VGG, attention-based architectures) for arrhythmia detection, achieving >85% F1 while reducing per-segment inference error to <0.05 MAE.
- Integrated an arrhythmia detection task into a multimodal healthcare foundation model, contributing reusable components to an internal ML platform used by multiple downstream teams.

Novo Nordisk – AI Analytics, Digital Data Science Dept.
Data Scientist Intern — *Boston, MA Jun.–Aug. 2024*

- Built reusable data analysis pipelines in Python to process and analyze large-scale clinical wearable datasets (300+ patients, 30+ days per subject).
- Implemented feature extraction and aggregation workflows for continuous glucose monitoring and activity data, supporting downstream ML modeling.
- Developed statistical and ML-based analyses to identify key behavioral and physiological factors affecting glucose control, generating insights adopted by digital therapeutics teams.

Research Experience

Enabling Data-driven Health Management with Mobile Technology — Dartmouth College
Student Researcher, HealthX Lab — *Sept. 2021 – Present*

- Developed cross-platform mobile clients supporting both Android (Kotlin) and iOS (Flutter), enabling secure data synchronization from Dexcom CGM APIs and Fitbit APIs for continuous wearable data collection.
- Built and deployed backend services on Google Cloud Platform using Go, with Firebase for user management and scalable storage of high-frequency time-series sensor data.
- Lead a benchmarking project evaluating state-of-the-art time-series foundation models, LLM for time series, and deep learning architectures, including Chronos, TimesFM, TimeLLM, LSTM, and Transformers for blood glucose prediction using uni- and multimodal data from medical- and consumer-grade wearables.
- Built and maintained end-to-end research data pipelines covering study design, data collection, preprocessing, feature engineering, and statistical analysis for multiple digital health studies.

Digital Biomarkers Processing Pipeline from Wearable Devices Duke University

Research Assistant, Big Ideas Lab *Jan. 2020 – June 2021*

- Developed a data compression and transformation pipeline for digital biomarkers from wearable sensors, implementing data processing and transformation method, such as Wavelet Transforms, Autoencoders, and Huffman Encoding, to achieve up to 130× compression with under 2% information loss (published in *Sensors*, 2021).
- Collaborated with clinicians and behavioral scientists on an IRB-approved study assessing patient perspectives on wearable data sharing, contributing to study and survey design and manuscript preparation (published in *npj Digital Medicine*, 2021).

Depression Level Classification Using Three-Channel EEG Data Shandong University

Undergraduate Research Assistant, Licai Yang Lab *Jan. 2015 – June 2015*

- Implemented a EEG preprocessing pipeline using band-pass Butterworth filtering, Hilbert–Huang Transform, and Empirical Mode Decomposition to extract α-band activity.
- Engineered time- and frequency-domain features and applied SVM and Random Forest models to classify depression levels using clinical three-channel EEG recordings from real-world patients.

Selected Publication

B. Lu, ..., C. Kang, "Feasibility of Ectopic Beat Detection and Count Estimation from Smartwatch-Based Photoplethysmography," Accepted by ICASSP, 2026.

B. Lu, ..., X. Zhou, G. Forlenza, T. Prioleau, "Mealtime Prediction using Wearable Insulin PumpData to Support Diabetes Management," Scientific Reports - Nature, 2024.

B. Lu,..., S Tang, T. Prioleau, "GlucoFM-Bench: Benchmarking Time-Series Foundation Models for Blood Glucose Forecasting," Under Review, KDD, 2026.

B. Lu, T. Prioleau, "Deep learning for blood glucose prediction: reproducibility challenges and factors affecting differential performance," Under Review, Plos Digital Health, 2026.

P. Belsare, **B. Lu,** A. Bartolome, T. Prioleau, "Investigating Temporal Patterns of Glycemic Control around Holidays," IEEE Eng. in Medicine and Biology Conference, July 2022.

B. Bent, **B. Lu,** J. Kim, J.P. Dunn. "Biosignal Compression Toolbox for Digital Biomarker Discovery". Sensors, 2021.

Peer-Review and Academic Service

Reviewer: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), International Conference on Acoustics, Speech, and Signal Processing (ICASSP), PLOS Digital Health, American Medical Informatics Association (AMIA)

Review Scope: Evaluated manuscripts in machine learning for digital health, including glucose prediction models, wearable physiological sensing, AI for healthcare management, deep learning reproducibility, dataset benchmarking studies, and time-series modeling using wearable data.

Skills

Programming: Python, Kotlin (Android), Go, Flutter/Dart, React Native, C, C++, Java, JavaScript, SQL, Bash, Git, MATLAB, Linux.

Machine Learning & Data: PyTorch, TensorFlow, Scikit-Learn, SciPy, cuBLAS, cuDNN, NCCL architectures (LSTM, TCN, CNN, Transformer, time-series foundation model, time-series LLM), GPU computing, CUDA, data preprocessing & analysis, biosignal preprocessing&modeling, hyperparameter tuning, model evaluation & reproducibility

Cloud & Tools: AWS (S3, SageMaker), GCP, Firebase (Firestore, Cloud Functions), HuggingFace, Docker, REST APIs, real-time data streaming.

Experience Areas: Deep Learning, Machine Learning, Statistical Analysis, Artificial Intelligence, Algorithms, Data Science, Mobile App Development, Human–Computer Interaction.