

YIDONG ZHU

(781) 627-5858 | yidong.zhu@outlook.com | linkedin.com/in/yidongzhu | yidongzhu.us

Summary

- Clinical ML data scientist and Computer Science PhD with 7+ years applying Python, R, model evaluation, feature engineering, experimental design, and time-series analysis to healthcare and wearable datasets.
- Built ML-ready EHR, SDOH, wearable, survey, and sensor features across 350,000+ NIH participants and validated wearable-sensor ML methods in first-author publications.
- Combines healthcare ML, reproducible evaluation, clinical-data discipline, and scientific communication for applied machine learning and digital-health teams.

Core Skills

Programming, Cloud & Database: Python, R, SQL, BigQuery, AWS (Lambda, DynamoDB, S3, EC2, Cognito)

Machine Learning & Statistics: scikit-learn, PyTorch, TensorFlow, model evaluation, experimental design, feature engineering, time-series analysis

Data Engineering & Pipelines: reproducible ETL, validation workflows, data-quality checks, evaluation datasets, cohort logic

Healthcare & Research Domain: EHR, SDOH, Fitbit and wearable data, clinical datasets, sleep logs, surveys, HIPAA-aware workflows

Visualization & Communication: Tableau, ArcGIS, scientific reporting, stakeholder dashboards, cross-functional communication

Professional Experience

Graduate Research Assistant | University of Massachusetts Lowell

Jan 2020 - May 2026

Advanced Analytics, ML-Ready Features & Healthcare Data Products

- Built reproducible Python, SQL, and BigQuery ETL pipelines on the All of Us Researcher Workbench, filtering 350,000+ NIH participants into EHR, wearable, and SDOH cohorts for downstream analytics.
- Engineered BigQuery workflows with CTEs and TIMESTAMP_TRUNC aggregation to extract, clean, and pivot 1,440-point daily Fitbit arrays, reducing ad-hoc query time by approximately 60% and creating reproducible features for predictive-health modeling.
- Implemented Python-based address normalization and a QA workflow that achieved 85.5% match accuracy across 460+ community sample points, improving reliability of spatial data linking and location-based reporting.

Experiment Design, Behavioral Data & Visualization

- Developed and validated a human-activity-recognition approach using mixed time- and frequency-domain representations with MixUp augmentation, achieving 95.3 +/- 1.8 overall performance on the ADL dataset and 94.72 +/- 2.0 on the ASTRI dataset.
- Led technical development for a 200+ participant sleep-app study, implementing three-arm A/B data-collection and validation workflows to support engagement and habit-formation analysis.
- Built AWS Cognito, Lambda, and DynamoDB workflows to authenticate participants, validate sleep logs and surveys, flag duplicate or irregular entries, and improve reliability of downstream behavioral analysis.

Sensor-Based Monitoring & Real-World Health Data - NSF Award #2 230 180

- Translated study requirements into an AWS ETL pipeline using EC2, S3, Lambda, and DynamoDB to ingest, validate, and maintain high-frequency sensor readings from portable E-Tongue devices across Massachusetts communities.
- Integrated the AWS backend with ArcGIS to publish a real-time geospatial dashboard, surfacing threshold violations and anomalous patterns for technical and non-technical stakeholders.
- Designed a community-engaged citizen-science workflow supported by scalable IoT infrastructure for water-contamination monitoring.

Additional Experience

Marketing Analyst | Boston Software Group Inc.

Jul 2019 - Sep 2019

- Built Python ingestion and Tableau reporting workflows for 16,000 Massachusetts company profiles, loaded structured fields into MongoDB, and identified 80 high-potential target companies through exploratory segmentation analysis.

Data Analyst (Research Project) | Viacom

Apr 2019 - Jun 2019

- Analyzed 200,000+ Facebook advertising records, built a CPM prediction model for 443 page IDs, and delivered conversion-funnel recommendations for audience targeting and content optimization.

Informatics Analyst (Intern) | Southern New Hampshire Health, Nashua, NH

Nov 2016 - Mar 2017

- Developed the organization's first 56-item HIPAA compliance audit checklist, identifying 2 critical risks with data-visualization dashboards and 3 evidence-based mitigation proposals.
- Managed an intranet content migration from a legacy to modern platform, completing delivery three weeks early with zero downtime; trained 15+ content owners weekly.

Selected Research Presentations

- Wearable-based Fair and Accurate Pain Assessment Using Multi-Attribute Fairness Loss in CNNs, MobiQuitous 2024, Oslo, Norway (November 2024).
- Enhancing Wearable Sensor Data Classification Through Novel Image Representation and MixUp Augmentation, AMIA Annual Symposium 2024.
- Enhancing Wearable-Based Real-Time Glucose Monitoring via Phasic Image Representation Learning, IEEE EMBC 2024.

Teaching Experience

- Machine Learning (CS 5100): delivered lab materials and supervised hands-on sessions for 50+ students per semester. Computing II & III (C++): conducted labs and mentored 60+ students. Computing in Health and Medicine (CS 4150): supported student learning on healthcare ML applications.

Education

PhD, Computer Science | University of Massachusetts Lowell

2020 - 2026

- Defended April 2026; graduated May 15, 2026. Research: machine learning for healthcare, wearable computing, statistical modeling, time-series data, and cloud data pipelines.

MS, Analytics | Northeastern University

2018 - 2019

MS, Health Informatics & Management | University of Massachusetts Lowell

2014 - 2017

BS, Business Administration | Nanjing University of Finance and Economics, China

2000 - 2004

Completed Certifications

- Hugging Face Large Language Models Specialization (2025)
- AWS Certified AI Practitioner (2025)
- Tableau Desktop Specialist (2022)