

ZIMENG (SIMONE) LYU

✉ zimenglyu@mail.rit.edu 🌐 <https://zimenglyu.com> ☎ (315) 560-9401 🌐 github.com/zimenglyu

Professional Summary

Ph.D. Candidate in Computer Science specializing in time series forecasting, AutoML, Neural Architecture Search, and Transformers. Delivered \$7.3M in power plant cost savings and beat stock market index with evolved RNNs. Proficient in C++, Python, PyTorch, Docker, Spark, and distributed computing for large-scale AI solutions. Serve as course designer and instructor for a graduate core course and graduate student research mentor. Strong written and verbal communication abilities.

Education

Rochester Institute of Technology, Rochester, NY 2019 – 2025 (expected)
Ph.D. in Computing and Information Sciences GPA: 3.9

Syracuse University, Syracuse, NY 2016 – 2018
M.S. in Computer Engineering GPA: 3.7

Selected Top Publications

Zimeng Lyu, Alexander Ororbia, Travis Desell. “Online Evolutionary Neural Architecture Search for Multivariate Non-Stationary Time Series Forecasting,” *Applied Soft Computing*, 2023. (IF: 8.7, h5-index: 133)

Zimeng Lyu, Travis Desell. “ONE-NAS: An Online NeuroEvolution based Neural Architecture Search for Time Series Forecasting,” *GECCO 2022*. (h5-index: 47)

Research Impact

Full publication list can be found at my personal website

Neural Architecture Search for Time Series Modeling

- Developed distributed, robust evolutionary NAS frameworks using MPI/C++, deployed on large-scale high performance clusters.
- **Stock forecasting**: Evolved RNNs with **100× fewer parameters** vs. SOTA Transformers while maintaining same performance.
- **Coal-fired power plant**: Achieved a **\$7.3M reduction in revenue losses** by installing evolved RNNs in the management system.
- **Portfolio trading**: Successfully beat the market using the evolved RNNs for stock forecasting, on highly liquid DJIA stocks.

Online Neural Architecture Search & Adaptive ML

- Invented ONE-NAS, the first online evolutionary NAS system enabling real-time model evolution and adaptation.
- Designed novel mechanisms for handling non-stationary data streams with minimal latency, crucial for production deployment.
- Implemented a distributed training framework supporting continuous model updates in production environments.

Minimally Supervised Learning with Topological Projections

- Pioneered a minimally supervised learning approach using Self-Organizing Maps for data projection.
- **Regression**: Predicted 13 coal properties from 512-dimensional spectral data using only 67 labeled samples (0.3% of the dataset).
- **Classification**: Achieved high accuracy on imbalanced flight-phase data with as few as 30 labeled samples per class.

Industry Experience

Machine Learning Research Engineer 2018 – 2019

Quantifly, Syracuse, NY

- Led development of deep learning image segmentation system for large-scale aerial image analysis.
- Architected end-to-end ML pipeline for urban planning insights from drone imagery.
- **US Utility Patent** (US 16/800959): Vehicle Parking Data Collection System.

Teaching & Mentoring

Instructor, Software Engineering for Data Science (MS Core Course), RIT 2024 – Present

- **Designed** industry-focused data engineering curriculum featuring scalable data pipelines (Spark, MongoDB, MySQL, Airflow).
- Implemented DevOps practices requiring students to use GitLab CI/CD, automated testing, and code review via merge requests.
- **Pioneered 4 advanced, real-world projects**—Medallion (stock market analytics), Lambda (IoT streaming), ETL (order processing), and database normalization (retail)—and allowed students to use any large language models.

Graduate Capstone Mentor, Rochester Institute of Technology 2022 – Present

- Guided 11 graduate students across 6 capstone research projects, 4 conference proceedings.
- Oversaw **Transformer-based** stock return prediction and **evolutionary NAS** hyperparameter optimization.

Core Skills

Languages & Frameworks: Python, C++, PyTorch, TensorFlow, JAX, Hugging Face, DEAP, PySpark, CUDA, Hadoop, Hive

ML Infrastructure: Distributed (SLURM, MPI), Spark, Docker, AWS, Ray, GitLab CI/CD, MySQL, MongoDB, PostgreSQL

Tooling & Data Science: Statistical Modeling, A/B Testing, Time Series Analysis, Bayesian Inference, Scikit-learn, Pandas, NumPy