

Yinyu Yao

Chapel Hill, NC 27514 | yyao2@unc.edu | +1 (984) 261-5738 | yinyu-yao.github.io/personal/

EDUCATION

University of North Carolina at Chapel Hill (*BS–MS 5-year Dual Degree*)

M.S. in Statistics, Analytics and Data Science

Sept. 2025 – May 2026

University of North Carolina at Chapel Hill

B.S. in Statistics & Analytics and Mathematics double major with **Distinction**

Sept. 2021 – May 2025

— Cumulative GPA: 3.62/4.0; **Major GPA: 3.75**

— Honors: *Highest Honors in Statistics Dept (2/178); Honors Carolina Laureate (Top 10% of class); Dean's List*

— Awards: *Summer Research Award (\$5,000+ research funding in Summer 2024)*

RESEARCH INTERESTS

Optimization Algorithms; Machine Learning for Operation; Data-driven Operations

WORKING PAPERS

1. “Block Coordinate-Descent BFGS Method for Convex Non-Smooth Problems”. With Michael O’Neill.

Manuscript in preparation for submission to *Computational Optimization and Applications*. [[Link](#)]

— We develop a block-coordinate BFGS method for convex composite objectives with overlapping-group regularizers, prove almost-sure global convergence under random block updates, and show substantial runtime reductions versus inexact proximal-gradient and FoGLASSO baselines at matched accuracy.

2. “Data-Driven EMS Stroke Triage under Missed-Stroke Constraints”. With Ali Parlaktürk, Chudi Zhong.

Manuscript in progress.

— Using linked EMS–hospital data from NC, we standardize structured and free-text triage fields and learn dual-stage ML policies that cut missed strokes by >60% (2.3%→0.9%) while keeping false-positive rates comparable to current practice.

RESEARCH EXPERIENCE

Student Researcher – Optimization Algorithms

Aug. 2023 – Present

Advisor: Prof. Michael O’Neill | Dept. of Statistics & OR, UNC

— **Method design:** Proposed a block–coordinate BFGS for overlapping-group regularizers that reuses within-block curvature and integrates proximal handling of overlap.

— **Theory:** Proved global convergence under a relaxed smooth-extension framework on bounded level sets; derived per-block complexity accounting for sparsity and overlap size.

— **Computation:** Built MATLAB solvers and a reproducible harness; benchmarked vs Proximal Gradient, InexactPG, FoGLASSO, and full BFGS on synthetic, LIBSVM, and gene-expression datasets (Longleaf/Slurm).

— **Writing:** Wrote the manuscript and packaged full reproducibility (seeds, scripts, configs) for submission.

Student Researcher - Data-Driven Operations

Oct. 2024 – Present

Advisor: Prof. Ali Parlaktürk | Dept. of Operations, UNC Kenan-Flagler

— **Problem framing:** Cast EMS stroke triage as Neyman–Pearson–style constrained classification: learn dual-stage ML triage rules that minimize missed strokes subject to an upper bound on the false positive rate.

— **Data & feature engineering:** Linked NC EMS and hospital encounters; standardized structured and free-text triage fields into clinically curated binary indicators, interaction terms, and hierarchical diagnosis groupings to build a high-fidelity supervised dataset.

— **Modeling:** Trained calibrated gradient-boosting and interpretable models (XGBoost, EBM, etc.) with stratified CV, partial-AUC–weighted scoring, and ROC-based threshold search focused on the low-FPR region.

— **Reproducibility:** Designed a dual-stage “confirm-and-rescue” policy that reduces missed strokes by > 60% (from 2.3% → 0.9%) on held-out evaluation while keeping false positive comparable to current EMS practice.

ACADEMIC PROJECTS & CONTESTS

Kaggle Machine Learning Competition

Jan. 2024 - May 2024

Bronze Medal (top 9% among 3856 teams)

- Built stacked gradient boosting ensembles (XGBoost, LightGBM) for predicting credit card loan default risk, enhancing model accuracy and interpretability in financial decision-making contexts.
- Engineered advanced feature families (interaction terms, rolling-window statistics, ratio variables) while systematically controlling data leakage with stratified K-fold CV and fold-wise target encoding.
- Conducted hyperparameter optimization and calibration (Platt/Isotonic) to maximize leaderboard performance; ablation studies isolated gains attributable to feature engineering vs. model choice.

COMAP Mathematical Contest in Modeling

Feb. 2023

Honorable Mention (top 11% among 20000+ teams)

- Designed an Adjusted Climate-Focused Green GDP model using analytic hierarchy process (AHP) and entropy weight methods (EWM) to balance environmental and economic factors.
- Applied polynomial regression models to analyze relationships between GDP, CO₂ emissions, energy waste, and resource depletion across 10 countries; demonstrated the Environmental Kuznets Curve empirically.

COMMUNITY SERVICE & LEADERSHIP

Volunteer Tutor, UNC Math Help Center

Jan. 2023 - Oct. 2024

- Tutored first & second year students in calculus and linear algebra weekly

Co-Founder, UNC Culture Diversity Club (a(C)c)

Aug. 2021 - Oct. 2024

- Led 40+ cultural-diversity events with 10+ student organizations

- Recognized among UNC's most-viewed organizations; featured in UNC newspaper

oSTEM Active Member

Aug. 2021 - Oct. 2024

SKILLS

Programming Languages: MATLAB, Python, R

ML and Statistical Packages: MATLAB (Optimization Toolbox; Statistics and Machine Learning Toolbox); Python (NumPy, pandas, scikit-learn, XGBoost, LightGBM, CatBoost); R (tidyverse, data.table, ggplot2, caret, glmnet, randomForest, xgboost)

Databases and Tools: Git, Unix shell, Docker, SLURM cluster, IBM SPSS, SAS, LaTeX, Tableau, CPLEX

REFERENCES

Dr. Ali Parlaktürk (Ali_Parlakturk@kenan-flagler.unc.edu)

Professor of Operations, UNC Kenan-Flagler Business School

Dr. Michael O'Neill (mikeoneill@unc.edu)

Assistant Professor of Statistics & Operations Research, UNC-Chapel Hill

Dr. Zhengwu Zhang (zhengwu_zhang@unc.edu)

Associate Professor of Statistics & Operations Research, UNC-Chapel Hill