Yinyu Yao

1300 Fordham Blvd, Chapel Hill, NC 27514 | yyao2@unc.edu | +1(984)-261-5738

EDUCATION

University of North Carolina at Chapel Hill *(BS–MS Dual Program)*

M.S. in Statistics & Operations Research   Sept. 2025 – May 2026

— GRE: 328 (158V, 170Q, 3.5AW), Aug. 16, 2025

University of North Carolina at Chapel Hill

B.S. in Statistics 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 (2/178); Honors Carolina Laureate (Top 10% of graduating class); Summer Research Award ($5,000+ research funding); Dean's List*

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. Submitted to *Computational Optimization and Applications*.

   — *We proposed a block-coordinate quasi-Newton method for overlapping-group regularizers with a global convergence guarantee, significantly reducing runtime versus inexact proximal baselines at comparable accuracy.*

2. **"A Machine Learning Approach to Identify Missing Stroke Cases in EMS Triage"**. With Ali Parlaktürk. In progress.

   — *Using linked EMS–hospital data from NC, we train and evaluate ML triage policies that cut missed strokes by >30% (2.3%→1.4%).*

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 FN-minimizing decision rules (stable FP) via thresholding ML risk scores mapped to operational actions

— **Data & feature engineering:** Linked EMS and hospital systems by harmonizing schemas (age, sex, race) to build a supervised dataset with reliable labels; added clinically motivated interaction features;

— **Modeling:** Implemented pipelines for logistic regression, SVM, RF, XGBoost, and other ML Models with stratified CV, cost-sensitive metrics, and ROC/PR-based threshold selection.

— **Reproducibility:** Achieved >30% reduction in false negatives with no increase in false positives on held-out evaluation; documented code and provenance for full auditability.

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, CO_2 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
Technologies: IBM SPSS, LaTeX, Tableau, SAS, 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