

Zishan Shao

1000 McQueen Drive · Durham, NC 27106 · (336) 391-4963 · bruceshao5418@gmail.com

Education

Duke University, Durham, NC

M.S. in Statistics, May 2026

Wake Forest University, Winston-Salem, NC

B.S. in Computer Science (honored), May 2024

B.S. in Statistics, May 2024

Specialization: High-Performance Computing & Machine Learning

Dissertation: "Communication Avoiding Coordinate Descent Methods On Kernel And Regularized Linear Models"

Advisor: Dr. Aditya Devarakonda

GPA: 3.97 Summa Cum Laude

Research Interests

- Statistical & Machine Learning
- High-Performance Computing
- Numerical Linear Algebra & Methods

Related Courseworks

Multivariable Calculus (MST 113), Discrete Mathematics (MST 117), Linear Algebra I (MST 121), Fundamentals of Computer Science (CSC 112), Data Structures & Algorithms I (CSC 201), Computer Systems I (CSC 250), Computer Systems II (CSC 251), Programming Languages (CSC 231), Probability (STA310), Introduction to Statistical Learning (STA 363), Multivariate Statistics (STA 362), Computer Vision (CSC 391), Statistical Inference (STA 311), Linear Models (STA 312), Algorithm Design and Analysis (CSC 301), Network Analysis (STA 352), Time Series Forecasting (STA 368)

Research Experience

Undergraduate Researcher, Department of Computer Science, Wake Forest University *Spring 2022 – Spring 2024*

Advisor: Dr. Aditya Devarakonda

Communication-Avoiding Dual Coordinate Descent Methods for Kernelized Machine Learning

- Developed and implemented scalable DCD and BDCD methods, reducing communication overhead in distributed systems for kernel support vector machine (K-SVM) and Ridge Regression (K-RR).
- Introduced s-step variants that enhanced numerical stability and achieved up to 9.8× speedup on a Cray EXcluster.
- Performed theoretical analysis and empirical evaluations, demonstrating efficiency gains and robustness in convex optimization.
- GitHub Repository: <https://github.com/Zishan-Shao/s-step-method-comm-avoiding>

Communication Avoiding Coordinate Descent Methods on Regularized Linear Models

- Formulated communication-avoiding Coordinate Descent variants for Logistic and Multinomial Regression considering Elastic-Net and LASSO, optimizing for reduced data exchanges in distributed environments.
- Integrated the s-step method to diminish communication costs while preserving algorithmic robustness and effectiveness.
- GitHub Repository: <https://github.com/Zishan-Shao/s-Step-Coordinate-Descent-GLMNet>

Efficient s-Step Method via Low Rank Matrix Approximation

- Formulated communication-avoiding Coordinate Descent variants for Logistic and Multinomial Regression considering Elastic-Net and LASSO, optimizing for reduced data exchanges in distributed environments.
- Integrated the s-step method to diminish communication costs while preserving algorithmic robustness and effectiveness.
- GitHub Repository: https://github.com/Zishan-Shao/nystrom_approximation

Undergraduate Researcher, Department of Computer Science, Wake Forest University Spring 2022 – Present

Advisor: Dr. Victor P. Pauca

- Developed PalmProbNet, a probabilistic model using transfer learning and deep neural networks for the automated detection of palm trees in UAV-derived orthomosaic imagery of the Ecuadorian Rainforest.
- Employed image processing techniques and Python libraries to extract and label distinct palm and non-palm image patches, which improved the training of the ResNet-18 and MLP models.
- Achieved outstanding model accuracy of 97.32% and Cohen's κ of 94.59%, demonstrating the effectiveness of the approach in dense and challenging tropical forest environments.
- GitHub Repository: <https://github.com/Zishan-Shao/palmdetection>

Publications

- **Zishan Shao**, Aditya Devarakonda. “*Scalable Dual Coordinate Descent for Kernel Methods*”. Preprint. [PDF]
- Kangning Cui, **Zishan Shao**, Gregory Larsen, Victor Pauca, Sarra Alqahtani, David Segurado, João Pinheiro, Manqi Wang, David Lutz, Robert Plemmons, Miles Silman. “*PalmHeatNet: A Probabilistic Approach to Understanding Palm Distributions in Ecuadorian Tropical Forest via Transfer Learning*”. Proceeding of 2024 ACM-Southeast (ACM-SE' 24). [PDF]

Laboratory and Computer Skills

- **Proficient:** Python, C, Java, R, pandas, numpy, scikit-learn, tensorflow, keras, seaborn, opencv.
- **Extensive project experience:** Exploratory Data Analysis (EDA), Unsupervised Learning (Clustering, Principal Component Analysis, P-norms), Missing Value Handling (Imputation, Complete Case Analysis), using tools such as Matlab, OpenMPI, cblas, mkl.lapacke, sparseblas.
- **Experienced with:** SQL, Dynamic Modeling, Linux, HTML, CSS, JavaScript, Excel.
- GitHub Page: <https://zishan-shao.github.io/>

Professional Memberships

- Phi Beta Kappa Honorary Society – Member.

- Upsilon Pi Epsilon Honorary Society – Member.

Honors

- 2022 COMAP's Interdisciplinary Contest in Modeling (ICM) – ***Meritorious Winner***
- Dean's List Scholar (all semesters)
- 2023 Wake Forest Research Fellowship (WFRF) – Awarded \$4,000.
- George Washington Greene Scholarship – \$350, awarded to one of seven recipients with either a 4.0 average or an average extremely close to 4.0.
- Upsilon Pi Epsilon Scholarship Nomination – Awarded \$1,000.

Activities

Center for Learning, Access, and Student Success (CLASS)

Peer Tutor (Paid)

Winston-Salem, NC

Oct. 14th, 2021 – Dec. 10th, 2023

- Provided on-site peer tutoring in computer science for the Center for Learning, Advising, and Student Success (CLASS) at Wake Forest University.