Sambit Panda

Baltimore, MD 21218 | US Citizen

919-637-6272 | spanda3@jhu.edu | linkedin.com/in/sampan501 | github.com/sampan501 | sampan.me

SUMMARY

- Graduating PhD candidate with experience in data science, machine learning, neuroscience, and genomics
- Author of 14 publications (h-index: 6); see details at https://sampan.me/research/
- 5+ years experience in developing data science solutions in academic and industry settings
- Interdisciplinary collaborator with experience working in clinical and translational settings

WORK EXPERIENCE

NeuroData, Johns Hopkins — Jan 2019 – Present
Graduate Researcher — *Baltimore, MD*

- Wrote 9 publications related to hypothesis testing, random forest, early cancer detection, causal inference, and neural networks.
- Developed and maintained open-source **Python** software packages such as hyppo (over 50 users and 200 stars) and scikit-tree (over 50 stars) and ported algorithms from my package into **SciPy**.
- Developed and tested **Python** packages using **Git**, **Docker**, cloud environments like **AWS EC2/S3** and **Azure VMs**, CI such as **TravisCI**, **CircleCI**, and **GitHub Actions**, and coverage tools like **codecov** and **coveralls**; code used packages like **pandas**, **NumPy**, **SciPy**, **matplotlib**, **seaborn**, **scikit-learn**, **pytest**, **Sphinx**, **plotly**
- Presented work at top conferences like the BRAIN Initiative meeting and chaired/reviewed for the SciPy conference.
- Peer-reviewed a paper for the SoftwareX journal
- Advised the venture capitalist firm A-Level Capital investing in life sciences companies and led grad and undergrad students as a TA to complete research projects for NeuroData Design I & II.

Neurobehavioral Core, NIEHS — May 2023 – Jul 2023
Data Science Intern — *RTP, NC*

- Applied algorithms I developed on hypothesis testing, standard workflows in classification and exploratory data analysis using **Python 3** and packages like **pandas**, **scikit-learn**, **matplotlib**, **seaborn**, **NumPy** and **SciPy**; got 1st place at the poster conference for work.
- Wrote 2 publications related to neuroscience and a **R** package; helped develop tutorials using the **datatools** package and **MySQL 8.0**

Sombers Lab, NC State — Jan 2015 – May 2018
Research Assistant — *Raleigh, NC*

- Created a new electrochemical sensor and wrote paper about it in ACS Analytical Chemistry.
- Investigated the chemical basis of abnormal involuntary movements (AIMs) during Parkinson's Disease.
- Presented research at top conferences like society of neuroscience (SfN) and Pittcon.
- Analyzed data and engineered solutions for numerous additional projects using **MATLAB** and **Graphpad PRISM**.

Burleson Research Technologies — May 2015 – Sep 2015
Intern — *RTP, NC*

- Tested pharmaceutical drugs on rats and mice through various methods such as oral gavage, *i.p.*, and *i.v.*
- Helped lab run under good laboratory practices.

Developmental Neurobiology Group, NIEHS — Jun 2013 – Jan 2014
Research Assistant — *RTP, NC*

- Trained in several basic genetics and neuroscience techniques such as PCR, gel electrophoresis, etc.

PROJECTS

treeple | *Python, Cython, GitHub Actions, Pytest* — **2023 – Present**
Helped develop the package for extensions of scikit-learn decision trees (50+ stars).

hyppo (originally mgcpy) | *Python, TravisCI, CircleCI, Netlify, Codecov, Coveralls, AWS, Azure, Pytest* **2018 – Present**
Developed and maintained a comprehensive multivariate hypothesis testing package in Python (50+ users & 200+ stars).

sampan.me | *HTML/CSS, Coffee/JavaScript, TeX, Cloudflare* **2018 – Present**
My personal website.

FiPhA | *R, Shiny* **2023**
Helped develop one of the most robust and user-friendly packages for fiber photometry analysis.

scipy.stats.multiscale_graphcorr | *Python, Cython, Pytest* **2019**
Added Multiscale Graph Correlation, a powerful multivariate independence test, to SciPy (the first such test).

SKILLS

Languages: Python, R, MATLAB, Shiny, Coffee/JavaScript, HTML/CSS, SQL, Familiarity in C/C++ and Java
Developer Tools: Git, Docker, CircleCI, TravisCI, Codecov, Coveralls, AWS, Azure, VS Code, TeX, Netlify, Cloudflare, Graphpad PRISM
Databases: MySQL with datatools
Libraries: pandas, NumPy, SciPy, matplotlib, seaborn, scikit-learn, datatools, pytest, Sphinx, plotly

EDUCATION

Johns Hopkins Medical Institute **Baltimore, MD**
PhD, Biomedical Engineering *Jul 2020 – Present*

Johns Hopkins University **Baltimore, MD**
MSE, Biomedical Engineering *Aug 2018 – May 2020*

NC State University & UNC Chapel Hill **Raleigh & Chapel Hill, NC**
BS, Biomedical Engineering & Biology *Aug 2014 – May 2018*

PUBLICATIONS (Highlighting 5 of 14)

1. Curtis, S.*, **Panda, S.***, Li, A.*, Xu, H., Bai, Y., Ogihara, I., O'Reilly, E., Wang, Y., Dobbyn, L., Popoli, M., Ptak, J., Nehme, N., Silliman, N., Tie, J., Gibbs, P., Ho-Pham, L., Tran, B., Tran, T., Nguyen, T., Goggins, M., Wolfgang, C., Wang, T., Shih, I., Fader, A., Lennon, A. M., Hruban, R., Bettegowda, C., Gilbert, L., Kinzler, K., Papadopoulous, N., Vogelstein, B., Vogelstein, J. T.^, Douville, C.^ (2024). *Detecting and Combining Useful Sets of Predictive Variables.* Manuscript submitted for publication.
2. **Panda, S.***, Shen, C.*, & Vogelstein, J. T. (2023). *Learning Interpretable Characteristic Kernels via Decision Forests* (arXiv:1812.00029). arXiv. https://doi.org/10.48550/arXiv.1812.00029
3. Shen, C., **Panda, S.**, & Vogelstein, J. T. (2022). The Chi-Square Test of Distance Correlation. *Journal of Computational and Graphical Statistics*, *31*(1), 254–262. https://doi.org/10.1080/10618600.2021.1938585
4. **Panda, S.**, Palaniappan, S., Xiong, J., Bridgeford, E. W., Mehta, R., Shen, C., & Vogelstein, J. T. (2021). *hyppo: A Multivariate Hypothesis Testing Python Package* (arXiv:1907.02088). arXiv. https://doi.org/10.48550/arXiv.1907.02088
5. **Panda, S.***, Shen, C.*, Perry, R., Zorn, J., Lutz, A., Priebe, C. E., & Vogelstein, J. T. (2021). *Nonpar MANOVA via Independence Testing* (arXiv:1910.08883). arXiv. https://doi.org/10.48550/arXiv.1910.08883

PRESENTATIONS (Highlighting 3 of 21)

1. **Panda, S.**, Wilson, L. R., Stallone, J., Kendricks, D., Stevanovic, K., & Cushman, J. D. (2023, July). *Elucidating Relationships within Neurological Screening Batteries via Random Forest-Based Hypothesis Testing* [Poster Presentation] RTP, NC, USA.
2. **Panda, S.**, Shen, C., Perry, R., Zorn, J., Lutz, A., Priebe, C. E., & Vogelstein, J. T. (2022, January). *Nonparametric MANOVA via Independence Testing* [Oral Presentation]. Global Young Scientists Summit, Virtual. https://www.youtube.com/watch?v=rJyuTwkgfjQ
3. **Panda, S.**, Shen, C., Perry, R., Zorn, J., Lutz, A., Priebe, C. E., & Vogelstein, J. T. (2021, June). *Nonparametric MANOVA via Independence Testing* [Poster Presentation] BRAIN Initiative Meeting, Virtual.

AWARDS & HONORS

Computational Biology Fellowship, *Johns Hopkins University*   **2020**
AWS IMAGINE Grant, *Amazon Web Services* (Supported the mgcpy (now hyppo) package)   **2018**
Magna Cum Laude, *NC State University*   **2018**
University Honors Program, *NC State University*   **2018**
Dean's List, *NC State University*   **2014 – 2018**
Goodnight Scholarship, *NC State University*   **2014**
National Merit Corporate Scholarship, *National Merit Scholarship Corporation*   **2014**