Naveen Kannan

RESEARCH ASSOCIATE

Cleveland, Ohio | United States Permanent Resident

+1 973 519 5397 | naveen.kannan12@gmail.com | naveenk2022.github.io/quarto_blog | naveenk2022 | naveen-kannan-177120118

Data Scientist and Bioinformatician with 2+ years of research experience applying advanced data analytics and machine learning to human genomic data. Experienced in automated deployment of scientific workflows with cutting edge machine learning, data analytics, and bioinformatics tools, with a strong passion for enabling easy access to Big Data tools and containers.

Work Experience

Research Associate *Cleveland, Ohio*
PQHS, SCHOOL OF MEDICINE, CASE WESTERN RESERVE UNIVERSITY *MAY 2023 - Present*

- **Automated Large-Scale Cluster Deployment and Management**
  - Deployed fully automated, unattended and **remote OS provisioning** and **DHCP-based stable IPv4 assignment** across a high performance computing cluster via **Ansible scripts** integrated with a **PXE server** and local iDRAC.
- **PostgreSQL Database Administration**
  - Installed, configured, and deployed a secure multi-core PostgreSQL server instance, created and defined 3 databases, and utilized **PostgreSQL JDBC integration with Spark** to automate schema definition and loading of **terabytes of large genomic flatfiles** into databases.
- **Automated HDFS and Spark Deployment**
  - Developed Ansible scripts to automate the installation and configuration of **Apache Hadoop HDFS, YARN, MapReduce, and Spark** on a HPC cluster with 18 worker nodes and 1 master node, while also configuring the allocation of cluster resources to the applications.
- **Docker/Singularity Container Deployment**
  - Developed Mamba-based Docker/Singularity containers for reproducible scientific pipelines, reducing container build times by 300 percent, and transferred Docker storage cache to a physical hard drive.
- **Genomic Data Analysis**
  - Collaborated with the **Alzheimer's Disease Sequencing Project**, annotating and scoring **362 million genetic variants** associated with Alzheimer's; presented statistical analysis findings at ASHG 2023 conference.

Graduate Research Assistant *Cleveland, Ohio*
PQHS, SCHOOL OF MEDICINE, CASE WESTERN RESERVE UNIVERSITY *JANUARY 2023 - MAY 2023*

- Engineered Docker containers for streamlined deployment of specialized R packages and Python modules, optimizing analyses for single-cell RNA datasets.

Graduate Research Assistant *Cleveland, Ohio*
GIS HEALTH AND HAZARDS LAB, SCHOOL OF MEDICINE, CASE WESTERN RESERVE UNIVERSITY *MAY 2022 - DECEMBER 2022*

- Developed Python scripts and YOLOv4 ML algorithms to automate the extraction of GPS coordinates from geospatial video feeds of refugee camps in the DRC after Mt. Nyiragongo eruption to generate of heat maps of tent distribution in refugee camps.

Junior Resident Doctor *Chennai, India*
DEPARTMENT OF PSYCHIATRY, SAVEETHA MEDICAL COLLEGE *JULY 2020 - SEPTEMBER 2020*

- Evaluated and diagnosed psychiatric illnesses, and met the assigned goal of management of five patients per day.

Junior Resident Doctor *Chennai, India*
MADRAS MEDICAL COLLEGE *MARCH 2019 - MARCH 2020*

- **Rotated through the following departments:** Internal Medicine, General Surgery, Obstetrics and Gynecology, Community Medicine, Psychiatry, Emergency Trauma Ward, Labor Ward.

Skills

Operating Systems	Programming Languages	Software	Geospatial Analytics	Markup Languages	Version Control	Machine Learning	Bioinformatics Tools
Unix, Debian Linux, Fedora Red Hat Linux, Windows 11	Python, R, Bash, SQL	Hadoop/Spark, Hive, Jupyter, Docker, Singularity, SLURM, Ansible	Geopandas, sf(R), QGIS, ArcGIS	LaTeX, Markdown, RMarkdown	Git, GitHub, gh client	PySpark, Tensorflow, Pytorch, pandas, scikit-learn, numpy, scipy	samtools, BCFtools, HTSlib, Tabix, Hail, Glow

Education

MSc in Biomedical and Health Informatics *Cleveland, Ohio*
CASE WESTERN RESERVE UNIVERSITY *JANUARY 2022 – AUGUST 2023*

- GPA-3.9

Bachelor of Medicine and Bachelor of Surgery   *Chennai, India*

MADRAS MEDICAL COLLEGE   *SEPTEMBER 2016 - MARCH 2020*

- GPA-4.0

Leadership

Steiner House International Co-op   *Cleveland, Ohio*

VICE PRESIDENT, ADMISSIONS MANAGER, AND ORDERING MANAGER   *MAY 2022 - OCTOBER 2022*

- Spearheaded integration initiatives for new residents, fostering community through monthly social events, while screening and interviewing prospective residents.
- Managed inventory for Steiner House, ensuring a continuous supply of consumables and non-consumables.

Personal Projects

Mapping the Environmental Impact of the Norfolk Southern Train Derailment in Ohio.   *Case Western Reserve University*

HTTPS://NAVEEN-KANNAN.SHINYAPPS.IO/FINAL_EAST_PALESTINE_DASHBOARD_2   *2023*

- Developed scripts to scrape air and water quality data from USGS and AirNow AQI for counties within a 30-mile radius of a train derailment, creating an interactive Shiny dashboard in R for users to analyze and visualize temporal and spatial trends in water and air quality across affected counties in Ohio, West Virginia and Pennsylvania.

Data Science Blog   *Cleveland, Ohio*

HTTPS://NAVEENKANNAN.NETLIFY.APP/BLOG_MAIN.HTML   *2022*

- Created beginner-friendly tutorials on Docker containerization, Spark/Hadoop big data engineering, GPU programming, and MLOps. Content geared towards demystifying complex data science concepts for students and early career professionals.

Publications

- Ajayakumar, J., Curtis, A. J., Maisha, F. M., Bempah, S., Ali, A., **Kannan, N.**, Armstrong, G. & Morris Jr, J. G. (2024). Using spatial video and deep learning for automated mapping of ground-level context in relief camps. International Journal of Health Geographics, 23(1), 23.

- Ruksakulpiwat, S., Thongking, W., **Kannan, N.**, Wright, E., Niyomyart, A., Benjasirisan, C., . . . & Still, C. H. (2024). Understanding the Relationship Between Comorbidities, Medication Nonadherence, Activities of Daily Living, and Heart Condition Status Among Older Adults in the United States: A Regression Analysis and Machine Learning Approach. Journal of Cardiovascular Nursing, 10-1097.

- **Naveen Kannan**[1], Nicholas Wheeler[1], Genome Center for Alzheimer's Disease, Li-San Wang[2], Yuk Yee Leung[2], William S. Bush[1],[1]Cleveland Institute for Computational Biology, Department for Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA;[2] Department of Pathology and Laboratory Medicine, Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, Quality Control metrics of individual genetic variants in the Alzheimer's Disease Sequencing Project are associated with FAVOR annotations.

- **N. Kannan**[1], N. Wheeler[1], Genome Center for Alzheimer's Disease, L-S. Wang[2], Y. Leung[2], W. S. Bush[1]; [1]Cleveland Inst. for Computational Biology, Dept. for Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Pathology and Lab. Med., Penn Neurodegeneration Genomics Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA., Annotation and scoring of the deleteriousness of individual genetic variants in the 4th release of the Alzheimer's Disease Sequencing Project;(PB4451). Presented at the Annual Meeting of The American Society of Human Genetics, November 3, 2023, in Washington DC.

- Segamalai, D., Abdul Jameel, A. R., **Kannan, N**., Anbalagan, A., Duraisamy, B., Raju, P., & Devy Gounder, K. (2017). Mediastinal pseudocyst: varied presentations and management—experience from a tertiary referral care centre in India. *HPB Surgery, 2017*.

- Vellaisamy, R., **Kannan, N**., Anbalagan, A., Raju, P., Duraisamy, B., Murugesan, C. S., & Gounder, K. D. (2016). Endoscopic access to hepatic duct through duodenum during follow up–after primary surgery for hepatolithiasis. *HPB, 18*, e530.

- Ramasamy, V., Vellaisamy, R., **Kannan, N**., & Gounder, K. D. (2016). Refined technique of access loop in hepatobiliary surgery. *HPB, 18*, e593-e594.

- **Kannan, N.**, Vellaisamy, R., Govindarajan, M., & Gounder, K. D. (2016). Pellagra following pancreaticoduodenectomy for malignant pancreatic carcinoid with pluripotent hormonal potential. *HPB, 18*, e381-e382.