Naveen Kannan

RESEARCH ASSOCIATE

Cleveland, Ohio | United States Permanent Resident

☐ +1 973 519 5397 | ✉ naveen.kannan12@gmail.com | 🌐 naveenkannan.netlify.app | 📷 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 bioninformatics tools, with a strong passion for enabling easy access to Big Data tools and containers.

Work Experience

Research Associate   *Cleveland, Ohio*

POPULATION HEALTH AND QUANTITATIVE HEALTH SCIENCES, SCHOOL OF MEDICINE, CASE WESTERN   *MAY 2023 - Present*
RESERVE UNIVERSITY

- **Postgres Database Administration**
  - Configured and deployed a **PostgreSQL server** instance on Linux system with multiple CPU cores.
  - Created and managed 3 databases, handling **schema design** and data modeling.
  - Utilized **Spark's InferSchema tool** and automated loading of genomic flatfiles into tables within different databases.
  - Managed lab user roles, privileges, and access controls for secure database operations.
  - Implemented firewall rules and network security measures for database connections.
- **Docker/Singularity Container Deployment**
  - Developed **Mamba-based Docker/Singularity containers** for reproducible scientific pipelines, facilitating efficient workflows for colleagues, reducing container build times by **300 percent**.
  - Deployed **RStudio and JupyterLab servers** (with pre-installed Bioinformatics tools) through Docker containers, circumventing the need to install these tools on the server directly.
  - Transferred the Docker storage cache to a physical hard drive while ensuring existing images were preserved, eliminating all future storage errors while freeing up root filesystem overhead.
- **Automated Cluster Configuration and HDFS Deployment**
  - Utilizing Ansible scripts, automated the deployment,configuration and initiation of **HDFS, YARN, Hadoop and Spark/PySpark with a JupyterLab kernel** across a 4 node computing cluster.
  - Automated the deployment and configuration of **Bioinformatics (Hail, Glow)** tools on these clusters, interfacing with Spark and JupyterLab, using Ansible scripts.
- **Genomic data Analysis**
  - Collaborated with the **Alzheimer's Disease Sequencing Project** by annotating and scoring the biological effects of 362 million structural genetic variants associated with Alzheimer's Disease in the R4 release of the ADSP.
  - Performed statistical analysis on the Data and presented findings in the form of a poster at the ASHG (American Society of Human Genetics) 2023 Conference in Washington DC.
  - Developed **generative adversarial network (GAN)** via an automated pipeline to simulate a chunk of realistic human genomic data by training the model on 1000 Genomes data, while incorporating automated quality checks for LD, HWE, and more, as part of a pending journal publication.

Graduate Research Assistant   *Cleveland, Ohio*

POPULATION HEALTH AND QUANTITATIVE HEALTH SCIENCES, SCHOOL OF MEDICINE, CASE WESTERN   *JANUARY 2023 - MAY 2023*
RESERVE UNIVERSITY

- 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*

- **Mapping human movement patterns after a natural disaster in the Democratic Republic of Congo.**
  - Developed Python scripts automating GPS coordinate extraction from GPX files of geospatial video feeds from refugee camps in the Democratic Republic of Congo post the volcanic eruption of Mt Nyiragongo.
  - Processed and transformed data into a machine-readable format, enabling streamlined and automated map generation.
  - Applied automated scripts to parse video feeds into frames, implementing Machine Learning Algorithms for tent detection.
  - Created heat maps of tent distribution according to GPS coordinates, allowing for the mapping of the human movement pattern in the aftermath of the eruption.
  - Wrote up the findings as a part of a manuscript for a peer-reviewed publication.

Junior Resident Doctor   *Chennai, India*

DEPARTMENT OF PSYCHIATRY, SAVEETHA MEDICAL COLLEGE   *JULY 2020 - SEPTEMBER 2020*

- **Diagnosing and managing patients with psychiatric illnesses**
  - Learned to evaluate and diagnose psychiatric illnesses in stressful environments.
  - Assigned and achieved a 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.

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

Skills

Operating Systems
UNIX, KUBUNTU DEBIAN LINUX, FEDORA RED HAT LINUX, RASPBIAN DEBIAN LINUX, WINDOWS 11

Programming Languages
PYTHON, R, BASH, SQL

Software
HADOOP/SPARK, HIVE, JUPYTER, DOCKER, SINGULARITY, SLURM, ANSIBLE

Geospatial Analytics/Visualization
GEOPANDAS, SF (R), QGIS

Markup Languages
LATEX, MARKDOWN, RMARKDOWN

Version Control
GIT

Python-based skills
PYSPARK FOR HADOOP/SPARK, TENSORFLOW, PYTORCH, PANDAS, SCIKIT-LEARN, NUMPY, SCIPY

Bioinformatics tools
SAMTOOLS, BCFTOOLS, HTSLIB, TABIX, PLINK

Teaching Experience

Teaching Assistant *Case Western Reserve University*
A DATA DRIVEN INTRODUCTION TO GENOMICS AND HUMAN HEALTH (PQHS 451) *AUGUST 2023 - DECEMBER 2023*
- Held weekly office hours, developed and graded assignments.

Leadership

Vice President *Cleveland, Ohio*
STEINER HOUSE INTERNATIONAL CO-OP *MAY 2022 - OCTOBER 2022*
- **Community Leadership**
    - Facilitated seamless integration of new residents, fostering community through monthly social events.
    - Collaborated with the executive board to ensure effective communication and daily management of the house.
    - Organized monthly events, contributing to a vibrant and inclusive living environment.

Admissions Manager *Cleveland, Ohio*
STEINER HOUSE INTERNATIONAL CO-OP *MAY 2022 - SEPTEMBER 2022*
- **Residential Admissions and Selections**
    - Screened and interviewed prospective residents, facilitating transparent communication of results within the house.
    - Successfully admitted diverse graduate students globally, including those from India, Bangladesh, China, Vietnam, Thailand, Nigeria, Ireland, and Iran.
    - Oversaw the onboarding process for 8 graduate students, ensuring a smooth transition to life in the United States at Steiner House.

Ordering Manager   *Cleveland, Ohio*
Steiner House International Co-op   *MAY 2022 - SEPTEMBER 2022*

- **Inventory Management**
    - Managed inventory for Steiner House, with a capacity of 21 students.
    - Responsible for ensuring a steady supply of consumables and non-consumables for the smooth functioning of the House.

Projects

Interactive Geospatial Dashboard Creation   *Case Western Reserve University*
Mapping the Environmental Impact of the Norfolk Southern Train Derailment in Ohio.   *2023*

- `https://naveen-kannan.shinyapps.io/final_east_palestine_dashboard_2`
    - Coded scripts to scrape publicly available air and water quality data from USGS(United States Geological Survey) and AirNow AQI for counties within a 30-mile radius of the Norfolk Southern train derailment in February 2023.
    - Implemented an interactive Shiny dashboard in R, offering users the ability to explore and analyze daily trends in Ohio, West Virginia, and Pennsylvania counties throughout January and February.
    - The dashboard includes dynamic time series plots for comprehensive visualization of air and water quality metrics during the specified timeframe.

Data Science Blog   *Cleveland, Ohio*
Tutorials on Data Engineering and Architecture   *2023*

- `https://naveenkannan.netlify.app/blog_main.html`
    - Authored tutorials on topics ranging from Docker/Singularity container creation to data engineering tutorials on setting up Spark and Hadoop.

Data Analytics and Visualization   *Cleveland, Ohio*
Analyzis and visualization of Public Health Datasets.   *2022*

- `https://naveenkannan.netlify.app/portfolio.html`
    - Performed statistical analysis of large public health datasets, including the Population Assessment of Tobacco and Health (PATH) Study, and the NHANES (National Health and Nutrition Examination Survey).
    - Visualized the distributions of variables and outcomes.
    - Interpreted and presented the findings in a scientific manner, alongside data visualization to further enhance the impact of the findings.

Publications

- Jayakrishnan Ajayakumar [a], Andrew J. Curtis [a], Felicien Maisha [b], Sandra Bempah [c], Afsar Ali [b,d], J. Glenn Morris Jr. [b,e], **Naveen Kannan** [a], and Grace Armstrong [a], [a]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA; [b]Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA; [c]Department of Geography, Kent State University, Kent, Ohio, USA; [d]Department of Environmental & Global Health, College of Public Health and Health Professions, University of Florida, Gainesville, Florida, USA; [e]College of Medicine, University of Florida, Gainesville, Florida, USA, Using Spatial Video and Deep Learning for Automated Mapping of Relief Camps. [Manuscript under review]

- **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.

- Suebsarn R., W. T., **Naveen K.**, Ellis W., Atsadaporn N., Chitchanok B., Chantira C., Christine S., Heba A. Using regression analysis and machine learning approach to understand the association between comorbidities, medication nonadherence, activity of daily living, and heart condition status among older adults in the United States. [Manuscript under review]

- 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.