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 Reserve University *MAY 2023 - Present*

- **Automated Large-Scale Cluster Deployment and Management**
  - Orchestrated the redeployment of a high-performance computing cluster with 18 data nodes, 2 master nodes, 1 service node, and 1 head node.
  - Configured a **PXE server** on the service node, enabling **fully automated, unattended OS provisioning** across the cluster.
  - Automated network configuration by matching NIC MAC addresses to IPs using a **DHCP server**, ensuring consistent IP address allocation across the cluster nodes.
  - Developed **Ansible scripts** to intergrate with **iDRAC**, facilitating remote, automated PXE-based cluster reprovisioning with a single command.

  PostgreSQL Database Administration
  - Configured, installed, and deployed a multi-core PostgreSQL server instance utilizing Ansible, while creating and managing **3 databases**, along with schema design and data modeling.
  - Utilized **PostgreSQL JDBC integration with Spark** to automate schema definition and loading of **terabytes of large genomic flatfiles** into databases.
  - Implemented **comprehensive security measures**, including user role management, privilege controls, firewall rules, and network security for secure database operations.

  Automated HDFS and Spark Deployment
  - Developed **Ansible scripts** to automate installation and configuration of **Apache Hadoop HDFS, YARN, MapReduce, and Spark** on a HPC cluster with 18 worker nodes and 1 master node.
  - Optimized YARN with 80 GB of RAM, 6 vCPUs per worker node, and 48 vCores per vCPU, facilitating **efficient resource management** and scheduling for Hadoop workloads.
  - Installed and configured Apache Spark, a powerful engine for large-scale data processing, and configured specific launch scripts for specific tasks.

  Docker/Singularity Container Deployment
  - Developed Mamba-based Docker/Singularity containers for reproducible scientific pipelines, reducing container build times by 300 percent, and deployed **RStudio and JupyterLab servers** through **Docker** containers.
  - Transferred Docker storage cache to a physical hard drive, preserving existing images and eliminating storage errors and root filesystem overhead.

  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.
  - Developed **GAN-based pipeline** to simulate realistic human genomic data by training on 1000 Genomes data, incorporating quality checks for LD, HWE, etc., as part of a pending journal publication.

Graduate Research Assistant *Cleveland, Ohio*

Population Health and Quantitative Health Sciences, 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*

- **Mapping human movement patterns after a natural disaster in the Democratic Republic of Congo.**
  - Developed Python scripts to automate the extraction of GPS coordinates from geospatial video feeds of refugee camps in DRC after Mt. Nyiragongo eruption, enabling automated map generation.
  - Implemented ML algorithms (YOLOv4) for tent detection by parsing video feeds into frames, and generated heat maps of tent distribution based on GPS coordinates.
  - 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, and was 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-based systems, Kubuntu, Debian, 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

Database Administration

HDFS(Hadoop Distributed File System), PostgreSQL server

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 the integration of new residents, fostered community through monthly social events, and collaborated with the executive board to ensure effective communication.

Admissions Manager *Cleveland, Ohio*

Steiner House International Co-op *MAY 2022 - SEPTEMBER 2022*

- **Residential Admissions and Selections**
  - Successfully admitted and onboarded a diverse graduate student cohort, including students from 8 different countries.

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.

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.

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 DevOps to data engineering tutorials.

Data Analytics and Visualization *Cleveland, Ohio*

Analysis 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).
  - Interpreted and presented the findings in a scientific manner, alongside data visualization to further enhance the impact of the findings.

Publications

- **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.

- 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.