

# Naveen Kannan

RESEARCH ASSOCIATE

Cleveland, Ohio | United States Permanent Resident

+1 973 519 5397 | [naveen.kannan12@gmail.com](mailto:naveen.kannan12@gmail.com) | [naveenk2022.github.io/quarto\\_blog](https://naveenk2022.github.io/quarto_blog) | [naveenk2022](https://github.com/naveenk2022) | [naveen-kannan-177120118](https://www.linkedin.com/in/naveen-kannan-177120118)

Experienced Bioinformatics Data Engineer with 3+ years of expertise in designing scalable ETL pipelines, integrating large-scale biological datasets, and building robust data infrastructure to support cutting-edge research. Proven ability to integrate complex omics data and establish best practices that drive technological innovation in therapeutic research.

## Work Experience

### Lead Data Engineer

Cleveland, Ohio

TRAILHEAD BIOSYSTEMS

JULY 2024 - Present

#### • Organizational Data Architecture Modernization

- Met with teams across the company to design an operational data model that integrates **6+ previously siloed data sources**, such as banked cell inventory, cell vial stocks, cell maintenance media, project management, cell source, and vial storage.
- Deployed a **PostgreSQL 17** database on an **Azure VM**, and configured **pgBackRest** for automated daily backups, supporting 99.9% system uptime.
- Developed the database schema utilizing **Alembic** for migration version control, and automated loading of data into the tables via **SQLAlchemy** after validating the data with **Pydantic**.
- Built and maintained a FastAPI server to expose CRUD APIs for scientists, improving system scalability.
- Transformed data architecture from legacy Excel-based tracking to a **comprehensive, centralized relational database system** from the ground up.

#### HD-DoE Database Design and Implementation

- Automated the extraction of structured data from **20GB of HTML reports** into a PostgreSQL database, reducing manual processing time by **200%**.
- Designed a database schema, enabling complex aggregate queries across **3 years' worth** of historical experimental data.
- Deployed Python scripts in reproducible Docker containers to **automate extracting data from reports** placed in a server by scientists.

### Research Associate

Cleveland, Ohio

PQHS, SCHOOL OF MEDICINE, CASE WESTERN RESERVE UNIVERSITY

MAY 2023 - JULY 2024

#### • Alzheimer's Disease Sequencing Project: Release 4 Annotation

- With the **Alzheimer's Disease Sequencing Project**, annotating and scoring **362 million structural genetic variants** associated with Alzheimer's disease, harmonizing datasets and disseminating data to collaborators for downstream analysis.

#### Data Architecture and Pipeline Development

- Designed and implemented fully automated pipelines for configuring and deploying an 18 node computing cluster, using Ansible and PXE servers for OS provisioning, stable IP management, and seamless remote cluster administration.

#### PostgreSQL Database Administration

- Deployed a multi-core PostgreSQL server instance and utilized **PostgreSQL JDBC integration with Spark** to automate schema definition and loading of **4 terabytes of large genomic flatfiles** into databases, reducing database setup and data processing time by 40%.

#### Cloud-Based Data Engineering and Informatics Pipeline

- Utilized a **Docker container** within an **AWS EC2 instance** to connect to an S3 bucket containing **several hundred gigabytes of genomic data**, and ran **bioinformatics pipelines** on the data for annotation and analysis via the Docker container.

#### Apache Hadoop Infrastructure Optimization for Biological Data

- Automated the deployment of **Apache Hadoop HDFS, YARN, MapReduce, and Spark** across an 18-node HPC cluster, achieving 95% resource utilization efficiency and reducing manual deployment time by 80%.

### 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 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	Data Engineering	Pipeline Tools	Cloud VM Tools	Version Control	Machine Learning	Bioinformatics
Linux (Debian, Fedora Red Hat, Raspbian), Windows 11	Python, R, Bash, SQL	Hadoop, Spark, Hive, PySpark, SQLAlchemy, Alembic, FastAPI, PostgreSQL, Pydantic	Docker, Singularity, Ansible, SLURM	AWS EC2, AWS S3, Azure VMs	Git, GitHub, gh client	TensorFlow, PyTorch, pandas, scikit-learn, numpy, scipy	Samtools, BCFtools, HTSlib, Tabix, Plink, Hail, Glow

## Education

### MSc in Biomedical and Health Informatics

CASE WESTERN RESERVE UNIVERSITY

- GPA-3.9

Cleveland, Ohio

JANUARY 2022 – AUGUST 2023

### Bachelor of Medicine and Bachelor of Surgery

MADRAS MEDICAL COLLEGE

- GPA-4.0

Chennai, India

SEPTEMBER 2016 - MARCH 2020

## 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**<sup>1</sup>, Nicholas Wheeler<sup>1</sup>, Genome Center for Alzheimer's Disease, Li-San Wang<sup>2</sup>, Yuk Yee Leung<sup>2</sup>, William S. Bush<sup>1</sup>, <sup>1</sup>Cleveland Institute for Computational Biology, Department for Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA; <sup>2</sup>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**<sup>1</sup>, N. Wheeler<sup>1</sup>, Genome Center for Alzheimer's Disease, L-S. Wang<sup>2</sup>, Y. Leung<sup>2</sup>, W. S. Bush<sup>1</sup>; <sup>1</sup>Cleveland Inst. for Computational Biology, Dept. for Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, <sup>2</sup>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.