

# Kasthuri Kannan, PhD

Associate Professor, UT MD Anderson Cancer Center

✉ [kskannan@mdanderson.org](mailto:kskannan@mdanderson.org) ☎ (551) 284 9106 📍 2130 W Holcombe Blvd. Houston. TX 77494

🔗 <https://kannan-kasthuri.github.io/> 🇺🇸 US Citizen 🌐 [linkedin.com/in/kasthuri-kannan-0a05b43](https://www.linkedin.com/in/kasthuri-kannan-0a05b43)

## PROFILE

As a seasoned data scientist with 15 years in life sciences, I excel in managing complex, large-scale data analysis and modeling projects. I am recognized for a strong leadership, guiding a multi-institutional team of 6 scholars and 7 faculty members. Actively seeking to leverage my expertise in an industrial setting, where I can drive innovation and deliver impactful results.

## EDUCATION

**Doctor of Philosophy (Computer Science)** 2002 – 2008  
Texas A&M University College Station, Texas

**Master of Science (Mathematics)** 2000 – 2002  
Texas A&M University College Station, Texas

**Master of Science (Mathematics)** 1998 – 2000  
Indian Institute of Technology, Madras Chennai, India

**Bachelor of Science (Mathematics)** 1995 – 1998  
University of Madras Chennai, India

## SKILLS

**Leadership** ● ● ● ● ●  
Managing data science and bioinformatics projects

**Software (Representative)** ● ● ● ● ●  
Python, R, Java, Cypher, SQL, Bioinformatics tools, Unix/HPC/Bash, Neo4j, MySQL, PostgreSQL, HTML, Javascript, Spatstat, Pytorch, Tensorflow

**Data Science/Machine Learning/AI** ● ● ● ● ●  
Math, CS, Statistics and Programming  
Ethical hacking, Domain knowledge, LLMs

**Bioinformatics & Image Processing** ● ● ● ● ●  
Pipeline development and data engineering  
Geospatial analysis

## PROFESSIONAL EXPERIENCE

**UT MD Anderson Cancer Center** 02/2020 – present  
Associate Professor Houston, TX

- Leading graph database development for multi-omics data integration
- Establishing spatial modeling of pathology images using R spatstat
- Supervising post-doctoral fellows and students

**New York University** 11/2013 – 08/2019  
Assistant Professor New York, NY

- Established bioinformatics pipelines and investigated cancer datasets
- Directed data science courses - see Data Science Education section

**Memorial Sloan-Kettering Cancer Center** 04/2011 – 10/2013  
Research Fellow/Associate New York, NY

- Established mutation pipeline for Brain/Head & Neck cancers
- Discovered ATRX mutations in lower grade gliomas
- Provided directed insights in various cancer studies through data analysis

## **Pennsylvania State University**

Research Associate

- Organized data analysis pipelines
- Offered bioinformatics consultation to researchers
- Managed sequencing tasks

10/2010 – 03/2011

State College, PA

## **Stowers Institute for Medical Research**

Research Specialist II

- Developed image processing methods for worm and fly tracking
- Implemented an automated data workflow to process yeast cell images
- Responsible for high quality image acquisition and microscope maintenance

01/2008 – 09/2010

Kansas City, MO

## **Knowledge Based Systems, Inc**

Internship

- Delivered data driven directed insights for aircraft movement operations
- Proposed cost and time saving measures for managing air force logistics

01/2007 – 12/2008

College Station, TX

## **ACTIVE DATA SCIENCE PROJECTS**

**Geospatial modeling of tissues** — Spatial point processes are powerful statistical frameworks for studying point patterns. By representing cells as points and annotating the measurements taken on those cells, such as gene expression at single cell level, it is appropriate to study their interactions using point processes. We extensively use spatstat, an R package, to model interactions and derive directed insights in cancers.

**Biomarker discovery using graph database** — Graph databases can help identify biomarkers by efficiently representing complex biological networks. By enabling powerful queries and community detection algorithms, graph databases make exploring the relationships between multiple genes easier, thus facilitating the discovery of potential biomarkers critical for diagnostics or therapeutic targets. We use Neo4j, a property graph database, and algorithms from the Graph Data Science Library (GDS) to derive insights and propose actionable biological targets.

**Biomarker validation using Graph Neural Networks (GNNs)** — Insilico validation of biomarkers is critical for providing actionable biological targets for experiments and prognosis. We apply GNNs (convolutional, attention etc.) to validate critical biomarkers discovered through the graph database. GNNs enhance biomarker discovery by leveraging the structure of biological networks, such as gene-gene interactions, and learning meaningful node/edge representations. By aggregating information from a node's neighbors, GNNs capture local and global patterns, allowing the model to predict relationships between genes and diseases more effectively.

## **DATA SCIENCE EDUCATION**

Developed and taught the following data science courses and lectures

- Programming for Data Analysis
- Machine Learning and AI
- Methods in Quantitative Biology

## **INVITED TALKS (REPRESENTATIVE)**

Mayo Clinic (2023), National University of Singapore (2022), Texas A&M University (2021), Courant Institute of Mathematical Sciences (2019), Institute of Mathematical Sciences, w/ honorarium (2018)

## **PUBLICATIONS**

Authored/co-authored 32 peer reviewed articles, that includes very high-impact journals.

Please refer [https://pubmed.ncbi.nlm.nih.gov/?term=kasthuri+kannan&sort=date&sort\\_order=desc](https://pubmed.ncbi.nlm.nih.gov/?term=kasthuri+kannan&sort=date&sort_order=desc) ☑