

Anurendra Kumar

217-819-8065 • Palo Alto, California, 94306 • anurendk@stanford.edu • anurendra.github.io

Research Interests: AI for Genomics. Causal and representation learning for perturbation and spatial omics; Robust inference under confounding and domain shift.

EDUCATION

- | | |
|--|-------------------|
| PhD, Computer Science (<i>Minor in Biology</i>)
<i>Georgia Institute of Technology (Transferred from UIUC)</i> | Aug '21 - Aug '25 |
| MS, Bioinformatics, Computer Science
<i>University of Illinois at Urbana-Champaign</i> | Aug '19 - May '21 |
| B.Tech-M.Tech (Dual Degree), Electrical Engineering (<i>Minor in Artificial Intelligence (AI)</i>)
<i>Indian Institute of Technology (IIT), Kanpur</i> | Jul '12 - May '17 |

CURRENT POSITION

- | | |
|--|-------------------|
| Postdoctoral Researcher, Stanford Cardiovascular Institute (Wu Lab) | Aug '25 - Present |
|--|-------------------|
- Developing causal and mechanism-aware models that leverage large-scale perturbation resources (LINCS/Tahoe/Parse) to infer transportable intervention effects in cardiomyopathies; grounding predictions using sparse iPSC-cardiomyocyte anchors and targeted experimental readouts.

PUBLICATIONS

- A. Kumar et al., *CellWHISPER: Inference of direct cell-cell communication from spatial transcriptomics*, **Under Review, Nature Computational Science, Second award at EWSC MIT symposium** [Paper] [Code] [Media]
 - Developed CellWHISPER, a scalable and confounder-aware toolkit to infer contact-mediated cell-cell communication, accounting for factors such as cell-type abundance, spatial proximity, and neighborhood composition.
 - Designed novel evaluation schemes showing CellWHISPER achieves <5% false positive rate (FPR) while existing methods have >90% FPR.
 - Enabled the first large-scale analysis of Gap Junction signaling, generating known and novel hypotheses on electrical synapses and glial communication in mouse brain ; also extended to ligand-receptor interaction analysis.
 - Introduced a novel latent variable model that better learns cell type–ligand/receptor affinities, improving interpretability and robustness of inferred communication patterns.
 - Performed differential signaling analysis across regions and conditions, identifying spatially constrained cell-cell programs relevant to Alzheimer’s disease.
- A. Kumar et al, *Intra-cellular Spatial Transcriptomics Toolkit (InSTanT)*, **Nature Communications 2024** [Paper] [Code] [Media]
 - Designed InSTanT, a statistically robust toolkit to analyze subcellular RNA co-localization from single-molecule ST data, addressing challenges like absence of a universal cell coordinate system and confounding effects.
 - Introduced the new statistical phenomenon of *d*-colocalization, developing robust and reproducible techniques to detect spatial proximities in transcriptomics data. InSTanT outperforms state-of-the-art by significant margin.
 - InSTanT’s differential co-localization module enables the detection of cell-type, phenotype, and region-specific co-localization patterns, while also identifying gene modules and spatially modulated gene pairs.
 - The statistical tests behind InSTanT pave the way for generating biologically insightful hypotheses, such as potential RNA interactions, ECM-cell interactions, and local translation.
- A. Kumar et al, *CoVA: Context-aware Visual Attention for Webpage Information Extraction*, **ACL Workshop’22** [Paper] [Code]
Proposed a Graph attention network that yielded interpretable results and achieved ~10% improvement over SOTA for Webpage Information Extraction. Created largest public labeled dataset of **7.7k** product webpage screenshots.
- A.Kumar et al, *DLVM: A Dynamic Model Based on Dirichlet Prior for Audio Processing*, **IEEE Transactions SLP (2019)** [Paper]
Extended DLVM (below model) to incorporate multi-order bidirectional dependency and to newer applications.
- A. Kumar et al, *A Dynamic Latent Variable Model for Source Separation*, **ICASSP’18** [Paper]
Extended the existing NMF and Probabilistic Latent Component Analysis (PLCA) used in acoustic modeling to exploit temporal dependence. Developed a dynamic **latent variable model** for time-varying non-negative data. Proposed a dynamic Dirichlet distribution as a prior distribution. Proved the block-wise concavity of expected log-likelihood.

- A.Kumar et al, *Monoaural audio source separation using VAE*, Interspeech '18 [Paper]
Proposed a principled generative approach using VAE for audio source separation. VAE computes efficient Bayesian inference which leads to a continuous latent representation characterizing each source. Our method performed better than best of the relevant methods with 2 dB improvement in the source to distortion ratio.

INDUSTRIAL EXPERIENCE

Intern, BRAID, Genentech, San Francisco May '24 - Aug '24

- Designed six classes of unsupervised metrics to evaluate single-cell embeddings from various **foundational models**. Demonstrated the trade-off between information retention and clusterability, and showed that foundational models outperform baseline models on robustness.

Intern, Data & Methods, Vertex Pharmaceuticals, Boston May '23 - Aug '23

- Developed a framework to integrate variant aware off-target detection in Vertex's computational pipeline for gene editing.

Research Intern, IBM Research, Yorktown Heights [Poster] May '22 - Aug '22

- Proposed and validated a novel transformer based architecture which can capture long-term memory. Our model extends on vanilla transformer and adds cross attention with memory defined as set of key and value tensors. This allows us to train same transformer on different datasets but with different memory. Developed a synthetic benchmark to validate our model.

Research Intern, IBM Research, Yorktown Heights [Poster] May '21 - Aug '21

- Adapted a GAN framework to sample parameters from the posterior distribution of mechanistic models represented by systems of ODEs. Developed a custom library in TensorFlow using custom gradients to enable batch computation of numerical methods. Validated the approach using monkey HIV viral data with known ODE dynamics and unknown latent cell populations. Implemented unit tests to ensure model accuracy and robustness.

Machine Learning Engineer Intern, Service Now May '20 - Aug '20

- Developed a DNN using BERT for content extraction from documents that yielded ~85% accuracy on SROIE dataset

Lead, Data & Artificial Intelligence, Startup Project Jan '18 - Mar '19

- Developed algorithms and benchmark metrics for various finance services e.g. credit scoring & risk estimation.

Research Intern, IBM Research, Bangalore [Presentation] Jun '17 - Aug '17

- Proposed a tree structured prior for representation learning of KB. Used proximal gradient to deal with non-smoothness.

Extreme Blue Intern, IBM, Bangalore [Video] May '16 - Jul '16

- Crop health monitoring from NIR & RGB images and various sensors. Awarded **AEGIS GRAHAM BELL AWARD**.

Research Intern, Data Science, Rivigo May '19 - Aug '19

Research Intern, Samsung May '15 - Jul '15