

# Anurendra Kumar

217-819-8065 • Champaign, Illinois 61820 • ak32@illinois.edu • anurendra.github.io

## RESEARCH INTERESTS: Computational Biology, Machine Learning

## EDUCATION

<b>University of Illinois at Urbana-Champaign</b> <i>M.S., Ph.D. in Computer Science</i> (Concentration in Bioinformatics)	Aug '19 - May '21 GPA: 4/4
<b>Indian Institute of Technology (IIT), Kanpur</b> <i>M. Tech. in Electrical Engineering</i> <i>B. Tech. in Electrical Engineering</i> (Minors in Artificial Intelligence, Linguistics)	Jul '12 - May '17 GPA: 9.67/10 GPA: 8.0/10

## EXPERIENCE

<b>Graduate RA, Prof. Saurabh Sinha, UIUC</b> <i>Single cell genomics</i> <ul style="list-style-type: none"><li>• <i>Spatio-temporal Transcritomics</i>: Proposed statistical approaches to understand the spatio-temporal regulation of genes</li><li>• <i>Neurogenomics</i>: Proposed an attention-based deep neural network (DNN) to decode combinatorial logic of enhancers</li><li>• <i>Gene Regulatory Network (GRN)</i>: Developed a parameter estimation for single cell simulator with an underlying GRN</li></ul>	Dec '19 - Present
<b>Machine Learning Engineer Intern, Service Now</b> <ul style="list-style-type: none"><li>• Developed a DNN using BERT for content extraction from document images which yielded ~85% accuracy on SROIE dataset</li></ul>	May '20 - Aug '20
<b>Graduate RA, Prof. Kevin Chang, UIUC</b> <i>Context-aware Webpage Object Detection [Submitted to CVPR '21]</i> <ul style="list-style-type: none"><li>• Proposed an attention based DNN which yielded interpretable results and achieved ~10% improvement over SOTA</li><li>• Created largest public labeled dataset of 7.7k product webpage screenshots</li></ul>	Aug '19 - May '20 [Code]
<b>Lead, Data &amp; Artificial Intelligence, Startup Project</b> <ul style="list-style-type: none"><li>• Developed algorithms and benchmark metrics for various finance services e.g. credit scoring &amp; risk estimation</li></ul>	Jan '18 - Mar '19
<b>Research Associate, IIT Kanpur-ISRO Collaboration</b>	Aug '17 - Dec '17
<b>Research Intern, IBM Research Lab</b> <i>Hierarchical sparse representation of Knowledge base (KB)</i> <ul style="list-style-type: none"><li>• Proposed a tree structured prior for representation learning of KB. Used proximal gradient to deal with non-smoothness</li></ul>	Jun '17 - Aug '17 [Presentation]
<b>Extreme Blue Intern, IBM</b> <i>Internet of Things (IOT) in Agriculture</i> <ul style="list-style-type: none"><li>• Crop health monitoring from NIR &amp; RGB images and various sensors. Awarded AEGIS GRAHAM BELL AWARD</li></ul>	May '16 - Jul '16 [Video]
<b>Research Intern, Samsung</b>	May '15 - Jul '15

## PUBLICATIONS

- A.Kumar, T.Guha, P.Ghosh, *Dirichlet Latent Variable Model: A Dynamic Model Based on Dirichlet Prior for Audio Processing*, IEEE Transaction on speech and language processing (2019) [Paper]  
Extended below model to incorporate multi-order bidirectional dependency and to newer applications
- A. Kumar, T. Guha, P. Ghosh, *A Dynamic Latent Variable Model for Source Separation*, Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'18) [Paper]  
Developed a dynamic latent variable model for time-varying non-negative data. Proposed a novel prior distribution which is particularly suitable for dynamic non-negative data and yields elegant update equations. It also lead to connecting our model to the two popular latent basis learning methods - PLCA and NMF
- A.Kumar<sup>+</sup>, L.Pandey<sup>+</sup>, V.Namboodiri, *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
- A. Kumar<sup>+</sup>, K. Morabia<sup>+</sup>, W. Wang, K. Chang, *VAMWOD: Visual Attention-based Model for Webpage Object Detection*, (Submitted to CVPR '21)

## SKILLS

---

- **Programming:** Python, Matlab, C, C++, LaTeX, R, JavaScript
- **Other Technologies:** PyTorch, TensorFlow, Caffe, Scikit-learn, AWS, GIT

## RELEVANT COURSEWORK

---

Machine Learning for Computational Biology • Advanced Bioinformatics • Advanced Biochemistry • Computational Bio-engineering • Machine Learning • Computer Vision • Natural Language Processing • Convex Optimization • Algorithms

## PROJECT HIGHLIGHTS

---

### **Latent representation based gene regulatory network inference from multi-omic data**

Aug' 19 - Present

*Prof. Jian Peng, UIUC*

[Report]

- Developed a DNN architecture to characterize a gene with complex 3D structure and epigenomic features.
- Implemented a gene interaction ranking framework from ENCODE epigenomic features and L1000 gene expression

### **Visual grain quality estimation**

Aug'16 - Nov'16

*Prof. T. Guha, IITK, (Best project in Image Processing course)*

[Report]

- Developed a proof-of-concept for quality estimation of grain from image of a grain sample
- Created a labeled dataset with help of traders and farmers
- Proposed a two-level segmentation method to segment overlapped grains which were further classified as grains / impurities demonstrating performance of proposed technique

### **Multiple Word Vector Embedding for polysemous words**

Sep' 15 - Nov'15

*Prof. A. Mukherjee, IITK*

[Project Material]

- Extended word vector model to have multiple representations for polysemous words
- Proposed two parametric and two non-parametric solutions that beat state-of-the-art in several specific cases
- Our primary contribution was in developing a non-parametric approach to clustering for optimal number of senses

### **Visual Odometry in Self-Driving Car**

Sep' 15 - Nov'15

*Prof. G. Pandey, IITK*

[Project Material]

- Developed and implemented visual odometry in self-driving car using opencv library in C++
- SIFT and FAST features were extracted in each frames
- KLT tracker with RANSAC for outlier rejection was employed for tracking features

### **Multiple Kernel Learning (MKL)**

Sep' 15 - Nov'15

*Prof. H. Karnick, IITK*

[Project Material]

- Implemented and performed experiments to show that MKL automatically learns efficient weighted distribution of multiple kernels and have the potential to handle data coming from heterogeneous sources