

# Surag Nair

2nd Year PhD Student | Stanford University  
www.stanford.edu/~surag | surag@stanford.edu

## Research Interests

Deep Learning for Single-cell Epigenomics, Computational Biology, Natural Language Processing

## Education

<b>Stanford University</b> <i>PhD in Computer Science</i> Cumulative Grade Point Average: 4.14	SEP 2019 - PRESENT
<b>Stanford University</b> <i>MS in Computer Science</i>	SEP 2017 - JUN 2019
<b>Indian Institute of Technology, Delhi</b> <i>B.Tech in Electrical Engineering</i> Cumulative Grade Point Average: 9.35/10	JUL 2013 - JUN 2017

## Relevant Coursework

<i>Computer Science</i>	:	Artificial Intelligence, Machine Learning, Probabilistic Graphical Models, Natural Language Processing, Computer Systems and Organization, Databases, Analysis of Networks, Cryptography, Operating Systems
<i>Biology/Bio computation</i>	:	Deep Learning in Genomics and Biomedicine, Structure of Biomolecules, Chromatin Regulation of the Genome

## Research and Technical Projects

**Deep Learning Methods for Genomics and Single-cell Epigenomics** APR 2018 - PRESENT  
*Supervisor: Dr. Anshul Kundaje, Stanford University*

### Epigenetic Landscape of Somatic Cell Reprogramming

- Exploring the chromatin accessibility dynamics of somatic cell reprogramming at single-cell resolution.
- Performing integrative analyses with histone and TF ChIP-seq data to highlight salient epigenetic features, and comparative analyses with Heterokaryon system to reveal optimal reprogramming trajectories.
- Developing interpretability methods for base-pair resolution neural networks to identify DNA sequence motifs.

### Speeding Up In-silico Saturation Mutagenesis (ISM) for Convolutional Sequence Models

- ISM is an interpretability method for deep learning sequence models in which each position in input sequence is perturbed and propagated through a trained model to measure the effect of the perturbation on the output.
- Developed fastISM, an algorithm that speeds up ISM by over 10x for convolutional neural networks.

### Cis-trans Deep Learning Models for Chromatin Accessibility

- Developed methods for incorporating genetic sequence with RNA expression data to predict chromatin landscape of unique cell types in order to impute existing DNase-seq databases and make predictions on unseen cell types.
- Improved state-of-the-art by introducing new features, a ResNet model architecture, and training procedures.

**Timely Detection of Extreme Failure Cases for Siri** JUN 2018 - AUG 2018  
*Summer Internship, Apple Inc. (Siri International team), Cupertino*

- Studied causes for failure in the complex Machine Learning based pipeline for Siri, Apple's voice assistant.
- Devised and implemented a NLP pipeline for real-time detection of failure cases based on usage logs data.
- Workflow consisted of periodic PySpark and Python scripts running on terabytes of real-time data.

**Inferring Temporal Knowledge for Near-Periodic Recurrent Events** JAN 2016 - JUL 2017  
*Bachelor's Thesis, Supervisor: Dr. Mausam, CSE, IIT Delhi*

- Designed and implemented a Probabilistic Graphical Model based information extraction system for predicting temporal schedules of periodic events from unstructured text sources such as webpages, blogs and news articles.
- System outperformed humans on experiments on large scale corpora including Clueweb and NYT Corpus.

## Open Source Contributions

- Multi-framework Alpha Zero** [2000+ stars] WINTER 2017-18
  - Developed a package for self-play based learning following the Alpha Zero paper by DeepMind. Allows easy addition of new games and works with all major deep learning frameworks (PyTorch, TensorFlow, Keras).
- PyTorch Implementation of seqGAN Algorithm** [450+ stars] AUTUMN 2017-18
  - Implemented a LSTM based deep learning language model and trained it using a Generative Adversarial Network framework using Policy Gradients, in PyTorch, based on the paper by Lantao Yu et al. 2016.
- Stanford CS230 Deep Learning Starter Code** [1500+ stars] WINTER 2017-18
  - Developed starter code for deep learning projects in PyTorch with accompanying tutorials.

## Teaching Experience

- CS230: Deep Learning** WINTER 2017-18
  - Course Instructors: Dr. Andrew Ng & Kian Katanforoosh, Stanford University
  - Project mentor for 14 teams who applied deep learning to domains including space imagery, translation, genomics, and photography. Developed questions for midterm examination. Graded exams and projects for 400+ students.
- COL333: Artificial Intelligence** SEMESTER I 2016-17
  - Course Instructor: Dr. Mausam, CSE, IIT Delhi
  - Developed, implemented and graded a course assignment on Markov Decision Process for 125+ students.

## Publications and Preprints

- fastISM: Performant in-silico saturation mutagenesis for convolutional neural networks:** Surag Nair, Avanti Shrikumar, Anshul Kundaje. BioRxiv 2020 (preprint).
- Integrating regulatory DNA sequence and gene expression to predict genome-wide chromatin accessibility across cellular contexts:** Surag Nair\*, Daniel S. Kim\*, Jacob Perricone, Anshul Kundaje. Bioinformatics 2019.
- Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays:** Rajiv Movva, Peyton Greenside, Georgi K. Marinov, Surag Nair, Avanti Shrikumar, Anshul Kundaje. PLOS ONE 2019.
- Inferring Temporal Knowledge for Near-Periodic Recurrent Events:** Dinesh Raghu\*, Surag Nair\*, Mausam. International Joint Conference on Artificial Intelligence (IJCAI) 2018.
- Technical note on transcription factor motif discovery from importance scores TF-MoDISco v0.4.4.2-alpha:** Avanti Shrikumar, Katherine Tian, Anna Shcherbina, Žiga Avsec, Abhimanyu Banerjee, Mahfuza Sharmin, Surag Nair, Anshul Kundaje. arXiv 2018 (preprint)

\*equal contribution

## Technical Skills

Languages : Python, R, C++, SQL  
Softwares/Tools : PyTorch, TensorFlow, NumPy, PySpark

## Scholastic Achievements

2016 Department Rank 5 (top 6%) among students of Electrical Engineering Department, Batch of 2017  
2014 Certificate of Merit for being in the top 7% students of IIT Delhi in the first two semesters  
2013 All India Rank 280 in IIT Joint Entrance Exam-Advanced out of over 1 million students  
2013 Qualified for KVPY Fellowship

## Extra Curricular Activities

- Social Service** DEC 2015 - JAN 2016
  - Samarthanam Trust for the Disabled, Bangalore
  - Taught English and Maths to students of grades I-V and basics of computers to adults of ages 18-25.
- Student Mentor** JUN 2015- JUN 2016
  - IIT Delhi
  - Student Mentor to 5 freshmen of the Engineering Physics Department to ensure smooth transition into IIT Delhi.