Aditya Sarkar

+91-7042728328 asnov2k@gmail.com aditya-sarkar441.github.io aditya-sarkar441

Interests

Optimal Control, Non-convex Optimization, Speech Processing, Generative models

Education

Indian Institute of Technology (IIT) Mandi July 2019 – May 2023
Bachelor of Technology (Honors) in Electrical Engineering GPA : 9.1/10

- Institute Rank: 2, Department Rank: 1
- Advisor : Dr Dileep A.D.
- Thesis : Distributional robustness for reliable Language Identification

Publications and Preprints

3. Sarkar A., Giros A., Mockly L., Moore J., Moore A., Nagareddy A., Chhugani K., Sarwal V., Darci Maher N., Chang Y., Garmire L., Bao R., Chikhi R.*, Mangul S.* **Unlocking the microblogging potential for science and medicine**, Poster presentation at RECOMB '22, Under review at Nature Human Behavior, doi: https://doi.org/10.1101/2022.04.22.488804
2. Sarwal V, Niehus S, Ayyala R, Kim M*, Sarkar A*, Chang S, Lu A, Rajkumar N, Darfci-Maher N, Littman R, Chhugani K, Soylev A, Comarova Z, Wesel E, Castellanos J, Chikka R, Distler MG, Eskin E, Flint J, Mangul S. **A comprehensive benchmarking of WGS-based deletion structural variant callers.** Briefings in Bioinformatics. 2022 Jul 18;23(4):bbac221. doi: 10.1093/bib/bbac221. PMID: 35753701; PMCID: PMC9294411.
1. Sarkar A., Bhavsar A., **Virtual Screening of Pharmaceutical Compounds with hERG inhibitory activity**, Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 2: BIOIMAGING, pages 152-159, doi: 10.5220/0010267701520159

* means joint authorship.

Manuscripts in Preparation

3. Sarkar A., Mangul S., **Gene expression deconvolution using Expectation Maximization**, (Manuscript in preparation), Manuscript **[Here]**
2. Sarkar A., Pazokitroudi A., Sankararaman S., **Contrastive Canonical Correlations Analysis**, (Manuscript in preparation), Manuscript **[Here]**
1. Sarkar A., Bronson Jeong M., Sankararaman S., **Dimensionality Reduction in Single Cell data**, (Manuscript in preparation), Manuscript **[Here]**

Research Projects

Estimation of mortality of Sepsis patients from gene expression Oct '22 - Present
Guide: Sriram Sankararaman, Serghei Mangul University of California, Los Angeles

- Explored existing models on predicting mortality of sepsis patients over various gene expression datasets belonging to different cohorts. Used SHAP and Random Forests to identify the genes and cell types responsible for the condition of sepsis mortality.
- Employed a Siamese model for predicting the mortality using gene expression data with a Triplet-loss based cost function. **[Slides]**

Deep cross modality hashing using cycle consistent transformer Sept '22 - Present
Guide: Aditya Nigam IIT Mandi

- Developed a cyclic end-to-end transformers that can form text representations from image, and image representations from text so as to reduce the modality gap. Loss function used was L2-Norm.
- Used ArcFace loss function to perform unimodal hashing over image representations of both the images and text. Currently exploring the developed model over various datasets such as NUS-Wide, Mir-Flickr25K and MS-COCO, trying to get better results over a variety of metrics. **[Slides]**

Distribution shifts in spoken language identification Sep '22 - Nov '22
Guide: Dileep A.D. IIT Mandi

- Proposed and developed an adversarial network with generator being an Attentive "squeeze and excitation" block, and the discriminator being a BiLSTM model, that can minimize the effectof channel variations of the speech signals.
- Surveyed and implemented various existing algorithms such as X-vectors, U-vectors, XLSR-53 and WSSL, over a various datasets such as Prasar Bharati and YouTube datasets **[Manuscript]**[**Slides**]

Cell type decomposition using Expectation Maximization Jun '22 - Oct '22
Guide: Serghei Mangul University of Southern California

- Surveyed and implemented existing models on deconvoluting cell types proportions from gene expression data. Used GEDIT's feature selection process based on information content to select the most probable set.
- Developed an Expectation Maximization based algorithm for estimating the cell type proportion from bulk gene expression data and reference data. **[Manuscript]**

Key Technical Projects

Implementation of Reinforcement learning algorithms Aug '21 - Sept '21
Guide: Sreelakshmi PM (Control theory EE301) IIT Mandi

- Defined appropriate states, actions, transition probabilities, rewards for a 2D maze as an Markov Decision Process.
- Implemented various methods such as Howard's policy iteration, value iteration, and linear programming to find the optimum policy that minimises the number of steps between two specified places while adhering to the constraints.

Implementation of Deep learning algorithms Aug '22 - Oct '22
Guide: Dileep AD (Deep Learning CS671) IIT Mandi

- Benchmarked various optimization algorithms such as Adaptive Moments, Gradient Descent, Nesterov GD, SGD on estimating the cell type proportions from gene expression.
- Implemented various algorithms such as LSTMs, CNN, RNN, Neural Networks and Perceptron on a variety of datasets (images, texts, speech recordings). **[Repository]**

Digital Image Processing Aug '21 - Dec '21
Guide: Renu Rameshan (Image Processing EE608) IIT Mandi

- Performed Otsu Binarization and Connected Components analysis to form boundaries around the letters of a document image.
- Performed DBScan over the letters identified to identify paragraphs. Then Tesseract was used to predict the letters and form the HTML document. **[Repository]**

Mind driven wheelchair Feb '20 - Jun '20
Guide: Shubhajeet Roy Chawdhary (Design Practicum) IIT Mandi

- Designed and developed the pipeline of the signal processing in the Mind driven wheelchair.
- Implemented deep learning architecture based on 1D convolutional neural network (CNN) for classifying EEG signals.**[Report]**

Talks and Posters

2. *Systematic analysis of the microblogging potential for science and medicine* (Prof. Serghei Mangul) **Poster** at La Jolla, California, RECOMB Seq 2022.
1. *Virtual Screening of Pharmaceutical Compounds with hERG inhibitory* (Prof. Arnav Bhavsar) **Talk and Poster** at Vienna, Austria, BIOSTEC 2021.

Technical Skills

Languages: C++, Python, R, VHDL, MATLAB, Bash, HTML/CSS, Javascript
Deep Learning: Caffe, TensorFlow, Numpy, Pandas, OpenCV, Matplotlib, Keras, Pytorch
Software: Keil μ Vision, LaTeX, Git, Scilab, SolidWorks, Arduino, Raspberry Pi, Modelsim, UCLA Hoffman2 Cluster

Awards and Achievements

- Ranked first in the department (out of 60 students) with a CPI of 9.1/10.0 after 6th semester.
- Received IIT academic award for being department topper in Fall 2019 and Spring 2020.
- Awarded department change due to exceptional academic performance in the first year (9.8 CPI).
- Selected for attending the Winter School for Cognitive Modelling 2020 on Cognitive Modelling, organised by University of Waterloo, Canada and University of Groningen, Netherlands.
- Ranked in top 1% (amongst 2.2 Million students) in the entrance exam of Indian Institute of Technology (IIT-JEE) 2019 (99.3 percentile)
- Awarded National Talent Search Scholarship in 2017 by NCERT.

Relevant Coursework

Electrical Engineering: Signals and Systems, Control theory, Signal Processing*, Optimization theory*, Matrix Theory, Probability and Random processes*, Digital Image Processing, Communication theory.
Computer Science: Statistical Learning, Deep Learning, Data Structures and Algorithms, Data Visualization, Data Science
Mathematics: Probability and Statistics, Linear Algebra, Calculus, Ordinary Differential equations
***indicates courses to be completed by June '23.

Teaching Assistant and Extra-curricular Activities

6. **Teaching Assistant - CS671 :** (Prof. Dileep A.D.) Teaching assistant for the course Deep Learning in next semester at IIT Mandi. Preparing tutorials on Keras-Tensorflow, Pytorch and Python. Will be conducting coding assignment evaluations.

5. **Teaching Assistant - CXPT599 :** (Prof. Serghei Mangul) Teaching assistant for the course Biomedical Data Science at USC. Prepared the lecture slides on introductory probability and statistics.

4. **Teaching Assistant - IC272 :** (Prof. Aditya Nigam) Teaching assistant for the course Data Science 3 at IIT Mandi. Mentored 300+ students.

3. **Teaching Assistant - IC252 :** (Prof. Satyajit Thakur) Teaching assistant for the course Data Science 2 (Probability and Statistics) at IIT Mandi. Mentored 300+ students.

2. **Web Developer of Mangul Lab Website :** (Prof. Serghei Mangul) Developed the official website for the Mangul lab at USC. Currently it can manage new blog posts, news, publications, team members etc. Coded in Javascript, CSS and HTML.

1. **Teaching Assistant for Introduction to Bioinformatics :** (Prof. Serghei Mangul) Conducted Python tutorials for students of Moldova State University of Medicine and Pharmacy, Chisinau, Moldova on Spring 2019.