Aditya Sarkar

003, Hostel B-19, Indian Institute of Technology Mandi, Kamand, Mandi - 175005
+91-7042728328 asnov2k@gmail.com aditya-sarkar441

Research Interests

Deep generative models, Bayesian Non-parametric models, Natural Language Processing, AI-assisted Healthcare

Education

Technical University of Darmstadt — Frankfurt, Hesse, Germany
Exchange student in Electrical Engineering — *April 2022 – August 2022*

Indian Institute of Technology Mandi — Kamand, HP, India
Bachelor of Technology (Honors) in Electrical Engineering — *July 2019 – May 2023*

- Major GPA : 9.01/10.0
- Institute Rank : 2, Department Rank : 1

Publications, Preprints, and Working Papers

7. **Application of Artifical Intelligence in sepsis prediction**
   Submitted to **Genome Biology** [pdf]
   Aditya Sarkar*, A. Moore, R. Alomair, K. Chhugani, R. Ayyala, Anni Wong-Beringer, Eleazar Eskin, Sriram Sankararaman, Serghei Mangul
6. **Rigorous benchmarking of DNA methylation-based deconvolution tools for analysis of cell type composition**
   Submitted to **Genome Biology** [pdf]
   Zeyuan (Johnson) Chen*, **Aditya Sarkar***, A. Moore, R. Alomair, K. Chhugani, R. Ayyala, Eran Halperin, Serghei Mangul
5. **Unlocking the microblogging potential for science and medicine**
   Submitted to **Genome Biology** [pdf]
   Aditya Sarkar, A. Giros, L. Mockly, J. Moore, A. Moore, A. Nagareddy, K. Chhugani, V. Sarwal, N. Darci Maher, Y. Chang, X. Lin, Lana X Garmire, Sarah Mojarad, Riyue Bao, Rayan Chikhi*, Serghei Mangul*
4. **Low availability of code and high availability of raw omics data accompanying biomedical studies**
   Accepted as Talk in **ISCB/ECCB 2021** [pdf] [video]
   Dhrithi Deshpande, Ruiwei Guo, **Aditya Sarkar**, Serghei Mangul
3. **A comprehensive benchmarking of WGS-based structural variant callers**
   To be Submitted in **Briefings in Bioinformatics** [pdf]
   Varuni Sarwal, Sebastian Niehus, R. Ayyala, M. Kim*, **Aditya Sarkar***, S. Chang, A. Lu, N. Rajkumar, N. Darci-Maher, R. Littman, K. Chhugani, A. Soylev, Z. Comarova, E. Wesel, J. Castellanos, R. Chikka, Margaret G. Distler, Eleazar Eskin, Jonathan Flint, Serghei Mangul
2. **The systematic assessment of completeness of public metadata accompanying omics studies**
   To be Submitted in **Genome Biology** [pdf]
   Yu Ning, A. Rajesh, R. Guo, E. Ling, I. Nakashidze, M. Yee Wong, R. Ayyala, J. Hu, **Aditya Sarkar**, A. Nosov, Y. Chang, Annie Wong-Beringer, M. Love, M. Abedalthagafi, Serghei Mangul
1. **Virtual Screening of Pharmaceutical Compounds with hERG inhibitory activity (cardiotoxicity) using Ensemble Learning**
   Biostec Bioimaging International Conference 2021 Also presented at conference [pdf] [video]
   Aditya Sarkar, Arnav Bhavsar (2021)

* means joint authorship

Research Experience

Statistical Analysis of DNA Methylation-based Deconvolution methods — *April '20 - Ongoing*
Advised by Eran Halperin and Serghei Mangul — *University of California, Los Angeles*

- Read various papers on existing deconvolution tools and installed them on cluster.
- Working on writing scripts for evaluating tools using Pearson correlation and absolute error.

Statistical Analysis of WGS-based SV callers *October '20 - April '21*

Advised by Jonathan Flint, Eleazar Eskin, Serghei Mangul and Varuni Sarwal *University of California, Los Angeles*

- Compared the performance of Structural Variant detection tools on human data against a PCR verified gold standard to determine tools with the best match of precision and sensitivity.
- Studied the effect of the coverage of data and deletion length on the accuracy of each tool under various thresholds.

Analysis of Tweets using Web Text Pattern Matching and NLP *October '20 - Ongoing*

Advised by Rayan Chikhi and Serghei Mangul *University of Southern California, Los Angeles*

- Scraped the web using Text Pattern Matching method with Beautiful Soup and Requests API
- Working on developing a Transformer model (NLP) for analysis of tweets.

Image Classification with Deep Metric Learning using Siamese Network *April '20 - December '20*

Advised by Arnav Bhavsar *Indian Institute of Technology Mandi*

- Classified X-ray images using Transfer Learning with base models such as VGG16, ResNet50 and DenseNet121
- Worked on developing a pipeline which uses metrics learning for classification of images. Acheived an accuracy of 95%.

Key Technical Projects

Image Segmentation and Classification | *CS608: Digital Image Processing* [Code] [Report] **Fall '21**

- Used binary thresholding and circular hough transform to segment pupil and iris from a dataset of eyes.
- Implemented PCA for dimensionality reduction followed by classification models like decision trees, VGG16 and ResNet50 achieving maximum accuracy of 88%.

Style Transfer in Audio and Images | *DS403/CS669: Statistical Learning* [Code] [Report] **Fall '21**

- Implemented the StyleGan architecture for style transfer in audio signals
- Worked on developing a deep generative model based on a recent CVPR paper that performs style transfer on Images.

Technical Skills

Languages: C++, Python, R, VHDL, MATLAB, Bash, HTML/CSS, Javascript
Machine Learning: Caffe, TensorFlow, Numpy, Pandas, OpenCV, Matplotlib, Keras
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 9.01/10.0 after 4th semester.
- Selected for foreign exchange to TU Darmstadt, Germany for spring semester out of 260 students.
- Received IIT academic award for being department topper in Fall 2019 and Spring 2020.
- Awarded an option of department change due to exceptional academic performance in the first year.
- 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

Computer Science: Reinforcement Learning,Statistical Learning, Deep Learning, Data Structures and Algorithms, Digital & Advanced Image Processing, Data Visualization, Data Science
Mathematics: Probability and Statistics, Data mining, Linear Algebra, Calculus, Ordinary Differential equations, Optimization for Machine Learning
Electrical Engineering: Signals and Systems, Control Theory, Analog and Digital Circuits, Communication Systems, Digital Signal Processing, Network Theory, Power Systems, Electronic Devices

Teaching Assistant and Extra-curricular Activities

Undergraduate Teaching Assistant | *HS342: German1* **Fall '21 and Spring '20**

- Used binary thresholding and circular hough transform to segment pupil and iris from a dataset of eyes.
- Implemented PCA for dimensionality reduction followed by classification models like decision trees, VGG16 and ResNet50 achieving maximum accuracy of 88%

Core Team member | *Programming, Astronomy and Research Clubs* **Fall '19 and Spring '20**

- Implemented the StyleGan architecture for style transfer in audio signals
- Worked on developing a deep generative model based on a recent CVPR paper that performs style transfer on Images.

Research Group Member | *Serghei Mangul's Lab at USC Los Angeles* **Spring '20, Fall '21, Spring '21**

- Helped in writing his grant paper and preparing lecture slides for his class on Biomedical Data Science
- Presented various SOTA machine learning papers from Nature, ICLR, CVPR, Genome Biology etc. in lab meetings.