

(434) 284-3794
Charlottesville, VA
saurav.sen@virginia.edu

Saurav Sengupta

PhD Candidate

Portfolio: ssen7.github.io
github.com/ssen7
linkedin.com/in/saurav-sengupta

I am a PhD Candidate at the School of Data Science at the University of Virginia at Charlottesville. I focus on applications of machine learning for clinical data like Electronic Health Records, radiology images and text. My work has led to understanding of pre-existing risk factors associated Long COVID in my work with the NIH, and combining vision and language for analyzing high resolution radiology images and to improve model interpretability which is vital for building trustworthy models in healthcare.

EDUCATION

PhD in Data Science, *School of Data Science, University of Virginia* EXPECTED MAY 2024
Thesis: Combining and evaluating interpretable computer vision and natural language processing techniques for use in clinical contexts

Master of Science in Data Science, *School of Data Science, University of Virginia* JUL 2018 — MAY 2019

B.E. (Hons.) in Electrical and Electronics, *Birla Institute of Technology and Science, Pilani, India* JUL 2012 — MAY 2016

EXPERIENCE

Graduate Research Assistant, School of Data Science AUG 2020 — Present
University of Virginia
Charlottesville, VA

- Developed interpretable deep learning based method using LSTMs for EHR for discovering potential risk factors for Long COVID.
- Analyzed Electronic Health Record data using PySpark for the NIH [National COVID Cohort Collaborative \(N3C\)](#) to detect Long COVID sub-phenotypes and generate geographic hotspots for Long COVID.
- Created novel deep learning method to generate automatic text descriptions for very high resolution histology images.

Data Scientist, Gastro Data Science Lab JUN 2019 — JUL 2020
University of Virginia
Charlottesville, VA

- Created Convolutional Neural Network (CNN) based models to detect gastro-intestinal diseases (Crohn's, Celiac) on high resolution biopsy images.
- [Gastroenterology Data Science Lab](#) received \$100k Litwin IBD Pioneers Program, Crohn's & Colitis Foundation grant based on work.

Associate Consultant, Development Bank of Singapore Project JUL 2016 — OCT 2017
Capgemini
Hyderabad, India

- Developed Java-based REST APIs for the Development Bank of Singapore (DBS) using Spring Boot.

PROJECTS

Automated report generation using Vision Transformers and BERT (code)
Created novel method for captioning extremely high resolution tissue images by fine-tuning pre-trained Vision Transformers and BioClinical BERT using HuggingFace Transformers and PyTorch Lightning, combining vision and language in histopathology. Achieved 0.59 BLEU-4 score and able to detect tissue type with 80% accuracy.

PAPERS

Automatic Report Generation for Histopathology images using pre-trained Vision Transformers (Machine Learning for Health (colocated with NeuRIPS 2023), Findings Track
Saurav Sengupta and Donald E. Brown

Determining risk factors for Long COVID using Positive Unlabeled learning on Electronic Health Records data from NIH N3C, accepted at IEEE ICMLA 2023
Saurav Sengupta, Johanna Loomba, Suchetha Sharma, Donald Brown et al.

Analyzing historical diagnosis code data from NIH N3C and RECOVER Programs using deep learning to determine risk factors for Long Covid, IEEE BIBM 2022
Saurav Sengupta, Johanna Loomba, Suchetha Sharma, Donald E Brown et al.

Deep Learning for Visual Recognition of Environmental Enteropathy and Celiac Disease (BHI 2019)
Aman Shrivastava, Karan Kant, Saurav Sengupta, Sung-Jun Kang, Mariam Khan et al.

SKILLS

Tools and Languages Python, R, Java 7 (Oracle Certified Associate), C, Verilog
Packages/Tools PyTorch, PyTorch Lightning, TensorFlow, AWS, PySpark, Palantir Foundry, SQL

ACTIVITIES

2023 MIDAS Future Leaders Summit Cohort, University of Michigan Ann Arbor APR 2023

School of Data Science MSDS Capstone Mentor JUL 2020 — JUL 2022

ICMLA 2023, 2024 Reviewer 2023