Saurav Sengupta

CONTACT INFORMATION

21 Bonnycastle Drive
Data Science Institute
University of Virginia

Email (user id): ss4yd@virginia.edu
Web: **https://ssen7.github.io**
Github: **https://github.com/ssen7**

EDUCATION

University of Virginia, Charlottesville, Virginia USA

M.S. in Data Science, May 2019
Coursework: Machine Learning, Bayesian Machine Learning, Reinforcement Learning

Birla Institute of Technology and Science, Pilani, Goa, India

B.E. (Hons.) Electrical and Electronics Engineering, May 2016

PROFESSIONAL EXPERIENCE

University of Virginia, Charlottesville, Virginia USA

Research Scientist **August, 2019 - present**

Part of the **Gut Intelligence Lab** working on applying and developing deep learning methods for furthering medical research with focus on gastroenterology.

- Built a Resnet50 based CNN model written in PyTorch to classify high resolution biopsy images into Crohns disease and its subtypes.
- Extracted explanations using Gradient Class Activation Mappings and Locally interpretable model agnostic explanations (LIME) to help find tissue regions of interest.

Capgemini, Hyderabad, India

Associate Consultant **August, 2016 - October, 2017**

Responsible for developing REST based APIs in Java for Development Bank of Singapore (DBS) using distributed Pivotal Cloud Foundry framework.

- Built application to generate automated emails using Spring Boot and JPA.
- Successfully pushed code to production after creating Jenkins based CI/CD pipelines.
- Implemented BPMN2 based workflows on Camunda Modeler for business process automation of bank reporting processes.

STMicroelectronics, India

Trainee **August, 2015 - December, 2015**

Interned with team designing an embedded neural network based ASIC.

- Wrote code to interface Random Access Memory with I/O and buses.
- Worked on design verification and synthesis.

ACADEMIC PROJECTS

Classify duodenal biopsy images into diseases using Pytorch Child Health Research Center, UVA

- Built a CNN model to classify high resolution gastro-intestinal biopsy images into Diseased (Celiac Disease or Enteropathy) vs. Normal tissue with 98% biopsy level accuracy.
- Extracted features from last layer of trained model and used correlation to find medically relevant patterns between features learned from CNN and biomarkers collected from patients.

Reinforcement Learning Based Drawing Agent

- Developed a Reinforcement Learning Based Drawing agent capable of learning to replicate a given image in an OpenAI gym based environment in Python.
- A reward function based on pixel matching error was used and Q-learning was used to generate action value mapping.
- An average training time of 300 seconds per image was achieved compared to 480 seconds for SARSA.

Distracted Driver Recognition Using Bayesian Neural Networks using Pytorch

- Bayesian Neural Network model chosen to build classifier to detect if driver is paying attention while driving as CNNs do not give uncertainty or recognize out of class labels.
- We used three hidden layers and assumed that the parameters have a Gaussian distribution. We observed that training loss saturates after 40 epochs.
- Demonstrated that Bayesian neural networks skips classification of 98.5% of samples from outside of the training labels which is impossible with a CNN.

Deep Learning for detecting event-based stock price dependencies across multiple market sectors

- Used stock market time series data to generate events using changepoint detection.
- Built a 1D CNN model using Keras to predict stock prices of a sector an hour after event detection in another sector.
- Model was able to predict stock prices in the Financial sector with an error of ± 0.26 USD for event detected in the Information Technology sector.

Creating billboard music web-scraper using Beautiful Soup

- Created a web scraper to collect top 100 Billboard Music charts for every month with singer name and genre.

Publications

Deep Learning for Visual Recognition of Environmental Enteropathy and Celiac Disease (BHI 2019)
Shrivastava, Aman, Karan Kant, Saurav Sengupta, Sung-Jun Kang, Marium Khan, S. Asad Ali, Sean R. Moore et al.

Deep Learning for Detecting Diseases in Gastrointestinal Biopsy Images (SIEDS 2019)
Srivastava, Aman, Saurav Sengupta, Sung-Jun Kang, Karan Kant, Marium Khan, S. Asad Ali, Sean R. Moore et al.

Mo1992 Solving the Stain Dilemma: Computational Image Analysis to Address Differential Tissue Staining Color Bias in Duodenal Biopsies (Gastroenterology)
Syed, Sana, Aman Shrivastava, Karan Kant, Saurav Sengupta, Luke Kang, Marium N. Khan, Najeeha T. Iqbal et al.

Technical Skills

- Languages: Python, R, Java 7 (Oracle Certified Associate), C, Verilog
- Packages/Tools: PyTorch, Keras, pandas, numpy, scikitlearn, scipy, caret(R), RShiny, TensorFlow, GitHub, Jenkins, AWS EC2, PySpark, Pivotal Cloud Foundry
- Database: MySQL, MariaDB
- Operating Systems: Windows, Linux

Volunteer Work

- Participated in projects to generate employment and self-help services for women from impoverished neighbourhoods in Zuarinagar, Goa as part of Nirmaan Organization.
- Deputy Event Manager-Joy of Giving Week 2013 BITS Pilani, Goa Campus.