Saurav Sengupta

ss4yd@virginia.edu | https://www.linkedin.com/in/saurav-sengupta/ | https://ssen7.github.io | +1 (434)-284-3794

EDUCATION

University of Virginia **July 2018-May 2019**
Masters of Science in Data Science, 3.96/4
Birla Institute of Technology and Science, Pilani **August 2012-May 2016**
Bachelors of Engineering in Electrical and Electronics Engineering, 7.2/10

PROJECTS

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

- Used Convolutional Neural Networks (**CNN**) to classify high resolution digitized biopsy images into Celiac Disease, Environmental Enteropathy and Normal tissues.
- Used pretrained **Resnet50** in our analysis, performs well on diverse set of images.
- Achieved close to **98%** biopsy level accuracy. Paper accepted in **IEEE BHI 2019 conference**.
- Backed by the **Bill and Melinda Gates Foundation** and **Aga Khan University**.

Music Genre Classification

- Classification of songs into one of 13 genres like rock/pop and country etc. using **ensemble models** for song features and lyrical data.
- Captured data from Million Songs Database. Created **AWS EC2 instance**, mounted snapshot, ran **Jupyter notebooks on EC2** to get data. Used genre data from different Kaggle datasets.
- Used **Random Forests**, **SVM**, **Naïve Bayes** for model using song features. Used tf-idf and topic modeling for lyrical data.
- Around 58.06% accuracy, close to the original paper on music classification of 61% but we classify more genres.

Reinforcement Learning Based Drawing Agent

- Developed a Q-Learning Based Drawing agent capable of learning to replicate a given image in an OpenAI gym based environment.
- Also, developed a Deep Q Network Learning based drawing agent.

Distracted Driver Recognition Using Bayesian Neural Networks

- Used Bayesian Neural Networks to classify images from Statefarm distracted driver dataset to classify images into one of 10 classes.
- Used Bayesian Convolutional Neural Nets to get out of class errors and model uncertainty.

Modeling Brain Wave Activity using Muse™ headset

- Used Logistic Regression and Gradient Boosting for real time classification of brain wave signals like alpha waves into left-right, up-down motion. Tested model by moving objects in a game environment.

PROFESSIONAL EXPERIENCE

Capgemini (Cloud Foundation Services) - Associate Consultant **August 2016 – October 2017**

- Responsible for developing REST based APIs using for Development Bank of Singapore (DBS).
- Used cloud platform Pivotal Cloud Foundry® (VMWare) to deploy code to production.
- Worked on making CI/CD pipelines and setting up test environments.
- Wrote micro-service for sending scheduled emails to parameterized receivers based on the output of a database query using Spring JPA.
- Worked on implementing BPMN2 based workflows on Camunda Modeler for business process automation of bank reporting processes.

STMicroelectronics – Trainee **July 2015 - December 2015**

- Interned with team designing an embedded neural network based ASIC.
- Wrote code to interface Random Access Memory with peripherals like I/O and buses.
- Worked in Unix environment and wrote Bash shell and Tcl/Tk scripts for code synthesis and design verification.

CONFERENCE PAPERS/PUBLICATIONS

- ***Deep Learning for Visual Recognition of Environmental Enteropathy and Celiac Disease***
  A. Shrivastava*, K. Kant*, ***S. Sengupta****, S. Kang*, M. N. Khan, S. Moore, S. A. Ali, B. Amadi, P. Kelly, S. Syed, D. Brown (Paper accepted at IEEE-EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS 2019 Chicago, IL) (* equal contribution)

- ***Deep Learning for Detecting Diseases in Gastrointestinal Biopsy Images***
  A. Shrivastava*, ***S. Sengupta****, K. Kant*, S. Kang*, M. N. Khan, S. Moore, S. A. Ali, B. Amadi, P. Kelly, S. Syed, D. Brown (Paper accepted at Systems and Information Engineering Design Symposium (SIEDS) at Charlottesville, Virginia) (* equal contribution) (Best Poster Award in Data Science).
- ***Solving the Stain Dilemma: Computational Image Analyses to Address Differential Tissue Staining Color Bias in Duodenal Biopsies.***
  S. Syed, A. Shrivastava, K. Kant, L. Kang, ***S. Sengupta***, M. Naveed Khan, N. Talat Iqbal, K. Sadiq, C. A. Moskaluk, B. Amadi, P. Kelly, S. Moore, D. Brown. (Abstract accepted for poster presentation at Digestive Disease Week (DDW), May 20th 2019).

TECHNICAL SKILLS

- **Languages**: Python, R, Java 7 (Oracle Certified Associate), C, Verilog
- **Database**: MySQL, MariaDB
- **Packages/Tools**: pandas, numpy, sklearn, caret(R), RShiny, TensorFlow, PyTorch, GitHub, Jenkins, AWS EC2, Apache Spark and EMR (Spark Cluster), Pivotal Cloud Foundry®
- **Operating Systems**: Windows, Linux