

## Experience

---

### Data Scientist

*Children's Hospital of Philadelphia Research Institute*

Philadelphia, PA

Jul 2018 - Present

- Led multiple NIH funded efforts from ideation to dissemination to develop machine learning solutions for predicting various clinical outcomes.
- Led efforts towards project conceptualization, data acquisition, exploratory data analysis and model development using structured and unstructured data on high performance computing (CPU/GPU) environments.
- Collaborated with diverse stakeholders and presented findings to technical and non-technical audiences.
- Managed summer interns and provided technical mentorship to fellow data scientists.

### Graduate Research and Teaching Assistant

*University of Pennsylvania*

Philadelphia, PA

May 2016 - May 2018

- Developed feature selection methods for high dimensional datasets and made [open-source contributions](#).
- Developed C++ modules to estimate molecular properties using variations to the MCMC algorithm.
- TA for Theoretical Machine Learning and Mathematical Statistics.

### Data Science Intern

*Avis Budget Group*

Parsippany, NJ

May 2017 - Aug 2017

- Built one of their earliest in-house rental demand prediction models using an ensemble approach.
- Automated the ground truth rental demand and model performance reporting processes through extraction, wrangling and visualization of large siloed corporate data.

## Skills

---

**Programming:** Python, R, SQL, C++, MATLAB

**Tools:** PyTorch, scikit-learn, pandas, NumPy, SciPy GCP, Git, Docker, Tableau

## Education

---

### University of Pennsylvania

Dual Master of Science (M.S.), Computational Science and Engineering  
*Concentration in Statistics, Machine Learning and Numerical Analysis*

Philadelphia, PA

2015 – 2018

### National Institute of Technology Karnataka

Bachelor of Technology (B.Tech.), Engineering  
*Concentration in Numerical Analysis and Simulation*

Surathkal, India

2011 – 2015

## Selected Publications and Projects

---

### Personalized prediction of asthma persistence

*PLoS One*

- Developed machine learning models (XGBoost, Random Forest, Logistic Regression) and training pipelines consisting of feature selection, class balance, Bayesian hyperparameter tuning and model evaluation.
- Addressed clinical explainability needs through permutation analysis and achieved ROC-AUC of 0.86 (95% precision, 82% recall at 70% specificity) for XGBoost, successfully demonstrating the utility of machine learning for a novel task.

### Learning to detect rib fractures in chest X-rays

*The Society for Pediatric Radiology Annual Meeting 2022*

- Developed CNNs with ResNet backbones to detect rib fractures in X-rays using PyTorch on a multi-GPU environment.
- Achieved ROC-AUC of 0.75 and model explainability with limited training data using patch-based transfer learning.

### Asthma Biomarker Detection

*International Conference on Health Informatics 2022*

- Developed XGBoost and Kernel SVM models with filter based feature selection to identify relevant asthma biomarkers from a very high dimensional dataset.
- Achieved ROC-AUC of 0.91 (93% TNR at 70% TPR), outperforming previously developed linear and rule based systems.

### Semantic Segmentation to automate radiological measurements

*Under Review*

- Built a U-Net model with Resnet-34 encoder to segment bones and automate geometric measurements on radiographs.
- Achieved dice coefficient of 0.9 and 40x improvement in diagnostic efficiency while preserving measurement accuracy.

### Personalized Recommendation of Diagnostic Resources

- Built an ALS matrix factorization based collaborative filtering system to recommend medical resources to assist radiologists in their diagnoses.
- Achieved an nDCG score of 77% and significantly improved clinical efficiency.