

SAKSHAM JAIN

sj305@uw.edu | [linkedin.com/in/saksham-jain01/](https://www.linkedin.com/in/saksham-jain01/) | +1 (984)-377-0779

EDUCATION

SEP 2022 - Present	Ph.D., Statistics - UNIVERSITY OF WASHINGTON, SEATTLE Research Interests: machine learning, deep generative models, causal reasoning
AUG 2020 - MAY 2022	M.S., ECE (concentration in ML) - DUKE UNIVERSITY Advisors: Prof.s Cynthia Rudin , Alex Volfovsky , Sudeepa Roy Grade: 3.9/4
AUG 2015 - JUN 2019	B.E., (concentration in Robotics & Automation) - NSIT, UNIVERSITY OF DELHI Advisor: Dr. Umang Soni Grade: 7.7/10

PREPRINTS & PUBLICATIONS (* DENOTES EQUAL CONTRIBUTION)

- Jain, S.*, Seale-Carlisle, T. M.*, Lee, C., Levenson, C., Ramprasad, S., Garrett, B., Roy, S., Rudin, C., & Volfovsky, A. (2024). "Evaluating pre-trial programs using machine learning matching algorithms". **AAAI '24 (oral)**. ([Preprint](#))
- Kornfein, C., Jain, S.*, Willard, F.*, Tang, C.*, Long, Y.*, Malof, J.*, Ren, S.*, & Bradbury, K.* (2022). "Closing the Domain Gap - Blended Synthetic Imagery for Climate Object Detection". **NeurIPS '22 Workshop on Tackling Climate Change with Machine Learning**. ([Publication](#)). Extended: *Environmental Data Science*. (2023). ([Publication](#))
- Nayak, G. K., Mopuri, K. R., Jain, S., & Chakraborty, A. (2021). "Mining Data Impressions from Deep Models as Substitute for Unavailable Training Data". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. ([Publication](#))
- Jain, S., Seth, G., Paruthi, A., Soni, U., & Kumar, G. (2020). "Synthetic data augmentation for surface defect detection and classification using deep learning". *Journal of Intelligent Manufacturing*, 1-14. ([Publication](#))
- Nayak, G. K., Jain, S., Babu, R. V., & Chakraborty, A. (2020, September). "Fusion of Deep and Non-Deep Methods for Fast Super-Resolution of Satellite Images". *IEEE International Conference on Multimedia Big Data (BigMM)* (pp. 267-271). ([Publication](#))
- Jain, S., Jain, S., & Komanduri, A. P. (2018, October). "A cascade predictive control strategy for active suspension systems". *IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA)* (pp. 1-5). ([Publication](#))
- Jain, S., & Sreedevi, I. (2018, December). "Robust Detection of Iris Region Using an Adapted SSD Framework". *Workshop on Computer Vision Applications* (pp. 51-64). Springer, Singapore. ([Publication](#))

PROFESSIONAL & RESEARCH EXPERIENCE

JUL 2023 - SEP 2023 LA, USA	AI/ML Engineering Intern , involve.ai <ul style="list-style-type: none">Built a research-and-revise approach for correcting LLM outputs based on retrieved evidence, implemented as a 'co-pilot' for tasks like market research.Built a DistilBERT-based multi-task text classifier for job titles, for a 4x improvement in recall for search and retrieval of sales leads based on natural language queries.Developed an ML-based approach for data matching under uncertainty, for record linkage and enrichment in/from large databases.
AUG 2021 - MAY 2022 NC, USA	Graduate Team Lead , Bass Connections Project, Duke Energy Initiative <ul style="list-style-type: none">Built a novel synthetic imagery-based domain adaptation technique for energy infrastructure detection models, improving AP by ~8.23% in cross-domain pairings across regions/terrains.Worked with aerial images of wind turbines, including data collection and annotation, and the YOLOv3 model for their detection.
MAY 2021 - APR 2022 NC, USA	Graduate Research Assistant , Almost Matching Exactly Lab, Duke University <ul style="list-style-type: none">Developed a novel interpretable approach for impact evaluation - at the sub-group level - of (judicial) pre-trial services on post-trial recidivism, using FLAME-DAME, a machine learning-based matching method for causal inference.
MAY 2021 - AUG 2021 NC, USA	Research Intern , RAILabs, Duke University <ul style="list-style-type: none">Enhanced a YOLO-based single-shot lesion detection pipeline with a patch classifier to further reduce false positives by ~80% at negligible cost to inference speed.Extended it to an end-to-end trainable pipeline for joint ipsilateral lesion matching and lesion detection in paired screening mammogram images.

JAN 2020 - JUL 2020 KARNATAKA, INDIA	Research Project Assistant , Visual Computing Lab, IISc Bangalore <ul style="list-style-type: none"> Developed a novel statistical approach for constructing ‘impressions’ of training data using only the parameters of a pre-trained CNN. Improved upon existing SOTA performance in several tasks in the training data-free ML setup. Built a novel model-agnostic technique for fast intelligent selective superresolution with comparable performance to deep superresolution models and nearly the speed of bicubic interpolation.
JUN 2019 - JAN 2020 UP, INDIA	Machine Learning Engineer , Million Sparks Foundation <ul style="list-style-type: none"> Built an unsupervised domain adaptation net for real-time user outcome prediction outperforming existing SOTA AUC scores by ~4.3%. Built models for text classification using BERT, recommendation using neural collaborative filtering, and explainable (with SHAP) churn prediction using a dense neural net classifier optimized for high recall.

SELECT PROJECTS

Variable Selection and Causal Inference (ongoing)

- Designed simple low-dimensional linear problems to test for presence of ‘bad’ controls (i.e., confounders that bias average causal effect estimates) in regression adjustment sets.
- Working on designing variable selection algorithms that select only ‘good’ and ‘neutral’ controls.

Out-of-Distribution (OOD) Detection using Likelihood Ratios ([Project](#))

- Implemented likelihood ratio-based approaches that construct a background-contrastive statistic, to correct for spurious background information that can confound deep generative models into assigning high likelihood to OOD image inputs.

Window Level Optimization for Pectoral Muscle Segmentation in MLO-view Mammograms ([Project](#))

- Formulated the window-levelling operation (the typically deterministically-set function for mapping a 12-16 bits/pixel grayscale DICOM image to an 8-bit grayscale display) as a convolution layer.
- Jointly optimized it with a U-Net, to obtain an improvement of ~4.5% on pectoral muscle segmentation.

An ‘Adaptive’ Methodology: Is this novel an adaptation? ([Project](#))

- Compared Tf-idf, BERT, Universal Sentence Encoder (USE), longformer (circa 2020) for investigating semantic textual similarity of (chunked) novels. Implemented a USE-based classifier (F1-score of 0.98) for detection of texts/plots that draw from *Robinson Crusoe*.

Bachelor’s Thesis: Synthetic Data Augmentation for Surface Defect Detection and Classification using Deep Learning ([Project](#))

- Proposed a synthetic data augmentation approach for (typically small) manufacturing defect datasets to offset the high opportunity cost of data acquisition.
- Compared DCGAN-based & ACGAN-based augmentations on a CNN, surpassing the classic augmentation performance by ~5.5%.

TECHNICAL SKILLS

Languages: Python, C/C++, Matlab, R

Frameworks/Libraries: TensorFlow, PyTorch, scikit-learn, Keras, OpenCV, numpy, pandas

Technologies: GCP, AWS, Apache Airflow, Git, Flask, REST, FastAPI, \LaTeX

AWARDS & RECOGNITION

SEP 2022	Coursera Department Fellowship by UW, for outstanding promise for graduate work
AUG 2021	CVSPK Talent Incentive Scheme by NSIT, Cash Award for SCI-indexed research
JUL 2019	Cash Award for Merit , both computer vision and machine learning summer schools at IIIT Hyderabad, for scoring in the top-20 out of 280 participants
MAY 2019	Bachelor’s Thesis Ranked 2 in the Dept. of Manufacturing Processes and Automation Engg.
JUN 2018	International Rank 20 , CanSat Competition 2018

TEACHING EXPERIENCE (PRE-DOC TEACHING ASSOCIATE)

SPRING 2023	Introduction to Statistical Machine Learning (STAT 435), Dept. of Statistics, UW Seattle
WINTER 2023	Statistical Methods in Engineering and Science (STAT 390), Dept. of Statistics, UW Seattle
FALL 2022	Elements of Statistical Methods (STAT 311), Dept. of Statistics, UW Seattle
SPRING 2022	Probabilistic Machine Learning (STA 561D), Dept. of Statistics, Duke University
FALL 2021	Vector Space Methods with Applications (ECE 586), Dept. of ECE, Duke University