

# ABHIBHA GUPTA

abg96@pitt.edu   ♦   +1 4129547510   ♦   Website   ♦   Linkedin

## EDUCATION

- University of Pittsburgh   August 2022 - Present  
MS in Information Science   GPA: 4/4
- Indian Institute of Information Technology, Nagpur   July 2017 - June 2021  
BTech in Computer Science and Engineering   GPA: 3.45/4

## TECHNICAL SKILLS

**Languages** - Python, R, MySQL, C++, C, Bash   **ML/DL** - PyTorch, Tensorflow, Fastai, Keras, scikit-learn, spaCy, Hugging Face, Langchain, Caret, glmnet  
**Data Processing** - Pandas, NLTK, CoreNLP, Gensim, Tidyverse (Dplyr, Tibble, Purrr), igraph   **Visualisation** - Matplotlib, Seaborn, ggplot  
**Coursework** - Machine Learning in R ([Code](#)), Databases, Algorithms, Data Mining, Artificial Intelligence, Information Storage and Retrieval, Natural Language Processing, Deep Learning, Computer Vision, Neuro-Fuzzy Techniques, Probability, Graph Theory, Data Science, Bio-informatics

## PUBLICATIONS

- Enhancing Visual Perception in Novel Environments via Incremental Data Augmentation Based on Style Transfer Paper  
Abhibha Gupta, Mansur Maturidi Arief, Rully Agus Hendrawan- International Conference on Robotics and Automation (ICRA) *Under Review*
- Towards Accurate and Clinically Meaningful Summarization of Electronic Health Record Notes: A Guided Approach. Slides  
Zhimeng Luo, Yuelyu Ji, Abhibha Gupta, Zhuochun Li, Adam Frisch, Daqing He, IEEE International Conference on Biomedical and Health Informatics
- Neural architecture search for pneumonia diagnosis from chest X-rays. Paper, Article  
Abhibha Gupta, Parth Sheth, Pengtao Xie, Journal of Nature Scientific Reports.
- Disambiguating spatial prepositions: The case of geo-spatial sense detection. Paper  
Mansi Radke, Abhibha Gupta, Kristin Stock, CB Jones, Transactions in GIS Journal.

## RESEARCH EXPERIENCE

University of Pittsburgh (IRiS Lab)   📍 Pittsburgh, US  
Graduate Student Researcher, Advisor: Daqing He   📅 May 2023 – Present

- Guided summarization of clinical notes: We developed a comprehensive summary template consisting of important medical entities covering patient demographics, chief complaint, OPQRST assessment, diagnostics, treatment, etc.
- Fine-tuned a BART-based 'guided' summarization model with textual guidance extracted from a sentence classifier trained to determine the importance of sentences within clinical notes.
- Implemented a fact-checking tool for validating the 'factual correctness' of predicted summaries against the ground truth by fine tuning a Bio-ClinicalBERT-based Named Entity Recognition (NER) model
- Currently experimenting with medically focused prompt engineering techniques by running multi-GPU inference using Huggingface Accelerate on large language models (LLMs) like LLaMA and Alpaca to improve the task of clinical note summarization.

Stanford University (SISL Lab)   📍 Remote  
Independent Student Researcher   📅 May 2023 – September 2023

- Improving rare traffic sign recognition via data augmentation: Propose a novel approach to enhance visual perception in autonomous agents by tackling the uncharted territory of "unknown unknowns" in real-world scenarios, particularly focusing on degraded signs not seen during training.
- Leveraged the Variational Prototyping Encoder (VPE) combined with neural style transfer to incrementally augment data, resulting in a significant boost in model robustness when faced with novel inputs.
- Demonstrated the effectiveness of the proposed methodology on the German Traffic Sign Recognition Benchmark (GTSRB) dataset, achieving an F1 score of 86% (36% increment over baseline) , underscoring the model's adaptability and resilience to previously unseen data variations. [Code](#)

University of California, San Diego   📍 Remote  
Research Intern, Advisor: Pengtao Xie   📅 May 2021 – April 2022

- Neural Architecture Search (NAS) for Pneumonia Diagnosis: Implemented Neural Architecture Search (NAS) for improving pneumonia diagnosis from Chest X-Ray images by Leveraging the 'Learning By Teaching' framework, inspired by teacher-student learning paradigm that outperforms previous NAS methods like DARTS and PC-DARTS by 5.1%.
- The searched model attained a 97.6% ROC-AUC score for pneumonia detection, while being 4% smaller than DARTS. [Code](#)
- Reading by Translating: Implemented the 'Reading by Translating' framework that improves the task of 'Machine Reading' i.e extracting meaningful instances from the dataset. Involves 2 transformer based encoder-decoder models, that are trained mutually on the task of Machine Translation to learn importance weights assigned to the dataset instances.

## PROJECTS

- Argumentative Stance Detection - 10th Multimodal ArgMining Workshop (EMNLP 2023)  
Secured 4th place among 9 teams by developing a weighted ensemble of multimodal (ViLT, FLAVA) and text models for tweet 'argumentative stance' prediction. Achieving an F1 score of 0.81, a 10% increase from baseline.
- Predicting Corrosion in Surface Coatings - PPG Paints  
Conducted exploratory data analysis and feature selection in R with Ggplot library. Employed Bayesian models (splines, XGBoost, Random Forests, MARS, Neural networks) to predict corrosion percentage, achieving a 0.94 ROC-AUC score. [Code](#)