

ABHIBHA GUPTA

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

EDUCATION

University of Pittsburgh 2022 - Present
MS in Information Science GPA: 4/4

Indian Institute of Information Technology, Nagpur 2017 - 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

- **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.

EXPERIENCE

University of Pittsburgh 📍 Pittsburgh, US
Student researcher, Advisor: Daqing He 📅 May 2023 – Present

- **Guided summarization of clinical notes:** Developed a comprehensive summary template covering patient demographics, chief complaint, OPQRST assessment, diagnostics, treatment, etc
- Developed a sentence classifier to determine sentence importance within clinical notes, providing guidance to the BART-based summarization model. Additionally, fine-tuned a Bio-ClinicalBERT-based Named Entity Recognition (NER) system, enabling the implementation of a fact-checking metric for validating predicted summaries against the ground truth.
- 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 📍 Remote
Independent student researcher 📅 May 2023 – September 2023

- **Improving rare traffic sign recognition via data augmentation:** Addressed challenges of out-of-training distribution images (e.g., Rusty traffic signs), crucial for vehicle perception systems. Under review at ICRA.
- Augmented the German Traffic Sign Recognition Benchmark dataset through neural style transfer, introducing rusty sign images. Constructed a pipeline with Variational Prototyping Encoders to identify and classify rare image class, achieving a **0.86 class-weighted F1 score**. [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.

Tata Research, Design and Development Centre 📍 Remote
Research Intern, Advisor: Sagar Sunkle 📅 May 2020 – Nov 2020

- **Named Entity Recognition (NER):** Worked on information retrieval using NER on domain-specific scientific corpora.
- Implemented incremental learning for NER by adapting classifiers from open-source libraries such as Adaptive Random Forests, AdaBoost, Pretrained Spacy NER, Conditonal Random Fields (CRF) and Seq2Seq model.
- Developed an 'Extractive search system for entity retrieval' similar to AllenAI's Spike cord search system that enables one to perform selective information extraction and annotate results by querying. [Code](#)

PROJECTS

Argumentative Stance Detection - 10th ArgMining Workshop (EMNLP 2023)

Developed weighted ensemble of multimodal and text models to predict tweet 'argumentative stance' achieving a 4th place finish among 9 teams.

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](#)