+1 805 637 9110
New York City
yw4196@columbia.edu

Youzhi Wang

GitHub: youzhiw
LinkedIn: youzhi-wang/

Education

Columbia University **New York, NY**
MS Computer Science **May 2024**
GPA: 3.96

University of California, Santa Barbara **Santa Barbara, CA**
BS (double major) Applied Mathematics, Statistics and Data science **June 2023**
Graduated with Honors and Major(Applied Math) distinction

Projects

NLP project

Creating robust datasets for LLM evaluation (New York, NY) **May 2024 – August 2024**
- Coauthored a paper for EMNLP conference
- Parameterized 1/3 of a math questions dataset (GSM8K) to introduce randomness to better evaluate LLM's reasoning ability
- Used python code to generate 400 questions and answers for evaluating LLMs

Deep Learning projects

Schizophrenia classification (New York, NY) **Jan 2024 – May 2024**
- Trained a VGG11 CNN model from scratch on 3D MRI brain images that achieved AROC of 0.87
- Utilized GradCAM from VGG11 model to visualize regions of the brain that cause schizophrenia

Cancer cell detection and classification (New York, NY) **Jan 2024 – May 2024**
- Fine tuned SWIN T transformer model to detect position of cancer cell in a image that achieves precision of 0.63
- Trained Neural Network model using AWS nodes

Multiple Sclerosis classification using Brain MRI (New York, NY) **Sep 2023 – Jan 2024**
- Wrote Python scripts using PyRadiomics to run radiomics analysis for 100 brain MRIs using clusters
- Wrote Bash scripts that runs brain MRI segmentation using FreeSurfer
- Wrote Python scripts using PyTorch and MONAI for classifying Multiple Sclerosis based on radiomics analysis that achieves 70% accuracy

Earthquake data collection and classification (Santa Barbara, CA) **Jul 2022 – Jun 2023**
- Received Applied math Major Distinction award from the University of California, Santa Barbara for my work
- Collected and cleaned large collections of seismic signals (over 100,000 waveforms) using Python and ObsPy
- Improved high-quality signal identification time by 5 times through restructuring the previous data processing pipeline in JupyterNotebooks
- Labeled 200+ high-quality Earthquake signals that can be used to train ML algorithms
- Tested 5 different Machine Learning algorithms (KNN, LDA/QDA, SVMs and DCNN) for classifying high quality seismic wave signals

other coding projects

HTTP 1.0 Web server **Sep 2023 – Dec 2023**
- Wrote a HTTP 1.0 web server using C and socket APIs with image hosting and an interactive search module that uses a linked list to search through a given database.

Skills

Proficient with: Python, PyTorch, sklearn, Java, Bash, SQL, AWS, Linux, Git, JupyterNotebook, R, RStudio
Proficient languages: Mandarin