YIHENG LI

Background: Biomedical Informatics @ Stanford University

★ li.terry710@gmail.com | (650)-665-3654 | linkedin.com/in/yiheng-li | terryli710.github.io

EDUCATION

Stanford University M.Sc. in Biomedical Informatics 09/2019-06/2021

- Research interests: deep learning application in medical imaging; multimodal data fusion; head impact analysis and deep learning brain injury prediction; meta-learning application in patient outcome prediction; natural language processing for clinical free text; genomic sequence embedding and activator prediction.

Shanghai Jiao Tong University B.Sc. in Resource and Environmental Science 09/2015-06/2019

University of California, Berkeley International Exchange Program 01/2018-05/2018

PUBLICATIONS AND MEETINGS

Deep Learning-Based Multi Contrast MRI Registration Model with a Realistic Flow Field and Reduced Over-Smoothing Effect - Role: first author	2023 Subtle Medical	ISMRM 2023
Deep Learning based Image Co-Registration Quality Control - Role: first author	2022 Subtle Medical	ISMRM 2022 ASNR 2022 RSNA 2022
A 3D Lung Lesion Variational Auto-Encoder - Role: first author	2022 Stanford	RSNA 2022 Cell Press Community (in-process) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4342994
AI-based analysis of CT images for rapid triage of COVID-19 patients - Role: main programmer, main author	2021 Stanford	npj Digital Medicine https://doi.org/10.1038/s41746-021-00446-z
Predictive Factors of Kinematics in Traumatic Brain Injury from Head Impacts Based on Statistical Interpretation - Role: first author, experiment designer, main programmer	2021 Stanford	Annals of Biomedical Engineering https://link.springer.com/article/10.1007/s10439-021-02813-z
The Relationship between Brain Injury Criteria and Brain Strain Across Different Types of Head Impacts Can Be Different Role: first author, experiment designer	2021 Stanford	Journal of the Royal Society Interface https://pubmed.ncbi.nlm.nih.gov/34062102/
Photosynthesis and Related Physiological Parameters Differences Affected the Isoprene Emission Rate among 10 Typical Tree Species in Subtropical Metropolises - Role: experimenter	2019 Shanghai	International journal of environmental research and public health https://doi.org/10.3390/ijerph18030954

EXPERIENCES

Deep Learning Research Scientist at Subtle Medical Inc. Menlo Park, CA 08/2021-07/2023

- Deep Learning based Image Co-Registration Quality Control — 08/2021-02/2022
  - Built a self-supervised classification model to QC the co-registration performance for pairs of 3D MRI images.
  - Developed a data augmentation pipeline including affine and deformable movement to generate pseudo data pairs to tackle the data challenges of obtaining labelled well-registered MRI data pairs.
  - To achieve multi-contrast prediction, compared two technical approaches: augmented contrasts vs. modality invariant representation.

Meeting accepted:
ISMRM 2022
ASNR 2022
RSNA 2022

- Deep Learning-Based Multi Contrast MRI Registration Model with a Realistic Flow Field and Reduced Over-Smoothing Effect — 02/2022-09/2022
  Meeting accepted: ISMRM 2023
  - Developed a deep learning based co-registration model that can be applied to multi-contrast MRI images of multiple anatomies.
  - Novel attempts, using “Jacobin loss” and “cycle consistent loss”, to deal with unrealistic flow field and over-smoothing effect of deep learning-based registration methods, especially in the “VoxelMorph” and “SynthMorph” framework.
  - Converted TensorFlow based “SynthMorph” code and trained PyTorch versions of “SynthMorph” and other variations of the model.
  - Improved the SSIM and PSNR of the registered image on BraTS and Lumbar-Spine Open Dataset by ~40% and ~50% respectively.

- A Self-Supervised Key Point Detection Framework For Multiple Applications — 09/2022-07/2023
  - Explore, adapted, developed and assessed multiple technical workflows for self-supervised key point detection in medical images.
  - Developed a two-step rule-based pipeline for brain MRI auto-formatting, using ANTs affine registration and SIFT key point matching.
  - Guiding a intern to adopted and optimized the performance of the “KeyMorph”, an automatic key point generator with registration training mechanism, by customizing the loss with distance and applying additional mask.
  - Start with the replication of a real-time self-supervised key point detector paper in PyTorch Lightning. Optimize over the original paper’s result by ~30% on a private self-curated test dataset with modified training strategy and grid loss search. A multi-purpose training framework is introduced by adding a tail to provide binary prediction of the key point existence. Transformer-based encoder replaced the CNN based encoder. Reinforced learning and iterative predictions are both tested for finer prediction.

- Creation and Maintenance of a PyTorch-Lightning and MONAI based Deep Learning Training, Logging and Inference Helper Package: Lumos-ToolKit — 09/2022-07/2023
  - Created and maintained a toolkit which includes pipelines for the key steps and pain point in deep learning model development for medical imaging: dataset curation and testing; image affine, deformable and other spatial transformations; logging and documentation of the model hyper-parameters, figures and performances; management of the training settings and configurations; image preprocessing pipeline; complicated loss settings and combinations of losses.
  - The whole package dynamically integrates the following packages: “pydantic”, “argdantic”, “PyTorch-Lightning”, “MONAI”, “rich”, etc.

Machine Learning Researcher at Gevaert Lab at Stanford DBDS — Stanford, CA — 03/2020-12/2022

- Unsupervised Lung Nodule Encoder Based on 3D Variational Auto-Encoder — 01/2021-present
  - Implemented a 3D Beta-VAE network based on PyTorch. Trained the network with cropped lung nodule patches in public datasets and evaluated on Stanford lung cancer dataset.
  - With demonstrated ability to accurately reconstruct lung nodule patches, the model’s bottleneck embeddings were extracted and used to train models for downstream tasks predictions.
  - Analysis results showed that VAE models trained combinedly on public datasets and target datasets can do best in downstream tasks, including classification of T/N stage etc.

- Patient Treatment Evaluation and Prediction Based on Multimodal Data Fusion of NSCLC — 07/2020-12/2022
  - Implemented patient information data extract pipeline from clinical texts, pharmacy tables and CT headers, and constructed a scalable visualization tool using seaborn and matplotlib for patient timeline.
  - Built SNP translation and annotation pipeline using web APIs, tools and databases including myvariant, wANNOVAR, dbSNP, REVEL, etc.
  - Extracted lung cancer lesion CT segmentations radiological features using PyRadiomics in batches.
  - Performed association analysis over extracted features and patient treatment responses with statistical tests including one-way ANOVA and visualizations including PCA and Manhattan plot.

- AI-Based Analysis of CT Images for Rapid Triage of COVID-19 Patients — 03/2020-07/2020
- Cleaned, queried, matched various data types (CT images, lab results, clinical notes) with Python and R for data fusion.
  - Applied statistical methods for data resampling (SMOTEENN) and feature selection (univariate selection methods based on F-test or mutual information). — Published: npj digital medicine
  - Implemented multiple statistical learning models with extensive cross-validated hyperparameter tuning. Evaluated the models with bootstrapping.
  - Performed survival analysis with Kaplan-Meier estimator, Log-rank test, and Cox proportional hazards model with R, to illustrate efficacy of model performance.
  - Visualized various data types using ComplexHeatmap with customized clustering.

Research Assistant at Chaudhari Lab at Stanford AIMI — Stanford, CA — 07/2020-07/2021

- Osteoporosis Risk Prediction Model Based on Abdominal and Pelvis CTs — 07/2020-07/2021
  - Adopted and implemented deep models from recent research for L3 vertebral level prediction and segmentation that are based on TensorFlow on custom datasets.
  - Cropped, reoriented, and visualized Dicom images with ITK.
  - Independently implemented preprocessing and segmentation pipeline for trabecular bone abdomen-pelvis CTs based on UNet.

Research Assistant at Camarillo Lab at Stanford Bioengineering — Stanford, CA — 11/2020-04/2021

- Predictive Factors of Kinematics in Traumatic Brain Injury from Head Impacts Based on Statistical Interpretation — 09/2020-01/2021
- Implemented 6 types of linear regression feature importance interpretation methods in Python, either from theory or transplanted from R. — Published: Annals of Biomedical Engineering
  - Crafted a novel perspective of interpreting contributors of linear statistical models by feature groups (defined as factors) without the difficulty of collinearity.
  - Built pipeline for visualization of input data in heatmaps and then visualized the results in various types of bar plots and tables using R and LaTeX.

- The Relationship between Brain Injury Criteria and Brain Strain across Different Types of Head Impacts Can Be Different — 11/2020-02/2021 Published: Royal Society

Researcher at Prof. Yin Shan's Lab at Department of Resource and Environment at Shanghai Jiao Tong University — Shanghai, China — 04/2018-06/2019

- Photosynthesis and Related Physiological Parameters Differences Affected the Isoprene Emission Rate among 10 Typical Tree Species in Subtropical Metropolises — 04/2018-06/2019
  - Established the continuous observation platform, which monitored biogenic VOC and secondary organic aerosol; 9 categories of meteorological data and 6 categories of air contamination data. — Published: International journal of environmental research and public health
  - Optimized various data collecting methods to achieve the automation of data collecting system to achieve continuous observation.
  - Selected a tree-based model to provide predictions to multi-dimensional data's influence on BVOC as well as secondary organic aerosol emission and transformation patterns; hyperparameter tuning is conducted using MATLAB.

Research Intern at Shanghai Academy of Environmental Sciences — Shanghai, China — 07/2018- 07/2019

- Responsible for calibrating PTR-ToF-MS instrument, making research on how sensitivity, humidity influence, sampling method could influence the sensitivity and detection limit. 04/2018-07/2019
- Learned the knowledge about proton-transfer reactions, principle of time-of-flight mass spectrometry, physical and chemical properties of biogenic volatile organic compounds.
- Used linear regression with Excel and R to fit the data; analyze the reasons for the bad fitting; calculated various values.

SKILLS

Technical: Python, R, PyTorch, TensorFlow, Bash, git, GCP, AWS, Azure, C++, SPSS, MATLAB, SQL
Language: English (proficient), Chinese (native)

OTHER PROJECTS

Research Assistant at Camarillo Lab at Stanford Bioengineering Stanford, CA 11/2020-04/2021

- Predictive Factors of Kinematics in Traumatic Brain Injury from Head Impacts Based on Statistical Interpretation 09/2020-01/2021
- Implemented 6 types of linear regression feature importance interpretation methods in Python, either from theory or transplanted from R. Published: Annals of Biomedical Engineering
  - Crafted a novel perspective of interpreting contributors of linear statistical models by feature groups (defined as factors) without the difficulty of collinearity.
  - Built pipeline for visualization of input data in heatmaps and then visualized the results in various types of bar plots and tables using R and LaTeX.

- The Relationship between Brain Injury Criteria and Brain Strain across Different Types of Head Impacts Can Be Different 11/2020-02/2021 Published: Royal Society

Researcher at Prof. Yin Shan's Lab at Department of Resource and Environment at Shanghai Jiao Tong University Shanghai, China 04/2018-06/2019

- Photosynthesis and Related Physiological Parameters Differences Affected the Isoprene Emission Rate among 10 Typical Tree Species in Subtropical Metropolises 04/2018-06/2019 Published: International journal of environmental research and public health
  - Established the continuous observation platform, which monitored biogenic VOC and secondary organic aerosol; 9 categories of meteorological data and 6 categories of air contamination data.
  - Optimized various data collecting methods to achieve the automation of data collecting system to achieve continuous observation.
  - Selected a tree-based model to provide predictions to multi-dimensional data's influence on BVOC as well as secondary organic aerosol emission and transformation patterns; hyperparameter tuning is conducted using MATLAB.

Research Intern at Shanghai Academy of Environmental Sciences Shanghai, China 07/2018- 07/2019

- Responsible for calibrating PTR-ToF-MS instrument, making research on how sensitivity, humidity influence, sampling method could influence the sensitivity and detection limit. 04/2018-07/2019
- Learned the knowledge about proton-transfer reactions, principle of time-of-flight mass spectrometry, physical and chemical properties of biogenic volatile organic compounds.
- Used linear regression with Excel and R to fit the data; analyze the reasons for the bad fitting; calculated various values.