Zimeng Yu

Personal Website
Google Scholar

Email: Outlook OR Yale
Mobile: +1 203-410-8140
ORCiD: 0009-0000-4467-198X

Research Interests

General Discipline: Computational Biology, Bioinformatics, Quantitative Biology, Deep Learning
Biological Field: Genetics, Genomics, Molecular Biology, Synthetic Biology, Computer Vision
Detailed Interests: CRISPR/Cas9, Gene Sequencing, Protein Design, Molecular and Drug Discovery, Cell-Cell & Protein Interactions

Education

- **School of Public Health, Shanghai Jiao Tong University** — Shanghai, China
  Preventive Medicine; CGPA: 3.64(3.91/4 for last 3.5 years, TOP 20%) — Sep 2020 – Jun 2025
- **KCL 2021 Summer School of Public Health** — London, UK
- **The Global Summer School 2022 of Tsinghua Vanke School of Public Health** — Beijing, China
- **A Courses**: Over 43 of all 90 courses, including Statistics (Score: 91), R programming(93), Basic Calculus(96), Elementary Chemistry C(A-), Ogranci Chemyrsti C(A-), Machine Learning(89.1), Markov and Decision-trees(90)

Publications

Journal Papers

- **In-Submitting**
  - Third Author (2024). A prognostic model for head and neck squamous cell carcinoma based on a ceRNA network to predict radiotherapy response. *Journal of Radiation Research and Applied Sciences.*

- **Preprint Published**
  - Co-Author (2024). BEdeepon: an in silico tool for prediction of base editor efficiencies and outcomes. C Zhang, Z Yu, D Wang, T Qi, Y Zhang, L Hou, F Lan, J Yang, L Shi, et al. *bioRxiv.*

Research Experience

- **New Drug and Molecular Discovery Tool Development** — Shanghai Jiao Tong University
  Undergraduation Thesis — Oct 2024 – Present

  - **Evaluation and Benchmarking**: Document and arrange previous related research works, then evaluate and benchmark the performance to ensure the network structure and molecular structural charac-terization
  - **Model Construction**: The primary plan is to construct feature augmentation and model architecture via contrastive learning.

- **Deep Learning for Detect Cell-Cell Communication** — Yale University
  Research Assistant, Core Member — Feb 2024 – Present

  - **Making Protocols and Strategies for this Project**: Collect relative thesis and datas, then helping design some of the detailed experiments and prospectives of this project. Also helping other students who are interseted in it with their questions and problems.
  - **Preprocessing Datasets and Databases**: Pytorch and R, preprocessing our raw datasets of single cells from plantforms such as CosMx, STARmap and Visium HD, and handling on public databases to find LR pairs and cell-type information, then embedding them into a graph structure for further analysis.
  - **Hyperparameter Detecting**: Pytorch and Linux, figuring out the assumed range of cells' diffusion diameter and threshold of activated communication, and via finetuning hyperparameters of models with the least validation loss to provide evidences.
  - **Model Construction modifying and Evaluation**: Pytorch, launching reasonable and precise evaluation and benchmark criterias such as choosing which correlative coefficient and how, modifying the most suitable model structure to achieve the best performance in them.
  - **Visualization and tool development**: CSS, html, JavaScript, visualizing the results and providing interactive tools for users to explore both cell-level and spot-level communication data and predictions.

- **Predict Base Editor Editing Outcomes with Deep Learning** Shanghai Jiao Tong University
  Co-Leader *Jun 2023 – Present*
    - **Data Preprocessing**: Linux, mapping designed library with raw mutated sequencing datas with tool bwe-mem2, calculating editing efficiency and outcomes proportion to generate processed data for training and testing.
    - **Data Analysis and Descriptive Statistics**: Python and Matlab, describing and summarizing processed datasets including editing windows across 20bp protospacer region, mutation trends of single mutated position, effects of 1-bp and whole region context of targeted editing efficiency.
    - **Benchmark and Discuss Model Development**: Jupyter Notebook, TensorFlow and Linux, processing and analysising raw datasets from 4 other works and configuring benchmark models in HPC for comparison, including AUC scores, Spearman and Pearson correlation.
    - **Defining Class and Embedding**: Python, defining classes of somple, processing embedded inputs with embeddingbag method, then generating batchs.
    - **Building and Finetuning Model**: Python, Linux, TensorFlow, building LSTM model mainly with additional attention layers, then use ray and cross-validation K-Fold and stratifiedkfold to finetune best parameters until the best performance of prediction.
    - **Manuscript Handwriting**: Main writer, writing and editing the manuscript, including the graphic abstract plotting, methods, results, discussion and supplementarysections.
- **Radiotherapy Response Models in Cell Carcinoma** Shanghai Jiao Tong University
  Team Member, published *Sep 2023 – Oct 2024*
    - **Data Visualization and Analysis**: Utilized R to create flowcharts and heat maps for data visualization, contributing to the analysis of The Cancer Genome Atlas (TCGA) data.
    - **Manuscript Preparation**: Assisted in drafting the manuscript, focusing on the Introduction and Conclusion sections for publication.
- **Generate Chat-AI of Radiation Biology by Multi-model** Shanghai Jiao Tong University
  Team Member *May 2024 – Present*
    - **Annotation**: Label-Studio, configuring plantform for image labeling and annotation, which help doctors labeling radiology images manually.
    - **Prompt Engineering**: Python and Markdown, designed and ensure exact and clear prompt for chat-bot to generate radiation biology text information according to input image
- **The First Global AI Drug Development Competition** Shanghai Jiao Tong University
  Team Member *Apr 2023 – Jul 2023*
    - **Preprocessing**: Processed and curated provided datasets to ensure high-quality inputs for predictive modeling.
    - **Benchmark Model**: Collaborated with team members to analyze and compare model performance metrics.
    - **Competition Organization**: Successfully uploaded and submitted competition results, achieving TOP 20%
- **Cancer Integrative Analysis Via Multimodal Machine Learning** Shanghai Tong Ren Hospital
  Student Researcher *Nov 2022 – Mar 2023*
    - **Reproduce Research Paper**: Utilized TCGA and cBioPortal databases for comprehensive genomic data retrieval and integration, reproduct the datasets in the research paper and get accurate and consistent results
    - **Visualization**: Executed Whole Slide Image (WSI) analysis, contributing to the large-scale pathology datasets.

Skills

- **Programming Languages**: Python & R(Professional), C, /LaTeX, Matlab, Linux Operation System
- **Languages**: Chinese (Native), English (TOEFL 101, R30 L24 S23 W24), Japanese (Limited)

Honors

- **2022-2023 A Student Scholarship, School of Medicine (No.1 Student)**
- **2023-2024 B Student Scholarship, School of Medicine (No.4 Student)**
- **Overseas Research Training Support Program Scholarship, 2024**