YUNRUI LU

yunrui.lu@dartmouth.edu | https://lutzzzlu.github.io | 929-666-2186

EDUCATION

Dartmouth College   Hanover, NH
MS in Quantitative Biomedical Sciences, Health Data Science Track; research advisor Prof. Levy   08/2021 – 12/2022
Xiamen University   Xiamen, China
Bachelor of Economics in Mathematical Economics; thesis advisor Prof. Yue   09/2017 – 05/2021
Bachelor of Arts in German Language and Literature; thesis advisor Prof. Mao   09/2017 – 05/2021

Honors: Dartmouth QBS MS Scholarship; Wang Yanan Institute for Studies in Economics Dean's List of All Semesters; Wang Yanan Institute for Studies in Economics Top Students Scholarship

PUBLICATIONS[1]

1. M. Fatemi*, **Y. Lu***, C. Sharma, E. Feng, Z. Azher, A. Diallo, G. Srinivasan, G. Rosner, K. Pointer, B. Christensen, L. Salas, et al. (2023) Feasibility of inferring spatial transcriptomics from single-cell histological patterns for studying colon cancer tumor heterogeneity. *(submitted to Genome Biology, under review)*
2. **Y. Lu**, G. Srinivasan, S. Preum, J. Pettus, M. Davis, J. Greenburg, L. Vaickus, and J. Levy. (2023) Assessing the Impact of Pretraining Domain Relevance on Large Language Models Across Various Pathology Reporting Tasks. *(submitted to Nature Scientific Reports, under review)*
3. **Y. Lu***, R. Hamilton*, J. Greenberg, G. Srinivasan, P. Shah, S. Preum, J. Pettus, L. Vaickus and J. Levy. (2023) Dendrite: A structured, accessible, and queryable pathology search database for streamlined experiment planning. *(submitted to Journal of Pathology Informatics, under review)*
4. M. Fatemi*, **Y. Lu***, A. Diallo, G. Srinivasan, Z. Azher, B. Christensen, L. Salas, G. Tsongalis, S. Palisoul, L. Perreard, F. Kolling IV, L. Vaickus, J. Levy. (2023) The overlooked role of specimen preparation in bolstering deep learning-enhanced spatial transcriptomics workflows. *(submitted to Briefings in Bioinformatics, under review)*
5. Z. Zhang, **Y. Lu**, S. Vosoughi, J. J. Levy, B. C. Christensen, and L. A. Salas. (2023) HiTAIC: hierarchical tumor artificial intelligence classifier traces tissue of origin and tumor type in primary and metastasized tumors using dna methylation. NAR cancer, 5(2):zcad017, 2023
6. J. Greenburg, **Y. Lu**, S. Lu, U. Kamau, R. Hamilton, J. Pettus, S. Preum, L. Vaickus, and J. Levy. (2023) Development of an interactive web dashboard to facilitate the reexamination of pathology reports for instances of underbilling of CPT codes. Journal of Pathology Informatics, page 100187, 2023
7. Z. Azher, M. Fatemi, **Y. Lu**, G. Srinivasan, A. Diallo, B. Christensen, L. Salas, F. Kolling IV, L. Perreard, S. Palisoul, et al. (2023) Spatial omics driven crossmodal pretraining applied to graph-based deep learning for cancer pathology analysis. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 2024
8. J. Levy, **Y. Lu**, M. Montivero, O. Ramwala, J. McFadden, C. Miles, A. G. Diamond, R. Reddy, R. Reddy, T. Hudson, et al. (2023) Artificial intelligence, bioinformatics, and pathology: Emerging trends part i—an introduction to machine learning technologies. Advances in Molecular Pathology, 2023
9. J. Levy, **Y. Lu**, M. Montivero, O. Ramwala, J. McFadden, C. Miles, A. G. Diamond, R. Reddy, R. Reddy, T. Hudson, et al. (2023) Artificial intelligence, bioinformatics, and pathology: Emerging trends part ii—-current applications in anatomic and molecular pathology. Advances in Molecular Pathology, 2023
10. G. Srinivasan, M. Davis, M. LeBoeuf, M. Fatemi, Z. Azher, **Y. Lu**, A. Diallo, M. Montivero, F. Kolling IV, L. Perrard, et al. Potential to enhance large scale molecular assessments of skin photoaging through virtual inference of spatial transcriptomics from routine staining. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 2024
11. A. Suvarna, R. Vempati, R. Chacko, G. Srinivasan, **Y. Lu**, B. Hunt, V. Torres, K. Samkoe, M. Davis, L. Fu, et al. (2023) DeltaAI: Semi-autonomous tissue grossing measurements and recommendations using neural radiance fields for rapid, complete intraoperative histological assessment of tumor margins. bioRxiv (submitted to Nature Scientific Reports, now under review), pages 2023–08, 2023

POSTERS

1. **Y. Lu**, M. Fatemi, R. Reddy, R. Reddy, C. Sharma, E. Feng, Z. Azher, F. W. Kolling, B. C. Christensen, L. A. Salas, R. E. Barney, S. M. Palisoul, G. J. Tsongalis, L. J. Vaickus, and J. J. Levy. VirtualRNA and VirtualProtein: Virtual staining tools to infer spatial molecular information from h&e wsi. In Computational Quantitative Biology Annual Meeting Poster Session 2023

---

[1] ***denotes equal contribution**

2. M. Chan, S. Han, **Y. Lu**, F. Kolling, R. Steiner, T. Punshon, B. Jackson, L. Vadhat, and J. J. Levy. Cell-type specific spatial molecular, elemental and histological signatures associated with tumor metastasis. In Computational Quantitative Biology Annual Meeting Poster Session 2023

3. Z. Azher, A. Suvarna, M. Fatemi, J.-Q. Chen, Z. Zhang, B. Christensen, L. Salas, **Y. Lu**, L. Vaickus, and J. Levy. Interpretable multimodal and crossmodal deep learning for improved cancer pathology analysis. In Computational Quantitative Biology Annual Meeting Poster Session 2023

GUEST LECTURES, TALKS, TEACHING

Guest Lecture
- **Dartmouth QBS177 Methods for Statistical Learning for Big Data**: Introduction to Neural Networks – 2022

Talks
- Dartmouth Computational & Quantitative Biology Annual Meeting: VirtualRNA and VirtualProtein: Virtual Staining Tools to Infer Spatial Molecular Information From H&E WSI –2023
- Dartmouth Computational & Quantitative Biology Annual Meeting: Cell-Type Specific Spatial Molecular, Elemental and Histological Signatures Associated with Tumor Metastasis –2023
- Dartmouth Computational & Quantitative Biology Annual Meeting: Interpretable Multimodal and Crossmodal Deep Learning for Improved Cancer Pathology Analysis –2023
- EDIT National Summer Research Program: Cell Type Specific Spatial Molecular, Elemental, Histological Signatures Associated with Tumor Metastasis – 2023
- EDIT National Summer Research Program: Deep Learning, Image Processing and Natural Language Processing – 2023
- EDIT National Summer Research Program: Introduction to Natural Language Processing – 2022

Teaching
- **Dartmouth QBS181 Data Wrangling**: Teaching Assistant – 2022
- Mentored NLP group of National Summer Research Program

RESEARCH EXPERIENCE

Dartmouth Health, Department of Pathology and Epidemiology   Lebanon, NH
Projects[2]: Cell Type Specific Spatial Molecular, Elemental, Histological Signatures Associated with Tumor Metastasis
- Utilized Laser Ablation Inductively Coupled Plasma Time of Flight Mass Spectroscopy (LA-ICPTOF-MS) for high-resolution elemental imaging and 10X Genomics Visium Spatial Transcriptomics assay for high multiplexing at high spatial resolution.
- Developed web application tool to manage the database, annotate WSI and integrate co-register, statistical and machine learning methods to study the tumor immune microenvironment and metastatic cancer using molecular and metallic biomarkers.
- Applied machine learning techniques, including MaskRCNN for cell detection, non-linear co-registration, developed unsupervised models including VAE to analyze gene-metal-cell-type associations and develop spatially-aware factor models for comprehensive profiling.

Projects[3]: Spatial Transcriptomics, Tumor Immune Microenvironment, GNN, Deep Learning
- Developed a novel cell graph neural network algorithm for inferring spatial mRNA expression patterns from whole slide images by integrating single-cell histological and transcriptomic data.
- Outperformed traditional patch-based computer vision methods, enabling effective prediction of gene expression patterns and revealing significant relevant cellular markers.
- Investigated the impact of improved specimen processing workflow on deep learning-based spatial transcriptomics assessment. Tested that Autostaining workflow considerably outperformed traditional methods.
- Spatial transcriptomics technologies enhance understanding of cellular and transcriptional heterogeneity and offer objective evaluation of photoaging and potential for skin cancer prevention.
- Potential to enhance cancer diagnosis, prognosis, and treatment strategies by unraveling the complex interplay between gene expression, cellular architecture, and spatial organization within tumors.

---

[2] On going projects, Poster 2
[3] Publication 1, 4, 10, Poster 1

Projects[4]: Pathology Reports, Computational Solutions, Web Application Tool, Natural Language Processing (NLP)
- Conducted in-depth assessment of NLP models in pathology-related applications, including CPT code classification, pathologist classification, report sign-out time regression, and text generation.
- Investigated the impact of pretraining strategies on the performance of deep learning-based NLP models, highlighting the importance of task-specific considerations.
- Developed Dendrite, a web application that transforms free-form pathology reports into structured formats, enhancing accessibility for research and clinical applications.
- Demonstrated Dendrite's efficiency in data organization, record linkage, text searches, and structured information filtering, providing valuable support for pathology research and quality improvement efforts.
- Designed a dashboard to accelerate the validation of misassigned Current Procedural Terminology (CPT) codes in pathology reports, leveraging machine learning algorithms and user-friendly features.
- Collaborated with pathologists to utilize the validation dashboard, improving the accuracy of CPT code assignments and contributing to revenue enhancement efforts in the pathology department.

Project[5]: DNA Methylation, Tracing Primary and Metastasized Tumors, Machine Learning, Multilayer Perceptron
- Assisted to develop HiTAIC, a novel DNA methylation-based algorithm with traditional machine learning methods and Multilayer Perceptron
- Achieved 99% accuracy in tracing primary cancer origin and 96% accuracy in identifying metastatic cancers
- Developed a user-friendly web-based application with potential clinical applications for identifying cancers of unknown origin.

Project[6]: Spatial Omics, Multimodal, GNN, Deep Learning
- Utilized spatial transcriptomics data to enhance graph-based deep learning in cancer histopathology analysis.
- Assisted to Implement a contrastive crossmodal pretraining model to extract molecular and histological information.
- Improved performance in various histopathology tasks, such as cancer staging, lymph node metastasis prediction, survival prediction, and tissue clustering analyses, showcasing the potential of leveraging spatial omics data in pathology workflows.

Project[7]: Mohs Micrographic Surgery (MMS), Neural Radiance Fields (NeRF)
- Developed DeltaAI workflow using NeRF for rapid 3D tissue modeling in Mohs Micrographic Surgery (MMS).
- Achieved more accurate and complete 360-degree, 3D tissue renderings with faster processing times compared to previous models.
- Potential to enhance MMS specimen processing, automate health record data export, and improve cancer excision procedures while promoting AI adoption in clinical settings.

Dartmouth College, Department of Computer Science   Hanover, NH
- Developed workflow to collect social media text data from Reddit using PRAW and PushShift APIs
- Participated in annotation of reddit drug related text data
- Advised undergraduate senior thesis psychological text analysis project, implemented NLP topic models including BERTopic, employed NLP models interpretation and validation

WORK EXPERIENCE

Dartmouth Health   Hanover, NH
Full time, Clinical Research Coordinator   *12/2022 – present*
- Led multiple research projects, demonstrating strong project management and research skills.
- Collaborated effectively with co-authors, fostering strong communication and teamwork, ensuring the successful execution of research projects and timely paper submissions.
- Conducted experiments and gathered data, showcasing a hands-on approach to research and a high level of self-reliance in achieving project objectives.

Deep Wisdom   Xiamen, China
Intern, Research Scientist   *10/2020 – 01/2021*

---

[4] Publication 2, 3, 6
[5] Publication 5
[6] Publication 7, Poster 3
[7] Publication 11

- Designed a new feature search framework using Monte Carlo tree search and genetic algorithms, significantly reducing search complexity.
- Managed dynamic price adjustment for company Rookie using algorithms and Bayesian optimization, enhancing collaboration with the company.
- Organized tutorial reference documents for the aforementioned statistical methods used.
- Independently explored machine learning framework: Autoseries, evaluating it for stock time series forecasting against mainstream models.

Youdao Dictionary Overseas Edition Department, Netease   Guangzhou, China
Intern, Data Scientist   *06/2020 – 10/2020*
- Developed automated data report production, reducing time from 4-5 hours to 5 minutes, enhancing work efficiency.
- Analyzed user engagement with articles, predicted clicks by machine learning, validating findings with A/B testing.
- Increased article clicks by 30% using predictive models.
- Conducted NLP-based sentiment analysis, calculated text semantic similarity, and optimized product updates based on user feedback.

Business-to-Business (B2B) Data Analysis Intern, Baidu   Beijing, China
Intern, Data Analyst   *02/2020 – 05/2020*
- Led process design and iterative development of B2B product search, as research, data analysis, and scheme design.
- Developed Python-based selenium crawler scripts, collected 100k high-quality data points.
- Analyzed data with Python, contributing to product insights and collaborating on project operations and reporting.

LEADERSHIP AND COMMUNITY ACTIVITIES

Teaching and Turtle Conservation   Galle, Sri Lanka
Leader of the Teaching Team and Member of the Turtle Conservation Team   *08/2018 – 09/2018*
- Organized team members to bring small English games to kindergartens.
- Provided guidance and supervision to students in kindergarten and primary school settings over four weeks.
- Played a vital role in the transportation of sea turtles to the Sea Turtle Conservation Center, actively participating in this mission twice a week.

Teaching Department, Xiamen University   Xiamen, China
Head of the Teaching Department and German Teacher   *09/2017 – 08/2018*
- Organized university students' innovation and entrepreneurship project.
- Led Xiamen University's language teaching program, teaching basic grammar, vocabulary, and national culture in multiple languages.
- Led a team of over 30 people, planned the teaching content and the use of classroom locations for the language program, which had more than 300 participants.
- Responsible for the vocabulary and cultural components of German teaching, completing a 12-class teaching plan in one semester of volunteer teaching activities.

Student Union, Xiamen University   Xiamen, China
Head of Sports Department   *08/2018 – 09/2018*
- Assisted in hosting school sports games, "Super Cup" school basketball league, school football league and many other school level events.
- Promoted the training of students in the department's history one month before the first sports meeting, planned to execute the pre-match training activities of nearly 80 people in the undergraduate and graduate schools, found and planed the training content and the work plan of the responsible personnel, and finally obtained the school-level sports meeting foreign language college Historical best award.
- Promoted the organization of large-scale school-level activities led by the student union of the college, e.g. "The God of Achievement", the competition attracted more than 150 teams, making this event a traditional college activity.

SKILLS

Python • R • Shell • High Performance Computing • Machine Learning • Dataviz • PyTorch • NLP • Sklearn • Plotly • SQL • Deep Learning • Docker • AWS • LaTeX • Nextflow • Computer Vision • Graph Neural Networks