

AAKASH TRIPATHI, Ph.D.

Machine Learning Engineer | Multimodal AI for Oncology

Aakash.Tripathi0304@gmail.com | (404) 234-8483

tinyurl.com/3xdd5h5n | orcid.org/0000-0001-7231-0487 | linkedin.com/in/aakash-tripathi | aakash-tripathi.github.io

PROFESSIONAL SUMMARY

Machine Learning Engineer at the H. Lee Moffitt Cancer Center and Research Institute and Ph.D. in Electrical Engineering specializing in multimodal foundation models for oncology. Creator of open-source platforms HONeYBEE and MINDS for unified cancer-data embeddings and large-scale data aggregation, with first-author work published in *npj Digital Medicine* (IF 15.1) and *Laboratory Investigation*. Experienced in large-scale model training and inference, digital pathology, multi-omics integration, and multi-agent clinical NLP systems deployed at institution scale.

AREAS OF EXPERTISE & RESEARCH FOCUS

- **Multimodal Foundation Models for Digital Biology (HONeYBEE & MINDS):** Lead developer of open-source multimodal AI platforms — HONeYBEE for unified cancer-data embeddings and MINDS for scalable data aggregation — plus a HuggingFace dataset with 3,500+ downloads enabling community research.
- **Large-Scale Training & Inference for Biomedical AI:** Adapting Transformers and Graph Neural Networks to process thousands of cancer cases, accelerating drug-discovery pipelines and identifying novel therapeutic targets across multi-omics data at scale.
- **Self-Normalizing Multi-Omics Platform for Precision Medicine (SeNMo):** Built a deep-learning system reaching 85% accuracy in survival prediction and treatment response by integrating incompatible molecular data formats for personalized treatment recommendations.
- **AI-Powered Medical Image Analysis & Digital Pathology:** Building automated gigapixel pathology and radiology analysis with 95% diagnostic accuracy, reducing pathologist workload by 60% while maintaining clinical-grade performance across cancer types.
- **Clinical NLP with Multi-Agent AI Systems (CLEVER):** Developing a consensus-based multi-agent system reaching 92% accuracy in extracting structured data from unstructured pathology reports, transforming clinical text into research-ready databases.

PROFESSIONAL EXPERIENCE

H. Lee Moffitt Cancer Center & Research Institute — Tampa, FL, USA

Sept 2025 – Present

Machine Learning Engineer I, Department of Machine Learning

- Published HONeYBEE in Springer Nature *npj Digital Medicine* (IF 15.1); developed the open-source multimodal processing framework with 40+ GitHub stars and 400+ PyPI downloads.
- Released a multimodal dataset on HuggingFace built with HoneyBee, reaching 3,500+ downloads and 11 likes.
- Developed CLEVER, an agentic document-extraction tool providing grounded, auditable extraction of clinical variables, patient timelines, and ICD billing codes; deploying it at institution scale for medical-record processing and document quality control.
- Developing a multi-agent system for automated extraction of neuro-oncology biomarkers from pathology reports.

University of South Florida — Tampa, FL, USA

Sept 2022 – Aug 2025

Graduate Research Assistant, Department of Machine Learning — Supervisors: Dr. Yasin Yilmaz & Dr. Ghulam Rasool

- Developed the open-source data-aggregation solution MINDS, with 11+ GitHub stars and 2,000+ PyPI downloads.
- Published in *npj Digital Medicine*, *International Journal of Molecular Sciences*, *Frontiers in Artificial Intelligence*, *MDPI Sensors*, and *Cancer Informatics*.

EDUCATION

University of South Florida — Tampa, FL, USA

Aug 2022 – Aug 2025

Ph.D., Electrical Engineering — Supervisors: Dr. Yasin Yilmaz & Dr. Ghulam Rasool

Dissertation: *Embedding-Based Deep Learning Frameworks for Multimodal Oncology Data Integration*.

Rowan University — Glassboro, NJ, USA

Sep 2018 – May 2022

B.S., Electrical and Computer Engineering

- Published in IEEE International Symposium on Circuits and Systems (ISCAS) and *Circuits, Systems, and Signal Processing*.
- Filed provisional patents on sEMG sensors for AI/VR controls.

PEER-REVIEWED PUBLICATIONS

1. **Tripathi, A.**, Waqas, A., Schabath, M. B., Yilmaz, Y., & Rasool, G. (2025). HONeYBEE: Enabling scalable multimodal AI in oncology through foundation model-driven embeddings. *npj Digital Medicine*, 8(1), 622. DOI: [10.1038/s41746-025-02003-4](https://doi.org/10.1038/s41746-025-02003-4)
2. **Tripathi, A.**, Waqas, A., Venkatesan, K., Ullah, E., Khan, A., Khalil, F., Chen, W.-S., Ozturk, Z. G., Saeed-Vafa, D., Bui, M. M., Schabath, M. B., & Rasool, G. (2026). Using consensus-based reasoning and large language models to extract structured data from surgical pathology reports. *Laboratory Investigation*, 106(2), 104272. DOI: [10.1016/j.labinv.2025.104272](https://doi.org/10.1016/j.labinv.2025.104272)
3. Flack, D., **Tripathi, A.**, Waqas, A., Rasool, G., & Dera, D. (2025). Robust multimodal fusion for survival prediction in cancer patients. *Cancer Informatics*, 24, 11769351251376192. DOI: [10.1177/11769351251376192](https://doi.org/10.1177/11769351251376192)
4. Waqas, A., **Tripathi, A.**, Ahmed, S., Mukund, A., Farooq, H., Johnson, J. O., ... & Rasool, G. (2025). Self-normalizing multi-omics neural network for pan-cancer prognostication. *International Journal of Molecular Sciences*, 26(15), 7358. DOI: [10.3390/ijms26157358](https://doi.org/10.3390/ijms26157358)
5. Waqas, A., **Tripathi, A.**, Ramachandran, R. P., Stewart, P. A., & Rasool, G. (2024). Multimodal data integration for oncology in the era of deep neural networks: A review. *Frontiers in Artificial Intelligence*, 7, 1408843. DOI: [10.3389/frai.2024.1408843](https://doi.org/10.3389/frai.2024.1408843)
6. **Tripathi, A.**, Waqas, A., Venkatesan, K., Yilmaz, Y., & Rasool, G. (2024). Building flexible, scalable, and machine learning-ready multimodal oncology datasets. *Sensors*, 24(5), 1634. DOI: [10.3390/s24051634](https://doi.org/10.3390/s24051634)
7. Ahmed, S., Nielsen, I. E., **Tripathi, A.**, Siddiqui, S., Ramachandran, R. P., & Rasool, G. (2023). Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, 42(12), 7433–7466. DOI: [10.1007/s00034-023-02454-8](https://doi.org/10.1007/s00034-023-02454-8)
8. Epifano, J. R., Silvestri, A., Yu, A., Ramachandran, R. P., **Tripathi, A.**, & Rasool, G. (2023). A comparison of feature selection techniques for first-day mortality prediction in the ICU. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1–5). IEEE. DOI: [10.1109/ISCAS46773.2023.10182228](https://doi.org/10.1109/ISCAS46773.2023.10182228)

DISSERTATION

Tripathi, A. G. (2025). *Embedding-Based Deep Learning Frameworks for Multimodal Oncology Data Integration* (Doctoral dissertation, University of South Florida).

PREPRINTS & MANUSCRIPTS UNDER REVIEW

1. **Tripathi, A.**, Nielsen, I. E., Umer, M., Ramachandran, R. P., & Rasool, G. (2025). Explainable AI in genomics: Transcription factor binding site prediction with mixture of experts. *arXiv:2507.09754*. [arXiv:2507.09754](https://arxiv.org/abs/2507.09754)
2. **Tripathi, A.**, Waqas, A., Schabath, M. B., Yilmaz, Y., & Rasool, G. (2025). EAGLE: Efficient alignment of generalized latent embeddings for multimodal survival prediction with interpretable attribution analysis. *arXiv:2506.22446*. [arXiv:2506.22446](https://arxiv.org/abs/2506.22446)
3. Garcia-Fernandez, C., Felipe, L., Shotande, M., Zitu, M., **Tripathi, A.**, Rasool, G., ... & Valdes, G. (2025). Trustworthy AI for medicine: Continuous hallucination detection and elimination with CHECK. *arXiv:2506.11129*. [arXiv:2506.11129](https://arxiv.org/abs/2506.11129)

4. Felipe, L., Garcia, C., El Naqa, I., Shotande, M., **Tripathi, A.**, Rudrapatna, V., ... & Valdes, G. (2025). TheBlueScrubs-v1: A comprehensive curated medical dataset derived from the internet. *arXiv:2504.02874*. [arXiv:2504.02874](https://arxiv.org/abs/2504.02874)
5. Waqas, A., **Tripathi, A.**, Stewart, P., Naeini, M., Schabath, M. B., & Rasool, G. (2024). Embedding-based multimodal learning on pan-squamous cell carcinomas for improved survival outcomes. *arXiv:2406.08521*. [arXiv:2406.08521](https://arxiv.org/abs/2406.08521)

INVITED TALKS, PRESENTATIONS & ABSTRACTS

Invited Talks

1. **Tripathi, A.** (2026). Running local LLMs behind institutional firewalls: A hands-on guide for secure clinical AI. Co-presented Learning Lab, Society for Imaging Informatics in Medicine (SIIM) Annual Meeting 2026, June 12, 2026.
2. **Tripathi, A.** (2026). Patient-Level Multimodal Data Fusion for Cancer Classification, Survival Prediction, and Similarity Retrieval: A Pan-Cancer Framework. 2026 Moffitt Scientific Symposium, Tampa, FL.
3. **Tripathi, A.**, Waqas, A., & Rasool, G. (2025). CLEVER: Multi-agent AI orchestration for temporal-aware clinical variable extraction from unstructured medical records. Dr. Robert Gillies Machine Learning Workshop in Cancer, October 30, 2025.
4. **Tripathi, A.**, & Rasool, G. (2025). HoneyBee: A framework for multimodal oncology datasets and models. NAIRR (National Artificial Intelligence Research Resource) Pilot Inaugural Annual Meeting, February 19–21, 2025, Washington, DC.

Conference Presentations & Abstracts

1. **Tripathi, A. G.**, Waqas, A., Davis, E. W., Permuth, J. B., Farinhas, J., Yilmaz, Y., ... & Rasool, G. (2026). Real-world evaluation of multimodal AI: Foundation model-driven multimodal AI for GBM, NSCLC, and PDAC. *Cancer Research*, 86(7_Supplement), 1251.
2. Waqas, A., **Tripathi, A. G.**, Bowles, K., Miner, B., Islam, J. Y., Coghill, A. E., Jones, A., ... & Rasool, G. (2026). Multi-agent AI orchestration for temporal-aware extraction of social determinants of health from unstructured clinical records in cancer populations. *Cancer Research*, 86(7_Supplement), 26.
3. **Tripathi, A.**, Elzaafarany, O., Mokhtari, S., Vogelbaum, M., Bui, M., & Rasool, G. (2026). 521 Multi-agent system for automated extraction of neuro-oncology biomarkers from pathology reports. *Laboratory Investigation*, 106(3).
4. Koutsoubis, N., **Tripathi, A.**, Waqas, A., Yilmaz, Y., Neill, B., Schabath, M., & Rasool, G. (2026). Multi-modal survival prediction on breast cancer mammograms and analysis of the prognostic ability of the contralateral breast. USF AI+X Symposium, April 17, 2026.
5. Waqas, A., **Tripathi, A.**, Naeini, M., Stewart, P., Schabath, M. B., & Rasool, G. (2025). Using patient embeddings from foundation models for enhanced survival analysis in lung squamous cell carcinoma. *American Journal of Respiratory and Critical Care Medicine*, 211(Abstracts), A3108.
6. Waqas, A., **Tripathi, A.**, Naeini, M., Stewart, P., Schabath, M. B., & Rasool, G. (2025). PARADIGM: An embeddings-based multimodal learning framework with foundation models and graph neural networks. *Cancer Research*, 85(8_Supplement_1), 991.
7. **Tripathi, A. G.**, Waqas, A., Yilmaz, Y., Schabath, M. B., & Rasool, G. (2025). Predicting treatment outcomes using cross-modality correlations in multimodal oncology data. *Cancer Research*, 85(8_Supplement_1), 3641.
8. **Tripathi, A.**, Waqas, A., Venkatesan, K., Ullah, E., Schabath, M. B., Bui, M., & Rasool, G. (2025). AI-driven extraction of key clinical data from pathology reports to enhance cancer registries. United States and Canadian Academy of Pathology (USCAP) 114th Annual Meeting, March 22–27, 2025, Boston, MA.
9. **Tripathi, A.**, Waqas, A., Venkatesan, K., Ullah, E., Bui, M., & Rasool, G. (2025). 1391 AI-driven extraction of key clinical data from pathology reports to enhance cancer registries. *Laboratory Investigation*, 105(3). DOI: [10.1016/j.labinv.2024.103629](https://doi.org/10.1016/j.labinv.2024.103629)
10. **Tripathi, A.**, Waqas, A., Yilmaz, Y., & Rasool, G. (2025). HoneyBee: Developing multimodal AI-ready datasets from public cancer repositories. NVIDIA GTC 2025, March 20, 2025.

11. Rasool, G., **Tripathi, A.**, Waqas, A., Ullah, E., & Bui, M. (2024). Extraction of discrete information from pathology reports using local and private LLMs. Oral presentation, Digital Pathology Association, Pathology Visions 2024.
12. Waqas, A., **Tripathi, A.**, Mukund, A., Stewart, P., Naeini, M., & Rasool, G. (2024). BIO24-031: Hierarchical multimodal learning on pan-squamous cell carcinomas for improved survival outcomes. *Journal of the National Comprehensive Cancer Network*, 22(2.5). DOI: [10.6004/jnccn.2023.7137](https://doi.org/10.6004/jnccn.2023.7137)
13. Waqas, A., **Tripathi, A.**, Ahmed, S., Mukund, A., Stewart, P., Naeini, M., Farooq, H., & Rasool, G. (2024). SeNMo: A self-normalizing deep learning model for enhanced multi-omics data analysis in oncology. *American Association for Cancer Research*, 84(6_Supplement), 908. DOI: [10.1158/1538-7445.AM2024-908](https://doi.org/10.1158/1538-7445.AM2024-908)
14. **Tripathi, A.**, Waqas, A., & Rasool, G. (2024). Unifying multimodal data, time series analytics, and contextual medical memory: Introducing MINDS as an oncology-centric cloud-based platform. *Journal of the National Comprehensive Cancer Network*, April 2024. DOI: [10.6004/jnccn.2023.7305](https://doi.org/10.6004/jnccn.2023.7305)
15. **Tripathi, A.**, Waqas, A., Yilmaz, Y., & Rasool, G. (2024). Multimodal transformer model improves survival prediction in lung cancer compared to unimodal approaches. *American Association for Cancer Research*, 84(6_Supplement), 4905. DOI: [10.1158/1538-7445.AM2024-4905](https://doi.org/10.1158/1538-7445.AM2024-4905)
16. Waqas, A., **Tripathi, A.**, Mukund, A., Stewart, P., Naeini, M., & Rasool, G. (2023). Pan-cancer learning for survival prediction. USF AI+X Symposium, September 29, 2023.
17. **Tripathi, A.**, Waqas, A., Yilmaz, Y., & Rasool, G. (2023). Advancing cancer research through integrated multimodal data analysis: A comprehensive framework for precision oncology. USF AI+X Symposium, September 29, 2023.

TEACHING EXPERIENCE

Teaching Assistant — Mayo Clinic AI Summit Workshop July 8, 2025
“Scalable Multimodal AI in Oncology Using HONeYBEE: From Embeddings to Clinical Impact.”

Teaching Assistant — NVIDIA Deep Learning Institute Feb 23, 2024
Building Transformer-Based Natural Language Processing — Public Student Workshop, North America (Virtual).

Teaching Assistant — NVIDIA Deep Learning Institute June 29, 2022
Building Transformer-Based Natural Language Processing — Public Student Workshop, North America (Virtual).

Teaching Assistant — NVIDIA Sponsored Workshop, Rutgers University Feb 22, 2020
Fundamentals of Deep Learning for Computer Vision — Rutgers Business School, NJ, USA.

PEER REVIEW SERVICE

Reviewer (2025) for: Nature Communications; Springer Nature *Scientific Reports*; Springer Nature *npj Digital Medicine*; Springer Nature *Discover Artificial Intelligence*; Springer Nature *Journal of Medical Systems*; Springer Nature *BMC Medical Informatics and Decision Making*; and Springer Nature *Discover Applied Sciences*.

TECHNICAL SKILLS

Programming Languages: Python, MATLAB/Octave, JavaScript & TypeScript, Go

Frameworks & Tools: PyTorch, TensorFlow, AWS, REACT, SQL & NoSQL databases, Docker, Podman, Apptainer, SLURM, HuggingFace, vLLM, Ollama, SGLang, LangChain, LangGraph, DSPy, Deep Agents, NVIDIA RAPIDS, HPC, UNIX shell, Bash scripting

HONORS & AWARDS

- Named on multiple funded and applied research grants from Federal and State agencies.
- Best Poster Award — 2024 Dr. Robert Gillies Machine Learning Workshop in Cancer, for “HONeYBEE: Enabling Scalable Multimodal AI in Oncology Through Foundation Model-Driven Embeddings” (\$1,000 prize).
- 2nd Place — Moffitt Cancer Center Annual 2024 Bio-Data Club Hackathon, for “Patient Data Vectors: An Efficient Cancer Research Framework,” December 12–13, 2024.
- University of South Florida Graduate Assistantship Award, 2022–2025.

- Dean’s List for academic excellence, Sept 2018 – May 2022.

PROFESSIONAL MEMBERSHIPS

- IEEE Eta Kappa Nu (IEEE-HKN) Honor Society — Member since April 2024.
- Institute of Electrical and Electronics Engineers (IEEE) — Member since January 2019.

MENTORING & SUPERVISION

Siddharth Sivaram — Ph.D. student, Electrical Engineering, USF	Aug 2025 – Present
Hanieh Ajami — Ph.D. student, Computer Science, USF	Jan 2025 – Present
Nikolas Koutsoubis — Ph.D. student, Machine Learning, Moffitt	Aug 2024 – Present
Dominic Flack — Graduate student, Imaging Science, RIT	Aug 2024 – Feb 2025
Kavya Venkatesan — Research Intern, Machine Learning, Moffitt	May 2023 – Present
Asim Waqas — Ph.D. student, Machine Learning, Moffitt	Aug 2022 – Aug 2025