

Rahul Thapa

rthapa84@stanford.edu | 736 Serra St., Stanford, CA, 94305 | 682-294-5509

<https://rthapa84.github.io/> | <https://www.linkedin.com/in/rahul-thapa/>

RESEARCH INTEREST

My research interest revolves around training multi-modal foundation models and evaluating them within the realm of healthcare. I am also interested in ensuring these models are not just accurate, but also reliable and safe.

EDUCATION

Stanford University, Stanford, CA, USA

Sept 2023 - May 2028

- PhD in Biomedical Data Science
- *Knight-Hennessy Scholar*, 2023

Villanova University, Villanova, PA, USA

Aug 2017 - May 2021

- Bachelor of Science with Honors in Computer Science; Minors in Mathematics and Physics
- Cumulative GPA: 4.00/ 4.00; Class rank: 1st out of 48 CS students

LANGUAGES AND TECHNOLOGIES

Machine Learning: Python, TensorFlow, PyTorch, R, MATLAB, Jupyter Notebook, PyCharm, PySpark

Web Development: JavaScript, Node.js, React, Flask, HTML, CSS

Others: Java, Sphinx, MongoDB, PostgreSQL, AWS, GCP, Databricks, Kubernetes

INDUSTRY EXPERIENCE

Data Scientist & ML Engineer, Stanford Medicine, Palo Alto, CA

Sept 2022 – Present

- Trained a foundation model using Electronic Health Record (EHR) data, evaluated the model across multiple clinically relevant tasks, and subsequently released the dataset for wider use (EHRSHOT)
- Fine-tuned Large Language Models (LLMs), created diverse NLP evaluation datasets, and generated comprehensive evaluation metrics
- Engineered a prototype leveraging LLM to address clinician queries within Ear, Nose, and Throat (ENT) referrals, enhancing information retrieval efficiency
- Evaluated the efficacy of GPT-3.5 and GPT-4 in meeting real-world information demands within healthcare delivery
- Deployed ML models into clinical practice, utilizing Databricks infrastructure, and conducted thorough investigations into their practical utility and effectiveness

Sr. Data Scientist & ML Engineer, Forta, Remote

Feb 2022 – Aug 2022

- Developed machine learning algorithms for autism spectrum disorder (ASD) diagnosis and personalized therapy
- Engineered ML algorithm to predict ADHD from behavioral data
- Completed multiple FDA submissions, patents, and publications

Data Scientist & ML Engineer, Dascena, Remote

Feb 2021 – Feb 2022

- Developing a new deep learning algorithm for predicting sepsis in clinical trial supported by NIH funding
- Redesigned and redeveloped Dascena's flagship sepsis prediction algorithm called InSight. The new model uses new sepsis onset definition and improves InSight's ability to reduce false positives, minimizing alert fatigue at hospitals
- Created a multi-task learning algorithm with recurrent neural network (RNN) that helps predict 13 different outcomes including sepsis, COVID-19, and acute respiratory distress syndrome (ARDS) using electronic health record (EHR) data
- Developed an ML algorithm that accurately predicts short-term fall at nursing home using minimal EHR data, outperforming existing fall prediction algorithms and risk stratification tools such as Morse Fall Scale
- Designed statistical experiments and implemented gradient boosted machine learning (ML) algorithms for predicting acute pancreatitis and non-alcoholic steatohepatitis

RESEARCH EXPERIENCES

Research Assistant, Computer Engineering Department, Villanova, PA

Jan 2020 – May 2021

- Engineered MoleHD framework for drug discovery application using HDC encoding mechanism, outperforming graph/recurrent neural networks while significantly reducing computational cost and memory usage
- Engineered HDXplore framework to perform differential testing on Hyperdimensional Computing (HDC) for image processing applications. Increased the accuracy of HD models from 81% to 89% with a 40% reduction in discrepancies
- Designed and implemented SpamHD framework for automated and memory-efficient text spam detection using HDC, outperforming baseline methods with 30 to 115 times reduction in model size
- Minimized energy consumption on image processing applications by ~80% under a quality constraint of 0-5% by finding an optimal combination of approximate settings using the genetic algorithm on Python

Machine Learning Research Intern (REU), University of California Irvine

Jun 2019 - Aug 2019

- Modeled a recurrent neural network (RNN) to predict signal delay using a sequence-to-sequence based approach with an accuracy of ~80% for a highly skewed and limited dataset

Data Analytics Research Assistant, Villanova Physics Department, Villanova, PA

Jan 2018 – Aug 2018

- Created histogram of relative orientation between the magnetic field and the intensity gradient

PUBLICATIONS

Thapa, R., A. Garikipati, M. Ciobanu, N. P. Singh, E. Browning, J. DeCurzio, G. Barnes, F. A. Dinunno, Q. Mao, and R. Das.

"Machine Learning Differentiation of Autism Spectrum Sub-Classifications." *Journal of Autism and Developmental Disorders* (2023): 1-16.

Wornow, Michael, Yizhe Xu, **Rahul Thapa**, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A. Pfeffer, Jason Fries, and Nigam H. Shah. "The shaky foundations of large language models and foundation models for electronic health records." *npj Digital Medicine* 6, no. 1 (2023): 135.

Wornow, Michael*, **Rahul Thapa***, Ethan Steinberg, Jason Fries, and Nigam Shah. "Ehrshot: An ehr benchmark for few-shot evaluation of foundation models." *arXiv preprint arXiv:2307.02028* (2023). (Accepted to **NeurIPS D&B 2023**)

Rahmani, Keyvan*, **Rahul Thapa***, Peiling Tsou, Satish Casie Chetty, Gina Barnes, Carson Lam, and Chak Foon Tso. "Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction." *International Journal of Medical Informatics* 173 (2023): 104930.

Ma, Dongning*, **Rahul Thapa***, and Xun Jiao. "MoleHD: Efficient Drug Discovery using Brain Inspired Hyperdimensional Computing." In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 390-393. IEEE, 2022.

Lam, Carson*, **Rahul Thapa***, Jenish Maharjan, Keyvan Rahmani, Chak Foon Tso, Navan Preet Singh, Satish Casie Chetty, and Qingqing Mao. "Multitask learning with recurrent neural networks for acute respiratory distress syndrome prediction using only electronic health record data: model development and validation study." *JMIR Medical Informatics* 10, no. 6 (2022): e36202.

Thapa, Rahul, Anurag Garikipati, Sepideh Shokouhi, Myrna Hurtado, Gina Barnes, Jana Hoffman, Jacob Calvert, Lynne Katzmann, Qingqing Mao, and Ritankar Das. "Predicting falls in long-term care facilities: machine learning study." *JMIR aging* 5, no. 2 (2022): e35373.

Ghandian, Sina*, **Rahul Thapa***, Anurag Garikipati, Gina Barnes, Abigail Green-Saxena, Jacob Calvert, Qingqing Mao, and Ritankar Das. "Machine learning to predict progression of non-alcoholic fatty liver to non-alcoholic steatohepatitis or fibrosis." *JGH Open* 6, no. 3 (2022): 196-204.

Thapa, Rahul, Zohora Iqbal, Anurag Garikipati, Anna Siefkas, Jana Hoffman, Qingqing Mao, and Ritankar Das. "Early prediction of severe acute pancreatitis using machine learning." *Pancreatology* 22, no. 1 (2022): 43-50.

Thapa, Rahul, Bikal Lamichhane, Dongning Ma, and Xun Jiao. "Spamhd: Memory-efficient text spam detection using brain inspired hyperdimensional computing." In *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pp. 84-89. IEEE, 2021.

Thapa, Rahul, Dongning Ma, and Xun Jiao. "Hdexplore: Automated blackbox testing of brain-inspired hyperdimensional computing." In *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pp. 90-95. IEEE, 2021.

Ma, Dongning*, **Rahul Thapa***, Xingjian Wang, Xun Jiao, and Cong Hao. "Workload-aware approximate computing configuration." In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 920-925. IEEE, 2021.

PREPRINTS

Fleming, Scott L., Alejandro Lozano, William J. Haberkorn, Jenelle A. Jindal, Eduardo P. Reis, **Rahul Thapa**, Louis Blankemeier et al. "Medalign: A clinician-generated dataset for instruction following with electronic medical records." *arXiv preprint arXiv:2308.14089* (2023).

Dash, Debadutta, **Rahul Thapa**, Juan M. Banda, Akshay Swaminathan, Morgan Cheatham, Mehr Kashyap, Nikesh Kotecha et al. "Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery." *arXiv preprint arXiv:2304.13714* (2023).

Maharjan, Jenish*, **Rahul Thapa***, Jacob Calvert, Misty M. Attwood, Sepideh Shokouhi, Satish Casie Chetty, Zohora Iqbal et al. "A New Standard for Sepsis Prediction Algorithms: Using Time-Dependent Analysis for Earlier Clinically Relevant Alerts." *Available at SSRN 4130480* (2022).

LEADERSHIP & OUTREACH

Developer Student Club (DSC); Lead, Google Developers **Apr 2020 – May 2021**

- Train peers on Google technologies such as Google Cloud Platform, TensorFlow, and Firebase.
- Organized *Fireside chat with Dr. John Hennessy*, the former president of Stanford: <https://bit.ly/2G9OUYj>

Veritas: Villanova Research Journal; Peer Reviewer, Center for Research and Fellowships **Jun 2020 – July 2020**

- Reviewed computer science manuscripts for the first peer-reviewed undergraduate research journal at Villanova

Resident Assistant, Villanova University Residency Life **Aug 2019 – May 2020**

- Organized community-building events, promoted diversity and inclusion

HONORS & AWARDS

Knight-Hennessy Scholar **May 2023**

- Accepted as 2023 Knight-Hennessy Scholar; total funding of ~300k over 3 years

Upsilon Pi Epsilon Honors Society **Mar 2021**

- Inducted into a selective honor society based on outstanding intellectual accomplishments in computer science

Phi Beta Kappa Honors Society **Mar 2021**

- Inducted into a selective honor society based on outstanding intellectual accomplishments and excellence in research

Villanova Innovation, Creativity, and Entrepreneurship (ICE) Incubator Award **Sept 2020**

- Awarded \$3500 in funding, education, and mentorship for VuShares startup project

Sigma Pi Sigma Honors Society **Nov 2019**

- Inducted into a selective honor society based on outstanding intellectual accomplishments in physics