RAZI MAHMOOD

razi_mahmood@berkeley.edu | (408) 540-4031 | **https://razi-mahmood.github.io/** | https://www.linkedin.com/in/razi-mahmood/ | US Citizen

Research-oriented data scientist with problem-solving and critical thinking skills demonstrated through publications. Passion for applying Artificial Intelligence (AI), Deep Learning, Machine Learning, Natural Language Processing (NLP), Computer Vision techniques to real-world use cases that impact societal issues in health and other areas. Committed to developing innovative techniques and solutions for large-scale problems.

EDUCATION

Rensselaer Polytechnic Institute (Ph.D), Biomedical Engineering ***Aug 2023 – present***
1st year graduate student in Deep Image Analysis Lab (DIAL) under Prof. Pingkun Yan.
Relevant Coursework: Medical Imaging, Applied Data Science

University of California, Berkeley, B.A. in Data Science (Domain Emphasis: Cognition), 3.56 Cumulative GPA Graduated: ***May 2022***
Relevant Coursework: Intro to AI (CS188), Data Science & CS Principles (Data 8/100, CS61A/B/C), Data Structures, Cognitive Science, Discrete Math & Probability, Natural Language Processing, Deep Learning, Computer Architecture, Data Inference & Decision Making, AI Ethics

PUBLISHED RESEARCH EXPERIENCE

Fact-checking for Generative AI, RPI Pre-Graduate Research **May 2023-Sept. 2023**

- Pre-graduate school independent research on image-driven fact-checking of AI-generated textual reports for chest X-ray imaging under the guidance of Prof. Pingkun Yan from RPI and radiologist Dr. Mannudeep Kalra from Harvard/MGH.
- Developed a new fact-checking model (AUC=0.87) by training on pairs of real and fake report sentences with imaging to correct AI-generated reports using CLIP encodings transformed to a higher-dimensional space for classification. Showed 15% quality improvement over AI reports.
- Research paper accepted for oral presentation at Machine Learning in Medical Imaging (MLMI) Workshop at MICCAI, 2023.

Machine Learning Research Intern – IBM Almaden Research ***Sept. 2021 – April 2022, Jun. 2014 – Jul. 2016***

- Co-developed with another summer intern, a **novel architectural change to deep convolutional neural networks** to preserve spatial locality. It was applied to recognize spatially localized anomalies in chest radiographs achieving 10% improvement in AUC on 57 findings in the MIMIC dataset of chest radiographs. This work on spatially-preserving flattening was published in IEEE ISBI Conference in 2022. A patent was also filed by IBM.
- Mentored by IBM researchers also between 2014-2016 during high school on several medical imaging AI research projects. Contributed to the development of statistical machine learning algorithms for detection of cardiac aneurysms and dilated cardiomyopathy in echocardiography.
- Filed 2 patent disclosures, and 2 additional publications in ***international conferences*** (AMIA'14 PMID: 25954393, IEEE ISBI'15) at age 14.

Machine Learning Research Intern – IBM Watson (https://www.ibm.com/watson) ***Jan-May 2021***

- ***Developed a novel log anomaly classification algorithm*** combining BERT language modeling of IT logs with supervised contrastive learning working with the IBM Watson AI Ops team in 2021 under the mentorship of Dr. Rama Akkiraju and Xiaotong Liu. Paper published in IEEE Big Data Workshop on Knowledge Discovery in Data Mining on IT Operations, as part of IEEE Big Data Conference, Dec. 2022.

WORK EXPERIENCE

Data & Analytics (DNA) Associate Engineer – Infosys ***Aug 2022 – April 2023***

- Underwent training courses in Apache Spark, MapReduce, SQL, Database Management, Unix, Hadoop, Software Fundamentals.
- Worked on capstone assignments in Spark, Hadoop, MongoDB software, experimented with Scala programming environment.
- Received hands-on experience with Big Data sample projects, new tools to handle large HDFS datasets.

Data Science Summer Intern – Hyperfine (https://hyperfine.io/) ***Aug 2021 – Nov. 2021***

- Developed a novel ***data labeling algorithm*** based on natural language data analysis (NLP) of associated radiology text reports and recognition of disease-specific concepts. Assembled vocabulary for brain diseases by data mining of textual reports using sentence parsing and phrasal grouping algorithms using NLTK and Spacy libraries in Python. Enabled rapid machine learning model build cycles to accelerate product development. Nearly 7200 annotations were extracted from 600 brain MRI reports achieving 88% precision and 70% recall.
- Developed an ease-of-use interface in Jupyter Notebook to reliably record ground truth anomaly labels indicated by clinicians in companion MRI reports. It led to increase in productivity impact with a **10.0** fold decrease in annotation time. Wrote Python scripts to organize data.

Data Science Summer Intern – Xoran Technologies (https://xorantech.com/) ***June 2021 – Aug 2021***

- Wrote *a* 3D ***image segmentation algorithm*** for cone beam CT based on U-Net deep learning model achieving a Dice coefficient of 0.68 for 9 anatomical structures in head and neck. Developed Python code using SimpleITK, Numpy, Python, Keras, and Tensorflow libraries.
- Trained Xoran staff on the use of data labeling tools (ITKSnap) for machine learning.

Academic Development Committee Mentor – Data Science Society @ UC Berkeley ***Aug 2021 - April 2022***

- Mentored 2 student groups on Data Science capstone projects for a research symposium as Data Science Society mentor.
- Organized discussion in mini-lectures, led topical Jupyter notebook walkthroughs in Deepnote.

Content Gathering Research Intern - SWAYD (Mobile Visual Discovery) (https://www.f6s.com/swayd) ***Jan. 2020 – Mar. 2020***

- Worked as an intern in a team of 4 for the startup. Implemented ***deep learning classifier for foods/dishes*** using ImageNet-trained DL models and linking foods to their respective restaurants via hashtags and geo-tags in Instagram posts.
- Obtained hands-on experience of data preparation, cleaning, processing, algorithms development, researching APIs/platforms (Postman, ClarifAI, Google Maps API).

PUBLICATIONS & PATENTS

1. **R. Mahmood**, D. Machado, G. Wang, M. Kalra, P. Yan, "Fact-checking of AI-generated Reports," in Proc. Machine Learning for Medical Imaging (MLMI) Workshop at Medical Imaging and Computer-Assisted Intervention (MICCAI) Conference, Springer Nature, 2023.
2. **R. Mahmood**, X. Liu, A. Xu, R. Akkiraju, "Contrast BERT: Supervised Contrastive Learning of BERT-Encoded IT logs for Anomaly Classification," in Proc. IEEE Big Data Workshop on Knowledge Discovery in Data Mining on IT Operations, Osaka, Japan, Dec. 2022.
3. N. Shrivastava, **R. Mahmood**, T. Syeda-Mahmood, "Spatially-preserving flattening in deep learning for location-aware classification," in Proc. International Symposium on Biomedical Imaging (ISBI), Kolkatta, India, 2022. Patent Filed (2022), Ref: P202200366US01.
4. **R. Mahmood,** T. Syeda-Mahmood, "Automatic detection of dilated cardiomyopathy in cardiac ultrasound videos," in Proc. American Medical Informatics Association (AMIA) Annual Conference, Washington, D.C., Nov., 2014. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419944/
5. **R. Mahmood**, T. Syeda-Mahmood, "Automatic detection of cardiac aneurysms in cardiac ultrasound videos," in Proc. International Symposium on Biomedical Imaging (ISBI), New York, April 2015. https://ieeexplore.ieee.org/document/7164115

CODING PROJECTS

Deep Learning-based Anomaly Detection in IT Logs -IBM Watson AI Ops (https://github.com/Razi-Mahmood/LogP) ***Jan. 2021 – May 2021***

- Developed a novel supervised contrastive deep learning-based classifier for anomaly identification in IT system logs achieving an accuracy of 97.3% on a dataset of 10,000 HDFS system logs. Code used BERT sentence transformer model in PyTorch and Tensorflow/Keras libraries.

Machine learning-driven Contraceptive Use Prediction (https://github.com/Razi-Mahmood/Contraceptive_Project) ***Jan 2020 – Apr. 2020***

- Worked in a three-member team to find optimal predictor variables for the use of contraceptives in a survey dataset gathered for Indonesian women for purposes of family planning rollout measures. Experimented with logistic regression, decision trees, and random forest with PCA on features using Scikit-learn and achieved train-test accuracies of 97% and 58%. Dealt with data pre-processing, cleansing, and formatting.

Cal Hacks 6.0 Collegiate Hackathon: LateNight ***Oct. 2019 – Dec. 2019***

- Developed an app as part of a group project that used neighborhood crime data from local county to develop a safety index for the restaurants in neighborhoods in Berkeley. Involved web scraping, crime record analysis, map visualization. Programmed in Swift and Python.

GitHub Repository Management (https://github.com/Razi-Mahmood/java_programs) ***Oct. 2019 – Dec. 2019***

- Developed a full-fledged GitHub clone in Java that implements functions of Github for repository data management.

Enigma encryption (https://github.com/Razi-Mahmood/java_programs) ***Aug. 2019 – Oct. 2019***

- Built a Java-based simulator for a generalized Enigma machine (used during WWII) for encrypting messages & substitution ciphers.

SKILLS

- Deep Learning, Machine Learning, Data Science, Data Analysis, Statistical methods(A/B Testing, Regression), Data Mining, Data ETL, Data Visualization, Making technical presentations.
- Python, Java, R, data querying, modeling with SQL, MATLAB, Jupyter Notebook, Deepnote, Visual Studio, IntelliJ, Eclipse IDE, Sublime, Google CoLab.
- Pandas, Seaborn, Tensorflow, Keras, Pytorch, Scikit-learn, Numpy, Nltk, Gensim, Spacy, Matplotlib
- ML Models for images and text including U-Net, VGG16, ResNet50, CNN, Word2Vec, SBERT, CLIP

EXTRACURRICULARS & AWARDS

- Data Science Society Mentor at UC Berkeley, Data Science Society.
- Runner up in the 2020 Undergraduate HCE Essay Prize at UC Berkeley for doing independent, original research on social, cultural, policy of ethical issues at the intersection of data science, technology and society.
- Selected for the competitive California Summer School for Math & Science (COSMOS) at UC Santa Cruz for Video Game Cluster.
- Creator, razifilmblog.blogspot.com. Personal film blog, critiquing aspects such as motion capture, CGI, and cinematography in contemporary reviews. Shared film critiques in school magazine. https://elestoque.org/2017/12/11/magazine/print-special-report/film-lover-razi-mahmood/
- Gourmet cook (Indian foods) with self-made recipes.