RAZI MAHMOOD

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

EDUCATION

University of California, Berkeley, B.A. in Data Science (Domain Emphasis: Cognition)   Expected Graduation: *May 2022*

PROFESSIONAL EXPERIENCE

Deep Learning Summer Team Intern – Hyperfine (https://hyperfine.io/)   *Aug 2021 – Sept 2021*
- Developed an *automated labeling algorithm* for brain MRI by analyzing their companion textual reports using language models, NLP, and vocabulary-driven concept extraction. 7200 annotations were extracted from 600 brain MRI reports achieving 88% precision and 70% recall.
- Developed user interfaces to record ground truth anomaly labels indicated by clinicians in companion MRI reports that led to **ten-fold** decrease in annotation time.

Data Science Summer Intern – Xoran Technologies (https://xorantech.com/)   *June 2021 – Aug 2021*
- *Developed a* 3D *anatomical segmentation algorithm* for cone beam CT studies. Reconstructed volumes for 9 anatomical structures in head and neck including eyes, maxillary sinus, sphenoid, etc. using U-net-based deep learning architecture trained on 17 CT volumes achieving a Dice coefficient of 0.68. Work used SimpleITK, Numpy, Python, Keras, and Tensorflow libraries.

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 an algorithm in Python for *automatically classifying foods/dishes* using ImageNet-trained DL models and linking them 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).

Machine Learning Research Intern – IBM Watson, IBM Research   *Jan-May'2021, Jun. 2014 – Jul. 2016*
- *Developed a novel log anomaly classification algorithm* combining BERT language modeling of IT logs with supervised contrastive learning working with mentors in the IBM Watson NLP team in 2021.
- Earlier, from 2014-2016, was mentored by researchers at IBM and Stanford during high school on several medical imaging AI research projects. Contributed to development of algorithms for automatic detection of pulmonary embolism, cardiac aneurysms, and dilated cardiomyopathy in CT and echocardiography.
- Learned use of many pre-deep learning statistical machine learning packages & tools such as ImageJ, coding in Java and Matlab to process medical data in HL7 and DICOM during this experience. Filed a patent disclosure.
- Presented in Synopsis Science Fair, 2014-2016, *published in international conferences* (AMIA'14 PMID: 25954393, IEEE ISBI'15) at age 14.

Academic Development Committee Mentor – Data Science Society @ UC Berkeley   *Aug 2021 - present*
- Mentored 2 student groups on Data Science capstone projects for a research symposium as Data Science Society mentor.
- Facilitated discussion in mini-lectures, curated topical Jupyter notebook walkthroughs in Deepnote; received hands-on experience of mentoring students, expanded practical knowledge of EDA, modeling, visualization, data cleaning, machine learning, hypothesis testing

SELECTED PROJECTS (More Details on: https://razi-mahmood.github.io/)

IBM Watson AI Ops – Deep Learning-based Anomaly Detection in IT Logs   *Jan. 2021 – May 2021*
- Project involved classifying IT system logs consisting of free text phrases into normal and anomalous logs
- Developed a new approach using self-supervised contrastive learning on BERT-encoded log data achieving an accuracy of 97.3% on a 10,000 HDFS system logs dataset. Code used both TensorFlow and PyTorch libraries and DL models built on these platforms.

NeuroTech Project – Deep learning-based EEG Analysis   *Aug 2021 – Dec. 2021*
- The project was to study the trends in the sleep cycle stages across a population using PolySomnoGraphic sleep recordings. I adapted a 1D CNN architecture (8 CNN layers, 1 drop-out layer) to implement a 5-channel EEG signal classifier of 5 sleep stages with an accuracy of 0.76.

Machine learning-driven Contraceptive Use Prediction   *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 97-58%. Dealt with data pre-processing, cleansing, and formatting.

Cal Hacks 6.0 Collegiate Hackathon: LateNight   *Aug 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.

Full-fledged CPU design   *Aug 2020 – Nov. 2020*
- Developed a full-fledged CPU design for processing a full set of RISC-V instructions using Logism in CS61C Computer Architecture course.

Simulation of the Enigma Machine   *Aug 2019 – Dec. 2019*
- Built Java-based simulator for a generalized version of the Enigma machine used during WWII for encrypting messages & substitution ciphers.

SKILLS SUMMARY

- **Fluent:** Python (3+ years), Jupyter Notebook, Deepnote, Pandas, Build machine learning models using libraries and platforms (Tensorflow, Keras, Pytorch, Scikit-learn, Numpy, Nltk, Gensim, Spacy, Matplotlib), Java (2+ years), Environments (Visual Studio, IntelliJ, Eclipse IDE, Sublime, ITKSnap)
- **Data science tasks & models:** Data preparation, processing, cleaning, standardization, analysis and visualization. DL Models (U-Net, VGG16, ResNet50, DenseNet, BERT, Word2Vec, other OpenAI models)
- **Relevant coursework:** Intro to AI (CS188), Data Science & CS Principles (Data 8/100, CS61A/B/C), Cognitive Science, Discrete Math & Prob.

PUBLICATIONS

- 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., November, 2014.
- 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.