Arshitha Basavaraj

Email: arshithab@proton.me

github.com/Arshitha

Experience

- **National Institute of Mental Health (NIMH)** — Bethesda, USA
  Data Engineer — *Sept 2019 — Sept 2024*
  - Implemented data pipelines to enhance reproducibility of analyses in neuroimaging studies.
  - Curated and shared large Findable, Accessible, Reproducible and Interoperable (FAIR) neuroimaging datasets.
  - Conducted training sessions to improve adoption of FAIR data principles by research groups within the NIMH.
  - Built an automatic PDF generator of the Brain Imaging Data Structure (BIDS) specification, a community developed data management standard.
  - Contributor for the BIDS Extension Proposal (BEP) 036 for phenotype data.
- **dataxu** — Boston, USA
  Engineering Intern — *May 2018 – Aug. 2018*
  - Ported legacy data transfer program to an event-trigerred, cleaner and more efficient script.
  - Boosted data transfer efficiency by **95%** while also improving scalability.
- **Indian Statistical Institute** — Bangalore, India
  Research Assistant — *May 2015 – July 2015*
  - Recognized and classified six facial expressions using frontal face images of varied subjects.
  - Improved classification efficiency by **2%** overall with respect to existing literature.
  - Discovered a simple, novel method using a combination of Mathematical Morphological and Image Processing techniques for feature extraction.
  - Published at the 10th International Conference on Industrial and Information Systems (ICIIS).

Education

- **Boston University, College Of Engineering** — Boston, USA
  Master of Science in Electrical & Computer Engineering — *May 2019*
- **National Institute of Technology, Karnataka** — Surathkal, India
  Bachelor of Technology in Electrical & Electronics — *May 2016*

Publications

- Taylor Paul A., Glen Daniel R., Reynolds Richard C., **Basavaraj Arshitha**, Moraczewski Dustin, Etzel Joset A.,Editorial: Demonstrating quality control (QC) procedures in fMRI, Frontiers in Neuroscience, Sec. Brain Imaging Methods, Volume 17, 31 May 2023, doi: 10.3389/fnins.2023.1205928
- Allison C. Nugent, Adam G. Thomas, Margaret Mahoney, Alison Gibbons, Jarrod T. Smith, Antoinette J. Charles, Jacob S. Shaw, Jeffrey D. Stout, Anna M. Namyst, **Arshitha Basavaraj**, et al., (2022). The NIMH intramural healthy volunteer dataset: A comprehensive MEG, MRI, and behavioral resource. Scientific Data, 9(1). doi: 10.1038/s41597-022-01623-9
- A. Apte, **A. Basavaraj** and Nithin R K, Efficient Facial Expression Ecognition and classification system based on morphological processing of frontal face images, 2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS), 2015, pp. 366-371, doi:10.1109/ICIINFS.2015.7399039.

Projects

- **Anatomical Scans Defacing workflow** — National Institute of Mental Health
  Open Science — *Jan 2022 – Jan 2023*
  - Automated the process of de-identifying structural MRI scans by removing facial features.
  - Tested existing defacing programs on two neuroimaging datasets containing over **2000** scans.
  - Integrated visual inspection and rating tool with an existing de-identification program to flag and correct failures efficiently.
  - **Significance:** De-identification of MRI scans is a crucial and high-effort final step before datasets can be shared openly. Automating the process of defacing scans, visual inspection and correcting failures has significantly improved the speed and efficiency of public data sharing.
  - **Project repository**
- **Language Correction** — Boston University
  NLP, Software Development — *Feb 2019 – May 2019*
  - Built an efficient and highly scalable English language checker.
  - Achieved **99.9%** compression efficiency for crawled data.
  - Extended the package to support Arabic, Chinese & German languages.
  - **Project repository**
- **Toxic Comment Classification** — Boston University
  Machine Learning — *Oct 2018 – Dec 2018*
  - Identified and classified toxic comments in the Jigsaw Toxic Comments dataset using classical and deep models.
  - Achieved best performance of **90.62%** with Logistic Regression with word and char n-grams.
  - Evaluated performance variations with different word embeddings such as count vectorizers, tf-idf and word2vec.
  - **Project repository**

Skills

- **Programming Languages:** Python (6 years), MATLAB (4 years), Java (1 year), C/C++ (1 year)
- **Technologies:** git/GitHub, DataLad, Singularity/docker, high performance computing (HPC), AWS
- **Human Languages:** Kannada and English.

Relevant Coursework

- **Machine Learning:** Oxford Machine Learning Summer School, Learning from Data, Introduction to Artificial Intelligence
- **Computer Science:** Data Structures and Algorithms, Building Software, Introduction to Databases
- **Mathematics:** Linear Algebra, Probability theory & its applications
- Neuromatch Academy: Computational Neuroscience