3803 Darwin Dr
Fremont
CA 94555

FAHIM ANJUM

dr.fahim.anjum@gmail.com
sites.google.com/site/mdfahimanjum93
linkedin.com/in/dr-fahim-anjum
github.com/MDFahimAnjum

SUMMARY

Experience	9+ years of experience in designing ML algorithms, statistical models, and efficient features for time-series analysis, forecasting, and classification
Coding Languages	Python, SQL, MatLab, R, Bash Script, Java, C, C++
ML Skills	FFT, Spectrogram, Stochastic models, Transformer, CNN, GRU, LSTM, VAE
Libraries/Services	PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, HuggingFace, Docker, AWS, Jupyter

EDUCATION

Ph.D. in ECE, The University of Iowa Aug 2021

Research: Designing computationally efficient features and ML algorithms for time series analysis and classification [PDF]

Coursework: Convex Optimization, Simulation & Modeling, Pattern Recognition, Advanced Control Theory, High-Performance Computer Architecture, Statistical Modeling

B.Sc. in ECE, The University of Iowa Dec 2014

Coursework: Pattern Recognition, Machine Learning, Signal Processing, Control Theory, Algorithms

EXPERIENCE

University of California San Francisco (UCSF) San Francisco, CA

Postdoctoral Researcher *Sep 2021 - Present*

- **Self-Supervised Learning of Time Series via Language Models:** Designed a novel tokenizer that converts time series data of brain signals into a fixed vocabulary via stochastic modeling and utilized BERT for pre-training and fine-tuning for detecting neurodegenerative diseases tasks.
- **Python library for signal processing:** Developed *TurboLPC*, an efficient Python library for stochastic modeling of time series data for speech and audio analysis, data compression, and feature extraction. TurboLPC is ***1,000× faster*** and comes with an advanced variation of frequency-warping for non-uniform frequency resolution [PyPI] [CODE].
- **Efficient Feature Engineering for Time Series Classifier:** Developed manual features using signal processing (spectral power) from LFP brain activity time series (RCS brain-implant) for computationally efficient sleep stage classification and achieved **over 90% accuracy** with lightweight ML models (SVM), matching DL models (CNN, Transformer, RNN) with automatic feature engineering (***published in Nature Communications***) [PDF].
- **Python library to modify LLM Tokenizer:** Developed *HugTokenCraft*, a Python library to simplify the process of vocabulary modification (add/remove tokens from dictionary) of a pre-trained tokenizer (BertTokenizer) used in BERT [PyPI] [CODE].
- **Data-driven Hidden Pattern Discovery:** Employed unsupervised ML clustering (Multivariate Auto-regressive HMM) on time series data (LFP signals from RCS device) to discover new microstructure of sleep stages, and hidden structures in sleep patterns using network/modulatory analysis.
- **Statistical Evaluation Framework:** Developed complex statistical models and tests (Linear Mixed effects/t-test/MANOVA/Hierarchical Cluster Analysis) for time series data and features (via cross-correlation, spectral coherence, spectrogram, and wavelet analyses) to evaluate the stability of data-driven clustering (microstructure of sleep) and discover neurophysiological changes in brain networks (***published in Nature Communications***) [PDF].
- **Artifact Detection & Removal:** Developed algorithms to detect and rectify abnormal data and noise artifacts in time series data (LFP signals from RCS device) using signal processing tools (template matching, cross-correlation, Spectral power threshold, Kalman Filtering) and unsupervised ML clustering (Multivariate Auto-regressive HMM) [PDF].

- **Data Synchronization among Wearable Devices:** Designed signal processing algorithms for synchronizing time series data from multiple wearable and implanted devices while accounting for missing data, disconnections, mismatch of timestamps, and non-uniform sampling rates [PDF].
- **Data Processing Pipeline for Wearables:** Developed a data collection and processing pipeline for wearable devices for building ML models and conducting statistical analyses for automatic data collection from servers (via RESTful APIs), data quality checks, resampling, artifact rejection, manual data corrections, and pre-processing (filtering).

The University of Iowa Iowa City, IA
Graduate Student Researcher *Jan 2015 - Aug 2021*

- **Novel Feature Extraction Method:** Designed a computationally efficient method to obtain features (via stochastic modeling) and ML algorithm for supervised classification of time series data improving both performance (***+13% accuracy***) and computational cost (***5× faster***) [PDF].
- **Source Estimation from Time Series:** Utilized single/multi-channel sensor time series data to estimate the characteristics of the source for localization (finding radiation source) and classification (detecting neurodegenerative diseases with EEG) tasks [PDF].
- **Motion Segmentation using subspace clustering:** Utilized ML algorithms (Mixture of Probabilistic Principal Component Analysis, k-Subspaces and Sparse Subspace clustering) to recognize moving objects in video [PDF].
- **Cloud-based AI-assisted education platform:** Developed smartphone app (Android) and Server client for providing cloud-based AI-assisted education to students [APP]
- **Deploying Web Service:** Developed Server-side web application (via Ruby on Rails) and complex database (postgres) and deployed using open source PaaS (CapRover) for the AI-assisted education app that supported 10k+ users.

SELECTED PUBLICATIONS & PATENTS

Anjum, M. F., Smyth, C., Zuzuárregui, R., Dijk, D.J., et al., "Multi-night naturalistic cortico-basal recordings reveal mechanisms of NREM slow wave suppression and spontaneous awakenings in Parkinson's disease", *Nature Communications*, 2024. [PDF]

Anjum, M.F., Espinoza, A.I., Cole, R.C. et al., "Resting-state EEG measures cognitive impairment in Parkinson's disease", *Nature npj Parkinson's Disease*, 2024. [PDF]

Yin, Z., Yu, H., Yuan, T., Zhang, N., Smyth, C., **Anjum, M. F.**, et al., "Generalized sleep decoding with basal ganglia signals in multiple movement disorders", *Nature npj Digital Medicine*, 2024. [PDF]

Smyth, C., **Anjum M. F.**, Ravi, S., Denison, T., Starr, P., Little, S., "Adaptive Deep Brain Stimulation for sleep stage targeting in Parkinson's disease", *Brain Stimulation*, 2023. [PDF]

Espinoza, A. I., May P., **Anjum M. F.**, Singh A., et al., "A pilot study of machine learning of resting-state EEG and depression in Parkinson's disease", *Clinical Parkinsonism & Related Disorders*, 2022. [PDF]

Dasgupta, S., **Anjum, M. F.**, Narayanan, N., Mudumbai, R., "Apparatus, systems and methods for diagnosing parkinson's disease from EEG data" *US Patent Application*, 2020. [PDF]

Anjum, M. F., Dasgupta S., Mudumbai R., Singh A., et al., "Linear predictive coding distinguishes spectral EEG features of Parkinson's disease", *Parkinsonism & Related Disorders*, 2020. [PDF]

Anjum, M. F., Haug J., Alberico S., Dasgupta S., Mudumbai R., et al., "Linear Predictive Approaches Separate Field Potentials in Animal Model of Parkinson's Disease", *Frontiers of Neuroscience*, 2020. [PDF]

Uc, E., **Anjum, M.F.**, Dasgupta, S., Narayanan, N., "Resting-state EEG Predicts Cognitive Impairment in Parkinson's Disease (P6-11.015)", *Neurology Apr*, 2023. [PDF]

Anderson, B. D. O., Dasgupta, S., Baidoo-Williams, H. E., **Anjum, M. F.**, Mudumbai, R., "Unique Maximum Likelihood Localization of Nuclear Sources", *58th IEEE Conference on Decision and Control*, 2019. [PDF]