Bay area
San Francisco
CA 94555

FAHIM ANJUM

dr.fahim.anjum@gmail.com
mdfahimanjum.github.io
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 analysis, forecasting, and classification of time series and image
Coding Languages	Python, SQL, MatLab, R, Bash Script, Java, C, C++
ML Skills	Transformer, Diffusion Models, StyleGAN, CNN, GRU, LSTM, VAE, Statistical Model
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*

- **Advancing Diffusion Models via Resampling Techniques from StyleGAN3:** Developed an enhanced diffusion model architecture incorporating alias-free resampling techniques inspired by StyleGAN3, achieving improved rotational equivariance and stable training without adding new trainable parameters, ensuring architectural simplicity and efficiency. [Results] [Code]
- **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. [PDF] [Code]
- **Efficient CNN Architecture for Time Series Classification:** Designed a minimalist CNN using a single convolutional layer for EEG-based Parkinson's disease classification that outperformed complex state-of-the-art architectures (***+2.3% recall***, ***+4% F1-score*** and ***+3.3% accuracy***), while enhancing layer interpretability. [PDF] [Code]
- **Optimized Feature Engineering for Time Series Classifier:** Developed optimized features using signal processing for computationally efficient classification of brain activity time series 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] [Featured in ScienceDaily]
- **Python Library for Accelerated Signal Processing:** Developed *TurboLPC*, an efficient Python library (***1,000× faster***) for stochastic modeling of time series data which has an advanced variation of frequency-warping for non-uniform frequency resolution for analysis, compression, and feature extraction [PyPI] [Code].
- **Python Library to Modify LLM Tokenizer:** Developed *HugTokenCraft*, a Python library to simplify the process of vocabulary modification (add/remove tokens) of a pre-trained BertTokenizer [PyPI] [Code].
- **Data-driven Hidden Pattern Discovery:** Employed unsupervised ML clustering (Multivariate Auto-regressive HMM) on time series data brain activities to discover hidden structures in activity patterns using network/modulatory analysis.
- **Generalized Multi-Class Supervised Models for Time Series:** Created multi-class supervised models for time series data of brain activity (***85% accuracy***) with extensive feature extraction, including 74 features (18 frequency-domain, 10 time-domain and additional coherence features). Evaluated ridge regression, SVM, KNN, decision tree, random forest, XGBoost, LightGBM, and artificial neural network models. [PDF]

- **Real-time Classification of Time Series:** Developed supervised models using Linear Discriminant Analysis for binary classification of real-time incoming time series data of brain activity using frequency-domain features for adaptive brain stimulation. [PDF]
- **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 and discover neurophysiological changes in brain networks (***published in Nature Communications***) [PDF].
- **Automated Artifact Detection & Removal:** Developed algorithms to detect and rectify abnormal data and noise artifacts in time series data of brain activities using signal processing tools (template matching, cross-correlation, Spectral power threshold, Kalman Filtering) and unsupervised clustering (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].

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].
- **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 a smartphone app (Android) and server-side web application for an AI-assisted education platform, supporting ***10,000+ users***. The app, deployed via Ruby on Rails with a complex PostgreSQL database and hosted using open-source PaaS (CapRover).

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., "LiPCoT: Linear Predictive Coding based Tokenizer for Self-supervised Learning of Time Series Data via Language Models", *arXiv e-prints, arXiv:2408.07292*, 2024. [PDF]

Anjum, M. F., "Parkinson's Disease Classification via EEG: All You Need is a Single Convolutional Layer", *arXiv e-prints, arXiv:2408.10457*, 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]

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]