Taha ValizadehAslani, Ph.D.

Taha.Valizadeh@gmail.com | tahaaslani.com | LinkedIn | GitHub | Google Scholar

Summary

Machine Learning Engineer with Ph.D. in Electrical Engineering and experience at Comcast and academia. Expertise in large language models, optimization algorithms, and information theory. Skilled in building scalable ML pipelines, developing domain-specific NLP models, and applying metaheuristic and information-theoretic approaches to network design and data analysis. Passionate about bridging theoretical innovation and practical deployment.

Experience

- **Comcast** Philadelphia, PA
  Software Engineer *Sept 2023 – Present*
  - Developed metaheuristic optimization pipelines (PSO, GA, ACO, DE) for designing Layer-3 fiber network topology, improving resiliency and reducing implementation cost.
  - Applied Monte Carlo simulation to assess network capacity overload risks under different stochastic traffic scenarios.
  - Designed time-series forecasting pipeline (ARIMA, ETS, exponential smoothing, etc.) to predict traffic utilization at national, regional, and site levels.
  - Built Spark/Databricks pipelines to monitor network utilization and ensure data quality.
  - Modeled network link latency estimation as a linear programming optimization problem.

- **Drexel University** Philadelphia, PA
  Graduate Researcher, Machine Learning (Dissertation) *Sept 2018 – Mar 2024*
  - Developed **PharmBERT**, a domain-specific LLM for FDA, surpassing baseline models in classification and extraction tasks.
  - Created **Two-Stage Fine-Tuning**, a novel deep learning training strategy for learning class-imbalanced data.
  - Proposed parameter-efficient fine-tuning of **LayerNorm**, reducing computational complexity by 99%.
  - Predicted antimicrobial resistance using the novel amino-acid k-mer features, achieving performance and interpretability.

Selected Projects

- **PharmBERT (Developed for FDA)** – A pre-trained domain-specific large language model for pharmacological text mining, outperforming SOTA models on different pharmaceutical tasks, such as detecting adverse drug reactions, identifying drug–drug interactions, and classifying drug-label text.
- **Two-Stage Fine-Tuning** – A novel deep learning training strategy for improving the model performance on under-represented classes and generalizability on out-of-distribution samples.
- **Amino-Acid k-mer** – A new feature for predicting antibiotic resistance from the genome. The proposed method outperformed SOTA techniques and provided new insights into resistance mechanisms, including changing our understanding of the *tet(D)* gene.
- **Efficient Fine-Tuning for LLMs** – Proposed a novel LayerNorm-based parameter-efficient tuning strategy for the transformer-based deep learning models, which reduces the computational complexity by 99%.
- **Mathematical Analysis and Improvement of IS-LDPC Code** – Proved four mathematical theorems that explain performance degradation of IS-LDPC codes, then proposed a method that improves the error-control performance based on the theorems.

Technical Skills

Languages: Python, SQL, R, MATLAB, Bash, C, C++, Scala
ML Frameworks: PyTorch, TensorFlow, Keras, XGBoost, Random Forest, SVM, scikit-learn, statsmodels
NLP: Hugging Face Transformers, BERT, GPT, LLaMA, Mistral, Instructor, fast.ai, NLTK
Optimization: Particle Swarm Optimization, Genetic Algorithms, Ant Colony Optimization, Differential Evolution
Theory: Information Theory, Probability Theory, Graph Theory
Data science tools: Spark, Hadoop, Databricks, KNIME, t-SNE, PCA, Tableau, Matplotlib, seaborn, ggplot2
High-performance computing: Slurm Workload Manager, SGE, AWS, Microsoft Azure, Cerebras (CS-2), Google Colab
Bioinformatic packages: Biopython, BLAST, SAMtools, MAFFT, Ray assembly, MetaBAT 2, Bowtie
Miscellaneous: pandas, NumPy, SciPy, Git, Docker, Conda, HTML, Linux, LaTeX, Microsoft Office

Education

- **Drexel University** — Philadelphia, PA
  Ph.D. Electrical Engineering — *Sept 2018 – Mar 2024*
- **Iran University of Science and Technology** — Tehran, Iran
  M.Sc. Electrical Engineering – Communication Systems — *Sept 2014 – Jan 2017*
- **Lorestan University** — Khorramabad, Iran
  B.Sc. Electrical Engineering – Power — *Feb 2009 – May 2013*

Publications

Dissertation

Taha ValizadehAslani. "Using Large Language Models for Analyzing Drug Label Data". Drexel University. 2024.

Journal and Conference Papers

Taha ValizadehAslani, Yiwen Shi, Jing Wang, Ping Ren, Yi Zhang, Meng Hu, Liang Zhao, Hualou Liang. "Two-Stage Fine-Tuning with ChatGPT Data Augmentation for Learning Class-Imbalanced Data". *Neurocomputing*, 127801. 2024.

Taha ValizadehAslani, Hualou Liang. "LayerNorm: A Key Component in Parameter-Efficient Fine-Tuning". arXiv:2403.20284.

Taha ValizadehAslani, Yiwen Shi, Ping Ren, Jing Wang, Yi Zhang, Meng Hu, Liang Zhao, Hualou Liang. "PharmBERT: a domain-specific BERT model for drug labels". *Briefings in Bioinformatics*, 24(4), bbad226. 2023.

Yiwen Shi, Ping Ren, Jing Wang, Biao Han, **Taha ValizadehAslani**, Felix Agbavor, Yi Zhang, Meng Hu, Liang Zhao, Hualou Liang. "Leveraging GPT-4 for Food Effect Summarization to Enhance Product-Specific Guidance Development via Iterative Prompting". *Journal of Biomedical Informatics*, 148, 104533. 2023.

Yiwen Shi, Jing Wang, Ping Ren, **Taha ValizadehAslani**, Yi Zhang, Meng Hu, Hualou Liang. "Fine-Tuning BERT for Automatic ADME Semantic Labeling in FDA Drug Labeling to Enhance Product-Specific Guidance Assessment". *Journal of Biomedical Informatics*, 138, 104285. 2023.

Yiwen Shi, **Taha ValizadehAslani**, Jing Wang, Ping Ren, Yi Zhang, Meng Hu, Liang Zhao, Hualou Liang. "Improving Imbalanced Learning by Pre-Finetuning with Data Augmentation". *Workshop on Learning with Imbalanced Domains*. 2022.

Waleed Iqbal, Elena V. Demidova, Samantha Serrao, **Taha ValizadehAslani**, Gail Rosen, Sanjeevani Arora. "RRM2B Is Frequently Amplified Across Multiple Tumor Types: Implications for DNA Repair, Cellular Survival, and Cancer Therapy". *Frontiers in Genetics*, 12, 628758. 2021.

Taha ValizadehAslani, Zhengqiao Zhao, Bahrad A. Sokhansanj, Gail L. Rosen. "Amino Acid k-mer Feature Extraction for Quantitative AMR Prediction by Machine Learning and Model Interpretation for Biological Insights". *Biology*, 9(11), 365. 2020.

Chiahui Chen, **Taha ValizadehAslani**, Gail Rosen, Carla Jungquist, Laura Anderson. "Healthcare Shift Workers' Temporal Habits for Eating, Sleeping, and Light Exposure: A Multi-Instrument Pilot Study". *Journal of Circadian Rhythms*, 18. 2020.

Taha ValizadehAslani, Abolfazl Falahati. "An Analysis and Improvement of Error Control Performance of IS-LDPC Codes with a Large Number of Subsets". *Physical Communication*, 31, 79–86. 2018.

Academic Peer Review

Reviewed for *PLOS Digital Health*