RAVITEJA VANGARA

LinkedIn | 505-488-1420 | ravitejavangara.com | rvangara.phd@gmail.com | GitHub | Gscholar

Skills

- **Languages**: Python, Bash, Matlab, R, Julia, C, C++ | **Cloud**: AWS, Azure, GCP | **Version Control:** Gitlab, GitHub
- **Frameworks**: PyTorch, TensorFlow, Docker, EC2, scikit-learn, OpenCV, Hugging Face
- **Machine Learning**: Matrix and Tensor Factorization, Dimension reduction techniques, classification and regression, Clustering,
- **Deep Learning**: Natural Language Processing, Computer Vision, Generative AI, GANs, VAEs, Diffusion Models, Transformers, Foundational Models, Large language models (LLM) applications.

Experience

ML Research Scientist — **Moores Cancer Center, UCSD** — *San Diego, CA, USA* — **09/2021 - Current**

- Spearheading the deep learning team in developing several AI models on advanced **histopathology high-resolution (Giga pixels)** imagery and **multiple instance learning CNNs** to identify cancer biomarkers and enhance prognostication accuracy with **Graph Attention Networks**.
- Instrumental in advancing the **SigProfiler Software** suite, analyzing large-scale cancer genomes through open-source repositories.
- Pioneered the development of **SPExtractor**, elevating cancer genome signature extraction accuracy by **20%** and reducing false positives via state-of-the-art pattern recognition, **optimization techniques**, and multi-GPU/CPU parallel processing integration. [Code]
- Developed **SPMatrixGenerator2**, which features **piecewise constant fitting** and integrated CNV/SV schema, eliminating the reliance on R wrappers and streamlining data processing and analysis. [Code].
- Conceived and executed **SPAssignment**, incorporating a custom forward stagewise algorithm and nonnegative least squares (**NNLS**) for enhanced sparse regression and numerical optimization in genomic analysis tools. [Code]

ML Research Scientist — **Los Alamos National Lab** — *Los Alamos, NM, USA* — **04/2018 - 08/2021**

Topic Identification and Evolution:

- Directed an **NLP** research team in scientific leadership identification and characterization within large-text corpora.
- Engineered **SeNMFk**, document clustering accuracy by 15% and achieving a 0.85 silhouette score through semantic-assisted NMF and benchmarked it in nine text corpora -BBC, BBC sport, Reuters, Guardian, and Newsgroups of over 20000 documents.[ICMLA2020]
- Developed a novel method for tracing the progression of scientific themes in superconductivity research over 11 years by developing non-negative tensor factorization models on a 54 x 13900 x 11 [Authors x Words x Time] datasets. This approach uncovered **seven** distinct evolutionary patterns in the subject matter, delineating both the temporal dynamics of topics and the individual contributions of each researcher within the field.

Tensor Networks:

- Automated latent dimensionality prediction in datasets using a Multi-Layer Perceptron integrated **with NMFk**, achieving **a 95%** success rate across **58,660** matrices. [MLST]
- Developed **pyNMFk**,a deconvolution approach for complex integral equations in anomalous diffusion problems, enabling precise identification of multiple diffusion sources based on sparse spatiotemporal data. [Code]
- Developed **SymmNMFk**, identifying correct cluster numbers in non-convex datasets by integrating symmetric NMF, consensus clustering, and PAC criterion, outperforming traditional methods. [Paper,talk]
- Developed the **DADm,** a multimodal fusion model for transcription factor binding prediction, improving classification AUC and robustness by integrating DNA sequence with biophysical properties.
- Optimized decoy selection in protein structure determination with **SNMF-DS**, reducing RMSD losses to below 1Å for 7 out of 17 proteins.[IEEE BIBM]

High Performance Computing:

- Key team member in developing distributed **DNMFk**, the first distributed model determination algorithm for large and dense non-negative large-scale matrices. [Code]
- Key team member in distributed non-negative tensor train decomposition algorithm, scaling to handle **256GB** of data with a maximum compression ratio of 1.47e4 and minimal error. [Paper]

Computational Scientist - Ph.D. — **University of New Mexico** — *Albuquerque, NM, USA* — **12/2015 - 04/2018**

- Enhanced electrochemical modeling precision by developing a shared-memory parallelized Python framework for classical density functional theory, addressing solvent structure and surface charge interactions.
- Gained new insights into electrostatics by studying the role of non-coulombic interactions in electric double layers, enhancing theoretical frameworks for charge regulation.

Education

4.04/4.0	**Ph.D. in Engineering [With Distinction]**, *University of New Mexico | Albuquerque, NM, USA*	*2015-2019*
4.04/4.0	**MS in Chemical Engineering**, *University of New Mexico | Albuquerque, NM, USA*	*2015-2017*
7.33/10.0	**B.Tech in Chemical Engineering**, *National Institute of Technology Warangal | Warangal, India*	*2011-2015*

RAVITEJA VANGARA

LinkedIn | 505-488-1420 | ravitejavangara.com | rvangara.phd@gmail.com | GitHub | Gscholar

Awards

- **R&D 100 Award (Information Technologies), R&D 100 Award Bronze (Market Disruptor)** : Part of winning team of prestigious 2021 R&D 100 Award [termed as Oscar of Inventions] for the team "Smart Tensors", Los Alamos National Laboratory **(2021) [Link1, Link2]**

Publications

Peer Reviewed Journal Publications

May 2024: Geographic variation of mutagenic exposures in kidney cancer genomes. *In Press, Nature* [link]
Dec 2023: Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics* [link]
Aug 2023: Visualizing and exploring patterns of large mutational events with SigProfilerMatrixGenerator. *BMC Genomics* [link]
Dec 2022: Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* [link]
Jun 2021: Classical density functional analysis of the ionic size effects on the properties of charge regulating electric double layers. *Mol. Phys* [link]
Aug 2021: Finding the number of latent topics with semantic non-negative matrix factorization. *IEEE Access* [link]
Feb 2021: A neural network for determination of latent dimensionality in non-negative matrix factorization. *MLST* [link]
Jan 2021: Improved protein decoy selection via non-negative matrix factorization. *IEEE/ACM TCBB* [link]
Feb 2020: Distributed non-negative matrix factorization with determination of the number of latent features. *J of Supercomputing* [link]
May 2020: Identification of anomalous diffusion sources by unsupervised learning. *Physical Review Research* [link]
Jun 2019: Coulombic and non-Coulombic effects in charge-regulating electric double layers. *Materials Research Express* [link]
Jan 2018: Solvophilic and solvophobic surfaces and non-Coulombic surface interactions in charge regulating electric double layers. *JCP* [link]
Dec 2017: Ionic solvation and solvent-solvent interaction effects on the charge and potential distributions in electric double layers. *JCP* [link]
Oct 2016: Electrolyte solution structure and its effect on the properties of electric double layers with surface charge regulation. *JCIS* [link]

Peer Reviewed Conference Papers

Mar 2022: Nonnegative tensor-train low-rank approximations of the Smoluchowski coagulation equation. *ICSLC* [link]
Apr 2021: Determination of the number of clusters by symmetric non-negative matrix factorization. *SPIE Big Data* [link]
Dec 2020: Decoy selection in protein structure determination via symmetric non-negative matrix factorization. *IEEE BIBM* [link]
Dec 2020: Semantic nonnegative matrix factorization with automatic model determination for topic modeling. *ICMLA* [link]
Sep 2020: Distributed non-negative tensor train decomposition. *IEEE HPEC* [link]
Nov 2019: Non-negative matrix factorization for selection of near-native protein tertiary structures. *IEEE BIBM* [link]

Dissertation

Dec 2019: Coulombic and non-Coulombic effects of single and overlapping Electric Double Layers with Surface Charge Regulation. UNM [link]

Patents/Applications

Nov 2023: Machine Learning Methods for Deconvolution of Integral Transformations and Their Application to Experimental Data Analysis. **App no. 63/424,835**, **App no. 18/508,168**