SANJEEVA REDDY DODLAPATI

+1-757-364-1561 | sdodl001@odu.edu | Google Scholar | LinkedIn | GitHub| X | Substack | Medium

SUMMARY

- Research Scientist with 6+ years of experience in deep learning, NLP, genomics, and drug discovery. Proven track record in leading multi-disciplinary research projects, publishing in peer-reviewed journals, and contributing to open-source ML frameworks. Specialized in uncertainty modeling, transfer learning, and scalable ML pipelines. Passionate about advancing fundamental research and translating innovations into real-world impact.
- Skilled in designing scalable ML pipelines and deploying models to production using CI/CD, Docker, MLflow, and Hugging Face on cloud platforms (AWS, Azure, GCP). Experienced in A/B testing, experiment tracking, and model performance evaluation aligned with business goals.
- **Collaborated** with multiple research teams resulted in **4 peer-reviewed publications** and 3 conference presentations
- **Continuous learning** through writing **blogposts** on AI for Science and earning **40+ ML course certifications**.
- **Areas of Interest:** Deep Learning, Genomics, Bioinformatics, Drug Discovery, Cheminformatics, AI for Healthcare.

EDUCATION

Ph.D. Computer Science (GPA: 3.9/4) | Old Dominion University, Norfolk, USA — Aug 2019 – July 2025

MS Computer Science (GPA: 3.5/4) | Georgia Institute of Technology, Atlanta, USA — May 2023 - Present

SKILLS

Programming: **Python** (expert), **R** (advanced), **Java** (intermediate), JavaScript, C, C++, **Bash**, SQL, Matlab

ML Packages: **PyTorch**, **TensorFlow**, **DeepSpeed**, NLTK, Deepchem, RDKit, Scikit-learn, SciPy

ML Techniques: **Deep Learning**, RL, NLP, Transfer Learning, Multi-task Learning, Graph Learning, Decision Trees, Experiment Design, Evaluation Metrics, Benchmarking, Meta Learnin

ML Architectures: LLMs, **Transformers**, **CNNs**, RNNs, LSTMs, Graph-NN, SSMs, Generative Models, Autoencoders

Bioinformatics: Bioconductor, DESeq2, Samtools, MEME-suite, caret, ggplot2, dplyr, ChemProp, Matplotlib

Database & Tools: SQL, Spark, Hadoop, Pandas, Dask, Numpy, Snowflake, AWS S3, MongoDB (basic)

MLOps & CI/CD: GitHub, Cloud-nativ Deployment, Docker, mlflow, Amazon Sagemaker, Azure ML, Hugging Face

Web Apps: HTML, CSS, PHP, flask, Django, fastAPI, Shiny, Quarto

OS & Cloud: Linux, HPC Cluster, macOS, Cloud (AWS, Azur, GCP)

RESEARCH EXPERIENCE

Graduate Research Assistant | Old Dominion University, Norfolk, VA — Aug 2019 - Present

Project I: Completing Single-Cell DNA Methylome Profiles via Transfer Learning Together With KL-Divergence

- Developed a Tensorflow-based **Transfer Learning framework** for DNA methylation prediction from genomic sequence.
- Applied Transfer Learning to impute missing CpG states, **boosting coverage from 1.5% to 50%** in sparse methylomes.
- Coupled KL divergence with Transfer Learning to optimize DNA methylation imputation, **increasing F1 score by 38%**.

Project II: Training Deep Neural Networks for DNA Methylation Prediction: A Data-Centric Perspective

- Designed a deep learning framework for data noise filtering, a rapid hyperparameter pre-screening, and model interpretability.
- Applied **adaptive noise filtering** to methylation data, enabling model **training with 50% less data** while preserving accuracy.
- Implemented **a novel hyperparameter pre-screening**, narrowed the search space and **reduced computational costs by 65%**.

Project III: Quantifying and Adapting the Uncertainty in Predictions of DNA Sequence-to-activity Models

- Developed a PyTorch **framework** for **uncertainty quantification** in genomic models to predict variant effect on functional activity.
- Enhanced variant effect prediction by integrating uncertainty estimation, improving accuracy in genomic sequence modeling.
- Developed a variant prioritization method to identify genetic variants with functional effects, reducing computational costs by 80%.

Project IV: Gene Expression and Chromatin Accessibility Analysis in Cardiac Fibroblasts and post-MI

- Collaborated with LSU on **RNA-seq analysis** of cardiac fibroblasts post-MI, identifying compensatory upregulation of non-Acta2.
- Developed **R pipelines** for RNA-seq and ATAC-seq analysis to identify differentially expressed genes and open chromatin regions.

Project V: Learning More Diverse Representations Through Hinge Loss Function

- Collaborated with UMich researchers to apply deep learning for identifying genome-wide variants influencing DNA methylation.
- Contributed through developing code to quantify uncertainty in variant effect on CpG prediction, identified credible set of variants.

Research Leadership & Impact Aug 2019 - Present

- L ed multiple independent research projects in genomics, NLP, and drug discovery, including ClinicalNormBERT, OmicsOracle, and UAVarPrior.
- Designed and implemented frameworks for uncertainty-aware modeling, transfer learning, and molecular optimization.
- Defined evaluation metrics and experimental protocols for genomic variant prioritization and methylation prediction.
- Mentored undergraduate teams at ODU, resulting in a winning app at the 2023 Speed Notes Competition.
- Influenced research direction through collaborations with LSU, UMich, and Boehringer Ingelheim on cross-disciplinary projects.

Independent Research Projects Aug 2019 - Present

- Developed **OmicsOracle**, an AI data agent that extracts genomic data from NCBI GEO, analyze, extract insights and visualize it.
- Developed scalable **pipelines** to preprocess, analyze, and visualize biological data using HPC clusters for GPU-accelerated DL.
- Developed a **drug-drug interaction prediction** model using a chemical knowledge graph, achieving near SOTA performance.
- Developing a comprehensive framework for molecular optimization and quantum machine learning for drug discovery.
- Trained a model to **predict selective borylation** of aromatic halides, achieving competitive performance despite limited data.
- **ML4Trading**: Built and backtested trading algorithms using Decision Trees and RL algorithms, optimizing risk-adjusted returns.
- **NLP**: Developed ClinicalNormBERT model for personalized clinical text normalization, improving data extraction in healthcare apps.
- **Health Informatics:** Designed healthcare apps for COVID-19 and HL7 FHIR data analysis, improving real-time clinical insights
- Built an **APT attack prediction** model using time-stamped cyber-attack data, achieving competitive threat forecasting accuracy.
- Trained a UNet model to **predict 3D protein structures** from cryo-EM data, performing on par with existing methods.
- Built personal portfolio website using quarto, flask and Django.

Collaborative & Service Experience | Old Dominion University, Norfolk, VA Aug 2019 – May 2023

- Actively contributed to the ML research community by reviewing papers for NeurIPS, ICML, ICLR, IJCAI (2021–2024).
- Shared research trends and best practices through blog posts and open-source contributions.
- Participated in academic collaborations and technical panels to shape research strategy and direction.
- Collaborated with LSU, UMich, and Boehringer Ingelheim on cross-disciplinary genomic and cheminformatics projects
- Collaborated with cross-functional teams including product managers, legal, and compliance to align ML solutions with business goals.
- **Courses**: Problem solving & programming I & II, Introduction to CS with Java and Introduction to Computer Architecture.
- **Responsibilities**: Leading the labs, holding recitation sessions, designing and grading projects and homework.
- **Mentored ODU undergrad student teams; winner of 2023 Speed Notes App Competition**

Research Intern Boehringer Ingelheim, Connecticut, USA May 2018 – Aug 2018

- Applied cheminformatics for property prediction to optimize synthesis protocols for chiral sulfinyl ketimine drug candidates.
- Developed two chiral sulfanilamide candidates with more than 99\% of enantio-selectivity, and made significant progress for third.

SELECTED PUBLICATIONS

1. **Dodlapati, Sanjeeva**, Z. Jiang, and J. Sun, "Completing single-cell dna methylome profiles via transfer learning together with kl-divergence," *Frontiers in Genetics*, vol. 13, p. 910 439, 2022
2. C. Li, J. Sun, Q. Liu, **Dodlapati, Sanjeeva**, H. Ming, L. Wang, Y. Li, R. Li, Z. Jiang, J. Francis, et al., "The landscape of accessible chromatin in quiescent cardiac fibroblasts and cardiac fibroblasts activated after myocardial infarction," *Epigenetics*, vol. 17, no. 9, pp. 1020-1039, 2022.
3. Y. Li, C. Li, Q. Liu, L. Wang, A. X. Bao, J. P. Jung, **Dodlapati, Sanjeev**, J. Sun, P. Gao, X. Zhang, et al., "Loss of acta2 in cardiac fibroblasts does not prevent the myofibroblast differentiation or affect the cardiac repair after myocardial infarction," *Journal of molecular and cellular cardiology*, vol. 171, pp. 117-132, 2022.
4. A. Chen, L. P. Samankumara, **Dodlapati, Sanjeeva,** D. Wang, S. Adhikari, and G. Wang, "Syntheses of bis-triazole linked carbohydrate based macrocycles and their applications for accelerating copper sulfate mediated click reaction," *European Journal of Organic Chemistry*, vol. 2019, no. 6, pp. 1189-1194, 2019.
5. **Dodlapati S**, Sun J. "Training Deep Neural Networks for DNA Methylation Prediction from DNA Sequence: A Data-centric Perspective". (*under preparation*)
6. **Dodlapati S**, Sun J. "Uncertainty-Aware Variant Effect Prediction for Genome-wide Prioritization of Non-coding Variants". (*under preparation*)
7. Du H, **Dodlapati S**, Parsons Z, Sun J & Lu J. "Learning more diverse genomic representations through hinge loss function". (*under preparation*)

CERTIFICATES & AWARDS

- **Best Mentor** Award from ODU for guiding student team in developing **Speed Notes** (summarization) app (Apr 2023)
- **CSIR-INDIA Junior Research Fellow** Scholar for Natural Product Drug Discovery (Mar 2008 – Dec 2008)
- **5+ Certificates in IPR** (basic courses) from World Intellectual Property Organization (2016 - 2017)
- **40+ Certificates in AI/ML courses** from online education platforms edx.org and coursera.org (2016 – Present)

Coursera Certificates:

- Agentic AI and AI Agents: A Primer for Leaders
- Introduction to Retrieval Augmented Generation (RAG)
- Google Prompting Essentials
- DevOps, DataOps, MLOps
- Python Essentials for MLOps
- MLOps Tools: MLflow and Hugging Face
- Introduction to Genomic Technologies
- Python for Genomic Data Science
- Introduction to Generative AI
- Generative AI: Elevate your Software Development Career
- Build Your Portfolio Website with HTML and CSS
- Spark, Hadoop, and Snowflake for Data Engineering

edX Certificates:

- DART.IMT.C.01: C Programming: Getting Started
- DART.IMT.C.02: C Programming: Language Foundations
- DART.IMT.C.06: Linux Basics: The Command Line Interface
- PH125.7x: Data Science: Linear Regression
- PH125.6x: Data Science: Wrangling
- PH125.4x: Data Science: Inference and Modeling
- PH125.5x: Data Science: Productivity Tools
- Data Science and Big Data Analytics: Making Data-Driven Decisions
- 6.00.1x: Introduction to Computer Science and Programming Using Python
- PH125.3x: Data Science: Probability
- PH125.2x: Data Science: Visualization
- PH125.1x: Data Science: R Basics
- CS1301x: Introduction to Computing using Python
- PH526x: Using Python for Research
- DS103x: Enabling Technologies for Data Science and Analytics: The Internet of Things
- DS102X: Machine Learning for Data Science and Analytics
- DAT201x: Querying with Transact-SQL
- DAT206x: Analyzing and Visualizing Data with Excel
- DAT101x: Data Science Orientation
- DAT210x: Programming with Python for Data Science
- DS101X: Statistical Thinking for Data Science and Analytics
- DAT203.3x: Applied Machine Learning
- DAT203.2x: Principles of Machine Learning
- DAT203.1x: Data Science Essentials
- DAT204x: Introduction to R for Data Science
- DAT208x: Introduction to Python for Data Science