SANJEEVA REDDY DODLAPATI

+1-757-364-1561 | [sdodl001@odu.edu](mailto:sdodl001@odu.edu) | [Google Scholar](https://scholar.google.com/citations?user=7FND004AAAAJ&hl=en&authuser=1) | [LinkedIn](https://www.linkedin.com/in/sanjeevardodlapati/) | [GitHub|](https://github.com/SanjeevaRDodlapati) [X](https://x.com/dodlapati_reddy) | [Substack](https://sanjeevareddydodlapati.substack.com/) | [Medium](https://medium.com/%40sdodl001)

# SUMMARY

* Software Engineer & AI/ML Researcher with 6+ years of experience designing AI-powered systems integrating generative AI, agentic AI, retrieval-augmented generation (RAG), and LLM orchestration. Proven expertise in building intelligent data agents, scalable ML pipelines, and deploying models into production environments using CI/CD and MLOps workflows. Passionate about accelerating discovery through automation, feedback-driven model optimization, and interdisciplinary collaboration that resulted in [**4 peer-reviewed publications**](https://scholar.google.com/citations?user=7FND004AAAAJ&hl=en&authuser=1)and 3 conference presentations
* **Continuous learning** through writing [**blogposts**](https://sanjeevareddydodlapati.substack.com/)on AI for Science and earning [**40+ ML course certifications**](https://www.linkedin.com/in/sanjeevardodlapati/details/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
* Agentic & Generative AI: Agentic AI Systems, Retrieval-Augmented Generation (RAG), Multi-Agent Orchestration, LangChain, LlamaIndex, Vector Databases, Prompt Engineering, Automated Data Exploration & Summarization.
* **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 Learning
* **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 & AI Systems:** CI/CD Pipelines, Automated Workflow Orchestration, Docker, MLflow, FastAI-based model Serving, Hugging Face Hub, Amazon SageMaker, GCP Vertex AI, Kubernetes.
* **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**](https://github.com/SanjeevaRDodlapati/Dev-MetTL)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**](https://github.com/sdodlapa/UAVarPrior)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**](https://github.com/sdodlapa/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](https://github.com/sdodlapati3/QeMLflow?tab=readme-ov-file) 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](https://github.com/SanjeevaRDodlapati/Clinical-Text-Norm-BERT?tab=readme-ov-file) model for personalized clinical text normalization, improving data extraction in healthcare apps.
* **Health Informatics:** Designed [healthcare apps](https://github.com/SanjeevaRDodlapati/Synthea_fhir_App?tab=readme-ov-file) 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](https://github.com/SanjeevaRDodlapati/mysite?tab=readme-ov-file) 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](https://coursera.org/share/f1064f381e7feee9a452c30a873aa519)
* [Introduction to Retrieval Augmented Generation (RAG)](https://coursera.org/share/4d174925bb8cc357b9482f3266366503)
* [Google Prompting Essentials](https://coursera.org/share/ac86cc33d2949a28a1f4cb7044d74911)
* [DevOps, DataOps, MLOps](https://coursera.org/share/7e23b7837a734381816c52fd45486c4e)
* [Python Essentials for MLOps](https://coursera.org/share/92852cc5c04d93bf43d1c581c481104b)
* [MLOps Tools: MLflow and Hugging Face](https://coursera.org/share/3d12d8a335930a614bd02a97230ff4c4)
* [Introduction to Genomic Technologies](https://coursera.org/share/d91a4deee897131d9ed68e148aed7a60)
* [Python for Genomic Data Science](https://coursera.org/share/18288d6e3c0903763415052e3e960874)
* [Introduction to Generative AI](https://coursera.org/share/328b2318d434178db083e50c4874e51c)
* [Generative AI: Elevate your Software Development Career](https://coursera.org/share/27165f83c6fc850d14adb83c54c07fa5)
* [Build Your Portfolio Website with HTML and CSS](https://coursera.org/share/0caaf39f67a84b6286ee08894102d785)
* [Spark, Hadoop, and Snowflake for Data Engineering](https://coursera.org/share/8062e7b1ccd1996ac9657e1b699b91e8)

### edX Certificates:

* [DART.IMT.C.01: C Programming: Getting Started](https://courses.edx.org/certificates/0fea984b64094875aeee5e4214fcc283)
* [DART.IMT.C.02: C Programming: Language Foundations](https://courses.edx.org/certificates/cb6540f5c5914e88b8a03672abca08b3)
* [DART.IMT.C.06: Linux Basics: The Command Line Interface](https://courses.edx.org/certificates/c1798f1e80784ae1be2b2b5f9b055a8c)
* [PH125.7x: Data Science: Linear Regression](https://courses.edx.org/certificates/36af6500cc61433cbe643d7ae433bb58)
* [PH125.6x: Data Science: Wrangling](https://courses.edx.org/certificates/cd124c6d036a44e19f6cf749600c2e19)
* [PH125.4x: Data Science: Inference and Modeling](https://courses.edx.org/certificates/b9f1e82f59cc4cb5a3acf0678c1f4ab0)
* [PH125.5x: Data Science: Productivity Tools](https://courses.edx.org/certificates/22d11c4306a047d9b29e6517f8c20160)
* [Data Science and Big Data Analytics: Making Data-Driven Decisions](https://courses.edx.org/certificates/2d35e46fc10e40c29de252d9239adbad)
* [6.00.1x: Introduction to Computer Science and Programming Using Python](https://courses.edx.org/certificates/aeecbe5965f746c584f7d973b601fc91)
* [PH125.3x: Data Science: Probability](https://courses.edx.org/certificates/bd53b6a2fb0241ffa8c7c6ff123c2676)
* [PH125.2x: Data Science: Visualization](https://courses.edx.org/certificates/39dc36158cbe4e919c524cc5f97d20b3)
* [PH125.1x: Data Science: R Basics](https://courses.edx.org/certificates/906ca4db6fd64b29a8c80e6cba0e9d80)
* [CS1301x: Introduction to Computing using Python](https://courses.edx.org/certificates/be9fecdc84654b80b348edb1174b2d7e)
* [PH526x: Using Python for Research](https://courses.edx.org/certificates/e002f68d8d904ec2af1baf7f4fee3427)
* [DS103x: Enabling Technologies for Data Science and Analytics: The Internet of Things](https://courses.edx.org/certificates/87ff1798aafe47a6a72ed91badd7f249)
* [DS102X: Machine Learning for Data Science and Analytics](https://courses.edx.org/certificates/1737ae156505457098349253798b2f7d)
* [DAT201x: Querying with Transact-SQL](https://courses.edx.org/certificates/f230c2c687d440d69766a49053ef42df)
* [DAT206x: Analyzing and Visualizing Data with Excel](https://courses.edx.org/certificates/6c79617abd23424198aaa0d664243ebc)
* [DAT101x: Data Science Orientation](https://courses.edx.org/certificates/cec026c4384e469ba6f8871d4112b5bc)
* [DAT210x: Programming with Python for Data Science](https://courses.edx.org/certificates/85ae5ef6c573486c9882cdcda76dd7a0)
* [DS101X: Statistical Thinking for Data Science and Analytics](https://courses.edx.org/certificates/9ae97d3e0f734a7faaea991cd2af5e58)
* [DAT203.3x: Applied Machine Learning](https://courses.edx.org/certificates/bc505a4580874dcc974c6c6ecc774b78)
* [DAT203.2x: Principles of Machine Learning](https://courses.edx.org/certificates/fac90c23947e4bac87db373de4053a1b)
* [DAT203.1x: Data Science Essentials](https://courses.edx.org/certificates/c3b343839c1b4048b40b516ec05a0ec0)
* [DAT204x: Introduction to R for Data Science](https://courses.edx.org/certificates/8e50b358031f41f893206f3029942596)
* [DAT208x: Introduction to Python for Data Science](https://courses.edx.org/certificates/f02a29d95f084c948749c41d98a587c4)