

Abhishek Daundkar, PhD

Cleveland, OH | aad157@case.edu | +1 (216) 392-4075 | [linkedin](#)

Summary

Machine Learning Scientist specializing in **X-ray Diffraction (XRD)** data analysis using **high-performance computing (HPC)**. Expertise in deploying **deep learning models** (TensorFlow/PyTorch) for materials discovery, with a focus on **image processing** and **time-series forecasting**. Proven ability to design **end-to-end ML pipelines** and **physics-informed neural networks**. Skilled in collaborative, interdisciplinary R&D for data-driven material science solutions.

Skills

Technical Skills: Python, PyTorch, TensorFlow, SQL, SPARQL, AWS Lambda, Git/GitHub, CI/CD (Jenkins), NLP (RASA, spaCy), Computer Vision (OpenCV, YOLOv5), Statistical Modeling (XGBoost, Pandas), High-Performance Computing (CUDA, MPI)

Materials Science: XRD Analysis, Rietveld Refinement, Perovskite Characterization, Molecular Dynamics, Phase-Field Modeling, FAIR Data Principles

Education

PhD Candidate - Applied Data Science in Materials Science, 2021 – 2025
Case Western Reserve University, Cleveland, OH

- [Graduate Certificate in Applied Data Science](#) (2023)

M.Tech - Mathematical Modelling and Computational Simulation, 2017 – 2019
Centre for Modeling and Simulation, SPPU, India

Work Experience

Graduate Research Assistant, Case Western Reserve University, Cleveland, OH Aug 2021 - Present

- Spearheaded \$500K [NSF - CSSI OAC](#) funded collaborative project [CRUX](#): Architected sophisticated noise generator autoencoder (PyTorch, Keras) for XRD data imputation, achieving **97%** accuracy on **perovskite** datasets
- Designed [FAIRmaterials](#) ontology ([Protégé/RDF](#)) with semantic versioning, enhancing data reproducibility
- Engineered **GPU ML pipelines**, accelerating compute time by **30%** via **CUDA multithreading**
- Orchestrated **CI/CD workflows** using Jenkins for automated model testing and deployment
- Implemented version control (Git) and code review processes for collaborative materials ontology development

Data Scientist, VIT Infotech, Bengaluru, India Nov 2020 – Jul 2021

- Engineered **time-series forecasting models** (XGBoost, ARIMA) for Hyundai Australia, boosting sales prediction accuracy by **15%**
- Automated **PDF data extraction** (PyPDF2, Tesseract), reducing manual processing by **50%**

Data Scientist, Gaian Solutions, Hyderabad, India Aug 2020 – Oct 2020

- Deployed **NLP models** (spaCy, RASA) for query generation from unstructured text, cutting response time by **40%**
- Productionized models via **AWS Lambda**, ensuring seamless integration with client systems

Data Scientist, Quantela Inc., Bengaluru, India June 2019 – July 2020

- Led a team to develop **facial recognition systems** (MTCNN, OpenCV) for Odisha Police, achieving **88% accuracy** on low-resource hardware
- Reduced false positives in **object detection models** (YOLOv4) by 30% via synthetic data augmentation

Projects

Perovskite XRD Analysis

Materials Informatics

- Achieved 97% peak detection accuracy via autoencoder-based recurrence plot analysis
- Validated against ICDD PDF-5+ database using cosine similarity metrics, ensuring commercial-grade pattern matching reliability

Synthetic XRD Generator

Materials Informatics

- Created 50k+ synthetic patterns for data augmentation using pseudo-Voigt simulations
- Enabled detection of low-intensity phases (<5% concentration)

FAIRmaterials Ontology Development

Materials Informatics

- Implemented SPARQL queries to enable semantic search across 10k+ materials datasets
- Integrated SQL databases (PostgreSQL) with RDF triplestores for hybrid querying

PDF-to-Query Generation Pipeline

NLP / Generative AI

- Built NLP pipeline for semantic query generation from PDFs using RAKE, Sense2Vec, and WordNet, reducing manual extraction time by 40%
- Deployed on AWS Lambda with Flask API, handling 500+ daily requests

RASA Chatbot with Alexa Integration

NLP / Generative AI

- Developed dialogue management system using LSTM networks, achieving 92% intent recognition accuracy

Invoice Data Extraction System

Computer Vision

- Combined object detection + OCR to digitize scanned invoices, achieving 94% table recognition accuracy

Real-Time Facial Recognition

Computer Vision

- Optimized model for Odisha Police deployment, achieving 88% accuracy on remote server
- Implemented K-Means clustering for identity matching, reducing search time by 60%

Automotive Sales Forecasting

Time Series / MLOps

- Developed region-wise prediction model improving Hyundai Australia's inventory planning accuracy by 15%
- Automated feature engineering pipeline using Pandas UDFs

Achievements

Callahan Fellowship - Case School of Engineering

- Awarded for most promising and academically qualified PhD applicants to the Case School of Engineering

PUDX ETA of PMT BUS - 1st Prize

- National level hackathon conducted by Pune Smart City (PUDX) in collaboration with IISC, Bangalore

Publications

Daundkar, A., et al. "Predictive Modeling of hidden peak information in XRD Patterns Using Autoencoders." Departmental Review, Case Western Reserve University, 2025 (Under Review)

Mengying Wang, et al. 2022. "CRUX: Crowdsourced Materials Science Resource and Workflow Exploration." In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22). Association for Computing Machinery, New York, NY, USA, 5014–5018. <https://doi.org/10.1145/3511808.3557194>

Mengying Wang, et al. 2023. "Selecting Top-k Data Science Models by Example Dataset". In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23). Association for Computing Machinery, New York, NY, USA, 2686–2695. <https://doi.org/10.1145/3583780.3615051>

Mengying Wang, et al. 2024. "ModsNet: Performance-Aware Top-k Model Search Using Exemplar Datasets". Proc. VLDB Endow. 17, 12 (August 2024), 4457–4460. <https://doi.org/10.14778/3685800.3685899>