

Divyansh Pandey

Lucknow, UP, India — +91-9305425557 — divyanshpandey0108@gmail.com

[LinkedIn](#) — [Portfolio](#) — [Github](#)

EDUCATION

Manipal University Jaipur

B.Tech (Hons.) Computer Science Engineering (AIML) — CGPA: 8.89

Jaipur, Rajasthan

Sep 2022 – Jun 2026

EXPERIENCE

AI Research Intern

Indian AI Research Organization (IAIRO)

Remote, India

Mar 2026 – Present

- Researching **Neurosymbolic AI** and **knowledge graphs** under Dr. Sheth, with focus on integrating **Small Language Models (SLMs)** and **embodied robotics** to build scalable, grounded real-world AI systems.

AI/ML Engineer Intern

VIGIL-Labs (IIT-H) — [Link](#)

Hyderabad, Telangana, India

Apr 2025 – Jul 2025

- **Accelerated** model training convergence by **optimizing** communication protocols, **reducing** global communication round **time by 45%** and utilizing **55% fewer** resources than baselines.
- Engineered a **decentralized** Federated Learning model for medical image classification, **surpassing baseline test accuracy by 20%** on complex non-IID **real-world** data while ensuring strict distributed data **privacy**.

PROJECTS

RAGineer — Python, PostgreSQL, Ollama, Chainlit, ChromaDB, RAGAS — [Repo](#)

- **Boosted** SQL generation accuracy **from 42% to 84%** via **RAG** pipeline (**ChromaDB + Qwen2.5-Coder**), achieving **0.91 retrieval precision** and **80% end-to-end accuracy** at **~2.5s** latency; hardened for **production** with SQL injection prevention and rate limiting, validated by **64 tests** at **97% coverage** via **RAGAS**.

Real-Time Fraud Detection System — Python, FastAPI, Docker, AWS — [Repo](#)

- Engineered a **FastAPI + Dockerized** microservice **reducing latency by 90%** (50ms to 5.4ms); **maximized** False Positive precision **from 6% to 78%** via class-weighted **XGBoost** with **MLflow**, maintaining **83% Recall**, and established **100% production observability** with **Prometheus & Grafana** to detect **concept drift**.

GetAnime — Python, Langchain, Streamlit, Groq, ChromaDB, Docker, K8s, GCP — [Repo](#)

- Architected a **RAG** pipeline (**LangChain, Groq LLM, ChromaDB**) with automated **ETL** via Hugging Face **Sentence Transformers**, delivering **sub-second latency** and **95% relevance accuracy** across **12,000+ entries**; maintained **99.5% uptime** orchestrating **Docker/K8s/GCP** microservices with **Grafana** monitoring.

TECHNICAL SKILLS

Languages: Python, SQL, Java, C

AI & Foundation Models: Generative AI, LLMs, Hugging Face, RAG Pipelines, Fine-tuning

ML/DL: Computer Vision, Deep Learning, Machine Learning, Time Series, Statistics, NLP

Frameworks: PyTorch, TensorFlow, Keras, Transformers, LangChain, scikit-learn, Pandas, NumPy, Matplotlib

Cloud & MLOps: AWS (SageMaker), Azure, GCP, Kubernetes, CI/CD Pipelines, Git, GitHub, MLFlow, Grafana

Database: PostgreSQL, MongoDB, Vector Databases (FAISS, Pinecone)

Software Development: Flask, Streamlit, RESTful API, FastAPI, Object Oriented Programming, DSA

RESEARCH

DEXNet: Ensemble Model for Lung & Colon Cancer Classification — SN CS 2026 — Springer — [Q1] — [Link](#)

- Developed a composite deep learning framework integrating **DenseNet**, **EfficientNetB3**, and **XGBoost** for histopathological cancer classification, achieving **99.87% accuracy** on LC25000 with **Grad-CAM** interpretability; deployed via a **HIPAA/GDPR-compliant AWS** pipeline for scalable real-time clinical inference.

Barbell Exercise Classification and Repetition Counting — ICDEC 2024 — Springer Nature — [Link](#)

- Built an ML system for barbell exercise classification and rep counting from **MetaMotion sensor data**, achieving **90%+ accuracy** via feature engineering and outlier detection pipelines for precise **human activity recognition**.

HONORS & AWARDS

- **Dean's List** for Excellence in Academics (**highest GPA**) — [Link](#)
- **2 x Dean's List** for Excellence in Off-campus Achievements — [Link](#)