Piyush Gupta

Delhi, India | +91-7001341611 | gguptapiyush45@gmail.com
linkedin.com/in/piyush-gupta-200416309 | github.com/Pg1910 | pg1910.github.io/portfolio

B.Tech, IIITDM Kurnool | Expected May 2027 | Available full-time from June 2027

SUMMARY

Applied ML engineer with four internships shipping production computer-vision and remote-sensing systems, from multi-modal aerial detection to medical-imaging segmentation. Depth in PyTorch, Vision Transformers, multi-modal fusion, and inference optimisation under tight compute budgets.

TECHNICAL SKILLS

Languages	Python, C++, SQL, Bash
ML / DL	PyTorch, TensorFlow, Hugging Face, OpenCV, scikit-learn, NumPy, Pandas
Modelling	CNNs (ResNet, EfficientNet, U-Net), Vision Transformers (ViT, DINO, EOMT), YOLO detectors, semantic segmentation, RAG, structured pruning, INT8 quantisation
Systems / MLOps	Docker, FastAPI, Git, DVC, Linux, AWS SageMaker, GCP, A/B testing

EXPERIENCE

ML Engineering Intern **Nov 2025 – Present**
Iris Aerial Innovations Pvt. Ltd. *Hybrid, Delhi*

- Architected a 5-stage aerial intelligence pipeline (multi-modal detection → fine-grained vehicle subclassification → temporal change detection → risk scoring → LLM mission summarisation) now serving 2,500 images/day to 3 operators in production testing.
- Raised 11-class aerial detection to **87 mAP@50**, +31 pts over an off-the-shelf YOLO checkpoint evaluated zero-shot on the aerial domain; lifted rare-class recall from 66 to **91** on a 4:1 imbalanced set via focal loss, stratified sampling, and hard-negative mining.
- Unified RGB, SAR, and drone imagery from sources including xView and RarePlanes, reconciling 3 annotation schemas into one training pipeline of **100K+ labelled frames**; cut dataset onboarding from 96 to 25 hours.
- Compressed the detector into a 12 GB VRAM budget via structured pruning and INT8 quantisation while preserving 11-class detection quality; implemented and adapted 5+ papers on detection robustness and domain adaptation.

Associate Agent Ops Engineer — ML **Jul 2025 – Oct 2025**
AnyTechPros Infotech LLP *Remote*

- Built multi-temporal satellite segmentation and change-detection pipelines over 400+ GB of imagery, reaching **83 IoU** and **92 macro-F1** on change masks.
- Designed SAR–optical fusion preprocessing (radiometric correction, co-registration, augmentation) that corrected inter-sensor geometric and radiometric drift, improving downstream change F1 by **20 pts**.
- Deployed models with DVC-versioned data and reproducible retraining, cutting experiment-reproduction time from days to 5 hours.

ML Research Intern **Jun 2025 – Jul 2025**
Maulana Azad National Institute of Technology *Bhopal, MP*

- Designed **GlaucFusion**, a bi-directional glaucoma detection pipeline fusing supervised DINO ViT with encoder-only EOMT; reached **96% accuracy**, +4 pts over a GlaucNet baseline across 7 public datasets totalling 30K+ images.
- Ran a controlled CNN-vs-Transformer study (ResNet-50, EfficientNet vs ViT-B/16, DINO) across accuracy, cross-dataset generalisation, and compute; Transformers gained **6 pts AUC** at materially higher inference cost.
- Built a multi-stage OpenCV quality gate (Laplacian sharpness, CLAHE, blur detection, cup-to-disc ratio) that automatically rejected unusable fundus images and harmonised 7 ophthalmology datasets under one schema.

Research Intern — Samsung PRISM **Dec 2024 – Mar 2025**
IIITDM Kurnool (Samsung-sponsored) *Kurnool, AP*

- Designed a custom dilated U-Net for breast-cancer segmentation on terahertz imagery using a combined BCE + Dice + IoU loss, outperforming a vanilla U-Net on both Dice and IoU under identical training budgets.
- Built the full raw-EMF → RGB conversion, augmentation, and heat-map evaluation stack, converting a raw instrument format into a trainable RGB corpus; accepted by Samsung R&D mentors.

PROJECTS

Clinical-RAG — *RAG, hierarchical chunking, dual-stage retrieval, FastAPI* github.com/Pg1910

- Replicated and extended CLI-RAG into a hospital-grade pipeline converting unstructured clinical notes into auditable structured summaries.
- Hierarchical chunking with dual-stage retrieval raises precision over a flat-chunk baseline; generation is constrained so every emitted field traces to a source span. Served behind FastAPI in Docker.

EDUCATION

B.Tech, Mechanical Engineering — Design & Manufacturing **2023 – 2027**
IIITDM Kurnool — CGPA 7.65/10 *Kurnool, AP*
Relevant coursework: Linear Algebra & Matrix Decomposition, Probability & Statistics,Numerical Methods.

Senior Secondary (CBSE) **2020 – 2022**
Jawahar Navodaya Vidyalaya — 83.6% *Purulia, WB*

ACHIEVEMENTS & LEADERSHIP

- **Amazon ML Challenge 2025** — Top 500 of 25,000+ teams nationally. **Inter-IIIT Hackathon 2026** — Finalist. Participant, Flipr and Google DSC Solution Challenge. Placement Cell Coordinator; Technical Head, Mechanical Engineering Association (2025–Present).