

Pranav Jadhav

+91-7838309721 | pranavjadhav001@gmail.com | [website](#) | [github](#) | [linkedin](#)

Senior Machine Learning Engineer with 6+ years of experience building production AI systems in medical imaging and healthcare. Specialized in deep learning, LLMs, computer vision, and inference optimization for clinical-scale deployment.

TECHNICAL SKILLS

Tools: Python, TensorFlow, Pytorch, HuggingFace, OpenCV, SQL, AWS, Docker, Flask, FastAPI, MLOps, Tensorrt
Technologies: Deep Learning, Computer Vision, LLMs, RAG, Metric Learning, Semantic Search, Diffusion Modelling, Model Optimization, Traditional Image Processing, Unsupervised/Supervised/Self-Supervised DL, OCR, Object Detection / Segmentation, Triton Inference Server

EXPERIENCE

Senior Machine Learning Engineer

Dec. 2024 – Present

Tricog Health

Bangalore

- Researched and trained a DPR model to retrieve medical conditions, anatomical structures, and severities from thousands of unstructured diagnostic reports, achieving over a **50% improvement over the BM25 baseline**.
- Led end-to-end development of an **LLM-powered automatic report generation system**, evaluated through custom domain-specific methods, **saving 3-4 minutes per case** while matching human-level performance.
- Architected and trained a **multi-task model** comprising of cardiac anatomical cavities segmentation, diastole/systole regression model and landmark keypoint detection **reducing manual ejection fraction assessment time from 5 mins to <3 sec**.
- Designed, developed and deployed a **Triton Inference Server pipeline** for cardiac ultrasound analysis — serving the above models, processing **~750 frames across 5 videos per echo study in under 14s end-to-end**, with async pipeline parallelism and shared memory transport — scaling to **~50K cases/day on a single GPU**.
- Owned and built a **HIPAA-compliant de-identification pipeline** for echocardiography studies — anonymizing patient names, IDs, dates, and masking PHI burned into video/image pixels, deployed on **AWS Lambda and Kubernetes** with parallel processing.

Machine Learning Engineer

Jun. 2022 – Dec. 2024

Tricog Health

Bangalore

- Built a **hierarchical echocardiographic view** classification model with **91% accuracy** and co-authored a research paper detailing its design and performance.
- Developed a classification model to predict the reference line location in Doppler imagery, achieving **90% accuracy**.
- Built **DICOM SR infrastructure** standardizing cardiac measurement reporting across **1000s of centers and 100s of echo machine variants** — replacing manual per-measurement data entry with **automated encoding in <1 sec**, ensuring data compatibility and interoperability across third-party vendor ecosystems.

Computer Vision Engineer

Aug. 2019 – Jun. 2022

Switchon

Bangalore

- Deployed Surface Inspection product utilizing **unsupervised anomaly detection methods** capable of identifying mm level defects.
- Built a **0 to 1 OCR product**, designing and implementing models for character and word recognition tailored to real-world industrial environment challenges.
- Established edge deployment infrastructure for Nvidia Xavier/Nano and Intel NUC, optimizing inference pipelines to **under 10 ms using techniques** such as TensorRT engine, pruning, and quantization.

BLOGS, PUBLICATIONS & PROJECTS

Cohort Retrieval using Dense Passage Retrieval: [Paper Link](#)

Open Source Contribution to Keras-CV: Contributed and Kicked off new branch of augmentation feature.[PR](#).

Deep Learning Models for Echo View Classification with Diverse Labels: Under review

Leveraging open-source LLMs for domain specific problems: [Medium](#)

Metric Learning for Classification: [Medium](#)

Implementing "Attention is All you need" using TensorFlow: [Medium](#)

How to conduct an ML experiment: [Medium](#)

Implementation of Generalizing GradCam for Embedding networks: [Paper](#) - [Repo](#)

HairStyle Transfer using Diffusion: Modified repaint paper to generate random hairstyles for the face and transfer the hairstyle from the source to the target image -[Repo](#)

ArcFace Inversion: generate class representations from pretrained ArcFace Model for explainability -[Repo](#)

EDUCATION

Guru Gobind Singh Indraprastha University

B.Tech in Mechatronics

Delhi, India

July 2015 – July 2019