

# Arjun Pramod

+91 9746782585 | arjunpramod509@gmail.com | linkedin.com/in/arjun-pramod | github.com/ArjunPramod

## SUMMARY

---

AI/ML Engineer with hands-on experience building and deploying production systems in Computer Vision, Generative AI, and NLP. Strong in end-to-end model development, data pipelines, real-time inference, and cloud deployment using PyTorch, YOLO, LangChain, FastAPI, Docker, and AWS. Focused on shipping scalable AI solutions for practical business use.

## TECHNICAL SKILLS

---

**Programming & Databases:** Python, SQL, PostgreSQL

**Machine Learning & Data:** Scikit-learn, XGBoost, SHAP, Feature Engineering, Hyperparameter Tuning, Model Evaluation, EDA, Data Analytics, Data Science, NumPy, Pandas, Matplotlib, Seaborn

**Deep Learning & Computer Vision:** PyTorch, TensorFlow, CNNs, Neural Networks, Transfer Learning, OpenCV, YOLO (YOLOv11), ConvNeXt, Roboflow, Object Detection, Image Classification, Pose Estimation, Grad-CAM

**Generative AI & NLP:** LLMs, LangChain, Transformers, Hugging Face, Sentence Transformers, FAISS, NLTK, spaCy, RAG, AI Agents, Natural Language Processing (NLP)

**Deployment & MLOps:** FastAPI, REST APIs, Docker, AWS (EC2, S3), MLflow, Streamlit, Git, FFmpeg, Model Deployment, Real-Time Inference, Webhooks, n8n

**Software Development:** HTML, CSS, JavaScript, React, Tailwind CSS, PyQt, Multithreading

## EXPERIENCE

---

### AI Engineer

Mar 2026 – Present

*Nesa Software Pvt. Ltd.*

*Kerala, India*

- Architected and deployed production-grade computer vision systems for CCTV surveillance analytics, enabling automated real-time and historical suspicious activity detection across live DVR/NVR video streams.
- Engineered an end-to-end door state monitoring pipeline integrating YOLO-based detection with ConvNeXt-Tiny multiclass classification (Open, Closed, Semi-Closed), covering Roboflow annotation, automated ROI extraction, and multi-stage real-time inference.
- Built scalable data engineering pipelines with perceptual hashing (pHash) and clustering-based stratification for leakage-free splits; trained, evaluated, and deployed YOLOv11, ConvNeXt-Tiny classifiers, and XGBoost models for surveillance video analytics workflows.
- Built and deployed a full-stack AI video analytics platform using Python, FastAPI, React, and Tailwind CSS, supporting RTSP stream ingestion, DVR/NVR integration, real-time AI inference, video playback.
- Designed and integrated LLM-powered Generative AI workflows — including retrieval-augmented generation (RAG) pipelines and AI-assisted automation features — for internal business applications.

### ML Research Intern

Jul 2023 – Sep 2023

*SRM University AP*

*Andhra Pradesh, India*

- Developed a preprocessing and augmentation pipeline (Python, OpenCV, Keras) generating 224×224 inputs; designed and trained a custom CNN over 15 epochs, achieving 95% validation accuracy and 92.63% test accuracy.
- Evaluated binary classifier with Scikit-learn metrics (macro F1: 0.93, confusion matrix) and packaged model artifacts for reproducible inference deployment.

### Artificial Intelligence Intern

Jul 2022 – Sep 2022

*Zebo.ai*

*Bengaluru, India*

- Built a DNN for heart disease prediction and a CNN-based dog breed classifier using TensorFlow and Keras transfer learning with data augmentation, covering end-to-end pipelines from preprocessing through model evaluation.
- Improved model performance through systematic hyperparameter tuning and batch inference experiments, benchmarking configurations to identify optimal model architectures.

## PROJECTS

---

### IntelliCam – Suspicious Behavior Detection Platform

Mar 2026 – Present

*Computer Vision & Surveillance Analytics System*

*PyTorch, ConvNeXt, OpenCV, FastAPI, React, Tailwind CSS*

- Developed an AI-powered CCTV analytics platform for real-time and historical suspicious behavior detection, integrating YOLOv11 detection, pose estimation, ConvNeXt-Tiny classification, and XGBoost into a multi-stage inference workflow.
- Built end-to-end data engineering pipelines with perceptual hashing (pHash) and clustering-based stratification; implemented Grad-CAM explainability and batch video inference workflows producing per-frame annotated output videos.
- Developed RTSP recording, live inference, and video playback applications using FastAPI, OpenCV, FFmpeg, React, and Tailwind CSS for real-time CCTV monitoring, video playback, and interactive video analytics.

### DoorSense AI – Intelligent Door State Monitoring System

May 2026 – Present

*Computer Vision & Video Analytics Platform*

*PyTorch, YOLO, ConvNeXt-Tiny, Roboflow, FastAPI, FFmpeg*

- Built an end-to-end CCTV analytics solution for automated door state classification (Open, Closed, Semi-Closed) from DVR/NVR streams, integrating YOLO-based detection with ConvNeXt-Tiny classification in a multi-stage real-time inference pipeline.
- Collected and annotated surveillance data with Roboflow; engineered an automated ROI extraction pipeline using detector outputs with configurable padding to generate large-scale classification datasets.
- Built a production-ready FastAPI web application supporting RTSP validation, DVR/NVR integration, inference configuration, monitoring dashboards, and full pipeline deployment.

### Customer Retention Decision Platform

Feb 2026

*Applied Machine Learning System*

*Scikit-learn, XGBoost, SHAP, PostgreSQL, MLflow, FastAPI, Streamlit, Docker, AWS*

- Built an end-to-end ML platform predicting telecom customer churn with XGBoost, capturing 72.5% of churners while targeting only 35.5% of the user base.
- Developed an explainable AI system using SHAP and customer segmentation to generate persona-based retention recommendations served via a FastAPI endpoint.
- Deployed on AWS EC2 with Docker containerization, Streamlit monitoring dashboard, MLflow experiment tracking, and PostgreSQL prediction logging.

### Retrieval-Augmented Document QA System

Jan 2026

*Applied NLP / RAG*

*LangChain, FAISS, Hugging Face, Sentence Transformers, FastAPI, Docker, AWS*

- Built a Retrieval-Augmented Generation (RAG) API for document-grounded question answering using LangChain, FAISS vector search, and Sentence Transformer embeddings.
- Implemented semantic search with FAISS and Maximal Marginal Relevance (MMR) retrieval to reduce hallucinations and return answers with traceable source citations.
- Dockerized and deployed the FastAPI service on AWS EC2; diagnosed and resolved out-of-memory (OOM) issues during local LLM inference, improving production stability.

## EDUCATION

---

### SRM University AP

*B.Tech in Computer Science and Engineering (AI/ML specialization), CGPA: 7.88*

Andhra Pradesh, India

*Jul 2021 – Nov 2025*

### Saraswati Vidyaniketan Public School

*Senior Secondary (Class XII), CBSE: 90.4%*

Kochi, Kerala

*Jun 2020 – Jul 2021*

### Bhavans Adarsha Vidyalaya

*Senior Secondary (Class X), CBSE: 92.2%*

Kochi, Kerala

*Jun 2018 – Mar 2019*

## ACHIEVEMENTS

---

Best Paper Award (ICAIN 2025) — Lead author, “ESADN: Enhanced Spatial Attention Network for Road Accident Detection”.

Gold Medalist — 1<sup>st</sup> place, ProductKraft Expo 1.0.

Certifications — Artificial Intelligence, Database Management Systems.