

SUSHANT KOTHARI

AI/ML Engineer | Computer Vision Engineer | Generative AI Engineer | Data Scientist

sk619kothari@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

AI/ML Engineer with experience developing production-ready Computer Vision, OCR, and GenAI solutions using AWS Bedrock and Azure AI Services. Skilled in building scalable LLM workflows, deploying deep learning models, and delivering AI-driven solutions.

SKILLS

Programming: Python, SQL, Object-Oriented Programming

Libraries: PyTorch, TensorFlow, Scikit-learn, OpenCV, Hugging Face Transformers, Ultralytics, Torchvision, Albumentations, Keras, NumPy, Pandas, SciPy, NLTK, Matplotlib

Core Competencies: Machine Learning, Deep Learning, Computer Vision, Generative AI, Natural Language Processing (NLP), Multimodal AI, Object Detection, Image Classification, Recommendation Systems, Feature Engineering, Transfer Learning, Model Fine-Tuning, Hyperparameter Tuning, Predictive Modeling, Time Series Forecasting, Data Preprocessing, Data Analysis

LLM & GenAI: Claude, LangChain, Prompt Engineering, Retrieval-Augmented Generation (RAG), Semantic Search

Cloud & AI Services: AWS Bedrock, Azure Document Intelligence, Azure Cognitive Search, Microsoft Translator

Tools: Git, GitHub, Docker, Streamlit, Flask

EXPERIENCE

AI/ML Engineer at Psych X86 Technologies

- Architected an AI-powered legal document system using **Claude** via **AWS Bedrock**, achieving **>90% answer accuracy**.
- Drove a **5-stage** AI pipeline for OCR, translation, chunking, LLM extraction, and semantic search for **500+ documents/day**.
- Engineered OCR pipelines using **Azure Document Intelligence**, processing **5,000+ pages** with high efficiency.
- Integrated **Microsoft Translator** to enable multilingual document processing across **15+ languages**.
- Implemented **Azure Cognitive Search's semantic ranking** to power retrieval, improving allegation mapping accuracy by **60%**.
- Reduced manual review time by **75%** while maintaining an **<10% error rate** across legal document processing.

Data Science Intern at Lincode

- Combined **YOLOv8** with a custom **Ray Casting Algorithm** to detect zone/line-of-sight crossings, automating event recording.
- Proposed and led **YOLOv5 to YOLOv8** migration, optimizing training, inference, and **ONNX** conversion for fast detection.
- Implemented an **OCR pipeline** using **CRNN** and **CTPN**, improving text detection and recognition performance.
- Built a Python GUI annotation tool for polygon annotation, labeling & class mapping, reducing dataset preparation time by **50%**.

AI Intern at Infosys

- Developed a hybrid **recommendation system** using content-based and collaborative filtering, achieving a **93% match rate**.
- Designed a text-based fashion recommender using BoW, TF-IDF, and Word2Vec across **12K+ products**.
- Incorporated **visual search** with EfficientNetV2L, VGG16, ResNet, and MesoNet for similarity matching on **15K+ images**.
- Addressed **cold-start** challenges for new products via **multimodal similarity** (text + image), boosting recommendation coverage.

PROJECTS

Wildlife Detection using Custom CNN Model

- Architected a **custom CNN** from scratch (no pretrained backbone) for wildlife detection, achieving **95%+ accuracy**.
- Tuned **network depth, kernel sizing & receptive fields**, using dropout & batch normalization to offset lack of transfer learning.
- Deployed a low-latency inference pipeline achieving **90% precision** at **<100ms/prediction**, enabling real-time species detection.
- Built **scalable** preprocessing & augmentation pipelines (normalization, denoising, batching) to handle **50K+ noisy, diverse images**.

Deepfake Detection Model

- Built a **MesoNet**-based deepfake model, restructuring conv layers to capture compression artifacts, achieving a **94.5% F1-score**.
- Redesigned MesoNet's **feature extraction layers** for 256x256 RGB inputs, boosting **recall to 92%** on subtle forgeries.
- Deployed a Streamlit-based real-time deepfake classifier delivering **96% precision** across **10K+ image evaluations**.
- Applied loss re-weighting, threshold calibration & augmentation, cutting **false positives by 25%** vs. baseline.

Medical AI Assistant: Disease Prediction Platform

- Engineered an **AI-powered health screening system** achieving **92%+ average accuracy** across various conditions.
- Constructed end-to-end ML pipelines with **imputation & SMOTE**, feature engineering, model tuning, and model persistence.
- Developed a **Streamlit** application delivering real-time health predictions & personalized risk assessments across multiple diseases.

Text-Based Emotion Recognition System

- Designed an NLP-driven **Bidirectional LSTM** with pretrained GloVe embeddings, achieving **96% classification accuracy**.
- Benchmarked Bidirectional LSTM against multiple ML models, with it **outperforming SVM by 10%** on emotion classification.
- Minimized **overfitting by 15%** using L2 regularization, EarlyStopping & grid-searched LR, then deployed via Flask.

EDUCATION

08/24 – 06/26 Dr. D. Y. Patil's Ramrao Adik Institute of Technology – **M.Tech in Data Science** CGPA: **9.16**

08/20 – 06/24 Dr. D. Y. Patil's Ramrao Adik Institute of Technology – **B.Tech in CSE with DS** CGPA: **9.26**

RESEARCH PUBLICATIONS

- **Air Quality Index Prediction Using Hybrid LSTM-GRU and Residual BiLSTM Models** Published
International Research Journal of Multidisciplinary Scope (IRJMS) Vol. 7, Issue 2, 2026
- **Hybrid Wildlife Classification and Detection Using EfficientNet-B4, Custom CNN and YOLOv10** Accepted
International Research Journal of Multidisciplinary Scope (IRJMS) 2026