Sanjay Srinivasa

ssrin054@ucr.edu | +1(951)-347-8940 | linkedin.com/in/sanjay-srinivasa-31668915a/

EDUCATION

University of California, Riverside — Riverside, USA
Master of Science, Computer Science — *Sept 2024 – Jun 2026*

R.V. College of Engineering (RVCE) — Bangalore, India
Bachelor of Engineering, Computer Science — *Aug 2017 – Jun 2020*

EXPERIENCE

United Health Group - Optum Applied Research Group *Bangalore, India* — Jul 2020 – Aug 2024

Associate AI/ML Engineer *[Nov 2023 – Aug 2024]*

- Developed a stacked ensemble model combining **random forests**, **gradient boosting**, and **logistic regression** to prioritize patient collections, increasing hospital collection rates by **70%**.
- Implemented a multi-stage NLP pipeline using **GPT 3.5, Vicuna**, and **DistilBERT**, enhancing call center performance, reducing repeat calls, and boosting Net Promoter Score by **82%**.
- Researched and fine-tuned **Microsoft's small language model (SLM) Phi-2 (2.7B)** for question-answering tasks on clinical reports, achieving minimal precision loss compared to larger LLMs.
- Applied linear and per-channel **quantization** on **VGG (Visual Geometry Group) CNN model** with just <**0.2%** accuracy loss, enabling faster processing of medical claim images to accelerate insurance approvals and reduce delays.
- Developed **RAG (Retrieval-Augmented Generation)** system with semantic chunking to assist call agents in answering provider and patient queries, reducing LLM hallucination and improving response accuracy by **15%**.

Data Scientist *[Mar 2023 – Oct 2023]*

- Collaborated with **Google** to evaluate generative AI LLMs on Zero-Shot and Few-Shot learning using their Cloud **Vertex AI** platform.
- Designed a **Python-PySpark** pipeline on Databricks, achieving **98.6%** precision in PHI/PII de-identification. using **SparkNLP** models like **ClinicalNER** and **De-identification NER**.
- Developed a system to identify patterns in patient insurance claims, using the unsupervised clustering algorithm **DBSCAN**, achieving **98.5%** precision and improving processing efficiency.

Software Engineer *[Jul 2020 – Mar 2023]*

- Developed a **YOLOv5**-based OCR system to extract patient vitals from PDFs, increasing extraction accuracy by **15%** using OpenCV and TensorFlow.
- Automated audio-to-text transcription with **Microsoft's Azure ASR** (Automatic Speech Recognition), streamlining deployment with **Jenkins** and **Docker**

SKILLS

- Python, C++, PyTorch, SQL, Git, Pandas, Docker, FastAPI, Flask, NumPy, Linux, Azure, Databricks.

PUBLICATIONS

- Naveena K G, **Sanjay S**, Kushal K S, Manish M Naik, Sharadadevi S K, **"Indian License Plate Number Detection Using Convolution Neural Networks"**, 2020 International Research Journal of Engineering and Technology. [Link]
- **Sanjay S**, Revathi S A, **"Knee Osteoarthritis Progression & Classification Using Cartilage Damage Index & Machine LearningAlgorithms: A Literature Review"**, 2020 International Research Journal of Engineering and Technology. [Link]

PROJECTS

1.Real-Time Vehicle Identification | ***CNN, OpenCV, TensorFlow*** — Dec 2019

- Developed a vehicle license plate identification system. Processed 45,000 frames with **sub-2 second** detection, handling low light, motion blur, and occlusions and achieving precision of **92%**.

2. Lane Detection for Autonomous Vehicles | ***ResNet, TensorFlow*** — Mar 2020

- Engineered a real-time lane detection system using Canny edge detection, Hough Transform, and **ResNet-18**, achieving **95%** accuracy in lane classification and segmentation