Nikhil Chowdary Paleti

510-935-8895 | npaleti@ucsd.edu | linkedin.com/in/nikhil-paleti | nikhil-paleti.github.io

Education

University of California San Diego — San Diego, CA
Master of Science in Data Science — *Sep 2024 – Jun 2026 (Expected)*

Amrita Vishwa Vidyapeetham University — Coimbatore, India
B.Tech in Computer Science and Engineering (Artificial Intelligence) — *Oct 2020 – Jun 2024*

Experience

Machine Learning Engineer Intern — Jan 2024 – Jun 2024
TechProfuse — *Hyderabad, India*

- Implemented unstructured data extractor using Gemini 1.5 Pro LLM, processing 50k documents in 15 hours.
- Led 3 interns to convert English text to Dutch in scientific images using Google Vision API and CV2.
- Developed Vision API parser for TIFF images, reducing processing costs by 95% over AWS.
- Automated email classifier and summarizer using Gemini 1.5 Flash model, achieving 95% accuracy.

Research Assistant — Jan 2023 – Sep 2023
Dr. Sowmya's Lab, Amrita Vishwa Vidyapeetham University — *Coimbatore, India*

- Evaluated 12 pre-trained CNNs for PD classification using T1-weighted MRI, identifying VGG19 as the best.
- Optimized MRI slice selection, reducing slices from 182 to 87 while retaining key structures.
- Mitigated data leakage and used Grad-CAM to visualize PD regions, enhancing reliability & explainability.
- Co-authored a paper published in *Digital Signal Processing*, Volume 147, April 2024. [Link]

Projects

Indic Verse: Advanced Indic Language LLM System — Jan 2024 - Apr 2024
Tech Stack: Python, PyTorch, Hugging Face Transformers, PEFT, NLP — *[Hugging Face]*

- Engineered English-to-Indic translation and transliteration modules, processing 3 types of datasets for training.
- Processed 1M+ sentences from 6+ datasets to create a robust training dataset for the Indic language models.
- Fine-tuned 3 LLM models using PEFT techniques, improving performance on Indic language tasks.

Knee Osteoarthritis Severity Classification using Diffusion Augmented Images — Aug 2023 – Dec 2023
Proceedings of ICACECS, pp. 266-274, Hyderabad, India — *[Link]*

- Improved knee OA classification accuracy from 68% to 84% using EfficientNetB3.
- Applied CLAHE preprocessing, boosting accuracy by 8%.
- Augmented dataset by generating 200 images per class using DDIM diffusion models, reducing class imbalance.

Improving RL Agent Training using Text-Based Guidance — May 2023 – Sep 2023
Proceedings of DravidianLangTech, pp. 33–42, Varna, Bulgaria — *[ACL Anthology]*

- Developed methodology to train RL agents using text-based instructions in 4 languages.
- Trained embedding networks on 3,504 image-text pairs, enhancing multilingual instruction understanding.
- Trained SAC agent for 50 million steps, enabling generalization to unseen paths with mixed-language instructions.

Few-Shot Approach to Dysarthric Speech Classification Using Transformers — Mar 2023 – Jul 2023
Proceedings of ICCCNT, pp. 1-6, Delhi, India — *[Link]*

- Achieved 85% accuracy in dysarthria detection using Whisper-large-v2, improving prior results by 10%.
- Enhanced multiclass classification accuracy to 67% with 'words' data, outperforming 'letters & digits' data by 9%.
- Trained a 1.5 billion parameter model using PEFT and LoRA techniques, reducing computational costs.

Technical Skills

Languages: Python (Advanced), SQL (Advanced), C++ (Intermediate)
Libraries: NumPy, Pandas, Matplotlib, Scikit-Learn, PyTorch, TensorFlow, Transformers, PySpark
Machine Learning: Supervised/Unsupervised Learning, Deep Learning, Reinforcement Learning
Specializations: Natural Language Processing, Computer Vision, Generative AI (Large Language Models)
Tools & Cloud: Docker, Git, AWS, Google Cloud Platform