Harsha Koduri

harsha.koduri123@gmail.com | linkedin.com/harsha-koduri | github.com/HarshaKoduri123

PUBLICATIONS

- **Koduri Harsha, Ma Ming.** "LLM-TAMIS: Large Language Model based Text-Augmented Medical Image Segmentation", MICCAI 2025 (Submitted)
- **Koduri Harsha, Ma Ming.** "OTID: Optimal Transport-Based Low-Dose CT Image Denoising." *SPIE Medical Imaging*, 2025, pp. 13406-99
- **Koduri Harsha, Ma Ming.** "Diabetic and Hypertensive Retinopathy Classification from Retinal Images Using Dual Vision Transformer." *International Conference on Digital Data Processing (DDP 2024)*, 2024.
- **Ajay Vamsi Jalluri, Harsha Vardhan Garine, Harsha Vardhan Koduri, Kiran Khatter, Soharab Hossain Shaikh, Devanjali Relan.** "Automatic Classification of Diabetic and Hypertension Fundus Camera Retinal Images Using Deep Learning." *International Conference on Signal, Machines, Automation, and Algorithm*

EXPERIENCE

Research Assistant [Aug 2024 - Current]

Yeshiva University, New York

- **Conducted research in AI**, focusing on Machine Learning Deep Learning and Large Language Models(LLMs) techniques under **Professor Ming Ma**.
- Developed **LLM-TAMIS**, a **Large Language Model-based Text-Augmented Medical Image Segmentation** framework integrating clinical text and images, improving segmentation accuracy; submitted to **MICCAI 2025**.
- Proposed **OTID (Optimal Transport-based Image Denoising)**, Using **Wasserstein GAN** with attention mechanisms for **low-dose CT image denoising**, achieving state-of-the-art performance; accepted at **SPIE Medical Imaging 2024**.
- Designed a **Dual Vision Transformer model** for **diabetic and hypertensive retinopathy classification**, enhancing feature extraction and classification accuracy; accepted at **Digital Data Processing (DDP) 2024**.

Data Science Intern [Jan 2022 - Jun 2022]

Sabudh Foundations, Punjab

- My task is to develop a **deep learning mode**l to **improve Single Image Super-Resolution (SISR)** by generating high-resolution images from low-resolution inputs.
- Proposed a **GAN-based approach** with a generator using residual blocks, sub-pixel convolution for upscaling, and batch normalization, along with a **VGG-style discriminator** with LeakyReLU activation, while designing a perceptual loss that combines VGG-based feature reconstruction to preserve details and adversarial loss for realistic textures.
- The model outperformed conventional super-resolution techniques, producing sharper, more realistic images with improved texture details and perceptual quality.

Machine Learning Intern [May 2020 - July 2020]

Cigniti Technologies, Hyderabad

- Worked on a **Sentiment Analysis** project. My task is to analyze customers reviews and classify those reviews using Deep Learning Models.
- Built **Positive-Negative** and **Service-Product classifiers** using the datasets provided by the company. A **Word2Vec** pre-trained model is used for Text Embeddings, These Embeddings are then fed to an **LSTM model** and trained for 1000 epochs. Finally, the model was deployed in the **AWS cloud**.

EDUCATION

Yeshiva University, *New York City* [Sep 2022 - May 2024]
Masters, Artificial Intelligence
Massachusetts Institute of Technology, *Online,* [Certificate]
MicroMasters, Statistics, and Data Science
BML Munjal University, *Gurgaon, India* [Jul 2018 - May 2022]
Bachelors, Computer Science

SKILLS

Languages: Python, Java, R, C++.
Frameworks: Pytorch, Tensorflow, Apache Spark, LangChain.
Libraries: SK-Learn, Numpy, Pandas, Matplotlib, Transformers, vLLM.
Databases: Mysql, Postgresql, MongoDB, Neo4j.
Cloud: AWS, Microsoft Azure.

PROJECTS

LLM-TAMIS | Pytorch

- Developed a **multimodal segmentation model** that processes both medical images and clinical text, enhancing segmentation accuracy by integrating contextual information from radiology reports.
- Implemented **LLaMa 3.2** a **Large Language Model (LLM)** and **Vision Transformer (ViT)** to extract meaningful text embeddings and align them with image features, improving feature representation.
- Designed a **CNN-based U-Net architecture** for medical image segmentation and introduced a novel **Spatial Channel Driven Module (SCDM)** to refine spatial and channel feature extraction
- **Achieved state-of-the-art results** on the **MosMedData+ and QaTa-COV19** datasets, outperforming existing methods in **Dice score and mIoU** evaluation metrics.
- This research was submitted to the **MICCAI 2025** conference, **one of the top conferences** for **medical imaging using AI**

Voice Cloning | Pytorch

- Developed a **deep learning-based Text-to-Speech (TTS) system**, combining a sequence-to-sequence model for mel spectrogram generation with a WaveNet vocoder for waveform synthesis.
- Implemented a **CNN-based encoder** with character embeddings, followed by a **bi-directional LSTM** and **location-sensitive attention** to align input text with spectrogram frames dynamically.
- Designed an **autoregressive LSTM decoder** to predict mel spectrograms, incorporating a **convolutional post-net** to refine spectrogram outputs for improved synthesis quality.
- Trained a modified **WaveNet vocoder with dilated CNNs**, conditioned on mel spectrograms, and utilized **Mixture of Logistics (MoL) loss** for realistic speech waveform generation.
- **Achieved state-of-the-art speech** synthesis on dataset **LibriSpeech** and **custom recordings** (including my professor's voice).

Autonomous Driving | Python, Carla

- Develop a program for an autonomous vehicle in the **CARLA simulator**.
- The vehicle's Longitudinal and Lateral directions were controlled by a **PID** and **Stanley controller**.
- Implemented **Uncensored Kalman filters** for state estimation by taking data from the sensors Lidar Radar and Camera.
- Developed a **3D perception** pipeline using camera coordinate frames with downward y-axis for height estimation, including **SVD-based plane estimation**, **Hough transform-based** lane boundary detection with slope filtering, and **semantic-guided** object detection filtering via normalized pixel counts.
- Implemented **behavioral planning logic**, **static collision checking**, **path selection**, and **velocity profile generation** for path planning.