Sahil Dharod

Sahil Dharod | sahildharod28@gmail.com | sahildharod | +91 9870343700

Research Interests

Multimodal Learning, Multi-Armed Bandits, Natural Language Processing, Automatic Speech Recognition

Education

Indian Institute of Technology, Bombay (IITB) *Mumbai, India*
Bachelor of Technology in Electrical Engineering (CPI : 9.77/10) *(Nov'21 - May'25)*
Minor in Artificial Intelligence and Data Science

Publications

Pavan Kalyan Tankala, Piyush Pasi, **Sahil Dharod**, *et al.*, "WikiDO : Evaluating Out-of-Distribution Generalization of Vision-Language Models in Cross-Modal Retrieval," Poster presented at the 38th Conference on Neural Information Processing Systems (NeurIPS) and at the RBFM Workshop, NeurIPS 2024 at Vancouver, Canada

Research Experience

Domain Generalization for Multi-Modal Machine Learning *(Jan'24 - May'24)*
Guide : Prof. Preethi Jyothi, In collaboration with Google, DeepMind *R&D Project*

- Developed the **WikiDO** benchmark for **cross-modal retrieval** with **380K image-text pairs** scraped from Wikipedia
- Conducted **image** and **text filtering**, along with **human evaluation**, to remove poor-quality captions and images
- Conducted evaluations on vision-language models like **BLIP, CLIP** and **BLIP-2** for **image** and **text retrieval** using WikiDO, highlighting performance gaps of **8-10%** in **R@K** values between in-distribution and out-of-distribution test datasets

Constrained Best Arm Identification in Grouped Multi-Armed Bandits *(Aug'24 - Dec'24)*
Guide : Prof. Sharayu Moharir *B.Tech Thesis*

- Surveyed literature on **best arm identification algorithms** and their **bounds** in unconstrained multi-armed bandit settings
- Developed and tested an **iterative algorithm** for best-arm identification, where each arm has independent sub-arms constrained to have means above a threshold μ_{TH}
- Devised a **lower bound** $\left(\mathcal{O}\left(H_{\text{id}}\right)\right)$ using **alternate environments** and an **upper bound** $\left(\mathcal{O}\left(H_{\text{id}} \ln \frac{H_{\text{id}}}{\delta}\right)\right)$ on the expected sample complexity in terms of the **hardness index** $\left(H_{\text{id}}\right)$ of the problem in a fixed confidence setting

Low Resource Voice Transfer for Speech Generation *(Aug'24 - Present)*
Guide : Prof. Ganesh Ramakrishnan, Prof. Preethi Jyothi *R&D Project*

- Set up **text-to-speech (TTS)** models like **VALL-E** that view TTS as conditional modelling for Indian languages
- Exploring **zero-shot dialectal transfer** to Indian language dialects, focusing on improving cross-dialect adaptability
- Working on improving **tokenization schemes** and **audio codec** in VALL-E to enable better sharing across languages

Professional Experience

AI Engineer & Researcher Intern, AWL, Inc., Sapporo, Japan *(May'24 - Jul'24)*
Guide : MD Intisar Chowdhury (PhD) *Summer Internship*
Introduction : AWL, Inc. in collaboration with Sony is the market leader in deep learning based video analytics in Japan

- Integrated parallel branches into **YOLO-X** for **age, gender,** and **head direction** prediction in a single forward pass
- Adapted techniques like **SimOTA** and **Dynamic K Estimation** for multi-task label assignment to optimize training
- Generated **pseudo labels** with **prompt guidance** for action recognition and finetuned **LLaVA-2B** with **LoRA adapters**
- Leveraged VLM's generalization for a **multi-task setting**, improving precision on downstream tasks by **20-25%**

Scholastic Achievements

- Department Rank **4** among 210+ students in the department of Electrical Engineering *(2024)*
- Recipient of **Institute Academic Prize** for excellent academic performance in 3rd year *(2024)*
- Secured an **All India Rank of 322** in **JEE Advanced** out of 150 thousand candidates *(2021)*
- Achieved an **All India Rank of 131** in **JEE Mains** among 1 million+ students, with **100 percentile** in **Physics** *(2021)*
- Stood **1st** in Maharashtra Board's **SSC (Class X) Examination** in hometown among **2000+** students *(2019)*
- Awarded **silver medal** twice in **Prabhutva** exam conducted by BrihanMumbai Ganit Adhyapak Mandal *(2014, 2017)*

AI & ML Projects

Adversially Robust LLM Tuning via Subset Selection with SPIN *(Mar'24 - Apr'24)*
Guide : Prof. Ganesh Ramakrishan *CS769 : Optimization in Machine Learning*

- Developed fine-tuning methods for **LLMs** under computational constraints, achieving a **10%** increase in **BLEU** score
- Implemented **curriculum learning**, facility-coverage based **data subset selection** along with **self-play fine-tuning (SPIN)**
- Established a baseline for **adversarial prompts** and demonstrated model robustness with modified **FGSM attacks**

Stochastic Modelling for Stock Price Prediction *(Nov'23 - Dec'23)*
Guide : Prof. Amit Sethi *EE782 : Advanced Topics in Machine Learning*

- Implemented a **multi-step stock price prediction** model using **Hierarchical VAE** and **diffusion** probabilistic techniques
- Improved model accuracy through **multi-step sampling**, **adaptive noise modeling**, and **denoising score matching**
- Extracted critical financial features such as **RSI, MACD, volatility** and **stochastic oscillator** to improve **MSE** by **31%**

Automatic Speech Recognition *(Jan'24 - Apr'24)*
Guide : Prof. Preethi Jyothi *CS753 : Automatic Speech Recognition*

- Enhanced **Grad-TTS** by replacing convolutions with **depthwise separable** ones and adding **cosine noise scheduling**
- Improved Conformer model with **PowerConv** module and **CTC**, reducing WER on **Librispeech** from **0.74** to **0.68**
- Presented a paper on self supervised learning with random projection quantizer and summarized and led a discussion on the paper DASpeech : Directed Acyclic Transformer for speech-to-speech translation through a blog

Enhanced CGAN with Siamese Discriminator *(Oct'23 - Nov'23)*
Guide : Prof. Amit Sethi *EE782 : Advanced Topics in Machine Learning*

- Engineered a **Conditional GAN** to generate diverse images of the same individual capturing different facial variations
- Integrated a **Siamese network** with triplet loss as a discriminator, using data augmentation for precise person re-identification

Image Captioning using Transformers *(Jun'23)*
Self Project

- Developed an image captioning system based on an encoder-decoder model which uses pretrained **ViT embeddings**
- Constructed the **decoder from scratch**, utilizing TensorFlow to build its key components namely: Input and Positional Embeddings, Masked Multi-Head Self and Cross Attention Layers and achieved a **BLEU2 score** of **0.34** on Flickr8k dataset

Movie Recommender System *(Mar'23 - Apr'23)*
Guide : Prof. Abir De *CS419 : Introduction to Machine Learning*

- Implemented a **content-based filtering** model with **TF-IDF** and **cosine similarity** to find the top **30** related movies
- Created a user-movie matrix and used **collaborative filtering** with **Pearson correlation** to estimate ratings for movies
- Developed a **hybrid filtering** approach by integrating the collaborative and content-based recommendation techniques

House Prices Analysis and Prediction *(Nov'22 - Dec'22)*
Guide : Prof. Amit Sethi *DS203 : Programming for Data Science*

- Performed **Exploratory Data Analysis** and feature engineering on a dataset of house prices and related variables
- Implemented **Lasso, Ridge and Random Forest Regression** to predict prices and achieved an average $\mathbf{R}^2$ **score of 0.9**

Other Projects

Pipelined RISC Processor *(Mar'23 - Apr'23)*
Guide : Prof. Virendra Singh *EE309 : Microprocessors*

- Designed a **6 stage pipelined RISC**, a **16 bit** processor with the stages instruction fetch, decode, register read, execute, memory access and write back in VHDL to execute **26** different instructions and achieve a **CPI** close to **1**
- Implemented hazard mitigation techniques like forwarding logic and stalling unit to handle various types of data dependencies

Bubble Trouble - Game Development in C++ *(Jan'22 - Feb'22)*
Guide : Prof. Parag Chaudhuri *CS101 : Computer Programming and Utilization*

- Utilized **OOPs** principles to implement dynamic shooter and bubble classes with accurate controls and projectile motion
- Implemented multiple features in the game including **3 progressive levels** with dynamically increasing speed, size and quantity of bubbles, **timer** as well as other graphic features to make the game more interactive and user-friendly

Wireless Vibration Sensing Node for Structural Health Monitoring *(Jan'24 - Apr'24)*
Guide : Prof. Siddharth Tallur *EE344 : Electronic Design Lab*

- Developed a **vibration sensor** with **data logging** and **wireless transmission** using ESP-32 for real-time monitoring
- Built a **GUI** to visualize key parameters and conduct **frequency analysis** on data obtained from various locations

Teaching Experience

Teaching Assistant *(May'23 - Jun'23)*
Instructor : Prof. Prachi Mahajan *MA108 : Differential Equations I*

- Part of a **40** student team selected for teaching a batch of **1400 UG freshmen** in their course on **Differential Equations**
- Addressed students' course-related queries and conducted **weekly tutorials** for problem-solving and concept discussions

Technical Skills

Languages	Python, C++, Embedded C, SQL, LaTeX, VHDL
Libraries	NumPy, Pandas, SciPy, PyTorch, Transformers, TensorFlow, OpenCV, STL, LangChain
Software/Tools	MATLAB, LaTeX, Git, Jupyter, Intel Quartus, GNU Radio, KiCAD

Courses Undertaken

AI & ML	Programming for Data Science, Introduction to Machine Learning, Advanced Topics in Machine Learning, Image Processing, Optimization in Machine Learning, Automatic Speech Recognition, Learning and Inference in High Dimensions*
Mathematics & CS	Single and Multi-variable Calculus, Linear Algebra, Differential Equations, Complex Analysis, Probability and Random Processes, Markov Chains and Queuing Systems, Computer Programming with C++, Communication Networks, Games and Information*
Electrical	Intro to EE, Power Engineering I and II + Lab, Analog Circuits + Lab, Digital Systems + Lab, Signal Processing, Control Systems + Lab, Electronic Devices + Lab, Microprocessors + Lab, Communication Systems + Lab, Electronic Design Lab, Information Theory and Coding
Others	Organic and Inorganic Chemistry, Physical Chemistry, Biology, Engineering Drawing, Economics, Philosophy, Environmental Studies

** to be completed by November 2024*

Extracurricular Activities

- Designed a **RC Plane** with a BLDC motor, propellors from scratch, performed maneuvers and flew the plane *(2022)*
- Completed **Tinkering Bootcamp** under Technical Summer School (TSS) , organized by Career Cell, IIT Bombay *(2022)*
- Contributed **80+ hours** of community service under the **Educational Outreach** initiative of NSS, IIT Bombay *(2022)*
- Elected as the **school headboy** and led and organized various intra-school competitions, workshops and events *(2019)*
- Secured **'A' Grade** in Govt. of Maharashtra's **Elementary** and **Intermediate Drawing Grade Examinations** *(2015, 2016)*