Shukla Ashish Anand

+91 7984966743 | shukla24@iiserb.ac.in

Education

Indian Institute of Science Education and Research (IISER), Bhopal — Bhopal, MP

Bachelor of Technology (B.Tech) in Data Science and Engineering, ***CPI: 8.22 / 10.0*** — *Aug. 2024 – May 2028 (Expected)*

- **Relevant Coursework**: Data Structures & Algorithms, Probability & Statistics, Linear Algebra, Discrete Mathematics, Econometrics, Signals and Systems.
- **Achievements**: Secured All India Rank 1241 in the highly competitive IISER Aptitude Test (IAT).

Technical Skills

Research Interests: Multimodal ML, Speech & Audio Processing, Information-Theoretic Learning, Edge Deployment, Audio-Visual Speech Recognition.
Languages: Python, C/C++.
Libraries & Frameworks: PyTorch, OpenCV, NumPy, Pandas, Streamlit.
Developer Tools: Git, Linux (Bash), SQLite, Remote GPU Clusters.
Core Competencies: Machine Learning, Probability & Statistics, Computer Vision, State Space Models (Mamba), Dynamic Programming.

Experience

Speech Lab, SCSE @ Nanyang Technological University (NTU) — Remote

Research Intern (Automatic Speech Recognition) — *Aug. 2026 – Dec. 2026*

- Undertaking a six-month research internship at the Speech Lab, SCSE, Nanyang Technological University (NTU), focusing on Automatic Speech Recognition (ASR).

Speech Lab, LTRC @ IIIT Hyderabad — Hyderabad, India

Research Intern — *May 2026 – Present*

- Engineered an unsupervised pitch stylization pipeline for downstream speech modeling by leveraging Information-Theoretic concepts, including Shannon Entropy, Mutual Information (MI), and KL Divergence.
- Applied Dynamic Programming (DP) and a Weighted Least Squares (WLS) regression engine to compute optimal piecewise contour segmentation.
- Derived a Bayesian Information Criterion (BIC/MDL) to autonomously determine model complexity, achieving a 38% RMSE reduction over heuristic baselines.
- Currently extending the algorithmic foundation to extract high-fidelity prosodic features for training robust Deep Learning speech representation models.

Visual Data Computing Group (VisDom) — Bhopal, India

Research Intern — *2025 – May 2026*

- Worked under Prof. Vinod K. Kurmi on multimodal and speech research, contributing to PRISM (accepted at ACM CIKM 2026), Mamba-AVSR, and VaaniAdapt.

Publications & Projects

PRISM (First Author) | *Accepted at ACM CIKM 2026, Rome, Italy* — 2026

- **Topic**: Unsupervised Test-Time Adaptation for Audio-Language Models under Domain Shift.
- Developed a fully blind adaptation framework to improve audio-text alignment without relying on batch-level data or oracle prompts.
- Engineered a feature-smoothing algorithm that iteratively aligns corrupted audio embeddings with clean text representations to mitigate severe environmental noise.
- Paper available at arxiv.org/abs/2608.15037.

Mamba-AVSR | *PyTorch, State Space Models, Llama-3.2, Multimodal LLMs* — Dec. 2025 – Present

- Prototyped an end-to-end Audio-Visual Speech Recognition pipeline, integrating Whisper-v3 and VideoMAE encoders with a Llama-3.2 backend via early-fusion embeddings.
- Experimented with replacing $O(N^2)$ transformer cross-attention with a linear-time $O(N)$ Mamba-2 connector to handle the complex synchronization of 50Hz audio and 15Hz video streams.
- Investigated advanced alignment techniques, including Odds-Ratio Preference Optimization (ORPO), to penalize LLM hallucinations and enforce grounded transcriptions.

VaaniAdapt | *Continual Learning, Multilingual ASR* — 2026

- Proposed a rehearsal-free continual learning framework that lets a deployed speech recognition system learn new languages without storing or replaying older speech data.

- Designed a low-rank residual adapter for each new language, trained on a frozen base model with an orthogonality constraint to reduce interference with adapters learned earlier.

DermaScribe | *System Architect & AI Lead (Python, SQLite, OpenCV)* Nov. 2025

- Architected a hybrid offline-online database system using SQLite and FTS5 Indexing, achieving sub-millisecond search speeds for healthcare queries.
- Developed a cost-optimization engine that caches AI API responses locally, significantly reducing inference costs and latency for repetitive queries.
- Built the AI Face Check module using OpenCV to detect and classify facial skin conditions, integrating the backend into a Streamlit web interface.
- Awarded 1st Place (Healthcare Track) at IISER Bhopal DSE Day Hackathon for robust, applied system design.