

KAVYA RANJAN SAXENA

Ph.D. Researcher in Machine Learning for Audio Signal Processing | IIT Kanpur, India

(+91) 9990859330

kavya.ranjan.94@gmail.com

krsaxena

0000-0002-6590-2019

SUMMARY

I have a demonstrably strong research background in Music Information Retrieval (MIR), Domain Adaptation, and exploratory work in Speech LLMs. I have experience in designing scalable research prototypes and curating high-quality datasets. Skilled at translating complex research ideas into practical, implementable solutions that support experimentation, reproducibility, and real-world applicability. Enthusiastic about cross-disciplinary collaboration, mentoring, and advancing state-of-the-art in audio ML with a focus on measurable, practical outcomes. I have submitted my Ph.D. thesis and am currently awaiting my final defense.

WORK EXPERIENCE

Intern - Speech Research Scientist

Krutrim AI

08/2024 - 03/2025

Bangalore, India

- Explored a few SOTA Speech LLMs, and performed a comparative study on the datasets for ASR tasks.
- Explored various methods used for generating synthetic data, and curated a synthetic dataset called IndicST for AST tasks ([published in ICASSPW 2025](#))

Tutor

IIT Kanpur

12/2023 - 01/2024

Kanpur, India

- Delivered foundational concepts of machine learning to working professionals enrolled in the e-Masters program. Assisted with course content, coding assignments, and live Q&A sessions to clarify ML intuition and applications.

Teaching Assistant

IIT Kanpur

06/2022 - 09/2023

Kanpur, India

- Delivered foundational concepts of machine learning to working professionals enrolled in the e-Masters program. Assisted with course content, coding assignments, and live Q&A sessions to clarify ML intuition and applications.

Invited Speaker

Queen Mary University of London

London, United Kingdom

https://www.youtube.com/watch?v=l-0r2_q7Xxs

RESEARCH EXPERIENCE

- Bridged a key research gap in domain adaptation for melody estimation by introducing meta-learning techniques (eg, MAML) to MIR. Published supervised and confidence-based adaptation strategies in CODS-COMAD, MLSP and TASLP.
- Designed and developed a CPU-friendly annotation tool for singing melody extraction from polyphonic audio, significantly reducing manual labeling time and efforts. Built with Flask and HTML/CSS/JS. Currently under *double-blind* review.
- Proposed a novel regression-based approach for melody estimation with uncertainty quantification, addressing ambiguity in confidence prediction from the traditionally framed classification problem. Currently *under review* at IEEE TASLP.
- Proposed a novel nonconformity score for conformal prediction to enable reliable post-hoc calibration of epistemic uncertainty, leveraging an Epinet-based Bayesian framework to address ambiguity in confidence estimation. Currently *in preparation*.
- Identified gaps in the evaluation of speech LLMs for Indian languages and led the development of IndicST, a multilingual synthetic dataset for AST tasks. Enabled benchmarking of models such as Whisper-v2 and SALMONN across resource-scarce Indian languages. Published the dataset in ICASSPW.

PUBLICATIONS

- Kavya Ranjan Saxena, Vipul Arora, 'Tutorial on understanding meta-learning for fast adaptation', CODS-COMAD 2023.
<https://dl.acm.org/doi/pdf/10.1145/3570991.3571030>
- Kavya Ranjan Saxena, Vipul Arora, 'Meta-Learning-Based Supervised Domain Adaptation for Melody Extraction', 2024 IEEE MLSP, London, UK, 2024.
<https://ieeexplore.ieee.org/document/10734810>
- Kavya Ranjan Saxena, Vipul Arora, 'Interactive Singing Melody Extraction Based on Active Adaptation', IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), 2024.
<https://ieeexplore.ieee.org/document/10530096>
- Kavya Ranjan Saxena, Vipul Arora, 'Uncertainty Quantification in Melody Estimation using Histogram Representation', (Under review in TASLP).
<https://arxiv.org/pdf/2505.05156.pdf>
- Sanket Shah, Kavya Ranjan Saxena, Manideep Bharadwaj, Sharath Adavanne, Nagaraj Adiga, 'IndicST: Indian Multilingual Translation Corpus For Evaluating Speech Large Language Models', ICASSPW 2025.
<https://ieeexplore.ieee.org/document/11011192>

SKILLS

Python, Tensorflow, PyTorch, LaTeX, GitHub Actions, HuggingFace(Transformers/Datasets), Technical Writing, Teaching/Mentoring

EDUCATION

IIT Kanpur

Ph.D. (CGPA: 8.68)

- Discipline: Signal Processing
- Relevant Coursework: Machine Learning for Signal Processing, Audio Representation Learning, Neural Networks)

JIIT Noida

Dual Degree - B.Tech+M.Tech (CGPA: 7.5)

- Discipline: Electronics and Communications Engineering

SCHOLARSHIP/AWARDS

UGC-JRF Award Letter 2018