Tanmay Khandelwal

+1 (201) 268-9949 tk3309@nyu.edu /tanmayy24 /tanmayy24 Google Scholar

Education

New York University (NYU), Courant Institute of Mathematical Sciences **Sep 2023 - May 2025**
Master of Science in Computer Science (Concentration in AI) ***GPA:*** *4.0/4.0* *New York, USA*
Relevant Courses: Deep Learning *(Yann LeCun)*, Computer Vision, Advanced Data Structures and Algorithms, Programming Languages

Birla Institute of Technology and Science (BITS), Pilani **Aug 2017 - Jul 2021**
B.E. (Hons) in Electrical and Electronics Engineering ***GPA:*** *8.43/10.0 (First Division)* *Pilani, India*
Relevant Courses: Machine Learning, Data Mining, Object Oriented Programming (OOP), Information Retrieval, Operating Systems

Experience

Music and Audio Research Laboratory (MARL), NYU **Sep 2023 - Present**
Graduate Research Assistant *New York, USA*

- Developing **diffusion-based text-to-audio generation model** integrating **contrastive language-audio pretraining (CLAP)** with a **variational autoencoder (VAE)** decoder and **generative adversarial network (GAN)** vocoder.
- Contributed to an open-source **Python** library: **Soundata** [Link] with **32k downloads**, introducing data loaders, interactive visualization tools, and standardized usage for enhanced reproducibility; submitted to **Journal of Open Source Software (JOSS)**.

Fortemedia - Nanyang Technological University (NTU) **Sep 2021 - Jul 2023**
Machine Learning Engineer *Singapore*

- Engineered **low-complexity** acoustic event detection system for **speech recognition** embedded devices, integrating **attention modules** and **Bi-GRU**, resulting in a **34.1%** improvement in PSDS metrics while reducing model complexity by **27.6%**.
- Devised **multi-task learning (MTL)** framework with two-stage **semi-supervised learning (SSL)** system for **speech-to-text software**, employing **transformer** and **conformer** to model global and local sequences; Improved PSDS by **45.5%**.
- Developed and **deployed** a scalable, **real-time** infant cry detection system utilizing **depthwise-separable convolutions** and **Django REST APIs** for backend services; Achieved an F-score of **0.738** on the curated dataset.

Bajaj Finserv Health Limited **Jul 2020 - Aug 2021**
Software Developer Intern *Pune, India*

- Engineered personalized medication **recommendation system** with past history and patterns-based quick suggestions using **Elasticsearch** and **SQL**, reducing E-consult time with **response time** $<$ **50ms**.
- Pioneered highly scalable **microservices** for a doctor's practice management system in **SpringBoot** and **NodeJs**, integrated with **MongoDB** and **Redis**, using **Docker** and **Kubernetes** for **1000+ daily** E-consult users.
- Generated automated data scraping using **Selenium**, integrated third-party services, implemented **CI/CD system** with **Azure DevOps**, conducted **unit testing**, minimized code duplicity and set up **Elasticsearch, Logstash, and Kibana** stack monitoring.

HTIC, Indian Institute of Technology (IIT) Madras **May 2019 - Jul 2019**
Machine Learning Intern *Chennai, India*

- Created a **natural language processing (NLP)**-based **search tool** for the Brain Architecture Portal, utilizing **spaCy**, **word2vec**, and **cosine similarity** for summarizing neuroscience articles, saving **30 hours+** of manual review.
- Enhanced text comprehension and evidence detection using **fine-tuned BERT** for neuroscience-specific language with **73% NTC**.

Publications

- **T. Khandelwal**, and R.K. Das, "A Multi-Task Learning Framework for Sound Event Detection (SED) using High-level Acoustic Characteristics of Sounds", ***INTERSPEECH 2023*** [DOI]
- **T. Khandelwal**, and R.K. Das, "Cross-Dimensional Interaction with Inverted Residual Triplet Attention for Low-Complexity Sound Event Detection", ***DCASE Workshop 2023*** [DOI]
- **T. Khandelwal**, R. K. Das, A. Koh and E. S. Chng, "Leveraging Audio-Tagging Assisted Sound Event Detection using Weak Generated Labels and Frequency Dynamic Convolutions (FDY-CRNN)", ***IEEE SSP 2023*** [DOI]

Skills

Languages: Python, C++, Java, JavaScript, Typescript, Matlab, Scala, ML
Tools: GitHub, Apache Kafka, Spark, Hadoop, Kubernetes, Docker, Git, GPU, Nvidia CUDA, MySQL, Agile
Frameworks: PyTorch, TensorFlow, PyTorch Lightning, Keras, OpenCV, Hugging Face, SciKit-learn, Matplotlib/Seaborn, Numpy, Pandas, NLTK, spaCy, XGBoost, Kubeflow, Prometheus, Grafana, Microsoft Azure, AWS (Sagemaker, Lambda, EC2), Google Cloud Platform (GCP)

Projects

- Formulated an SED system for DCASE 2023 Task 4A leveraging **bi-directional encoder representation from audio transformers (BEATS)** embeddings with **frequency dynamic convolutions (FDY-CNN)** and **asymmetric focal loss (AFL)**; Utilized **data augmentation**, **adaptive post-processing**, and **ensemble** learning for elevated generalization; Achieved **4th** position. [Link]
- Established a multimodal **visual question answering (VQA)** system using **CNNs, LSTMs, and a stacked attention network (SAN)**, achieving state-of-the-art accuracy **80.5%** on the VQA-v2 dataset and the Radiology dataset. [Link]