Subash Khanal

✉ subash.khanal.cs@gmail.com   🌐 Personal Website   🎓 Google Scholar

EDUCATION

Washington University in St. Louis   **St. Louis, MO, U.S.**
Doctor of Philosophy in Computer Science   *2022 - August 2025*

University of Kentucky   **Lexington, KY, U.S.**
Master of Science in Electrical Engineering (GPA: 4.0/4.0)   *2018 - 2020*

Nitte Meenakhshi Institute of Technology   **Bengaluru, India**
Bachelors in Electronics and Communication Engineering (GPA: 9.54/10 - Gold Medalist)   *2012 - 2016*

WORKING EXPERIENCE

Multimodal Vision Research Lab   **Washington University in St. Louis**
Graduate Research Assistant   *August 2022 - Present*

○ Developed multimodal deep learning frameworks for a shared embedding space between images, audio, and text:
  - **GeoCLAP** (BMVC 2023): Proposed a tri-modal embedding space aligning satellite imagery, audio, and text for large-scale, zero-shot soundscape mapping.
  - **PSM** (ACM Multimedia 2024): Extended a contrastive learning framework with a probabilistic, multi-scale, and metadata-aware embedding space for improved zero-shot soundscape mapping.
  - **Sat2Sound** (Under Review 2025): Developed a state-of-the-art framework that incorporates vision-language model (VLM) outputs to enhance semantic understanding of a place, learns a shared codebook for fine-grained alignment, and supports retrieval-based, location-conditioned soundscape synthesis.
○ Built large-scale, ML-ready multimodal datasets:
  - **GeoSound**: A multimodal dataset combining multi-resolution satellite imagery, geotagged audio, and textual descriptions.
  - **EarthCaps**: A global dataset of ~2.5M satellite images covering the entire landmass of the Earth and ~10M audio and visual captions generated using a large multimodal language model.
○ Contributed to **RANGE** (CVPR 2025), a retrieval-augmented method for learning multi-resolution geolocation embeddings.
○ Worked on additional multimodal embedding methods:
  - **Sat2Cap** (EarthVision 2024) and **GeoBind** (IGARSS 2024): Frameworks for query-driven multimodal mapping.
  - **TaxaBind** (WACV 2025): A unified embedding space for ecological applications.
○ Contributed to models for ecological modeling and fine-grained visual classification:
  - **LD-SDM** and **BirdSAT** (WACV 2024) : Approaches for species distribution modeling and classification using LLM-driven learning and self-supervised embeddings.
○ Contributed to diffusion-based generative modeling projects:
  - **GeoSynth** (EarthVision 2024): A framework for controlled satellite image synthesis.
  - **MVPS** (TMLR 2025): A method for panoramic target view synthesis using satellite imagery as an additional cue alongside nearby panoramas.
○ Collaborated on the IARPA-funded **SMART** project:
  - Developed self-supervised learning methods (e.g., Masked Autoencoders) for remote sensing applications such as semantic change detection.

Valuation and Market Dynamics   **Zillow Group**
Applied Scientist Intern   *May 2024 - August 2024*

○ Explored directions to improve *Zestimate* by incorporating diverse features into predictive models.
○ Developed a multimodal model for home sales price estimation using structured tabular data and floorplan images.
○ Achieved approximately 3% improvement across all key metrics compared to existing baselines.

Lin Brain Lab   **University of Kentucky**
Graduate Research Assistant   *August 2020 - August 2022*

○ Developed machine learning models using heterogeneous biomedical data, including medical imaging, genetic data, and electronic health records, to support Alzheimer's disease research.
○ Designed CNN and Vision Transformer models trained on MRI/PET scans for early AD prediction.
○ Proposed **CAT-XPLAIN**, a causality-based interpretable ViT model that is explicitly trained to identify key image regions contributing to predictions.

Speech and Signal Processing Lab   **University of Kentucky**
Graduate Research Assistant   *August 2018 - August 2020*

○ Analyzed articulatory and acoustic patterns in native and non-native English speech.
○ Developed an ASR-based Mispronunciation Detection and Diagnosis (MDD) framework using RNNs trained on articulatory and acoustic features.

Kantipur Engineering College   Lalitpur, Nepal

Lecturer   *April 2017 - July 2018*

- Taught undergraduate courses on Microprocessors and Instrumentation.
- Designed and conducted lab sessions in Digital Signal Processing and Microprocessor programming.

KEY SKILLS

- **Core Expertise:** Deep Learning, Self-Supervised Learning, Generative Modeling, Multimodal Learning, Vision-Language Models, Contrastive Learning, Diffusion Models, Geospatial AI.
- **Technical Stack:** Python, PyTorch, Hugging Face, Scikit-learn, Git, Docker, QGIS (basic).
- **Modalities Worked With:** Satellite imagery, audio, text, medical imaging, speech, genomics, tabular data.
- **Applications:** Multimodal Retrieval, Generative AI, Predictive Modeling, Interpretability.

AWARDS AND ACHIEVEMENTS

- COMPEX Scholarship offered by the Indian Embassy in Nepal for undergraduate study in India, 2012–2016.
- ECE Gold Medal, 2016 for graduating with Rank 1 in the department.

PUBLICATIONS

Preprints:

- **Khanal Subash**, Sastry Srikumar, Dhakal Aayush, Ahmad Adeel and Jacobs Nathan, "Sat2Sound: A Unified Framework for Zero-Shot Soundscape Mapping," arXiv:2505.13777, 2025.
- Sastry Srikumar, Xin Xing, Dhakal Aayush, **Khanal Subash**, Ahmad Adeel, and Jacobs Nathan, "LD-SDM: Language-Driven Hierarchical Species Distribution Modeling," arXiv:2312.08334, 2024.
- **Khanal Subash**, Brodie Benjamin, Xing Xin, Lin Ai-Ling and Jacobs Nathan, "Causality for inherently explainable transformers: CAT-XPLAIN," **Spotlight Presentation** at XAI4CV: Explainable Artificial Intelligence for Computer Vision (IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops), arXiv:2206.14841, 2022.

Published Papers:

- Sastry Srikumar, Dhakal Aayush, Xing Eric, **Khanal Subash**, Jacobs Nathan, "Global and Local Entailment Learning for Natural World Imagery, " International Conference on Computer Vision (ICCV), 2025.
- Xiong Zhexiao, Xing Xin, Workman Scott, **Khanal Subash** and Jacobs Nathan, "Mixed-View Panorama Synthesis using Geospatially Guided Diffusion, " Transactions on Machine Learning Research(TMLR), 2025.
- Dhakal Aayush, Sastry Srikumar, **Khanal Subash**, Ahmad Adeel, Xing Eric and Jacobs Nathan, "RANGE: Retrieval Augmented Neural Fields for Multi-Resolution Geo-Embeddings, " IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
- Sastry Srikumar, **Khanal Subash**, Dhakal Aayush, Ahmad Adeel and Jacobs Nathan, "TaxaBind: A Unified Embedding Space for Ecological Applications," IEEE Winter Conference on Applications of Computer Vision (WACV), 2025.
- **Khanal Subash**, Xing Eric, Sastry Srikumar, Dhakal Aayush, Xiong Zhexiao, Ahmad Adeel and Jacobs Nathan, "PSM: Learning Probabilistic Embeddings for Multi-scale Zero-Shot Soundscape Mapping," ACM Multimedia, 2024, doi: 10.1145/3664647.3681620.
- Dhakal Aayush, Ahmad Adeel, **Khanal Subash**, Sastry Srikumar, Kerner Hannah and Jacobs Nathan, "Sat2Cap: Mapping Fine-Grained Textual Descriptions from Satellite Images," IEEE/ISPRS Workshop: Large Scale Computer Vision for Remote Sensing (EARTHVISION), 2024. **Best Paper Award**.
- Sastry Srikumar, **Khanal Subash**, Dhakal Aayush, and Jacobs Nathan, "GeoSynth: Contextually-Aware High-Resolution Satellite Image Synthesis," IEEE/ISPRS Workshop: Large Scale Computer Vision for Remote Sensing (EARTHVISION), 2024.
- Dhakal Aayush, **Khanal Subash**, Sastry Srikumar, Ahmad Adeel and Jacobs Nathan, "GeoBind: Binding text, image, and audio through satellite images," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2024. **Oral Presentation**.
- Sastry Srikumar, **Khanal Subash**, Dhakal Aayush, Di Huang and Jacobs Nathan, "BirdSAT: Cross-View Contrastive Masked Autoencoders for Bird Species Classification and Mapping," IEEE Winter Conference on Applications of

Computer Vision (WACV), 2024.

- **Khanal Subash**, Sastry Srikumar, Dhakal Aayush and Jacobs Nathan, "Learning Tri-modal Embeddings for Zero-Shot Soundscape Mapping," British Machine Vision Conference (BMVC), 2023.

- Xing Xin, Liang Gongbo, Zhang Yu, **Khanal Subash**, Lin Ai-Ling and Jacobs Nathan, "Advit: Vision transformer on multi-modality pet images for alzheimer disease diagnosis," IEEE International Symposium on Biomedical Imaging (ISBI), 2022, doi: 10.1109/ISBI52829.2022.9761584.

- **Khanal Subash**, Chen Jin, Jacobs Nathan and Lin Ai-Ling, "Alzheimer's Disease Classification Using Genetic Data," IEEE International Conference on Bioinformatics and Biomedicine (BIBM) Workshop, 2021, doi: 10.1109/BIBM52615.2021.9669730.

- Brodie Benjamin, **Khanal Subash**, Rafique Muhammad Usman, Greenwell Connor and Jacobs Nathan, "Hierarchical Probabilistic Embeddings for Multi-View Image Classification," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2021, doi: 10.1109/IGARSS47720.2021.9554405.

- **Khanal Subash**, Johnson Michael T., Soleymanpour Mohammad and Bozorg Narjes, "Mispronunciation Detection and Diagnosis for Mandarin Accented English Speech," International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2021, doi: 10.1109/SpeD53181.2021.9587408.

- **Khanal Subash**, Johnson Michael T. and Bozorg Narjes. "Articulatory Comparison of L1 and L2 Speech for Mispronunciation Diagnosis," IEEE Spoken Language Technology Workshop (SLT), 2021, doi: 10.1109/SLT48900.2021.9383574.

- **Khanal Subash**, "Mispronunciation Detection and Diagnosis for Mandarin Accented English Speech," Theses and Dissertations–Electrical and Computer Engineering, 156, 2020, doi: 10.13023/etd.2020.340.