Manish Dhakal

Computer Vision . Large Language Models

🌐 manishdhakal.com.np   in manishdhakal521
✉ mdhakal3@gsu.edu   ⌨ manishdhakal
🎓 Google Scholar   📍 Atlanta, GA, USA
Ⓜ manishdhakal

Research

Applied deep learning for computer vision and LLMs with the motive of *model compression and efficient learning.* Publications at premier venues, such as *WACV, MICCAI, ACCV, IJCNLP-AACL,* etc.

Education

Ph.D. in Computer Science
Georgia State University

Graduate School
August 2024 – 2029*

Supervisor: Yi Ding, Ph.D.

- *M.S. Degree* to be awarded in the spring of 2026.
- Deep learning projects on *2D/3D computer vision and large language models (LLMs),* funded by the US Department of Defense (DoD).
- Research in *efficient machine learning* and *model compression.*
- *Grade: 3.9*

Bachelor in Computer Engineering
Pulchowk Campus, Institute of Engineering, Tribhuvan University

Undergraduate School
November 2017 – April 2022

Thesis Supervisor: Prof. Subarna Shakya

- *Ranked 11th (top 0.1%)* in the engineering entrance exam, competing with *15,000+ candidates*, received *full scholarship* for undergraduate studies.
- Gained knowledge about *significant CS courses* like AI, Image Processing, Data Structure & Algorithm, DBMS, Software Engineering, and so on.
- *Thesis:* Automatic speech recognition for *low-resourced Nepali language* which was later presented at an IEEE conference.

Ongoing Projects

Knowledge Distillation
LLMs

- Developing a robust knowledge distillation method utilizing a *teacher-student paradigm* to transfer complex knowledge efficiently.
- Engineering a feature-transfer mechanism between the teacher and student models' intermediate layers to capture and leverage rich, internal representations, ensuring robustness and high fidelity in model compression and knowledge transfer.

LLM-Guided Adaptation for Vision-Language Models
LLMs, VLMs

- *Substituting visual prompts* with textual descriptions to enable pure LLMs to handle vision-language tasks, thereby mitigating the need for standard Vision-Language Models (VLMs).
- LLM-centric VLM adaptation method where the LLM itself provides preference feedback to optimize the textual descriptions.

Parameter-Efficient Fine-tuning
Point Cloud

- *Graph Features Tuning (GFT)*, a Parameter-Efficient Fine-Tuning (PEFT) method specifically for point cloud data, reducing computational and memory costs compared to earlier approaches (*accepted at WACV*).
- Achieved competitive performance on object classification and segmentation tasks while substantially *reducing the number of trainable parameters* by integrating graph features into deeper layers via skip connections and efficient cross-attention.

Publications

2022-Present

- *Dhakal, M.* , Dasari, V. R., Sunderraman, R., & Ding, Y. (2026, March). GFT: Graph feature tuning for efficient point cloud analysis. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV).* [Paper, Code]
- Mim, S. T., Morris, J., *Dhakal, M.* , Xiu, Y., Gorlatova, M., & Ding, Y. (2025, December). Can a unimodal language agent provide preferences to tune a multimodal vision-language model? In *Proceedings of the International Joint Conference on Natural Language Processing and the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL).*
- Budathoki, A., & *Dhakal, M.* (2025). Adversarial robustness analysis of vision-language models in medical image segmentation. *arXiv preprint arXiv:2505.02971.* [Paper, Code]
- Adhikari, R., Thapaliya, S., *Dhakal, M.* , & Khanal, B. (2024). TuneVLSeg: Prompt tuning benchmark for vision-language segmentation models. In *Proceedings of the Asian Conference on Computer Vision (ACCV).* [Paper, Code]
- *Dhakal, M.* , Adhikari, R., Thapaliya, S., & Khanal, B. (2024). Finetuning vision-language segmentation efficiently with lightweight blocks. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).* [Paper, Code]
- Adhikari, R.*, *Dhakal, M.** , Thapaliya, S.*, Poudel, K., Bhandari, P., & Khanal, B. (2023, October). Synthetic Boost: Leveraging synthetic data for enhanced vision-language segmentation in echocardiography. In *Proceedings of the International Workshop on Advances in Simplifying Medical Ultrasound.* [Paper, Code]
- Poudel, K.*, *Dhakal, M.** , Bhandari, P.*, Adhikari, R.*, Thapaliya, S.*, & Khanal, B. (2024). Exploring transfer learning in medical image segmentation using vision-language models. In *Medical Imaging with Deep Learning (MIDL).* PMLR. [Paper, Code]
- *Dhakal, M.* , Chhetri, A., Gupta, A. K., Lamichhane, P., Pandey, S., & Shakya, S. (2022, July). Automatic speech recognition for the Nepali language using CNN, bidirectional LSTM, and ResNet. In *Proceedings of the International Conference on Inventive Computation Technologies (ICICT).* IEEE. [Paper, Code]

Work Experience

Nepal Applied Mathematics and Informatics Institute for research (NAAMII)   *Lalitpur, Nepal*
Research Assistant   *April 2022 – June 2024*

Supervisor: Bishesh Khanal, Ph.D.

- Developed skills for *object detection and segmentation* tasks on 2D medical images and explored their multi-modal approach (esp. *vision-language models*); also worked on segmentation with *3D mesh* data.
- Demonstrated *strong skills in writing scientific manuscripts,* with multiple papers submitted for review, showcasing the ability to *communicate methodologies, results, and implications* effectively.
- Ensured *reproducibility and modularity in ML projects* by implementing robust methodologies and practices, allowing for the transparent and replicable programming of the projects.

Past Projects

Lower Limb Segmentation   *July 2023 – September 2023*
Medical Imaging

Supervisor: Taman Upadhaya, Ph.D.

- Conducted *training experiments* of different deep learning models on the remote server to segment three bones – knee, pelvis, and ankle – from CT scans of the lower limbs of patients.
- *Deployed a robust Python rest API* on the remote server for the segmentation request from a client, with a pipeline including pre-processing, inference, and post-processing steps.
- Ensured *interoperability, reproducibility, and understandability* of the deployed application using Docker, and well-structured documentation and comments.

Vision Language Segmentation Models (VLSMs) for Medical Images   *May 2023 – August 2024*
Medical Imaging

- Reported zero-shot and finetuned segmentation performance of *4 VLSMs* on *11 medical datasets* using *9 types of prompts* derived from *14 attributes*, prompts are given as text conditioning information.
- Worked with *encoder-decoder architecture* to generate binary segmentation masks for VLSMs.
- Tested the compatibility of the VLSMs (such as *CLIPSeg and CRIS*) pre-trained for open-domain images with medical images.

Object Detection in 2D Orthopantomogram (OPG) Images   *September 2022 – August 2024*
Dental Imaging

- Critically analyzed the *literature and state-of-the-art* models for different segmentation and detection tasks on radiology images of dentistry and their inadequacy.
- Designed and developed the *data annotation tool* for object detection over 2D OPG images.
- Working on identification and localization of dental *anatomical structures and abnormalities* while benchmarking with existing methods like *YOLO, RetinaNet, RCNN, and FastRCNN*.

Segmentation in 3D Teeth Scan   *Summer 2022*
MICCAI Challenge 2022

- Learned about the representation and preprocessing of *3D mesh and point cloud* data.
- Benchmarked with different 3D point cloud segmentation models such as *Pointnet/++ and DeltaConv*.

Nepali AutoComplete and LM   *August 2020 – October 2020*
Open Source Project

- Designed and trained *language model of Nepali (ie. Devnagari transcript)* for the text auto-complete system.
- Programmed the *pre-processing pipeline* to remove the non-Nepali characters from the dataset.

Super-Resolution with GAN (SRGAN)   *May 2020 – August 2020*
Open Source Project

- Implemented *open source model* of *SRGAN* with Keras/TensorFlow.
- Developed the *understanding of generator and discriminator* in GAN-based generative models.

Teaching

Community Eye, ENT & Rehabilitation Center (CEERS)   *Bhaktapur, Nepal*
Trainer   *June 2023 – June 2024*

- *Training a group of interns* to develop medical imaging applications with the use of ML.
- Instructing and guiding them about ML through activities like *paper reading sessions, lecture-lab sessions,* and *topic presentations*.

4th Annual Nepal AI School (ANAIS)   *Kathmandu, Nepal*
Teaching Assistant   *May 2023 – June 2023*

- *Guided participants* through a series of labs related to *neural networks, transformers, federated learning, graph neural networks, active learning,* and so on.
- *Mentored three groups* during the 10-day *machine learning hackathon* (namely, Hack-a-Dev).

Software Fellowship, Locus 2021   *Online*
Programming Instructor   *Summer 2021*

- Provided tutoring on *software development life cycle* and assisted participants with *software documenta- tion* and *library/framework installation*.
- Taught participants about *API development for web applications,* emphasizing its concepts, best practices, and usage.

Achievements/Services

2025	*AAAI Reviewer*
2024	*Presidential Fellowship* at GSU (Only 6% of the doctoral students).
2024	*LMIC Travel Grant* by the conference of MICCAI.
2017	*Undergraduate* funded by the Government of Nepal.

Technical skills

Deep Learning	*Unimodal and multimodal learning*, large language models (LLMs), model quantization, efficient learning, reinforcement learning fine-tuning, 2D/3D vision, convolution, and transformers.
Writing	*Knowledge synthesis* from the existing literature, *writing scientific documents* and *manuscripts* with *LaTex*, and *communicating the results* to the community with transparency.
Remote Server	Able to work with *remote Linux machines* for coding and project deployment using *SSH, shell script, tmux, Nginx, and Docker*.

References (Research Advisors)

Yi Ding, Ph.D.
Assistant Professor, *Department of Computer Science, Georgia State University*
yiding@gsu.edu

Bishesh Khanal, Ph.D.
Research Director, *Nepal Applied Mathematics and Informatics Institute for research (NAAMII)*
bishesh.khanal@naamii.org.np

Prof. Subarna Shakya
Professor*, Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University*
drss@ioe.edu.np

Taman Upadhaya, Ph.D.
Staff Scientist, *Cedars-Sinai* **||** Adjunct Research Scientist, *Nepal Applied Mathematics and Informatics Institute for research (NAAMII)*
taman.upadhaya@naamii.org.np