

THAI-HOC NGUYEN

AI Researcher

☎ (+84) 834.365.442 ✉ thaihocit02@gmail.com 🎓 Scholar 🐙 Github 🏠 Homepage

Education

University of Information Technology and Communication

2020 – 2025

Bachelor of Engineering in Data Science & Artificial Intelligence

GPA: 3.03/4.0

• **Core Coursework:**

- * Programming: Python, C/C++, R, Java.
- * Math for AI: Linear Algebra & Calculus, Probability & Statistics.
- * Core AI: Predictive Modeling, ML, DL, NLP, Image Processing, Big Data, Data Processing & Data Visualization.

• **Capstone Project:** *A Study on Ensemble Learning for Cervical Cancer Classification* (Grade: **9.4/10**).

- * Developed an ensemble framework using SOTA backbones to automate cervical cytology screening on real-world dataset **CC-VN2025** from Hospital A, Thai Nguyen.
- * Demonstrated that ensemble learning significantly outperformed single-model baselines in sensitivity and specificity for early cancer detection.

Work Experience

Institute of Applied Science and Technology

Aug 2023 – Aug 2024

Intern Computer Vision Researcher

Thai Nguyen, Viet Nam

- Developed and optimized SOTA deep learning models (CNNs, Transformers) for cervical cancer classification on imbalanced clinical datasets.
- Implemented ensemble strategies and advanced data augmentation, significantly improving sensitivity and specificity for early-stage tumor detection.
- Deployed and customized large-scale annotation platforms (CVAT, Label Studio) on local servers, establishing standardized data labeling workflows.
- Conducted literature reviews and benchmarks, contributing to the public CC-VN2025 dataset for cervical cytology research.

Institute of Applied Science and Technology

Aug 2024 – Present

AI Researcher

Thai Nguyen, Viet Nam

- Investigated and implemented self-supervised speech models (Wav2Vec 2.0, HuBERT, Whisper) for Mispronunciation Detection and Diagnosis (MDD).
- Engineered an end-to-end scoring and diagnostic pipeline for TNEDU AI project, providing multi-dimensional assessment (phoneme accuracy, prosody, fluency).
- Developed unified training strategy integrating diverse datasets (L2-Arctic, TIMIT, SpeechOcean) with domain adaptation for non-native speech.
- Optimized phonetic alignment and feature extraction to enhance sensitivity in detecting subtle mispronunciations, improving correlation with human evaluations.
- Architected and deployed scalable RESTful APIs for real-time MDD inference with low-latency feedback in interactive English learning apps.

Research Interests

My research focuses on architecting scalable and robust AI systems, bridging theoretical breakthroughs and industrial deployment. I specialize in multimodal learning (LLMs, VLMs), 3D/medical vision, and speech processing, supported by strong foundations in Big Data Engineering (Hadoop/Spark) and MLOps. My goal is to deliver optimized, end-to-end AI solutions for high-impact domains such as healthcare and education.

Publications

1. Thi-Xuan Tran[†] and Thai-Hoc Nguyen[†]. [An Automatic Machine Learning Based Customer Segmentation Model with RFM Analysis](#). In *International Conference on Advances in Information and Communication Technology (ICTA)*, 2024.
2. Van-Khanh Tran[†], Thai-Hoc Nguyen[†], Chi-Cuong Nghiem and Xuan-Lam Dinh. [A Study on Ensemble Learning for Cervical Cytology Classification](#). In *International Conference on Advances in Information and Communication Technology (ICTA)*, 2024.

[†] Equal contribution.

Technical Skills

🔗 **Languages:** Python, C/C++, SQL, L^AT_EX.

🧠 **ML/DL Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn.

📁 **Domain-Specific:** TorchAudio, Librosa, OpenCV, Albumentations, Pandas, NumPy, SciPy, Matplotlib.

🔧 **Engineering & MLOps:** Docker, Git, Hadoop, Spark, MLflow, WandB, FastAPI, Gradio, Streamlit, Linux.

👤 **Soft Skills:** Scientific Writing, Research Methodology, Cross-functional Collaboration, Problem-solving.

Honours and Awards

- **Outstanding Research Student Award.** Recognized for exceptional achievements in scientific research and dataset contributions. At *University of Information and Communication Technology (ICTU)*, 2024.
- **Student of 5 Merits.** Awarded to top students demonstrating excellence in: Morality, Academics, Physical Fitness, Volunteering, and Integration. Achieved at both *University level (ICTU)* and *Regional level (Thai Nguyen University)*, 2024.
- **Outstanding Student.** Maintained distinction GPA for consecutive academic years. At *University of Information and Communication Technology (ICTU)*, 2023 – 2025.

References

Lab Director

Email: tvkhanh@ictu.edu.vn

Dr. Tran Van Khanh

Institute of Artificial Intelligence (IAI)

Relationship: Internship Supervisor

Lab Director

Email: dxlam@ictu.edu.vn

Dr. Dinh Xuan Lam

Institute of Applied Science and Technology (IAST)

Relationship: Project Mentor