VU VIET ANH

Address: 12A/99 Trung Kinh Street, Yen Hoa Ward, Hanoi, Viet Nam

Mobile: 0989894650

Gmail: mrvietanh2@gmail.com

OBJECTIVES

As a graduate of Hanoi University of Science and Technology with a strong academic foundation and hands-on research experience in Machine Learning (ML), Deep Learning (DL), and Large Language Models (LLMs), I am seeking an opportunity to work in an academic environment where I can contribute to both teaching and research. I am deeply passionate about AI and ML algorithms, and I have a strong desire to inspire and educate future generations in this field. My long-term goal is to pursue advanced studies and research at higher levels, aiming to make meaningful contributions to the academic and scientific community.

IN BRIEF

- Experience in ML (Ensemble Learning, Time-series data), Deep Learning (Computer Vision), and Large Language Models (LLMs), especially in healthcare applications.
- Hands-on research and implementation of LLM-based systems for natural language understanding and question answering in the medical domain.
- Programming Languages: Python, Java, C++, SQL
- Machine Learning Frameworks: TensorFlow, Keras, PyTorch
- Data Analysis Tools: Pandas, NumPy, SciPy, Matplotlib
- Experience in Computer Vision algorithms and techniques, such as object detection, segmentation.
- Nearly 2 years of hands-on experience working in Machine Learning, Deep Learning and LLM-based projects across academia and industry.

EDUCATION & ACHIVEMENTS

- Hanoi University of Science and Technology, Vietnam — Octorber 2020 – July 2024
  BSc with Honours, in Computer Science (**CPA: 3.48/4.0**)
- **The Second – Prize Winner** in WIDS DATATHON 2023
  Adapting to climate change by improving extreme weather forecasts
- Certification: IBM python for Data Science

RESEARCH INTERESTS

1. Machine Learning and Deep Learning
2. Large Language Model

LANGUAGES

- **VSTEP B2** — March 25, 2025
- Ability to read and comprehend English documents
- Basic communication skills

WORK EXPERIENCE

❖ **Hanoi University of science and Technology** September 2022 – now

Computer Science Student

Project: **Fuzzy Learning for classification problems and Learning about the core AI model**

- Developed a new model for classification problems based on Intuitionistic Fuzzy Set (IFS) theory. The new model is named **CS-IFS**, a classification model based on fuzzy measurement scales.
- Utilized new fuzzy measurement scales, such as Ngân, Manhattan, and Euclid.
- The project included phases of data collection, algorithm development, algorithm implementation, evaluation, and user interface deployment.

Project: **Adapting to climate change by improving extreme weather forecasts (WIDS datathon 2023 - Kaggle)**

- This project involves developing a climate change prediction model based on machine learning models.
- Processed data in a time-series format and applied ensemble techniques to enhance the performance of the machine learning models.
- Specifically utilized the latest prediction models such as Gradient Boosting, Light Gradient Boosting, and CatBoost, and combined them.
- This project was part of a competition held at the end of 2022 on the Kaggle platform aimed at predicting climate change.

Project: **Vietnamese Speaker Verification (VLSP 2021 competition)**

- This project involves developing a Vietnamese speaker verification system using deep learning models.
- Processed voice samples to create speaker-specific acoustic features for enrollment and verification.
- Specifically utilized models such as ECAPA_TDNN, ECAPA_CNN_TDNN, and RawNet3, and experimented with loss functions like triplet, softmax, and proto.
- This project is part of the VLSP competition focusing on speaker verification, aiming to accurately determine if two voice samples are from the same person.

Project: **Aspect-Based Sentiment Analysis (ABSA - NLP LAB)**

- Built a Vietnamese Aspect-Based Sentiment Analysis system using SA-Transformer.
- Combined NLP techniques such as POS tagging and Biaffine Parsing to extract aspect–sentiment pairs.
- Collected and annotated domain-specific data to train and evaluate the model.
- Achieved improved accuracy in aspect-level sentiment classification compared to baseline methods.

❖ **FPT Information System Company** August 2023 – May 2024

AI Engineer Intern

Project: **Image captioning**

- This project involves developing an image captioning system using advanced deep learning models.
- Processed image data to generate descriptive captions that accurately reflect the content of the images.
- Specifically utilized state-of-the-art models such as Transformers and Convolutional Neural Networks (CNNs) integrated with attention mechanisms to enhance caption generation.
- This project aims to create an effective solution for automated image description, leveraging cutting-edge techniques in the field of computer vision and natural language processing.

Project: **OCR for Financial Report Recognition**

- Developed an OCR system to recognize and extract key information from financial reports of banks.
- Applied state-of-the-art OCR models such as PaddleOCR, Tesseract, and TrOCR (Transformer-based OCR) integrated with NLP for data structuring.
- Utilized LayoutLM / LayoutLMv3 for document layout understanding to improve extraction of tables, figures, and key-value pairs.
- Implemented pre-processing for document images (deskewing, binarization, noise reduction) to improve text detection and recognition accuracy.
- Built post-processing pipeline with Regex + Named Entity Recognition (NER) to normalize extracted information.
- Deployed the model as part of an AI pipeline to assist financial data analysts.

❖ **Vietsens Technology Group** September 2024 – now

AI Engineer

Project: **AI Healthcare Assistant for Hospital Management**

- Developed an AI-integrated solution for healthcare, deployed within the National Hospital Management Software, particularly adopted by Bach Mai Hospital.
- Utilized advanced Machine Learning and Natural Language Processing (NLP) techniques, including Large Language Models (LLMs), for real-time medical support.
- Designed and implemented modules for patient data analysis, case management assistance, and initial diagnosis support.
- Built interactive AI agents to assist doctors and medical staff with medical history inquiry, prescriptic diagnostics.
- Ensured seamless integration with existing hospital information systems (HIS) for improved operational efficiency and patient care.

Project: **Prescription Recommendation and Conflict Detection System**

- Developed a predictive system to support doctors in prescribing medications based on patient information.
- Utilized data analysis techniques to preprocess and extract features from electronic health records.
- Applied traditional machine learning classification models (Decision Trees, Random Forests, SVM) and experimented with advanced classifiers for improved accuracy.
- Integrated Large Language Models (LLMs) to evaluate prescriptions, detect duplicate active ingredients, and identify potential drug–drug interactions.
- Delivered a decision-support tool that helps healthcare professionals reduce prescription errors and improve patient safety.

RESEARCH EXPERIENCE

Student Research with Dr. Kate Han and Dr. Thanh Nguyen March 2023 – now

- Researched Machine Learning (Ensemble Learning, Data Stream Learning) and Deep Learning architectures
- Genetic Algorithms (GA)
- Evaluation and Hypothesis Statistical Tests

Project: **VISTA:Variable-Length Genetic Algorithm and LSTM-Based Surrogate Assisted Ensemble Selection algorithm in Multiple Layers Ensemble System**

- This project focused on developing the VISTA method, which integrates a Variable-Length Genetic Algorithm (VLGA) with an LSTM-based surrogate model for optimizing ensemble selection in Multiple Layers Ensemble Systems (MLES).
- Researched existing ensemble learning methods and surrogate-assisted evolutionary algorithms (SAEA).
- Proposed a new method combining VLGA with LSTM to transform variable-length encoding into fixed-size representations for fitness prediction.
- Implemented a multi-layer ensemble system where each layer's ensemble of classifiers (EoC) is trained on both original data and predictions from the previous layer.
- Conducted experiments on 15 popular datasets, demonstrating that VISTA outperforms benchmark algorithms.
- Successfully developed and validated the VISTA method, highlighting its efficacy in improving the performance of MLES.

Project: **Imputation Surveys**

- Researched Imputation Surveys in UCI Machine Learning Repository (15 datasets)
- Deployed some Imputation methods (SimpleImputer, KNN, MICE, GINN,…) for missing data and evaluation in the Classification domain

Student Research with Assoc.Prof. Pham Van Hai September 2023 – now

- Research Machine Learning Algorithm and Deep Learning models
- Computer Vision (CV)

Project: **Application of deep learning network in underwater image recognition and segmentation (Manuscript under review)**

- This project involves applying deep learning networks for underwater image recognition and segmentation.
- Processed underwater images to improve clarity and segment aquatic organisms accurately.
- Specifically utilized ResNet for image denoising, followed by Unet_v2 and V_UNet for segmentation, combining advanced architectures like Unet with Spatial Dropout and Vgg16 for enhanced performance.
- The project aims to address the challenges of underwater image analysis by leveraging modern deep learning techniques to achieve accurate recognition and segmentation of marine life.

Project: **Lightweight Facial Expression Recognition using Attention-based MobileNetV2 (Manuscript under review)**

- Designed a lightweight deep learning model for facial expression recognition optimized for deployment on edge devices with limited computational resources.
- Utilized a truncated MobileNetV2 backbone for efficient feature extraction from facial images.
- Proposed a Patch Extraction Block to divide the feature maps into non-overlapping regions, enabling the model to focus on local facial features even under occlusion or head pose variations.
- Integrated an Attention Classifier with self-attention mechanism to enhance feature discrimination and classification accuracy.
- Achieved competitive results on standard datasets (RAF-DB, FER2013, FERPlus) and a custom real-world challenge dataset with occlusion and pose variations.

PUBLICATION

- Kate Han, Truong Thanh Nguyen, **Viet Anh Vu**, Alan Wee-Chung Liew & Tien Thanh Nguyen. 2024. *"VISTA: A Variable Length Genetic Algorithm and LSTM-Based Surrogate Assisted Ensemble Selection algorithm in Multiple Layers Ensemble System"* – IEEE_SCCI 2024.