Thanh Tu, Do

MASTER STUDENT · FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vietnam National University, University of Science, Ho Chi Minh City, VN

✉ tu.dothanh1906@gmail.com | ⌂ www.iamtu.dev | ⊡ young1906 | +84 343 610 772

Research Interest

My research interest lies in the application of Machine Learning and Deep Learning techniques in Sciences. Particularly, I am interested in Probabilistic Machine Learning because it provides a language to describe and model the data-generating process, from which we can have a better understanding of the underlying process being studied.

Education

Vietnam National University, Ho Chi Minh University of Science   *Ho Chi Minh City*

MASTER PROGRAM IN DATA SCIENCE, FALCUTY OF MATHEMATICS AND COMPUTER SCIENCE.   *Jan 2022 - Current*

- **Advisor**: Dr. Thu Nguyen - Simula Met, Norway
- **Thesis**: Blockwise Principal Component Analysis for monotone missing data imputation and dimensionality reduction
- **GPA**: 8.85

Foreign Trade University of Vietnam, Hanoi Campus   *Hanoi, Vietnam*

BACHELOR OF INTERNATIONAL BUSINESS AND ECONOMICS   *July 2011 - May 2015*

Research Experience

School of Biomedical Engineering, International University   *Ho Chi Minh, Vietnam*

ADVISOR: DR. HA THANH HUONG   *2019 - 2021*

- **Stress Level Detection During Examination In College**
  - Proposed a method for automatic ocular artifact removal from EEG signal using machine learning technique, which resulted in a proceeding paper.
  - Designed and implemented data preprocessing pipeline with automated artifacts removal.
  - Conducted literature review to define a set of features on the acquired EEG signal, such as different entropy features, asymmetry index, and power spectral density.
  - Implemented feature extraction pipeline on preprocessed data and performed data analysis and visualization on acquired features.

Vietnam National University, University of Science   *Ho Chi Minh City, Vietnam*

CO-ADVISORS: DR. THU NGUYEN, DR. BINH T. NGUYEN   *Dec 2021 - Dec 2023*

- **Missing data imputation**
  - Performed literature review on current methods of handling missing data.
  - Implemented PCAM algorithm, a novel method to estimate lower dimension representation of randomly missing data.
  - Conducted experiment to measure the effectiveness of PCAM against other baseline methods.
  - Wrote manuscript of the proposed method to submit to peer review.

Vietnam National University, University of Science   *Ho Chi Minh City, Vietnam*

ADVISORS: DR. THU NGUYEN   *Dec 2021 - Dec 2023*

- **Imbalance learning problem**
  - Proposed a method of addressing the imbalance learning problem by *classifier guided MCMC* to oversample the minority class.
  - Proposed to use *Conditional Variation Autoencoder* to generate samples from minority class.
  - Implemented the proposed algorithms and compare results against other baseline methods.

Publications

ACCEPTED

(**BME 2020**) **Tu, Do Thanh**, Thuong Nguyen, Anh Tho Le, Sinh Nguyen, Huong Ha. *"Automated EOG removal from EEG signal using Independent Component Analysis and Machine Learning Algorithms"* at The 8th International Conference in Vietnam on the Development of Biomedical Engineering.

(**ICHST 2023**) **Tu, Do Thanh**, Luan Van Tran, Tho Anh Le, Thao Mai Thi Le, Lan-Anh Hoang Duong, Thuong Hoai Nguyen, Anh Minh Hoang An, Duy The Phan, Khiet Thu Thi Dang, Quyen Hoang Quoc Vo, Nam Phuong Nguyen, Huong Thanh Thi Ha. *"Stress prediction using machine-learning technique on physiological signal"*

(**KSE 2023**) Lien P. Le, **Tu T. Do**, Thu Nguyen *"Data Imputation for Multivariate Time-series Data"*

(**IJCNN 2024**) **Tu T. Do**, Mai Anh Vu, Hoang Thien Ly, Thu Nguyen, Steven A. Hicks, Michael A. Riegler, Pål Halvorsen Halvorsen and Binh T. Nguyen. *"Blockwise Principal Component Analysis for monotone missing data imputation and dimensionality reduction"*

SUBMITTED

Mai Anh Vu[*], Thu Nguyen[*], **Tu T. Do**[*], Nhan Phan, Nitesh V. Chawla, Pål Halvorsen, Michael A. Riegler and Binh T. Nguyen. *"Conditional expectation with regularization for missing data imputation"*

Tu T. Do, Mai Anh Vu, Hoang Thien Ly, Thu Nguyen, Steven A. Hicks, Michael A. Riegler, Pål Halvorsen Halvorsen and Binh T. Nguyen. *"Estimating lower-dimensional space representation in Principal Component Analysis under missing data condition"*

Awards and Certificates

2011	**Second prize, National Physics Olympiad**, VinhPhuc specialized high school
2018	**Certificate of completion Machine Learning Course**, Coder School, Ho Chi Minh City, Vietnam
2019	**Statement of accomplishment with merit, Machine Learning and Deep Learning Foundation**, VietAI, Ho Chi Minh City, Vietnam

Skills

Python Familiar with deep learning and machine learning frameworks such as Tensorflow, Pytorch, Numpy, and Scikit-Learn.

Database Common SQL dialects, such as PostgreSQL, BigQuery, and MySQL.

Linux Basic shell script.

Language proficiency IELTS 7.5, GRE 320