VIETNAM NATIONAL UNIVERSITY, HANOI
VNU UNIVERSITY OF ENGINEERING AND TECHNOLOGY

SOCIALIST REPUBLIC OF VIETNAM
Independence – Freedom – Happiness

INFORMATION ON DOCTORAL THESIS

1. Full name : Giang Thanh Trung 2. Sex: Male

3. Date of birth: 17/02/1986 4. Place of birth: Sonla

5. Admission decision number: 642/QĐ-CTSV Dated: 14/09/2014

6. Changes in academic process:

Changing the role of scientific supervised, that is reflected in the Decision No. 1126/QĐ-ĐT dated December 30, 2016 of the Rector of the University of Technology and Engineering.

Extending the study period, that is reflected in the Decision No. 881/QD-DT dated September 14, 2017 of the Rector of the University of Engineering and Technology.

Returning PhD students to the locality is reflected in the Decision No. 1131/QD-DT dated October 17, 2019 of the Rector of the University of Engineering and Technology.

7. Official thesis title: The study on Dimensionality Reduction Methods, Applying in Patients Classification.

8. Major: Information System 9. Code: 9480104.01

10. Supervisors:

1. Assoc. Prof. Tran Dang Hung - Hanoi National University of Education.

2. Dr. Le Nguyen Khoi - University of Engineering and Technology, VNU.

11. Summary of the **new findings** of the thesis:

In the research scope, there are three research objectives which are are completely achieved, the details are as follows:

Firstly, the thesis has analyzed the strengths and limitations of two effective dimensionality reduction methods in details including Multiple Kernel Learning for Dimensionality Reduction (MKL-DR) and Robust Principal Component Analysis (RPCA). The thesis recommends the improvement method for fMKL-DR by improve its performance in time-calculation to propose fMKL-DR method. The proposed method

significantly reduces its time, then increases its applicability while the number of observational samples is more and more increasing.

Secondly, the thesis has built two patient classification models for two diseases of cancer and Alzheimer. These two models with fMKL-DR and RPCA methods have helped reduce data dimensions. Furthermore, such models have integrated data from different datasets which helps take advantages of useful information in each set. These models have had great contributions because nowadays, each object has been observed in multiple ways, and each observed dataset has different useful information. Therefore, integrating information from different datasets is the tendency for present data analysis methods.

12. Practical applicability:

From our proposed patient classification model in the thesis, we can possibly easily build toolkits for classifying cancer patients and Alzheimer's patients. These toolkits will practically apply in the diagnosis and treatment of these diseases. Moreover, it is possible to easily extend to build a patient classification toolkit for other diseases.

13. Further research directions

Studying new dimensionality reduction methods and improving existing data dimensionality reduction methods, focusing on dimensionality reduction methods, which can be applied in biomedical problems. Moreover, focusing on research on deep learning methods, as it is a proven method to reduce data dimensionality.

Building a number of data dimensional reduction toolkits, which are suitable for processing molecular biology data sets. In particular, the theoretical basis is based on data dimensional reduction methods, which combine data from many different data types to create a unified data set. In particular, the combination of image data and microarray data in the diagnosis and treatment of diseases is a potential research direction because the nature of each data type contains a lot of useful information.

14. Thesis-related publications

1. "Stratifying cancer patients based on multiple kernel learning and dimensionality reduction", In 2017 9th International Conference on Knowledge and Systems Engineering (KSE), IEEE, pp. 106-111, IEEE Xplore, (2017). (Scopus, DBLP)
2. "fMKL-DR: A Fast Multiple Kernel Learning Framework with Dimensionality Reduction", In International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making, Springer, Cham, pp. 153-165, (2018). (DBLP)

3. "Stratifying Patients Using Fast Multiple Kernel Learning Framework: Case studies of Alzheimer’s Disease and Cancers", BMC Medical Informatics and Decision Making, 108 (2020). (ISI Q1, IF = 2.067)
4. "A Combination Model of Robust Principal Component Analysis and Multiple Kernel Learning for Cancer Patient Stratification", The Second International Conference on Artificial Intelligence and Computational Intelligence (AICI 2021), Springer.

This list includes 4 articles