



## To improve classification accuracy of Alzheimer's Disease diagnosis by handling missing data

Liu, S., Ye, Q., Coyle, D., Cao, Y., Yang, H., & Ding, X. (2019). *To improve classification accuracy of Alzheimer's Disease diagnosis by handling missing data*. Poster session presented at TMED 10, Londonderry, United Kingdom.

[Link to publication record in Ulster University Research Portal](#)

### Publication Status:

Published (in print/issue): 03/09/2019

### Document Version

Publisher's PDF, also known as Version of record

### General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

## TITLE

To improve classification accuracy of Alzheimer's Disease diagnosis by handling missing data

## AUTHORS

Shuo Liu<sup>1</sup>, Qinyong Ye<sup>2</sup>, Damien Coyle<sup>1</sup>, Junxiu Liu<sup>1</sup>, Yi Cao<sup>3</sup>, Hongqin Yang<sup>4</sup>, Xuemei Ding<sup>\*1,5</sup>

## AFFILIATIONS

<sup>1</sup>School of Computing, Engineering & Intelligent Systems, Ulster University, Derry~Londonderry, BT48 7JL, UK

<sup>2</sup>Department of Neurology, The Union Hospital of Fujian Medical University, Fuzhou 350005, China

<sup>3</sup>Business School, University of Edinburgh, Edinburgh, EH8 9YL, UK

<sup>4</sup>Fujian Provincial Key Laboratory for Photonics Technology, Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou 350007, China

<sup>5</sup>Faculty of Mathematics and Informatics, Fujian Normal University, Fuzhou 350108, China

\*Email: [x.ding@ulster.ac.uk](mailto:x.ding@ulster.ac.uk) (Dr. Xuemei Ding)

## Background

Increasing senior people suffer from Alzheimer's disease (AD), and it is significant to make accurate and early diagnosis for treatment and improvement of life quality. Many computer-aided systems have been widely used to classify dementia. While clinical data usually contain missing data, which may have an adverse effect on classification accuracy. In order to obtain more convincing classifiers, this work handled missing data using four computational algorithms and evaluated if the classification accuracy is improved with imputed data.

## Materials & Methods

The initial dataset was collected from a local hospital, containing 185 healthy control samples and 187 AD samples. We randomly replaced 20%, 40% and 60% of complete data with missing value from each feature. The mode substitution, K-nearest neighbours (KNN), multiple imputation (MI) and random forest (RF) algorithms were used to impute missing data. The RF classification approach and J48 algorithm were applied on original complete data and the complete data with imputing value.

## Results

The RF algorithm performed best for imputing missing value. The R squared declines with the increasing of the number of missing value. With 40% missing values decreasing to 20% in each feature, the R squared goes from 0.897 down to 0.892. Additionally, the overall classification accuracy was improved with imputed data. The performance of RF classification method, with 89.52% for classification accuracy, 0.954 for AUC and 0.790 for Kappa value, is better than that of J48 (88.63%, 0.944, and 0.718 respectively) for AD classification.

## **Conclusion**

Imputing missing data properly can offer more samples to refine the classifiers with respect to AD diagnosis, which in turn improves the classification accuracy. Future extensions will work on more imputation methods and classifiers.