

# Prediction of dementia risk with community health data using machine learning approaches

Prof. Kup-sze Choi<sup>1</sup>, Dr. Xiao Shen<sup>1</sup>, Dr. Ta Zhou<sup>1</sup>, Mr. Cho Lik Ho<sup>1</sup>, Dr. Guanjin Wang<sup>2</sup>

<sup>1</sup>The Hong Kong Polytechnic University, , Hong Kong, <sup>2</sup>Murdoch University, , Australia

## Objectives

Despite being a global health problem, people with dementia are largely unnoticed, while early detection of dementia is important for timely diagnosis and intervention. In this regard, prediction of dementia risk with health profile of community-dwelling elderly people was proposed by using machine learning techniques. Other than hospital-based data which usually record information of dementia at a later stage, community health data in primary care settings has the potential to render signs or hints of the disease to alert healthcare providers and the elderly. In this study, the machine learning techniques of k-nearest neighbours (KNN) and support vector machine (SVM) were applied to community health profile to classify between normal versus not-normal cases.

## Methods

A set of health data of 298 community-dwelling elderly people, collected during primary healthcare services in different districts of Hong Kong from 2008 to 2018, was used in the study. The dataset included demographic information, bio-measurements and data collected with questionnaires on mobility, nutrition assessment, depression assessment, happiness assessment, pain assessment, etc., constituting a total of 217 features. The dataset also included the scores of mini-mental state examination (MMSE) which served as benchmarks for model training. Two-class classification was concerned, i.e. normal cognition (MMSE score between 24 to 30 points) and not normal (MMSE score between 0 to 23 points). In model construction, the features were normalized to the interval of [0,1] and the dataset were split into training and test set at the ratio of 7:3, for both kNN and SVM.

## Results

Experiments were conducted on the dataset for 100 times and the accuracy of two-class classification with kNN and SVM was  $0.81 \pm 0.033$  and  $0.67 \pm 0.046$  respectively. The accuracy of kNN was better than that of SVM; the latter was also more spread out. In terms of accuracy, kNN was a better method in this study.

## Conclusions

The results suggest that machine learning algorithms can be applied to community health profile to predict dementia risk, which can then give alerts and draw the attention of healthcare providers. Advanced algorithms will be explored to improve classification performance. As the data are collected in primary care settings, the proposed approach has the potential to detect dementia at early stage. Nevertheless, the issue of data imbalance with the dataset may affect the performance since the proportion of the cases of normal cognition is larger. Future work will be conducted to counteract the issue with appropriate computational algorithms.