Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

Cronfa - Swansea University Open Access Repository

_____

This is an author produced version of a paper published in:
The Lancet

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa45957

_____

Paper:

Zhou, S., Tsang, G., Xie, X., Huo, L., Brophy, S. & Lyons, R. (2018). Mining electronic health records to identify
influential predictors associated with hospital admission of patients with dementia: an artificial intelligence approach.
The Lancet, 392, S9

http://dx.doi.org/10.1016/S0140-6736(18)32166-4

Mining Electronic Health Records to Identify Influential Predictors Associated with Hospitalisation of Dementia Patients: An Artificial Intelligence Approach

Shang-Ming Zhou[1], Gavin Tsang[2], Xianghua Xie[2], Lin Huo[3], Sinead Brophy[1], and Ronan A Lyons[1]

[1]The Health Data Research UK Wales and Northern Ireland Site, Swansea University Medical School, Swansea, SA2 8PP, UK.
[2]Department of Computer Science, Swansea University, Swansea, SA2 8PP, UK.
[3]China-ASEAN Research Institute, Guangxi University, Nanning, China.

Background: Early prediction of the outcomes of dementia is significant and challenging. This study aimed to identify influential predictors from primary care electronic health records that can accurately and reliably predict the outcomes of dementia: being admitted to hospital or remaining with the society.

Methods: The dementia patients were collected from general practice (GP) and hospital data in Wales between 1908 and 2015. These records were linked at the individual patient level via the Secure Anonymised Information Linkage databank. The GP records of each patient were selected one year before diagnosis up to hospital admission. An artificial intelligence technique, neural network with entropy regularisation, was used to automatically identify the most influential predictors from initial GP Read codes, gender and age. *K-fold* cross validation was used to assess the predictive performance of the identified signals.

Findings: 52.5 million individual records about 59,298 dementia patients were utilised. Among them, there were 30,178 patients admitted to hospital and 29,120 patients remaining within GP. The male patients had more cases of admissions than those of staying with GP (11233/9441), while the female patients have more cases of staying with GP than those of admissions (19679/18945). From the 54,649 initial event codes, 10 most important signals were identified, including 2 diagnostic events ("*Nightmares*", "*Essential hypertension*"), 5 medication events ("*Serc-16 tablet*", "*Ibugel gel 100g*", "*Simvastatin 40mg tablet*", "*Influvac sub-unit prefilled syringe 0.5mL*", "*Adcal-D3 1.5g/10ug chewable tablet*") and 3 procedural events ("*Third party encounter*", "*Social group 3 - skilled*", "*Blood glucose raised*"). They performed significantly above chance to predict the outcomes of dementia with the sensitivity of 0.758 ± 0.014, specificity of 0.759 ± 0.025, precision of 0.766 ± 0.016, and negative predictive value of 0.751 ± 0.005.

Interpretation: Our findings suggest that machine learning models provide an effective means of identifying influential clinical signals to predict the outcomes of dementia significantly above chance.

Declaration of interests:

We declare no competing interests.