High-dimensional machine learning to predict hospital readmission among older people with chronic kidney disease: abridged secondary publication

Q Zhang *, E Leung, JTF Lau, KKF Tsoi, HY So, WWS Ho, SCY Cheung

KEY MESSAGES

1. A novel factorisation-based machine-learning model, which is Hong Kong–weighted with fuzzy partition, is proposed to predict the risk of hospital readmission among patients with chronic kidney disease.
2. This is one of the first machine-learning models for readmission prediction based on territory-wide data in Hong Kong.
3. Application of this model may help reduce the cost of hospital management and improve the quality of life among patients with chronic kidney disease.

Hong Kong Med J 2023;29(Suppl 1):S24-6
HMRF project number: 16171991

[1] Q Zhang, [2] E Leung, [2] JTF Lau, [2] KKF Tsoi, [3] HY So, [4] WWS Ho, [5] SCY Cheung

[1] *School of Data Science, City University of Hong Kong, Hong Kong SAR, China*
[2] *The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong SAR, China*
[3] *Department of Quality and Safety, Prince of Wales Hospital, Hong Kong SAR, China*
[4] *Department of Medicine and Therapeutics, Prince of Wales Hospital, Hong Kong SAR, China*
[5] *Department of Medicine, Queen Elizabeth Hospital, Hong Kong SAR, China*

* Principal applicant and corresponding author: qingpeng.zhang@cityu.edu.hk

Introduction

Over 20% of older people experience unexpected readmission to accident and emergency department within 1 month of discharge. Identifying at-risk patients is important in geriatric medical care planning. However, in older people with multiple chronic diseases, traditional screening methods or basic machine-learning algorithms are difficult to identify disease risk patterns. We developed a Hong Kong–weighted factorisation machine with fuzzy partition (wFMFP) to predict the risk of readmission among patients with chronic kidney disease (Fig 1).

Methods

A total of 418 305 admission records of 19 457 older patients aged ≥65 years with chronic kidney disease as the primary diagnosis between 2008 and 2017 in Hong Kong public hospitals were retrieved. Of the 418 305 admission records, 392 863 (12 046 patients) were within 30 days of previous discharge to home. Of the 12 046 patients, 4523 (37.55%) had a single readmission and 7523 (62.45%) had multiple readmissions. Data collected included age, district of residence, admission specialty (medicine or clinical oncology), length of stay (LOS), source of admission (emergency department, other hospital, or others), triage category at the emergency department (critical, emergency, urgent, semi-urgent, non-urgent), subacute care (yes or no), LOS of the subacute care, total quantity of drug dispensed, and total prescribed dispensing duration (days).

Among 12 046 patients readmitted within 30 days, the mean patient age was 72.35 years; the mean LOS was 1.4360 days; the mean total quantity of drug dispensed was 14.4189 doses; and the mean total prescribed dispensing duration was 8.7560 days. There were 386 362 admissions to medicine specialties and 51 admissions to oncology specialties. The most common triage category at emergency departments was urgent (42.05%), followed by semi-urgent (26.07%), emergency (17.78%), non-urgent (13.48%), and critical (0.62%). The mean LOS of subacute care was 17.5550 days. The mean haemoglobin A1c was 52.5285 mmol/mol.

Factorisation machines provide a general predictor that can efficiently model high-order interactions among explanatory features in linear time complexity.[1] Matrix factorisation decomposes a matrix into matrices. Tensor factorisation is the high-order extension of matrix factorisation; it enables modelling of heterogeneous and multidimensional data.[2] Matrix and tensor factorisations can extract the latent components to enhance data-mining tasks. In this project, CANDECOMP/PARAFAC factorisation was used for prediction. The admission count was used to construct tensors in the proposed models, and non-negative tensor factorisation methods using generalised Kullback-Leibler divergence and multiplicative update rules were followed.[2,3]

Traditional regression methods usually add clinical attributes as independent variables, ignoring the high-dimensional interrelations and high-order interactions between clinical attributes and diseases. Tensor factorisation provides a powerful framework

to model such multi-aspect data by explicitly exploiting the high-dimensional structure to identify latent clusters of data.[2,3] In this project, we used tensor factorisation to evaluate the risks of all chronic diseases. For each patient, tensor factorisation produced a ranked list of chronic diseases according to predicted risk scores. Risk scores for different diseases in such a rank were used as predictors in a traditional machine-learning model to predict the risk of readmission within a certain period (30 days, 90 days, 1 year).

In the wFMFP model, step 1 is to initialise the number of clusters (equal to the number of sub-wFM models) and the overlap parameter η for training data segmentation. Step 2 is to use fuzzy c-means to obtain membership μis, cluster centres β_s in (1), and the spread width δ_s^q in (2). Step 3 is to repeat step 2 when the stop criterion is not satisfied; otherwise, continue to step 4. Step 4 is to obtain training subsets Ds in (3) according to the generated centres and the spread width through fuzzy c-means. Step 5 is to construct sub-wFM prediction models. Step 6 is to generate the overall prediction output of wFMFP by using (4) with fuzzy weighting mechanism in (5).

$$\mu_{is} = \frac{\left(1/\|\vec{x}_i - \vec{\beta}_s\|^2\right)^{1/(z-1)}}{\sum_{s=1}^{S}\left(1/\|\vec{x}_i - \vec{\beta}_s\|^2\right)^{1/(z-1)}}, \quad \vec{\beta}_s = \frac{\sum_{i=1}^{N}\mu_{is}^z \vec{x}_i}{\sum_{i=1}^{N}\mu_{is}^z}. \tag{1}$$

$$\delta_s^q = \sqrt{\frac{\sum_{i=1}^{N}\mu_{is}^z \|x_i^q - x_s^q\|^2}{\sum_{i=1}^{N}\mu_{is}^z}}, \; q = 1, 2, \cdots, Q \text{ and } s = 1, \cdots, S. \tag{2}$$

$$\mathcal{D}_s = \{X_s, Y_s\} = \{(\vec{x}_i, y_i) | \beta_s^q - \eta\delta_s^q \le x_i^q \le \beta_s^q + \eta\delta_s^q\}, \; s = 1, \cdots, S \tag{3}$$

$$\hat{y}(\vec{x}_i) = \frac{\sum_{s=1}^{S}\omega_s(\vec{x}_i)\cdot wFM_s(\vec{x}_i)}{\sum_{s=1}^{S}\omega_s(\vec{x}_i)}, \; i = 1, 2, \cdots, N. \tag{4}$$

$$\omega_s^q(x_i^q) = \max\left(\min\left(\frac{x_i^q(\beta_s^q - \eta\delta_s^q)}{\beta_s^q - (\beta_s^q - \eta\delta_s^q)}, \frac{(\beta_s^q + \eta\delta_s^q) - x_i^q}{(\beta_s^q + \eta\delta_s^q) - \beta_s^q}\right), 0\right)$$

$$i = 1, 2, \cdots, N, q = 1, 2, \cdots, Q, s = 1, 2, \cdots, S \tag{5}$$

Performance of wFMFP and other models (factorisation machine with fuzzy partition, factorisation machine, polynomial kernel-based support vector machine, sigmoid kernel-based support vector machine, radial basis function kernel-based support vector machine, and multilayer perceptron) was evaluated by coefficient of determination (R^2), mean squared error, mean absolute error, mean absolute percentage error, and median absolute percentage error. Higher R^2 and lower mean squared error, mean absolute error, mean absolute percentage error, and median absolute percentage error indicate better performance of prediction.

Results

Correlation analysis between readmission variables is shown in Fig 2. wFPFM outperformed other models in predicting readmission risk when the number of training subsets was >5 (Table).

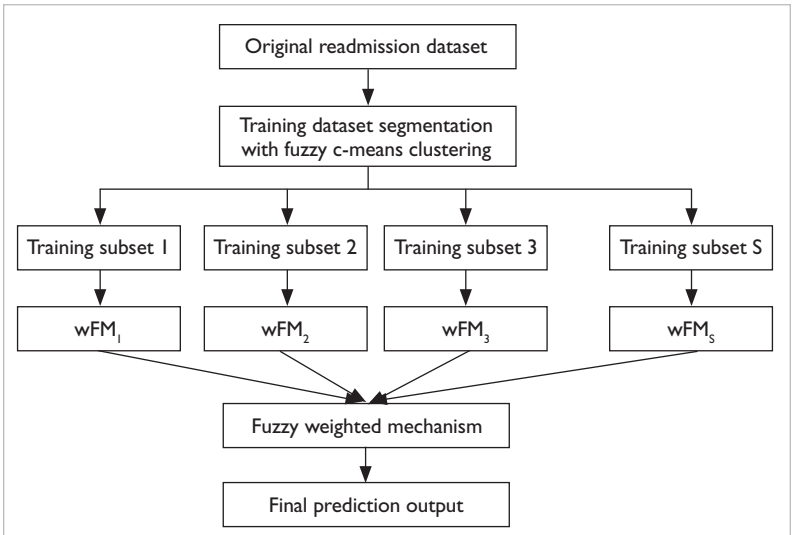

FIG 1. The architecture of weighted factorisation machine (wFM) with fuzzy partition

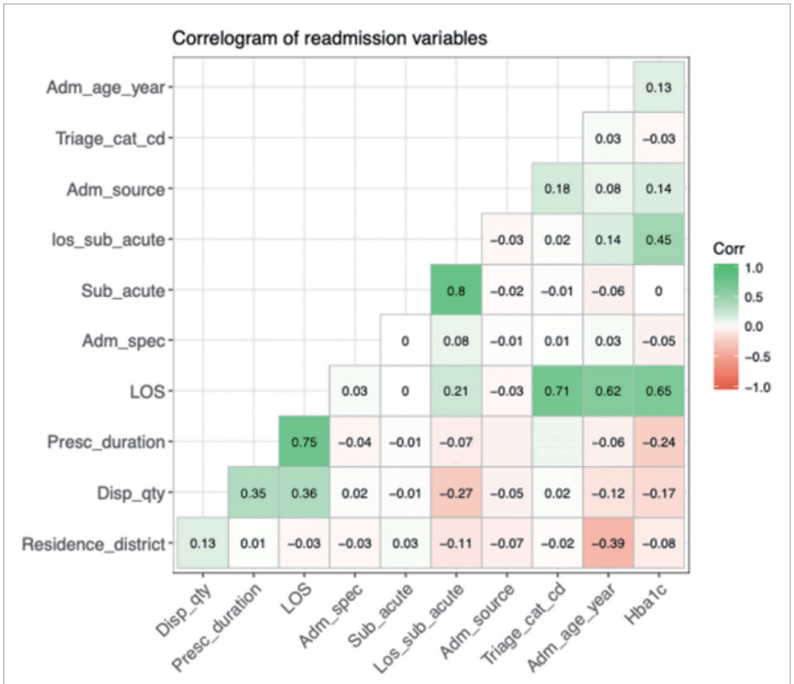

FIG 2. Correlation analysis between readmission variables

Discussion

The wFMFP can deal with the boundary effects and data sparsity issue. In the present study, the wFMFP was superior to other models in predicting readmission risk of patients with chronic kidney disease. The wFMFP can be used as an integral component of the decision support system to better characterise, forecast, and provide preventive

TABLE. Performance of weighted factorisation machine with fuzzy partition (wFMFP) and other machine-learning models in predicting the risk of hospital readmission among patients with chronic kidney disease

Model	R^2	Mean squared error	Mean absolute error	Mean absolute percentage error	Median absolute percentage error
wFMFP					
3 wFMs	0.8267	0.2764	0.2692	0.2458	2521
4 wFMs	0.8518	0.2256	0.2263	0.2390	0.2338
5 wFMs	**0.8865**	**0.2181**	0.2411	0.2311	**0.2130**
6 wFMs	0.8762	0.2512	**0.2142**	**0.2265**	0.2286
7 wFMs	0.8336	0.2292	0.2295	0.2389	0.2421
8 wFMs	0.8349	0.2277	0.2221	0.2459	0.2333
9 wFMs	0.8101	0.2305	0.2322	0.2474	0.2605
10 wFMs	0.7758	0.2471	0.2302	0.2528	0.2747
Factorisation machine with fuzzy partition					
3 FMs	0.7947	0.2633	0.2593	0.2836	0.2994
4 FMs	0.8019	0.2773	0.2667	0.2787	0.2856
5 FMs	0.8249	0.2628	0.2584	0.2924	0.3172
6 FMs	0.8383	0.2649	0.2854	0.2833	0.3174
7 FMs	0.8052	0.2734	0.2621	0.2881	0.3130
8 FMs	0.8495	0.2624	0.2973	0.2856	0.3127
9 FMs	0.7942	0.2735	0.3133	0.3000	0.3245
10 FMs	0.7538	0.2863	0.2973	0.3158	0.3182
Weighted factorisation machine	0.8318	0.2368	0.2656	0.2916	0.3276
Factorisation machine	0.7966	0.2490	0.2931	0.2984	0.3105
Polynomial kernel-based support vector machine	0.7718	0.2948	0.3172	0.2934	0.3123
Sigmoid Kernel-based support vector machine	0.7348	0.2817	0.3291	0.2942	0.3027
Radial basis function kernel-based support vector machine	0.7943	0.2771	0.3012	0.2987	0.3120
Multilayer perceptron	0.7943	0.2771	0.3012	0.2987	0.3220

guideline for readmission risk management. The wFMFP can be used for assessment of discharge and readmission risk. Outcomes from wFMFP can provide a prediction of the explicit risk factors of each patient, which helps decide the disease management strategy such as discharge or intensive medical care for a certain disease. Implementation of predictive readmission analytics provides hospitals an easy-to-use readmission management tool for healthcare cost reduction, efficiency improvement, and lowering LOS deviation. For patients, the model can assist in the domestic care and provide guidance in lifestyle. For example, patients with high risk of readmission because of diabetes should pay attention to daily glucose control. In addition, the wFMFP can be extensively used in various prediction and classification tasks in other domains (transportation, finance).

We plan to examine the prediction performance differences of wFMFP with different weighted strategies, especially when we assign weights through clustering algorithms. We also plan to conduct comparative analysis of the performance of wFMFP under different loss functions. We also plan to investigate solutions to improve the performance of wFMFP in parallel computing settings to practical use once the real-time model training is needed.

Funding

This study was supported by the Health and Medical Research Fund, Health Bureau, Hong Kong SAR Government (#16171991). The full report is available from the Health and Medical Research Fund website (https://rfs1.fhb.gov.hk/index.html).

Disclosure

The results of this research have been previously published in:

1. Zhou J, Li X, Wang X, Chai Y, Zhang Q. Locally weighted factorization machine with fuzzy partition for elderly readmission prediction. Knowl Based Syst 2022;242:108326.

References

1. Randle S. Factorization machines. 2010 IEEE International Conference on Data Mining.
2. Papalexakis EE, Faloutsos C, Sidiropoulos ND. Tensors for data mining and data fusion: models, applications, and scalable algorithms. ACM Trans Intell Syst Technol 2016;8:16.
3. Sidiropoulos ND, De Lathauwer L, Fu X, Huang K, Papalexakis EE, Faloutsos C. Tensor decomposition for signal processing and machine learning. IEEE Trans Signal Process 2017;65:3551-83.