

Comparison of Models for the Prediction of Medical Costs of Spinal Fusion in Taiwan Diagnosis-Related Groups by Machine Learning Algorithms

Ching-Yen Kuo, MSc^{1,2}, Liang-Chin Yu, PhD¹, Hou-Chaung Chen, PhD³, Chien-Lung Chan, PhD^{1,4}

¹Institute of Information Management, Yuan-Ze University, Taoyuan, Taiwan; Departments of ²Medical Administration and ³Orthopedics, Taoyuan General Hospital, Ministry of Health and Welfare, Taoyuan, Taiwan; ⁴Innovation Center for Big Data and Digital Convergence, Yuan-Ze University, Taoyuan, Taiwan

Objectives: The aims of this study were to compare the performance of machine learning methods for the prediction of the medical costs associated with spinal fusion in terms of profit or loss in Taiwan Diagnosis-Related Groups (Tw-DRGs) and to apply these methods to explore the important factors associated with the medical costs of spinal fusion. **Methods:** A data set was obtained from a regional hospital in Taoyuan city in Taiwan, which contained data from 2010 to 2013 on patients of Tw-DRG49702 (posterior and other spinal fusion without complications or comorbidities). Naïve-Bayesian, support vector machines, logistic regression, C4.5 decision tree, and random forest methods were employed for prediction using WEKA 3.8.1. **Results:** Five hundred thirty-two cases were categorized as belonging to the Tw-DRG49702 group. The mean medical cost was US \$4,549.7, and the mean age of the patients was 62.4 years. The mean length of stay was 9.3 days. The length of stay was an important variable in terms of determining medical costs for patients undergoing spinal fusion. The random forest method had the best predictive performance in comparison to the other methods, achieving an accuracy of 84.30%, a sensitivity of 71.4%, a specificity of 92.2%, and an AUC of 0.904. **Conclusions:** Our study demonstrated that the random forest model can be employed to predict the medical costs of Tw-DRG49702, and could inform hospital strategy in terms of increasing the financial management efficiency of this operation.

Keywords: Spinal Fusion, Machine Learning, Diagnosis-Related Groups, Taiwan, Costs and Cost Analysis

Submitted: December 11, 2017

Revised: January 16, 2018

Accepted: January 22, 2018

Corresponding Author

Chien-Lung Chan, PhD

Institute of Information Management, Yuan-Ze University, 135 Yuan-Tung Road, Jung-Li, Taoyuan 320, Taiwan. Tel: +886-3-4638800 (ext. 2605), E-mail: clchan@saturn.yzu.edu.tw

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2018 The Korean Society of Medical Informatics

I. Introduction

Spinal fusion is one of the most common procedures performed by spine surgeons. Between 1998 and 2008, the annual number of spinal fusion discharges increased 2.4-fold, and the national bill for spinal fusion increased 7.9-fold in the United States, while laminectomy, hip replacement, and knee arthroplasty showed relative increases of only 11.3%, 49.1%, and 126.8%, respectively [1]. Most recently, the annual number of lumbar spinal fusions has continued to increase, especially at high- and medium-volume hospitals in New York [2].

In many countries, to control the rising costs of healthcare, Diagnosis-Related Groups (DRGs) have been created. In

the late 1960s, the Yale Center for Health Studies developed DRGs to classify inpatient resource use. The goals were to motivate physicians to use hospital resources and other resources more economically, to document the relationship between medical and administrative decisions, and to define hospital products and services by diagnosis [3]. The DRG system was designed to control hospital reimbursements by replacing retrospective payments with prospective payments for hospital charges. Patients are assigned to a DRG based on their diagnosis, procedures, age, gender, discharge status, and the presence of complications or comorbidities [4]. For example, the Centers for Medicare and Medicaid Services (CMS) classify spinal fusions DRGs by anterior or posterior spinal fusions and with or without complications or comorbidities, with a total of 8 DRGs. However, according to previous research, there are significant cost variations between different types of spinal surgical procedures, based on the complexity and extent of the surgical procedures, as well as variations within a given DRG [5,6].

DRGs are linked to fixed payment amounts based on the average treatment cost of patients in the group, not based on costs actually incurred. Hospitals make financial gains by treating patients for whom the hospital costs are lower than the fixed DRG reimbursement rate. Conversely, hospitals suffer financial losses when treating patients whose costs exceed the fixed DRG reimbursement rate.

Taiwan launched a single-payer National Health Insurance program on March 1, 1995. As of 2014, 99.9% of Taiwan's population were enrolled [7]. To control rising medical costs, Taiwan has been implementing DRGs (Tw-DRGs) since January 2010. The Tw-DRG 3.4 version of spinal fusion, which classifies patients into anterior or posterior spinal fusions, is divided into five DRGs: Tw-DRG496 (combined anterior/posterior spinal fusion), Tw-DRG49701 (posterior and other spinal fusion with complications or comorbidities), Tw-DRG49702 (posterior and other spinal fusion without complications or comorbidities), Tw-DRG49801 (anterior spinal fusion with complications or comorbidities), and Tw-DRG49802 (anterior spinal fusion without complications or comorbidities) [8].

It is very important to monitor spinal fusion DRGs by constructing prediction models of medical costs for spinal fusion and identifying the potential relationship between patient attributes and medical costs. For patients with potential for high medical resource consumption, hospitals can adopt effective treatment plans to improve patient care and manage hospital resources in advance.

Machine learning techniques have recently been used in

various healthcare applications [9]; machine learning models are non-parametric in nature and do not need the assumptions that are made in traditional statistical techniques [10]. Various machine learning strategies were previously compared using field-specific datasets, of which several had significantly better predictive power than the more conventional alternatives [11].

The application of machine learning techniques can solve classification problems, develop prediction models, and identify high-risk patients, but to the best of our knowledge, no study has employed machine learning to predict medical costs in DRGs. A wide set of machine learning techniques has been employed to develop prediction models, such as naïve-Bayesian, support vector machines (SVM), logistic regression, C4.5 decision tree, and random forest methods. All five models are typical examples of supervised machine learning.

Therefore, the purposes of this study were to compare the performance of naïve-Bayesian, SVM, logistic regression, C4.5 decision tree, and random forest methods in predicting the medical costs of spinal fusion in terms of profit or loss effects according to patient characteristics in Tw-DRGs, and to apply these methods to explore the important factors associated with the medical costs of spinal fusion, enabling better management of these patients.

II. Methods

1. Data Collection and Preparation

In this study, a data set was obtained from a regional hospital containing data from January 2010 to December 2013, from which data of patients who underwent spinal fusion surgery were collected. The hospital is a public regional teaching hospital located in Taoyuan City in northern Taiwan, which currently has 1,545 employees, including 194 staff physicians and 972 beds.

The data was original claim data of inpatient admissions used for reimbursement. The database included basic characteristics of patients, admission dates, discharge dates, primary diagnoses, complications or comorbidity ICD-9-CM codes, medical orders, and costs. In the regional hospital, Tw-DRG49702 (posterior and other spinal fusion surgeries without complications or comorbidities) contained the largest number of cases among the five spinal fusion Tw-DRGs; therefore, we chose to use Tw-DRG49702 as the basis for analysis in this study.

According to a previous study, the factors affecting the costs of performing a spinal fusion surgery include patient

age, complications, gender, obesity, diabetic status, and depression [12]. The predictive variables obtained from the database included demographic factors, such as gender, age, primary disease, complications or comorbidities, number of complications or comorbidities, number of intervertebral cages, and length of stay. The class label was defined as 'loss' for those patients whose medical costs exceeded the Tw-DRG49702 fixed payment, meaning that the hospital took a loss, and 'non-loss' for patients whose medical costs fell below the Tw-DRG49702 fixed payment, in which cases the hospital did not take a loss.

There were 532 cases in total, of which the medical costs for 124 (23.3%) patients exceeded the Tw-DRG49702 payment ('loss'), and the costs for 408 (76.7%) patients fell below the Tw-DRG49702 fixed payment ('non-loss'). To redress the imbalance in the data distribution between loss and non-loss, we used the synthetic minority over-sampling technique (SMOTE), which is an important approach in which the positive class or the minority class is oversampled. The SMOTE approach can improve the accuracy of classifiers for a minority class [13].

2. Machine Learning Algorithm for Prediction

In this study, we assessed five classification models, namely, naïve-Bayesian, SVM, logistic regression, C4.5 decision tree, and random forest models.

1) Naïve-Bayesian algorithm

Naïve-Bayesian classifiers or simple-Bayesian classifiers based on Bayes' theorem assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence [14]. Naïve-Bayesian classifiers are among the simplest models in machine learning. Miranda et al. [15] detected cardiovascular disease risk factors using a naïve-Bayesian classifier.

2) Support vector machines algorithm

The SVM algorithm was proposed by Cortes and Vapnik [16] in 1995, and it has become the most influential classification algorithm in recent years. The SVM technique builds a maximum-margin hyper-plane that is positioned in transformed input space and divides the pattern classes, while the distance to the closest plainly divided patterns is maximum. SVM can be used to effectively perform non-linear classification. Kuo et al. [17] used SVM to predict the mortality of hospitalized motorcycle riders.

3) Logistic regression algorithm

Logistic regression is a regression model in which the dependent variable is categorical. Logistic regression is used to model the probability of some event occurring as a linear function of a set of predictor variables, and it is widely used in the medical field to predict the diseases or survivability of a patient [9].

4) C4.5 decision tree algorithm

A decision tree is a flow-chart-like tree structure in which each node denotes a test on an attribute value, each branch represents an outcome of the test, and the tree leaves represent classes or class distributions [14]. There are several different decision tree algorithms, such as Iterative Dichotomiser 3 (ID3), C4.5 decision tree, and Classification and Regression Trees (CART). C4.5 is an algorithm used to generate a decision tree developed by Quinlan [18], and is an extension of Quinlan's earlier ID3 algorithm. C4.5 is often referred to as a statistical classifier [19], and it is a widely used classifier to face real world problems [20].

Habibi et al. [21] used decision tree to find the features related to type 2 diabetes risk factors to help in the screening of diabetes patients.

5) Random forest algorithm

A random forest classifier, proposed by Breiman [22], is an ensemble classifier that produces multiple decision trees using randomly selected features. In classification, trees are voted by the majority. The final classification is obtained by combining the classification results from the individual decision trees. Raju et al. [23] used the random forest model to explore factors associated with pressure ulcers.

The performance of the models considered in this study was assessed by computing the accuracy, sensitivity, specificity, and the total area under the receiver operating characteristic (ROC) curve (AUC). Accuracy is the ability to differentiate between loss and non-loss cases correctly. Sensitivity is the ability to identify loss cases correctly. Specificity is the ability to identify non-loss cases correctly.

These measurements are expressed in terms of true positive (TP), false negative (FN), true negative (TN), and false positive (FP) values:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100,$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100,$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100.$$

The ROC curve is a plot between the true positive rate (TP/TP+FN) and the false positive rate (FP/FP+TN). The classification performance is represented by the total AUC. The closer the area is to 0.5, the less accurate the corresponding model is. A model with perfect accuracy will have an area of 1.0 [14]. The Waikato Environment for Knowledge Analysis (WEKA) 3.8.1 was used for prediction in this study. To avoid overfitting due to use of the same data for training and testing of different classification methods, a 10-fold cross-validation method was used to minimize the bias associated with the random sampling of the training data. In the 10-fold cross-validation, the data set was divided into 10 parts. Then 9 parts were used for training, and the remaining part was used for testing. The process was repeated until all parts had been tested [10]. The goal of this process was to determine which data mining algorithm performs best so we could use it to generate our target predictive model [24].

The process of data extraction and analysis in this study is shown in Figure 1.

III. Results

1. Spinal Fusion Patient Characteristics

Table 1 shows the demographic and clinical characteristics of patients in the Tw-DRG49702 group during the study period. There were 532 cases in total. The mean (standard deviation) total medical cost was US \$4,549.7 (SD = 1,581.7), and the mean age of the patients was 62.4 (SD = 12.5) years.

The average number of complications or comorbidities was 2.0 (SD = 1.22). The mean length of stay was 9.3 (SD = 3.9) days. Among the subjects, 41.4% of the subjects were male, and 58.6% were female. Major primary diseases included lumbar stenosis, spondylolisthesis, lumbar disc displacement, acquired spondylolisthesis lumbar atresia fracture, and scoliosis; complications or comorbidities included high blood pressure, sciatica, diabetes, and osteoporosis. There were significant differences in the actual medical cost, length of stay, number of intervertebral cages, lumbar disc displacement, lumbar atresia fracture and scoliosis between the loss and non-loss groups. However, no significant differences were noted in terms of gender, age, number of complications or comorbidities, lumbar stenosis, high blood pressure, sciatica, spondylolisthesis, diabetes, acquired spondylolisthesis, and osteoporosis between the loss and non-loss groups.

2. Performance of Models

Table 2 summarizes the performance of all five models analyzed in this study, with accuracies ranging from 76.68% for the naïve-Bayesian model to 84.30% for the random forest model. The random forest model achieved better predictive performance than the other methods, with the highest accuracy, sensitivity, specificity, and AUC. The model achieved an accuracy of 84.30%, with a sensitivity of 71.40%, a specificity of 92.20%, and an AUC of 0.904. The next best model was logistic regression, with an 82.16% accuracy, a 69.80% sensitivity, an 89.70% specificity, and an AUC of 0.860. The worst model in terms of predictive value was the naïve-Bayesian model, with an accuracy of 76.68%, a sensitivity of 56.90%, a specificity of 88.70%, and an AUC of 0.815 (see details in Appendix 1).

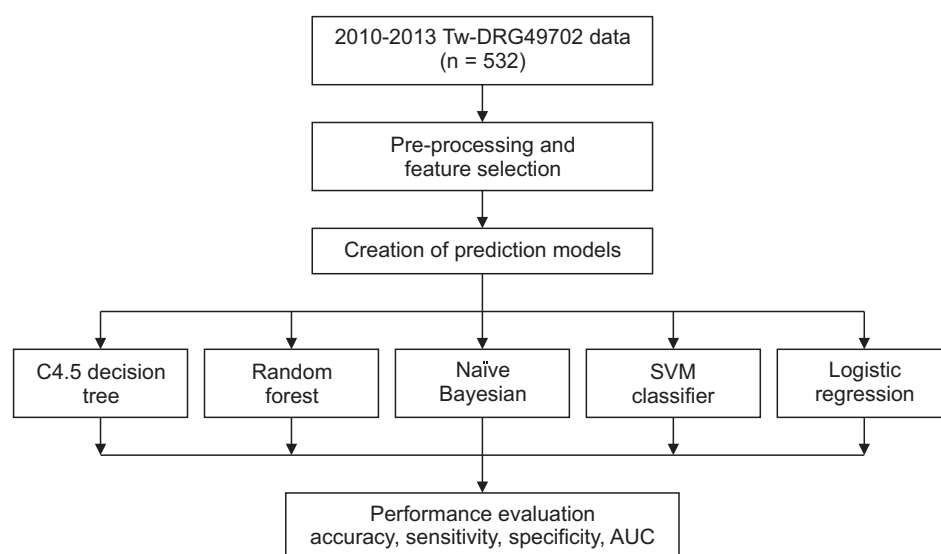


Figure 1. Procedure for data extraction and analysis. Tw-DRG: Taiwan Diagnosis-Related Group, SVM: support vector machine, AUC: area under the receiver operating characteristic curve.

Table 1. Spinal fusion Tw-DRG49702 patient characteristics

| Variable | Total (n = 532) | Loss (n = 124) | Non-loss (n = 408) | p-value |
|--------------------------------|-------------------|-------------------|--------------------|----------|
| Total medical cost (USD) | 4,549.7 ± 1,581.7 | 6,483.4 ± 1,514.6 | 3,997.9 ± 1,115.5 | 0.000*** |
| Gender | | | | |
| Male | 220 (41.4) | 46 (37.1) | 174 (42.6) | 0.272 |
| Female | 312 (58.6) | 78 (62.9) | 234 (57.4) | |
| Age (yr) | 62.4 ± 12.5 | 65.7 ± 12.9 | 61.3 ± 12.2 | 0.842 |
| Number of CCs | 2 ± 1.22 | 2.3 ± 1.2 | 1.8 ± 1.2 | 0.864 |
| Length of stay (day) | 9.3 ± 3.9 | 12.1 ± 5.2 | 8.4 ± 3.1 | 0.001** |
| Number of intervertebral cages | 0.4 ± 0.7 | 0.8 ± 0.9 | 0.3 ± 0.5 | 0.001** |
| Lumbar stenosis | | | | |
| Yes | 390 (73.4) | 90 (40.2) | 300 (73.5) | 0.834 |
| No | 242 (45.6) | 134 (59.8) | 108 (26.5) | |
| High blood pressure | | | | |
| Yes | 213 (40.0) | 56 (45.2) | 157 (38.5) | 0.184 |
| No | 319 (60.0) | 68 (54.8) | 251 (61.5) | |
| Sciatica | | | | |
| Yes | 205 (38.5) | 43 (34.7) | 162 (39.7) | 0.314 |
| No | 327 (61.5) | 81 (65.3) | 246 (60.3) | |
| Spondylolisthesis | | | | |
| Yes | 127 (23.9) | 27 (21.8) | 100 (24.5) | 0.531 |
| No | 405 (76.1) | 97 (78.2) | 308 (75.5) | |
| Diabetes | | | | |
| Yes | 97 (18.2) | 30 (24.2) | 67 (16.4) | 0.05 |
| No | 435 (81.8) | 94 (75.8) | 341 (83.6) | |
| Acquired spondylolisthesis | | | | |
| Yes | 68 (12.8) | 13 (10.5) | 55 (13.5) | 0.381 |
| No | 464 (87.2) | 111 (89.5) | 353 (86.5) | |
| Lumbar disc displacement | | | | |
| Yes | 68 (12.8) | 5 (4.0) | 63 (15.4) | 0.001** |
| No | 464 (87.2) | 119 (96.0) | 345 (84.6) | |
| Osteoporosis | | | | |
| Yes | 40 (7.5) | 14 (11.3) | 26 (6.4) | 0.069 |
| No | 492 (92.5) | 110 (88.7) | 382 (93.6) | |
| Lumbar atresia fracture | | | | |
| Yes | 24 (4.5) | 14 (11.3) | 10 (2.5) | 0.000*** |
| No | 508 (95.5) | 110 (88.7) | 398 (97.5) | |
| Scoliosis | | | | |
| Yes | 15 (2.8) | 9 (7.3) | 6 (1.5) | 0.001** |
| No | 517 (97.2) | 115 (92.7) | 402 (98.5) | |

Values are presented as mean ± standard deviation or number (%).

CCs: complications or comorbidities.

** $p < 0.01$, *** $p < 0.001$.

Table 2. Comparison of performance of various prediction models

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC |
|---------------------|--------------|-----------------|-----------------|-------|
| Naïve Bayesian | 76.68 | 56.90 | 88.70 | 0.815 |
| SVM | 82.16 | 66.10 | 91.90 | 0.790 |
| Logistic regression | 82.16 | 69.80 | 89.70 | 0.860 |
| C4.5 decision tree | 78.51 | 68.50 | 84.60 | 0.770 |
| Random forest | 84.30 | 71.40 | 92.20 | 0.904 |

AUC: area under the receiver operating characteristic curve, SVM: support vector machine.

IV. Discussion

To the best of our knowledge, this study was the first to use machine learning to analyze DRG medical costs. The medical costs of performing spinal fusion in Tw-DRG49702 (posterior and other spinal fusion without complications or comorbidities) in a regional hospital in Taoyuan city in Taiwan were predicted, and the factors associated with profit and loss in terms of medical costs in Tw-DRG49702 were analyzed, using various machine learning techniques. The results of the study showed that the length of stay, number of intervertebral cages, lumbar disc displacement, lumbar atresia fracture, and scoliosis were important factors associated with the medical costs of Tw-DRG49702. In addition, we found that the random forest model had the best predictive performance in comparison with the logical regression, SVM, C4.5 decision tree, and naïve-Bayes models. We were able to successfully predict 84.30% of the patients' medical costs of Tw-DRG49702 using the random forest method.

The length of stay was an important variable in terms of determining medical costs for patients undergoing spinal fusion, the loss group having a significantly longer length of stay. Future management leading to expected reductions in hospital stay will be based on continuous co-operative efforts to improve clinical guidelines or apply lean methods to produce standardized clinical pathways [25].

In our study, in comparison with the C4.5 decision tree classifier, the random forest model had better classification accuracy, their accuracies being 78.51% and 84.30%, respectively. The random forest algorithm, which is one of the most powerful ensemble algorithms, is an effective tool for prediction. Because of the law of large numbers it does not overfit [22]. Previous research has shown that an ensemble is often more accurate than any of the single classifiers in the ensemble [26]. Hu et al. [27] experimentally compared the performance of SVM, C4.5, bagging C4.5, AdaBoosting C4.5, and

random forest methods for the analysis of seven microarray cancer data sets. The experimental results showed that all ensemble methods outperformed C4.5. Masetic and Subasi [28] confirmed the superiority of the random forest method over the C4.5 and SVM methods for the detection of congestive heart failure.

This study also found the random forest model to be superior to traditional logistic regression, a result similar to those of previous studies. The random forest model was more accurate than logistic regression in predicting clinical deterioration. A study of the accuracy of mortality prediction for patients with sepsis at the emergency department found that the random forest model was more accurate (AUC = 0.86) than the logistic regression model (AUC = 0.76, $p \leq 0.003$), and the random forest model was more accurate in predicting mortality after elective cardiac surgery than the logistic regression model [29]. Raju et al. [23] also found that the random forest model had the highest accuracy when used to explore factors associated with pressure ulcers in comparison with decision tree and logistic regression models. These results implied that the random forest model is suitable for classification of the medical costs of Tw-DRG49702.

The strength of this study was that it explored spinal fusion medical cost predictive models and identified important factors; however, there were some limitations of our study. First, the accuracy of this model was 84.30%, meaning that there still are other potential factors that could affect the medical costs of spinal fusion. Second, the study was only performed at a single hospital and with small sample size. It is recommended that data from larger hospitals are analyzed in future study.

Our study demonstrated that the random forest model can be used to predict the medical costs of Tw-DRG49702 (posterior and other spinal fusion without complications or comorbidities), and based on the important factors identified, this study can inform hospital strategy in terms of increasing the efficiency of management of this type of operation in financial terms. Furthermore, methods of this type can also be used to address related problems, such as predicting the costs of other DRGs.

Conflict of Interest

No potential conflict of interest relevant to this article was reported

Acknowledgments

This study was supported by the Taoyuan General Hospital, Ministry of Health and Welfare, Taiwan (No. PTH10307).

References

- Rajae SS, Bae HW, Kanim LE, Delamarter RB. Spinal fusion in the United States: analysis of trends from 1998 to 2008. *Spine (Phila Pa 1976)* 2012;37(1):67-76.
- Jancuska JM, Hutzler L, Protopsaltis TS, Bendo JA, Bosco J. Utilization of lumbar spinal fusion in New York State: trends and disparities. *Spine (Phila Pa 1976)* 2016;41(19):1508-14.
- Rimler SB, Gale BD, Reede DL. Diagnosis-related groups and hospital inpatient federal reimbursement. *Radiographics* 2015;35(6):1825-34.
- Hsiao WC, Sapolsky HM, Dunn DL, Weiner SL. Lessons of the New Jersey DRG payment system. *Health Aff (Millwood)* 1986;5(2):32-45.
- Ugiliweneza B, Kong M, Nosova K, Huang KT, Babu R, Lad SP, et al. Spinal surgery: variations in health care costs and implications for episode-based bundled payments. *Spine (Phila Pa 1976)* 2014;39(15):1235-42.
- Wright DJ, Mukamel DB, Greenfield S, Bederman SS. Cost variation within spinal fusion payment groups. *Spine (Phila Pa 1976)* 2016;41(22):1747-53.
- National Health Research Institutes. Background of National Health Insurance Research Database in Taiwan [Internet]. Miaoli County, Taiwan: National Health Research Institutes; c2016 [cited at 2018 Jan 10]. Available from: <http://nhird.nhri.org.tw/en/index.html>.
- Taiwan National Health Insurance Administration. DRG payment system 2017 [Internet]. Taipei, Taiwan: National Health Insurance Administration; c2017 [cited at 2018 Jan 10]. Available from: https://www.nhi.gov.tw/Content_List.aspx?n=9261941716EB8070&topn=CA428784F9ED78C9.
- Tomar D, Agarwal S. A survey on data mining approaches for healthcare. *Int J Biosci Biotechnol* 2013;5(5):241-66.
- Moon M, Lee SK. Applying of decision tree analysis to risk factors associated with pressure ulcers in long-term care facilities. *Healthc Inform Res* 2017;23(1):43-52.
- Yahya N, Ebert MA, Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Statistical-learning strategies generate only modestly performing predictive models for urinary symptoms following external beam radiotherapy of the prostate: a comparison of conventional and machine-learning methods. *Med Phys* 2016;43(5):2040-52.
- Walid MS, Robinson JS Jr. Economic impact of comorbidities in spine surgery. *J Neurosurg Spine* 2011;14(3):318-21.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-57.
- Han J, Kamber M, Pei J. *Data mining: concepts and techniques*. 3rd ed. Amsterdam: Elsevier; 2011.
- Miranda E, Irwansyah E, Amelga AY, Maribondang MM, Salim M. Detection of cardiovascular disease risk's level for adults using naive Bayes classifier. *Healthc Inform Res* 2016;22(3):196-205.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning* 1995;20(3):273-97.
- Kuo PJ, Wu SC, Chien PC, Rau CS, Chen YC, Hsieh HY, et al. Derivation and validation of different machine-learning models in mortality prediction of trauma in motorcycle riders: a cross-sectional retrospective study in southern Taiwan. *BMJ Open* 2018;8(1):e018252.
- Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1(1):81-106.
- Archana S, Elangovan K. Survey of classification techniques in data mining. *Int J Comput Sci Mob Appl* 2014;2(2):65-71.
- Sanz J, Paternain D, Galar M, Fernandez J, Reyero D, Belzunegui T. A new survival status prediction system for severe trauma patients based on a multiple classifier system. *Comput Methods Programs Biomed* 2017;142:1-8.
- Habibi S, Ahmadi M, Alizadeh S. Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. *Glob J Health Sci* 2015;7(5):304-10.
- Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.
- Raju D, Su X, Patrician PA, Loan LA, McCarthy MS. Exploring factors associated with pressure ulcers: a data mining approach. *Int J Nurs Stud* 2015;52(1):102-11.
- Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform* 2008;77(2):81-97.
- Bradywood A, Farrokhi F, Williams B, Kowalczyk M, Blackmore CC. Reduction of inpatient hospital length of stay in lumbar fusion patients with implementation of an evidence-based clinical care pathway. *Spine (Phila Pa 1976)* 2017;42(3):169-76.
- Kulkarni VY, Sinha PK. Random forest classifiers: a

- survey and future research directions. *Int J Adv Comput* 2013;36(1):1144-53.
27. Hu H, Li J, Plank A, Wang H, Daggard G. A comparative study of classification methods for microarray data analysis. *Proceedings of the 5th Australasian Conference on Data Mining and Analytics*; 2006 Nov 29; Sydney, Australia. p. 33-7.
28. Masetic Z, Subasi A. Congestive heart failure detection using random forest classifier. *Comput Methods Programs Biomed* 2016;130:54-64.
29. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS One* 2017;12(1):e0169772.

Appendix 1. Performance results of five different models

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC | Confusion matrix | Weka model parameters |
|---------------------|--------------|-----------------|-----------------|-------|---|--|
| Naïve Bayesian | 76.68 | 56.90 | 88.70 | 0.815 | a b ← classified as 362 46 a = NO 107 141 b = YES | weka.classifiers.bayes.NaiveBayes |
| SVM | 82.16 | 66.10 | 91.90 | 0.79 | a b ← classified as 375 33 a = NO 84 164 b = YES | weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K |
| Logistic regression | 82.16 | 69.80 | 89.70 | 0.86 | a b ← classified as 366 42 a = NO 75 173 b = YES | weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4 |
| C4.5 decision tree | 78.51 | 68.50 | 84.60 | 0.77 | a b ← classified as 345 63 a = NO 78 170 b = YES | weka.classifiers.trees.J48 -C 0.25 -M 2 |
| Random forest | 84.30 | 71.40 | 92.20 | 0.904 | a b ← classified as 376 32 a = NO 71 177 b = YES | weka.classifiers.trees.RandomForest -I 100 -K 0 -S 1 |

AUC: area under the receiver operating characteristic curve, SVM: support vector machine.