Application of Machine Learning Methods in Nursing Home Research: Using Fall-Related Data in Nursing Homes

Soo-Kyoung Lee

Keimyung University

Jinhyun Ahn

Jeju National University

juh hyun shin (✉ juhshin@ewha.ac.kr)

Ewha Womans University College of Law and Graduate School of Law <https://orcid.org/0000-0001-8058-7195>

Ji Yeon Lee

Ewha Womans University

Research article

Keywords: machine learning, accidental falls, nursing homes

Posted Date: April 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-21878/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: To compare machine learning (ML) methods beyond traditional regression analysis using a real dataset on nurse staffing and falls in nursing homes.

Methods: We used a total of 6 ML algorithms to compare the superiority of ML over traditional regression methods. We applied three representative ML algorithms—random forest (RF), logistics regression, support vector machine (SVM): linear, polynomial, radial, and sigmoid—to the preprocessed dataset to develop a prediction model ($N = 60$). We used an accuracy measure to evaluate prediction models.

Results: RF was the most accurate model (0.883), followed by the logistic regression model (0.867). Using RF procedures, we identified a total of 6 variables as predictors of falls including hours per resident day of administrative staff, proportion of nursing home residents with psychiatric medications, aggressive behaviors and cognitive dysfunction, urinary incontinence, current number of residents in each nursing home, and the maximum capacity of each nursing home.

Conclusions: The appropriate choice of prediction model is quite challenging for nursing researchers. For effective fall management, researchers should consider organizational characteristics as well as personal factors. The examination of related factors on falls is quite meaningful in that preventing falls contributes to improving the quality of life and care of residents and decreases health care costs.

Background

The most frequently reported adverse event among nursing home (NH) residents, falls are associated with increased morbidity and mortality, and with reduced functioning [1–3]. The U.S. fall prevalence is 1.7 falls per resident yearly in NHs [1]; about 50% of NH residents in developed countries fall each year [4]. Accurate prediction of factors associated with a fall in NHs is important because nurses, health care professionals, researchers in practice, administrative staff, researchers, and politicians address fall issues. These stakeholders develop targeted fall-prevention management and assess residents based on the factors associated with a fall [5]. Prior studies identified a number of risk factors for falls, including age, sex, visual deficits, psychotropic medications, cognitive dysfunction [6–8], range of motion, urinary incontinence [9–12], hours per residents day (HPRD) of registered nurses (RNs), and skill mix.

Many prior studies on the prediction of factors associated with falls in NHs have mainly used traditional regression analysis data [13, 14]. However, traditional regression analysis has limitations in addressing the variability in data, nonlinear interactions of variables, and diverse distributions of datasets, which usually require basic assumptions [15]. To offset these limitations, researchers recently introduced a new approach: machine learning (ML). ML is a new computer-science study area using statistics, optimization methods, or artificial intelligence to gain statistical algorithms retrieved from a real dataset [15]. With input data, ML is able to construct algorithms that can continue improving prediction and generate automated knowledge through data-driven predictors or decisions. ML is especially beneficial when researchers are working to discern many relevant features that are not independent [16]. ML is a valid substitute for traditional statistical analysis in predicting conditional probabilities [17]. Using ML, researchers may report latent variables that were not usually reported using traditional data-analysis methods. ML also helps in providing clinical decision support [15]. Because of these advantages of ML, ML methods were recently introduced in the fields of health, medicine, and nursing to predict indicators such as mortality rates after radical cystectomy [15], mortality-risk predictors for patients with burn injuries [17], health-related quality of life for elders with chronic disease [19], and emergency department (ED) visits as predictors for residents in NHs [20]. In these studies, ML successfully identified mortality predictors, ED visits, and quality-of-life-related factors [15, 18, 19]. The dataset used in NH research and practice is becoming more dimensional, complex, and abundant, with advanced informatics and data management. Also, factors identified as fall-related factors in NHs are not independent of each other [21, 22]. Considering the characteristics of these NH datasets, it is necessary to introduce ML when analyzing data from NHs.

One form of ML, the random forest (RF) algorithm, is applicable when researchers have more predictors than observations, based on the theory of ensemble learning that allows an algorithm to learn simple and complex classification functions accurately. RF does not require fine-tuning of parameters. Default parameterization often leads to excellent performance [23]. The support-vector-machine (SVM) model looks for a super-plane set of a high- or infinite-level space and uses it to perform classification and regression. SVM can learn complex classification functions efficiently, and they employ powerful regularization principles to avoid overfitting [24]. One SVM, SVM linear, is a single parameter classification function where the parameter biases the test along the regression output values of the SVM. Researchers use SVM linear to handle large amounts of data vectors, such as text categorization. Also, researchers often use the SVM polynomial to process images and the SVM radial when they have no prior information about the data. Mainly, researchers use SVM sigmoid as a proxy for neural networks [25].

In this study, we introduced possible data analysis or interpretation in addition to using the beyond-traditional-regression method in the field of nursing. The purpose of this study was to compare six ML methods (random forest, logistics regression, linear support vector SVM, polynomial SVM, radial SVM, and sigmoid SVM) in six aspects (accuracy, sensitivity, specificity, positive predictor values, negative predictor values, and receiver operating characteristic curve) by demonstrating real data-based predicted factors related to resident falls. Further, we aimed to identify predicted factors related to falls in NHs according to ML methods. Our prediction of fall risks enables nurses in NHs to support clinical discussions of risk factors and prevention strategies to lower fall incidence.

Methods

Study Design

This study used an NH dataset to compare various ML methods beyond traditional regression analysis by demonstrating real data-based predicated factors related to resident falls.

Ethical Consideration

We obtained approval for this study from an Ewha Woamans University Institutional Review Board in Korea (IRB NO. ewha-136-4). We protected the confidentiality of participating organizations. The consent of the study was received in written by the participants.

Measurements/Instruments

Prediction Variables

Information on resident capacity, number of current residents, organization location, ownership, religious affiliation, affiliated hospitals, residents' age, residents' long-term-care insurance grade, residents' sex, and long-term-care facility grade came from the participating organizations. Administrators and directors of nursing (DONs) at each participating NH offered staff-related data. HPRD referenced mean hours worked by each type of staff (RNs, certified nurse aids, care workers, director, secretary general, social worker, dietician, and administrative staff) divided by the total number of residents [26]. The operational definition of skill mix was the proportion of RNs on the nursing staff and care workers. To calculate staff turnover, we calculated the total quantity of present staff at the end of the calendar year, the quantity of terminations in a calendar year, and turnover per year using the Nursing Facility Staff Survey.

Outcomes Variable

We selected the variable *falls* as the outcome variable, representing the number of NH residents having experienced any kind of fall in the recent 3 months. Administrators, DONs or nurses at each participating NH also provided fall-related data based on chart reviews.

Data Collection/Procedure

We retrieved the dataset employed in this study from a 3-year research project (2017–2020), Estimating Optimal Nurse Staffing for Nursing Home Residents Using an Optimization Model, funded by the Korea National Research Foundation. It received 300,000\$ in financial support from the funder. The principal investigator of this paper conducted the parent study on optimized nurse staffing HPRD on best quality-care outcomes. In the parent study, from 2017 to 2020, the principal investigator or researchers visited the NHs that agreed to participate, introducing the parent study to administrators or DONs and obtaining informed-consent forms. Administrators, DONs, or nurses filled out the survey by reviewing residents' charts, nurses' notes, and accident reports at the organizational level. From 2017 to 2020, researchers collected the data every 4 months. For this analysis, we retrieved only the first measured dataset: 60 NHs in 17 provinces of South Korea completed the survey from May 2017 to August 2017.

Data Analysis

Figure 1 depicts the two steps of workflow of this study: feature selection and data preprocessing. Because the collected dataset has 57 independent variables, it was infeasible to try every combination of variables. We used RF to single out nonsignificant variables, reducing the number of independent variables to 13. The data-preprocessing step normalized values to supply the dataset to ML algorithms. We applied three representative ML algorithms—RF, logistics regression, and SVM (linear, polynomial, radial, sigmoid)—to the preprocessed dataset from a prediction model.

Variable Selection

We selected the variable *falls* as the dependent variable, representing the number of NH residents having experienced any kind of fall in the past 3 months. Values ranged from 0 to more than 100 with most values smaller than 20 (see Fig. 2). In our prediction task, we distinguished between a residents' group having no falls (classified as 0) and another group having at least one fall (classified as 1). We had 57 independent variables. Because the combination of independent variables was 144,115,188,075,855,870, it was infeasible to create a model to evaluate every combination. Thus, we used RF to select significant variables, allowing us to draw importance scores for each variable. Table 1 lists significant variables, ordered by importance scores greater than 0.5 because we wanted to select about 10 variables to make the remaining process feasible.

Table 1
Organizational Characteristics of Nursing Homes (n = 60)

Variable	Frequency	%	<i>M</i>	<i>SD</i>	Min	Max
Number of Capacity			75.13	55.835	7	296
Number of current Residents			66.73	48.513	7	295
Location of organizations						
Metropolitan (over million)	29	48.3				
Urban location (over round half million)	13	21.7				
Local small city (5-50 thousand)	13	21.7				
Rural area (less than 50 thousand)	5	8.3				
Ownership						
Profit	10	16.7				
Not for profit	49	81.7				
Religious affiliation						
No religion	39	65.0				
Christianity	11	18.3				
Catholic	3	5.0				
Buddhism	3	5.0				
Others	3	5.0				
Affiliated hospitals						
No	5	8.3				
Yes	54	90				
Average age of residents			83.60	2.404	80	89
Determined Long-term care insurance grade by committee of Korean National Health Insurance Corporation (%)						
1st ^a			12.000	8.820	0	41
2nd ^b			22.403	11.466	0	60
3rd ^c			35.957	14.397	0	90
4th ^d			22.415	16.951	0	90
5th ^e			1.185	3.004	0	15
Unrated			5.983	19.485	0	100
Gender (%)						
Male			20.77	11.358	0	64
Female			78.95	11.300	36	100
Long-term care facility grade (%)						
Grade A (Superior)	23	38.3				
Grade B (Above average)	8	13.3				
Grade C (Average)	6	10.0				
Grade D (Below average)	7	11.7				
Grade E (Poor)	0	0				
Ungraded	16	26.7				
Number of Registered Nurses			1.48	2.891	0	17
Number of Certified Nurse Aides			1.95	1.443	0	6
Number of Care Worker			26.58	20.522	1	132

Variable	Frequency	%	<i>M</i>	<i>SD</i>	Min	Max
Hours Per Resident Day						
Registered Nurses			0.191	0.243	0.000	1.030
Certified Nurse Aides			0.367	0.261	0.000	1.425
Care Worker			3.829	1.633	1.325	9.769
Director			0.269	0.231	0.039	1.425
Secretary General			0.122	0.120	0.000	0.518
Social Worker			0.396	0.273	0.000	1.425
Dietician			0.093	0.092	0.000	0.365
Administrative Staff			0.146	0.226	0.000	1.368
Turnover (%)						
Registered Nurses			9.497	21.610	0.000	100.000
Certified Nurse Aides			12.540	22.490	0.000	66.666
Care Worker			19.064	18.500	0.000	86.179
Director			4.17	19.071	0.000	
Secretary General			12.2	33.129	0.000	
Social Worker			17.99	29.491	0.000	
Dietician			29.83	42.652	0.000	
Cook			38.46	36.311	0.000	
Administrative Staff			21.39	35.178	0.000	
Note. ^a Nursing Home Residents who are completely dependent for activities of daily living. ^b Nursing Home residents who are mostly dependent for activities of daily living. ^c Nursing Home Residents who are partially dependent for activities of daily living. ^d Nursing Home residents who have limited dependence for activities of daily living. ^e Nursing Home residents with dementia, M = mean; SD = standard deviation; Min = minimum; Max = maximum						

Data Preprocessing

In the data-preprocessing step, we discarded meaningless variables, encoded into integer numbers, and normalized these values. We discarded variables irrelevant to the prediction of falls in NHs: institution identification, name, evaluation time (the date of evaluation performance), year of institution establishment, and province. For categorical variables, we normalized values into integers to reflect the grade. For example, facility-grade variables had values, such as A, B, C, D, new, grading exception, and evaluation nontarget; we assigned integers 4, 3, 2, 1, 0, 0, and 0, respectively. We assumed Grade A was the best score and D the worst. We assigned values of zero when values indicated no evaluation was performed. Religion values had already been encoded into integers except for string values, such as “5 (Won Buddhism)”^c; it was trivial to convert the string to 5. Some variables represented values in ratio form. The variable Skill_mix_RN_CNA represents the ratio of the number of nurses to the number of nursing assistants. For example, the value 1:4 means the number of CNAs is four times the number of nurses. We converted the value to 0.25.

Predicting Modeling

We applied three ML algorithms to the preprocessed dataset. We measured baseline by applying the RF method first to optimize the matrix of the prediction model. RF is an ensemble ML method that constructs multiple decision trees at training time [27]. The output class rests on the mean prediction of individual decision trees. RF overcomes the overfitting problem of conventional decision-tree methods, generating multiple decision trees at training time.

Researchers widely use logistic regression; the ML method uses a logistic function to address two class-dependent variables [28]. SVM, also a widely used ML method, tries to find a support vector by correctly dividing given training data in a featured space, constructed by a kernel function [29].

We used R packages to run these ML algorithms. For RF, we set the parameter ntree to 100, representing the number of generated trees. We observed no significant improvement when the value of ntree was bigger. For logistic regression and SVM, we used default parameter-value settings. We used four well-known kernel functions in SVM: linear, polynomial, radial, and sigmoid. Based on this configuration, we had six models.

Evaluation of Prediction Models

To validate each prediction model, we used five-fold cross validation, dividing the data set into five parts, then training in four and testing in one, repeating the process until testing all parts. We built a confusion matrix to gauge each model’s performance, measured with six parameters: accuracy, sensitivity, specificity, positive predictor values, negative predictor values, and receiver operating characteristic curve. We assigned the accuracy measure the highest importance because the main goal of the study was to find a prediction model that best predicts the variables [30].

Table 2
Comparison of Performance in Prediction Models

Model	Accuracy	Sensitivity	Specificity	PPV	NPV
Logistic regression	0.867	1.000	0.467	0.849	1.000
Random forest	0.883	0.978	0.600	0.880	0.900
SVM linear	0.867	0.978	0.533	0.863	0.889
SVM polynomial	0.867	1.000	0.467	0.849	1.000
SVM radial	0.850	0.978	0.467	0.846	0.875
SVM sigmoid	0.850	0.933	0.600	0.875	0.750
Note. PPV = positive predictive values, NPV = negative predictive values, SVM = support vector machine.					

Table 3
Predicting Variables of Six Models

Model	Hours per Resident day of secretary general	Proportion of residents with psychiatric medication	Proportion of residents with urinary incontinence	Hours per resident day of care worker	Current number of resident in a nursing home	CNA_HPRD	Proportion of residents with aggressive behavior	Proportion of residents with cognitive decline	Turnover of care worker	Maximum capacity of residents	Total
Logistic regression	0	0	0			0					4
Random forest		0	0		0		0	0		0	6
SVM linear	0			0	0	0			0		5
SVM polynomial	0	0	0		0						4
SVM radial	0	0		0				0			4
SVM sigmoid	0		0	0		0	0				5
-	5	4	4	3	3	3	2	2	1	1	-
Note. CNA = certified nurse aide, HPRD = hours per resident day, SVM = support vector machine.											

Results

Characteristics of the Organizations (n = 60)

Table 1 summarizes descriptive statistics of participating NHs. Of NHs, 70% were located in metropolitan or urban areas (population over half million); about 82% were not for profit. About 83.3% affiliated with a religion and 90% affiliated with a hospital. Residents' average age was 83.6 (*SD*= 2.404). About 34.4% of residents had first and second ratings, most of whom were bedridden with severe activities of daily living limitations. Of NHs, 38.3% received a superior grade (Grade A) in the long-term care facility evaluation. The average number of nurses was 1.48, the HPRD for nurses was 0.191 (11 min and 30 sec), and the turnover rate was 9.497%.

Predictive Performance

Table 2 shows performance and Fig. 3 shows the corresponding area under the curve: the RF model had the greatest accuracy (0.883). After the logistic regression model (0.867), the SVM linear and polynomial (0.867), SVM radial, and sigmoid (0.850) models had decreasing orders of performance. Figure 4 shows the accuracy of the kernel function of SVM by number of input features.

Table 3 shows the association between 10 variables identified by ML with falls. Using RF procedures, we identified six variables as predictors of falls including HPRD of administrative staff; number of NH residents with psychiatric medications, aggressive behaviors, and cognitive dysfunction; urinary incontinence; current number of residents in each NH; and the maximum capacity of each NH. Logistic regression, SVM polynomial, and SVM radial identified only four predictors, whereas SVM linear and sigmoid identified five predictors.

The strongest predicting variable was HPRD of the secretary general, supported by logistic regression, SVM linear, SVM polynomial, SVM radial, and SVM sigmoid. The proportion of residents with psychiatric medications was supported by logistic regression, random forest, SVM polynomial, SVM radial, and SVM

sigmoid. The proportion of residents with urinary incontinence was supported by logistics regression, random forest, SVM polynomial, SVM radial, and SVM sigmoid.

Discussion

The proportion of residents prescribed psychiatric medications and urinary incontinence were top factors in the logistic regression, RF, SVM polynomial, SVM radial, and SVM sigmoid. Psychiatric medication was a very strong predictor of falls, consistent with previous research [6, 7, 9, 10, 31–33], but inconsistent with a couple of studies [34, 35]. The proportion of residents with urinary incontinence was an important factor in falls, supported in previous research [9–12], but not supported in other studies [7, 32]. Results confirmed that the incidence of resident falls relates to increased psychiatric medication and urinary incontinence [9, 12, 36, 37]. No practical guide exists for NH staff based on scientific evidence [38]. Related factors supported in this study should be included in guidelines: careful medication administration by professional RNs and a healthy environment to reduce falls considering the case mix of residents [1, 39].

Five ML algorithms predicted the HPRD of the secretary general, inconsistent with previous research [10]. The logistic regression, SVM linear, and SVM sigmoid also supported the HPRD of CNAs as a predictor of falls, supported in previous research [40–42] but inconsistent with other research [43, 44]. The number of care workers was another important factor for falls, supported in SVM linear, SVM radial, and SVM sigmoid, but not supported in Shin [43] and Shin and Hyun [44]. Remarkably, administrative staff was an important factor in resident falls using ML analysis. Very limited study [10] reported the stability of administrative staff related to resident falls. A scientific and sound interdisciplinary approach among NH staff requires teamwork among nurses and administrative staff, including shared governance [45].

Only RF reported that the total number of residents was an important factor in falls, consistent with Thomas et al. [42], but not with Van Doorn et al. [7]. The RF and SVM radial reported that the proportion of residents with cognitive dysfunction related to falls, consistent with preceding research [6–8], but inconsistent with Datta et al. [34].

Controversies persist about applying ML versus regression; researchers use regression-report outcomes of independent variables using odds ratios, charts, or nomograms, but ML can not report these values [46, 47]. The RF reported more accurate values than other algorithms including SVM linear/polynomial/radial/sigmoid and regression in this study. Previous studies reported the accuracy of memetic algorithms; a meta-analysis reported that SVM was more accurate than an artificial neural network decision tree, naïve Bayes, and K-nearest neighbor [48].

Nursing data are also increasing, and nursing researchers are using large-scale data on the quality, efficiency, and effectiveness of nursing [49]. Research on large-scale nursing data using data mining is important in improving nursing quality, supporting clinical professionals' decision-making, and determining health care administrators' policies [50].

Classical statistical methods may report only similar results because of difficulties in dealing with big data. However, ML was designed to analyze the raw dataset without specific interpretation or refinement of data, and can report related factors that were not usually reported in previous research. This research applied many ML methods to discern the best fit method. Future research also needs to evaluate many ML methods to investigate related factors on residents' outcomes in NH research.

Conclusions

This study applied six ML methods to predict factors related to falls of NH residents. Examination of related organizational factors on falls is quite meaningful in that preventing falls contributes to improving the quality of life and care of residents and decreasing health care costs. We propose RF as a powerful algorithm to discern predictors of falls in NHs. We validated the scientific data analysis by developing six ML algorithms, applied to a dataset from 60 NHs across Korea. In the future, it is necessary to demonstrate different ML algorithms to apply other data to confirm its potential.

Although we provided the definition of falls to data collectors in each NH, they may have underestimated or overestimated the number of falls; we could collect only self-reported data from each NH. Also, we did not collect the specific location or date of falls; rather, we asked how many slips or falls took place in the last quarter. Also, we did not differentiate between types of falls, slips, and fall-related injuries in this study; future studies should consider these factors. Study results should be understood with attention to the small sample size in Korea. However, we tried to offset this limitation by querying 60 NHs across Korea.

Another limitations was the use of traditional statistical methods of ML to learn the statistically significant variables. Thus, to confirm the superiority of ML in NH research, future studies are required to discern potential factors that were not studied in the past using newly introduced ML methods.

Abbreviations

CNA
Certified Nursing Assistant
DON
Director of Nursing
ED
Emergency Department
HPRD
Hours Per Resident Day

ML
Machine Learning
NH
Nursing Home
RF
Random Forest
RN
Registered Nurse
SVM
Support Vector Machine

Declarations

Ethis approval and consent to participate :

We obtained approval for this study from an Ewha Womans University Institutional Review Board in Korea (IRB NO. ewha-136-4). We protected the confidentiality of participating organizations. The consent of the study was received in written by the participants.

Consent to publish:

All authors agreed to publish this article

Availability of data and materials:

The datasets generated and analysed during the current study are not publicly available because it stated that it would be used only for the purpose of this study when participants' consent is obtained.

Competing interests:

The authors declared no conflict of interest.

Funding:

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(No. NRF-2018R1D1A3B07050652). It received 300,000\$ in financial support from the funder.

Authours' Contributions:

Substantial contributions to conception and design-JH,SK

Acquisition of data-JH

Analysis and interpretation of data-SK,JH

Drafting the article-JH,SK,JY

Revising the article critically for important intellectual content-JH,SK,JY

Final approval of the version to be published-JH,SK,JH,JY

Acknowledgements:

Not applicable

References

- 1 Becker C, Rapp K. Fall prevention in nursing homes. Clinics in geriatric medicine. 2010;26(4):693–704. <https://doi.org/10.1016/j.cger.2010.07.004>
- 2 Gray-Miceli D, de Cordova PB, Crane GL, Quigley P, Ratcliffe, SJ. Nursing home registered nurses' and licensed practical nurses' knowledge of causes of falls. J Nurs Care Qual. 2015;31(2):153–160. <https://doi.org/10.7282/T38054HV>

- 3 Gray-Miceli D, Ratcliffe SJ, Thomasson A. Elderly fallers at greatest risk for head injury. *Gerontologist* 2013;53(S1):489. <https://doi.org/10.1093/geront/gnt151>
- 4 Vandenberg AE, van Beijnum BJ, Overdeest VGP, Capezuti E, Johnson TM 2nd. US and Dutch nurse experiences with fall prevention technology within nursing home environment and workflow: a qualitative study. *Geriatr Nurs.* 2017;28(4):276–282. <https://doi.org/10.1016/j.gerinurse.2016.11.005>
- 5 Chaccour K, Darazi R, El Hassani AH, Andres E. From fall detection to fall prevention: A generic classification of fall-related systems. *IEEE Sensors J.* 2016;17(3):812-822. <https://doi.org/10.1109/jsen.2016.2628099>
- 6 Seijo-Martinez M, Cancela JM, Ayán C, Varela S, Vila H. Influence of cognitive impairment on fall risk among elderly nursing home residents. *Int Psychogeriatr.* 2016;28(12):1975–1987. <https://doi.org/10.1017/s1041610216001113>
- 7 Van Doorn C, Gruber-Baldini AL, Zimmerman S, Hebel JR, Port CL, Baumgarten M, et al. J. Epidemiology of Dementia in Nursing Homes Research Group. Dementia as a risk factor for falls and fall injuries among nursing home residents. *J Am Geriatr Soc.* 2003;51(9):1213–1218. <https://doi.org/10.1046/j.1532-5415.2003.51404.x>
- 8 Whitney J, Close JC, Jackson SH, Lord SR. Understanding risk of falls in people with cognitive impairment living in residential care. *J Am Med Dir Assoc.* 2012;13:535–540. <https://doi.org/10.1016/j.jamda.2012.03.009>
- 9 Hasegawa J, Kuzuya M, Iguchi A. Urinary incontinence and behavioral symptoms are independent risk factors for recurrent and injurious falls, respectively, among residents in long-term care facilities. *Arch Gerontol Geriatr.* 2010;50(1):77–81. <https://doi.org/10.1016/j.archger.2009.02.001>
- 10 Shin JH. Nursing staff characteristics on resident outcomes in nursing homes. *J Nurs Res.* (2019);27(1):1–9. <https://doi.org/10.1097/jnr.0000000000000269>
- 11 Kron M, Loy S, Sturm E, Nikolaus T, Becker C. Risk indicators for falls in institutionalized frail elderly. *Am J Epidemiol.* 2003;158(7):645–653. <https://doi.org/10.1093/aje/kwg203>
- 12 Zarowitz BJ, Allen C, O'Shea T, Tangalos E, Berner T, Ouslander JG. Clinical burden and nonpharmacologic management of nursing facility residents with overactive bladder and/or urinary incontinence. *The Consultant Pharmacist.* 2015;30(9):533–542. <https://doi.org/10.4140/tcp.n.2015.533>
- 13 Bor A, Matuz M, Csatornai M, Szalai G, Bálint A, Benkő R, et al. Medication use and risk of falls among nursing home residents: a retrospective cohort study. *Int J Clin Pharm.* 2017;39(2): 408–415. <https://doi.org/10.1007/s11096-017-0426-6>
- 14 Paliwal Y, Slatum PW, Ratliff SM. Chronic health conditions as a risk factor for falls among the community-dwelling US older adults: a zero-inflated regression modeling approach. *BioMed Res Int.* 2017;51146378. <https://doi.org/10.1155/2017/5146378>
- 15 Wang G, Lam KM, Deng Z, Choi KS. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. *Comput Biol Med.* 2015;63:124–132. <https://doi.org/10.1016/j.compbiomed.2015.05.015>
- 16 Quinlan JR. C4.5: Programs for machine learning. San Francisco, CA: Morgan Kaufmann, 1993.
- 17 McCarthy JF, Marx KA, Hoffman PE, Gee AG, O'Neil P, Ujwal ML, et al. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Ann NY Acad Sci.* 2004;1020:239–262. <https://doi.org/10.1196/annals.1310.020>
- 18 Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. *Burns.* 2015;41(5):925–934. <https://doi.org/10.1016/j.burns.2015.03.016>
- 19 Lee HY, Blegen MA, Harrington C. The effects of RN staffing hours on nursing home quality: A two-stage model. *Int J Nurs Stud.* 2014;51(3):409–417. <https://doi.org/10.1016/j.ijnurstu.2013.10.007>
- 20 Veyron JH, Friocourt P, Jeanjean O, Luquel L, Bonifas N, Denis F, et al. Home care aides' observations and machine learning algorithms for the prediction of visits to emergency departments by older community-dwelling individuals receiving home care assistance: A proof of concept study. *PloS One.* 2019;14(8):e0220002. <https://doi.org/10.1371/journal.pone.0220002>
- 21 Fu MR, Wang Y, Li C, Qiu Z, Axelrod D, Guth AA, et al. Machine learning for detection of lymphedema among breast cancer survivors. *Mhealth.* 2018;4,Art. 17. <https://doi.org/10.21037/mhealth.2018.04.02>
- 22 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York, NY: Springer, 2009. https://doi.org/10.1111/j.1751-5823.2009.00095_18.x
- 23 Breiman L. Random forests. *Mach Learn.* 2001;45,5–32. <https://doi.org/10.1012/A;1010933404324>
- 24 Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics.* 2000;16(10):906–914. <https://doi.org/10.1093/bioinformatics/16.10.906>

- 25 Center for Medicare and Medicaid. Staffing Data Submission PBJ. 2018 [cited 2019 April 20]. Available from: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Staffing-Data-Submission-PBJ.html>.
- 25 Boyle D. Countering compassion fatigue: A requisite nursing agenda. *Online J Issues Nurs*. 2011;16(1):2. <https://doi.org/10.3912/OJIN.Vol16No01Man02>
- 26 Ho TK. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* Piscataway, NJ: Institute of Electrical and Electronics Engineers; 1995:278–282. <https://doi.org/10.1109/ICDAR.1995.598929>
- 27 Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons, 2013. <https://doi.org/10.1002/9781118548387>
- 28 Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. 1999;9:293–300. <https://doi.org/10.1023/A:1018628609742>
- 29 Lee SK, Son YJ, Kim J, Kim HG, Lee JI, Kang BY, Cho HS, Lee, S. Prediction model for health-related quality of life of elderly with chronic diseases using machine learning techniques. *Healthc Inform Res*. 2014;20(2):125–134. <https://doi.org/10.4258/hir.2014.20.2.125>
- 30 Bozat-Emre S, Doupe M, Kozyrskyj AL, Grymonpre R, Mahmud SM. Atypical antipsychotic drug use and falls among nursing home residents in Winnipeg, Canada. *Int J Geriatr Psychiatry*. 2015;30(8):842–850. <https://doi.org/10.1002/gps.4223>
- 31 Kosse NM, de Groot MH, Vuillerme N, Hortobágyi T, Lamoth CJ. Factors related to the high fall rate in long-term care residents with dementia. *Int Psychogeriatr*. 2015;27(5):803–814. <https://doi.org/10.1017/s104161021400249x>
- 32 Milos V, Bondesson Å, Magnusson M, Jakobsson U, Westerlund T, Midlöv P. Fall risk-increasing drugs and falls: A cross-sectional study among elderly patients in primary care. *BMC Geriatr*. 2014;14(1):40. <https://doi.org/10.1186/1471-2318-14-40>
- 33 Datta A, Datta R, Elkins J. What factors predict falls in older adults living in nursing homes: A pilot study. *J Funct Morphol Kinesiol*. 2018;4(1):Art. 3. <https://doi.org/10.3390/jfmk4010003>
- 34 Hanlon JT, Zhao X, Naples JG, Aspinall SL, Perera S, Nace DA., et al. Central nervous system medication burden and serious falls in older nursing home residents. *J Am Geriatr Soc*. 2017;65(6):1183–1189. <https://doi.org/10.1111/jgs.14759>
- 35 Berry SD, Placide SG, Mostofsky E, Zhang Y, Lipsitz LA, Mittleman MA, et al.. Antipsychotic and benzodiazepine drug changes affect acute falls risk differently in the nursing home. *J Gerontol A Biol Sci Med Sci*. 2015;71(2):273–278. <https://doi.org/10.1093/gerona/glv091>
- 36 Delbaere K, Close JC, Menz HB, Cumming RG, Cameron ID., Sambrook PN, et al. Development and validation of fall risk screening tools for use in residential aged care facilities. *Med J Aust*. 2008;189(4):193–196. <https://doi.org/10.5694/j.1326-5377.2008.tb01980.x>
- 37 Cameron I, Kurlle S, Crotty M. Hip fracture rehabilitation for people living in nursing homes: Randomised trial. *Ann Phys Rehabil Med*. 2018;61(Suppl):e339. <https://doi.org/10.1016/j.rehab.2018.05.794>
- 38 Morris V, Wagg A. Lower urinary tract symptoms, incontinence and falls in elderly people: Time for an intervention study. *Int J Clin Pract*. 2007;61(2):320–323. <https://doi.org/10.1111/j.1742-1241.2006.01174.x>
- 39 Leland NE, Gozalo P, Teno J, Mor V. Falls in newly admitted nursing home residents: A national study. *J Am Geriatr Soc*. 2012;60(5):939–945. <https://doi.org/10.1111/j.1532-5415.2012.03931.x>
- 40 Spector W, Shaffer T, Potter DEB, Correa-de-Araujo R, Limcangco MR. Risk factors associated with the occurrence of fractures in US nursing homes: Resident and facility characteristics and prescription medications. *J Am Geriatr Soc*. 2007;55(3):327–333. <https://doi.org/10.1111/j.1532-5415.2007.01081.x>
- 41 Thomas KS, Hyer K, Castle NG, Branch LG, Andel R, Weech-Maldonado R. Patient safety culture and the association with safe resident care in nursing homes. *Gerontologist*. 2012;52(6):802–811. <https://doi.org/10.1093/geront/gns007>
- 42 Shin JH. Why do we require registered nurses in nursing homes? Using longitudinal hierarchical linear modeling. *H Nurs Scholarsh*. 2018;50(6):705–713. <https://doi.org/10.1111/jnu.12412>
- 43 Shin JH, Hyun TK. Nurse staffing and quality of care of nursing home residents in Korea. *J Nurs Scholarsh*. 2015;47(6):555–564. <https://doi.org/10.1111/jnu.12166>
- 44 Taylor JA, Parmelee P, Brown H, Strothers HS III, Capezuti E, Ouslander JG. A model quality improvement program for the management of falls in nursing homes. *J Am Med Dir Assoc*. (2007);8(3 Suppl):S26–S36. <https://doi.org/10.1016/j.jamda.2006.11.005>
- 45 Belle VV, Pelckmans K, Suykens JA, Huffel SV. Learning transformation models for ranking and survival analysis. *J Mach Learn Res*. (2011);12:819–862.
- 46 Steyerberg EW, van der Ploeg T, Van Calster B. Risk prediction with machine learning and regression methods. *Biometrical J*. (2014);56(4):601–606. <https://doi.org/10.1002/bmj.201300297>

47 Nindrea RD, Aryandono T, Lazuardi L, Dwiprahasto I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: A meta-analysis. Asian Pac J Cancer P. (2018);19(7):1747–1752. <https://doi.org/10.22034/APJCP.2018.19.7.1747>

48 Maas ML, Delaney C. Nursing process outcome linkage research: issues, current status, and health policy implications. Med Care. (2004);42(2):40–48. <https://doi.org/10.1097/01.mir.0000109291.44014.cb>

49 Eriksen LR, Turley J, Denton D, Manning S. Data mining: A strategy for knowledge development and structure in nursing practice. Stud Health Technol Inform. (1997);46:383–388. d <https://doi.org/10.3233/978-1-60750-890-8-383>

Figures

Figure 1
Process of data analysis.

Figure 2
Histogram of falls.

Figure 3

Area under the curve of prediction models.

Figure 4

Kernel function of support vector machine.