

Predicting 30-day readmissions in pneumonia patients using machine learning and residential greenness

DIGITAL HEALTH
Volume 11: 1–14
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076251325990
journals.sagepub.com/home/dhj



Seohyun Choi^{1,†} , Young Jae Kim^{2,†} , Seon Min Lee^{3,4}
and Kwang Gi Kim^{2,3,4,5}

Abstract

Introduction: Identifying factors that increase the risk of hospital readmission will help determine high-risk patients and decrease the socioeconomic burden. Pneumonia is associated with high readmission rates. Although residential greenness has been reported to have beneficial health effects, no studies have investigated its importance in predicting readmission in patients with pneumonia. This study aimed to build prediction models for 30-day readmission in patients with pneumonia and to analyze the importance of risk factors for readmission, mainly residential greenness.

Methods: Data on 47 risk factors were collected from 22,600 patients diagnosed with pneumonia. Residential greenness was quantified as the mean of normalized difference vegetation index of the district in which the patient resides. Prediction models were built using logistic regression, support vector machine, random forest, and extreme gradient boosting.

Results: Residential greenness was selected from the top 21 risk factors after feature selection. The area under the curves of the four models were 0.6919, 0.6931, 0.7117, and 0.7044. Age, red blood cell distribution width, and history of cancer were the top three risk factors affecting readmission prediction. Residential greenness was the 15th important factor.

Discussion: We constructed prediction models for 30-day readmission of patients with pneumonia by incorporating residential greenness as a risk factor. The models demonstrated sufficient performance, and residential greenness was significant in predicting readmission. Incorporating residential greenness into the identification of groups at high risk for readmission can complement the possible loss of information when using data from electronic health records.

Keywords

Pneumonia, hospital readmission, prediction, machine learning, residential greenness, NDVI

Submission date: 1 October 2024; Acceptance date: 18 February 2025

¹Department of Medicine, College of Medicine, Gachon University, Incheon, Republic of Korea

²Gachon Biomedical & Convergence Institute, Gil Medical Center, Gachon University, Incheon, Republic of Korea

³Medical Devices R&D Center, Gachon University Gil Medical Center, Incheon, Republic of Korea

⁴Department of Biohealth & Medical Engineering, Gachon University, Seongnam-si, Gyeonggi-do, Republic of Korea

⁵Department of Biomedical Engineering, Gil Medical Center, College of Medicine, Gachon University, Incheon, Republic of Korea

[†]Seohyun Choi and Young Jae Kim contributed equally to this work. Seohyun Choi and Young Jae Kim are the co-first (lead) authors.

Corresponding author:

Kwang Gi Kim, Department of Biomedical Engineering, College of IT Convergence, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam-si, Gyeonggi-do 13120, Republic of Korea.
Email: kimkg@gachon.ac.kr



Introduction

Unplanned 30-day readmission rate is a measure of the success of inpatient treatment,¹ and its increase poses a socioeconomic burden, making it an important parameter in public health.² According to the 2008 report by Medicare Payment Advisory Commission, Medicare can use approximately 12 billion dollars annually on preventable readmissions.³ Accordingly, the Centers for Medicare and Medicaid Services (CMS) Hospital Readmission Reduction Program (HRRP) has been tracking readmission rates for certain conditions and penalizing hospitals. Pneumonia has a readmission rate of 15–20% and is the third leading cause among tracked conditions.⁴ Therefore, predicting readmission risk of patients with pneumonia is important to allocate limited resources to high-risk patients, thereby reducing readmission rates.

Residential greenness promotes human health through outdoor activities.^{5–7} High residential greenness is associated with a lower risk of cardiovascular conditions, including acute myocardial infarction and heart failure (HF),^{8,9} diabetes mellitus,^{10–12} and mental conditions including depression and anxiety disorders.^{13–15} It has also been shown to alleviate the effect of air pollution on the human body.^{16–18} In this regard, the association between chronic obstructive pulmonary disease (COPD), asthma, and rhinitis is being actively researched, and the risks of such respiratory diseases have been shown to decrease.^{18,19}

The effects of residential greenness and pneumonia risk have recently been examined; however, data are scarce. To our knowledge, there have only been a cross-sectional and a cohort study.^{17,20} No studies have associated the readmission risks of patients with pneumonia with residential greenness, and nor have any presented prediction models incorporated residential greenness information.

Machine-learning models for predicting readmission of pneumonia patients have been presented in numerous studies,^{21–23} with demographic information and clinical information collectible from electronic medical records (EMR). Among the factors proven by the majority of studies to be significant in predicting readmission are clinical factors, including previous emergency room (ER) or outpatient visits and serum albumin concentration, which represents the severity of the condition during index hospitalization.²⁴

However, not all variables and data are sufficient for readmission prediction. Globally, as described by Wells et al., data available in clinical data warehouses (CDW) or EMR are not initially collected for clinical research, leading to a high chance of missing values. Failure to deal with them properly results in errors in data analysis.²⁵ Therefore, it is essential to identify the different factors that can complement possible missing values and simplify collection. Specifically, machine-learning models that utilize only previous test results, demographic information, or

other information not collected for the sole purpose of prediction minimize invasive procedures, making them highly useful.²⁶ Thus far, machine-learning models built for the prediction of readmission have only included data on air pollution and weather information among the environmental factors.²⁷ However, a model that includes residential greenness as a factor predicting readmission in patients with pneumonia has not yet been reported. If proven significant, residential greenness could serve as a novel, non-invasive predictor, enhancing prediction accuracy even when certain values are difficult to collect.

This study aimed to build prediction models of the 30-day readmission of patients with pneumonia including factors such as residential greenness. We hypothesized that residential greenness is significant in predicting readmission and attempted to develop prediction models with acceptable performance including the factor.

Materials and methods

Research environment

We used a processor with an 8-core CPU running at 2.90 GHz, with 32 GB RAM. The algorithms were executed using Python (version 3.8.0) on a 64-bit CPU. Model training was performed using a framework based on Scikit-learn (version 1.3.2).

Residential greenness

Residential greenness was quantified using the normalized difference vegetation index (NDVI). NDVI is the most commonly used index for analyzing green life and is based on the fact that areas with healthy greenness reflect infrared radiation and absorb red wavelengths better than areas with poor greenness.^{28,29} The NDVI was calculated using the reflectance levels of near-infrared and red radiation detected by the satellite, according to the following formula:

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$$

where NIR is the amount of near-infrared light reflected, and RED is the amount of red light reflected. Values can range from –1 to 1, where –1 indicates poor plant life and 1 indicates rich plant life in the area.²⁹

Assuming that the distribution of NDVI in a particular space does not change significantly over time,²⁸ we used the NDVI image of the Republic of Korea calculated using spectral data captured in 2019 from NASA's Landsat 8. The data were provided by the Korean Institute of Geoscience and Mineral Resources and downloaded from the Environmental Big Data Platform.³⁰ The NDVI values were restricted to >0 to minimize the impact of water bodies.³¹ The downloaded data were processed using open-source software QGIS 3.36.3-Maidenhead. The NDVI images corresponding to Incheon Metropolitan City were

divided by administrative district boundaries (Eup, Myeon, and Dong), and the average NDVI for each district was calculated. The boundaries of the legal administrative districts



Figure 1. NDVI image of South Korea (top) and near Incheon Metropolitan City (bottom) viewed via QGIS 3.36.3-Maidenhead. Green color indicates the Incheon Metropolitan City. Dark green lines indicate boundaries of the districts of Incheon Metropolitan City. Brighter colors represent high values (near 1), which indicates rich greenness. Used with permission from Korean Institute of Geoscience and Mineral Resources³⁰ and V-World.³² NDVI: normalized difference vegetation index.

were obtained from the Ministry of Land, Infrastructure, and Transport's Spatial Information Industry Promotion Agency's open platform (V-World).³² Figure 1 presents an NDVI image of Incheon Metropolitan City divided by these legal district boundaries.

Study population, data pre-processing, and statistical analysis

This study was conducted at Gachon University Gil Medical Center in Incheon, South Korea, with data collected retrospectively on patients. The inclusion criteria were male and female patients hospitalized at the Gachon University Gil Medical Center between 1 January 2014 and 30 June 2024, diagnosed with pneumonia (ICD-10 codes J12–J18), and residing in Incheon Metropolitan City at the time of hospitalization. Only patients with a hospital stay of one or more days were included. Patients who died during the study period, were transferred to another hospital or hospice after discharge, or had unclear discharge types (e.g., absconded) were excluded, as a subsequent rehospitalization cannot follow such cases. The first hospitalization of a patient during the study period was considered as an index hospitalization. The data have been accessed for research purposes since 1 September 2024, and were fully anonymized to ensure that individuals cannot be identified.

All clinical and non-clinical data, excluding residential greenness data, were collected from the Clinical Research Data Warehouse (CRDW). The clinical factors included demographic factors, patient history or status, physical measures, and other measurements or test results obtained during hospitalization. The clinical factors included in this

Table 1. Clinical and non-clinical factors of data collected in this study.

Clinical factors	Numerical data	Age, body mass index, duration of hospital admission, oxygen saturation, ³⁴ diastolic blood pressure, ³⁵ systolic blood pressure, ³⁵ temperature, ³⁴ pulse, respiratory rate, ³⁴ albumin, ³⁶ blood urea nitrogen, ³⁷ calcium, ³⁸ creatinine, ³⁷ potassium, ³⁹ protein, ⁴⁰ sodium, ³⁹ total bilirubin, basophils, ⁴¹ eosinophils, ⁴² hemoglobin, ⁴³ hematocrit, ⁴³ lymphocytes, ⁴¹ mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, ⁴⁴ mean corpuscular volume, mean platelet volume, ⁴⁵ monocytes, ⁴¹ platelet distribution width, ⁴⁵ platelet count, ⁴¹ red blood cell count, ⁴⁶ red blood cell distribution width, ⁴⁷ segment neutrophils, ⁴⁸ white blood cell counts ⁴⁹
	Categorical data	Sex, comorbidities or medical history (diabetes mellitus, hypertension, cancer, stroke, cardiovascular diseases, renal diseases, chronic obstructive pulmonary disease, asthma, dyslipidemia, obesity, immunodeficiency, tuberculosis)
Non-clinical factors	Numerical data	Mean of normalized difference vegetation index of the district the patient resides in
	Categorical data	Marital status, district of residence ^a

^aPatient residence information was only collected to incorporate residential greenness into the full data and was excluded from building prediction models.

study were sex, age at admission, duration of hospital stay, body mass index (BMI), vital signs, laboratory test results, and medical history. Conditions for which medical history was collected were identified according to the Charlson comorbidity index,³³ based on full or abbreviated disease names in the nursing records. Residential greenness was quantified as the mean NDVI of residential districts. The residential district was collected at the administrative district level (Eup, Myeon, and Dong) to align with the NDVI values but was not used in the prediction model. Detailed data are presented in Table 1. All procedures were approved by the Institutional Review Board (GBIRB2024-252). Because of the retrospective nature of the study, the need to obtain informed consent was waived for the individual participants by the Ethical Committee.

The final dataset, consisting of patient information collected from the CRDW and the mean NDVI of residential districts, contained missing values of 2.3%, and attributes including missing values of over 80% were not collected.⁵⁰ For attributes with missing values, the mean of the entire dataset for the attribute was used for imputation. All continuous variables were normalized to values between 0 and 1 using a min-max scaler (Scikit-learn).

Patients who were diagnosed with pneumonia as a primary or secondary condition during their index hospitalization and who (1) were re-hospitalized within 30 days and (2) were diagnosed with pneumonia again as the primary or secondary condition during readmission were considered readmitted. Patients who were re-hospitalized within 30 days but were not diagnosed with pneumonia during the re-hospitalization period were not considered readmitted.

Statistical analysis

Collected numerical data were statistically analyzed using an independent sample *t*-test. All statistical analysis was done using Python 3.8.0. (Pandas, Statsmodels).

Construction and verification of prediction model

Feature selection is the process of selecting a subset of important features to be included in the model construction, thereby increasing the model performance. Decreasing the number of features reduces model complexity, making the model easier to interpret and shortening the training time. In this study, we sequentially applied the filter and wrapper methods to 47 variables collected during the data collection phase.

First, we calculated Pearson's correlation coefficients among the variables. To reduce redundancy, variables with a correlation coefficient greater than 0.8 were compared, and the variable more highly correlated with 30-day readmission was retained. Next, multicollinearity was decreased by calculating the variance inflation factor.

Among the wrapper methods, backward elimination was used, and the model with the highest adjusted R^2 coefficient was selected for optimal performance.

The dataset was randomly split into training ($n = 40,000$) and test ($n = 10,000$) sets at a ratio of 80:20. Scaling, under-sampling, and feature selection processes were performed on the training set and subsequently applied to the test set.

Four machine-learning algorithms were employed to achieve the optimal predictive performance: logistic regression (LR), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGB). LR is used to analyze and predict the relationship between one or more independent variables and a single binary dependent variable. Because it is utilized when the dependent variable is binary, it is useful for predicting the probability of a certain event occurring in clinical and non-clinical fields.⁵¹ SVM is a method for finding a hyperplane for classification and is useful when there are evident classes. It is used in various fields including neurology.⁵² RF generates numerous decision trees and combines the results of each tree to retrieve the final classification result. In the clinical field, they are used for disease prediction and classification.⁵³ Its robustness to missing data and outliers makes it particularly valuable for datasets with such characteristics.⁵⁴ XGB is another type of tree-learning algorithm that is a type of gradient boosting. It is less time-consuming and more effective in handling missing data.⁵⁵

The hyperparameters were optimized using a Grid Search and five-fold cross-validation on each model. Model performance was evaluated using area under the curve (AUC), precision, recall, accuracy, and the *F1* index. Table 2 summarizes the hyperparameters used in each model.

Results

Subjects

The study population included 22,600 patients who were admitted after a diagnosis of pneumonia. Of them, 8961 cases among the index hospitalizations resulted in 30-day readmission for pneumonia, comprising 17.9% of all admission cases. Table 3 summarizes the clinical and non-clinical factors of the study population.

The NDVI was divided based on 166 legal administrative districts (Eup, Myeon, and Dong) in Incheon Metropolitan City, and the average NDVI for each district was calculated. Statistical analysis of the NDVI data segmented by district is presented in Table 4.

Construction and verification of prediction model

Among the 47 variables collected, pairs with Pearson correlation coefficients greater than 0.8 were evaluated. Three variables with a lower correlation with 30-day readmission were excluded. Sixteen additional variables

Table 2. Summary of parameter values in each model.

Models	Parameters	Values	Meaning of parameter
LR	C	100	Regularization parameter; complexity indicator
	penalty	l2	Regularization type; l2 penalizes large coefficients, preventing overfitting
	solver	lbfgs	Algorithm used to optimize the model
SVM	C	10	Regularization parameter; complexity indicator
	gamma	1	Kernel coefficient
	probability	True	Whether to enable probability estimates
RF	max_depth	14	Maximum depth of each tree
	min_samples_split	25	Minimum samples required to split an internal node
	n_estimators	600	Number of trees
XGB	learning_rate	0.01	Step size shrinking rate
	max_depth	6	Maximum depth of each tree
	n_estimators	600	Number of trees

LR, logistic regression; SVM, support vector machine; RF, random forest; XGB, extreme gradient boosting.

were removed to reduce multicollinearity. The remaining 28 variables were analyzed using a backward regression method, and the combination with the highest adjusted R^2 was selected. The final model consisted of 21 variables including NDVI, laboratory test results, sex, age, comorbidities, and medical history.

The performance of each model is represented by the area under the receiver operating characteristic (ROC) curve (AUC), precision, recall, accuracy, and $F1$ -score and compared. Figure 2 shows the ROC curves and AUC of the models, and Table 5 presents the precision, recall, accuracy, and $F1$ index of the models. Among models, RF demonstrated best performances with an AUC of 0.7117 ± 0.0036 with a statistically significant difference compared to LR ($p = 0.0083$).

Figure 3 shows the permutation importance of variables. Age, red blood cell distribution width (RDW), and cancer history were shown to be the most significant factors for readmission prediction, and NDVI ranked fifteenth among the 21 selected variables in terms of permutation importance. Table 6 shows the permutation importance in each model of the top three variables and NDVI.

Discussion

CMS defines hospital readmissions under the “all-cause” standard, meaning the reasons for readmission do not

need to be related to the cause of index hospitalization.⁵⁶ However, many studies argue that this definition may not fully reflect the hospital’s treatment capability or the severity of the condition and therefore should be used with caution.^{57–59}

In cases where a patient who was diagnosed with pneumonia was readmitted due to a cause unrelated to pneumonia or other respiratory diseases, it cannot be assumed that (a) environmental factors influenced readmission and (b) the readmission was of any kind related to the index hospitalization. Therefore, this study only considered readmissions when pneumonia was diagnosed during both initial and subsequent hospitalizations. In this study, 8961 (17.9%) of the 50,000 hospital admissions were 30-day readmissions. In a systematic review, Weinreich et al. reported that all-cause readmission rates for patients with pneumonia ranged from 11.8 to 20.8%, with a median of 17.3%.²³ An observational study by Jang and Ahn showed that 8.4% of patients with community-acquired pneumonia (CAP) were readmitted within 30 days, with 37.5% readmitted for reasons related to pneumonia.⁶⁰ Considering that previous studies defined readmission using an all-cause definition, this study did not show subjectively high readmission rates. This is likely due to the increased readmission rates during the COVID-19 pandemic and varying access to hospital care.

Table 3. Unweighted, baseline summary statistics of the study population.

Summary statistic	Readmitted (<i>n</i> = 5095)			Not readmitted (<i>n</i> = 17,505)			<i>p</i>
	Mean or <i>n</i> (%)	Min-Max	SD	Mean or <i>n</i> (%)	Min-Max	SD	
<i>Patient information</i>							
Age (years)	57.78	0-98	27.49	40.82	0-106	32.91	<0.01
Sex (male)	3090 (60.65%)			9664 (55.21%)			
Admission duration (days, discharge – admission)	11.68	0.00–130.00	14.13	8.02	0.00–298.00	10.69	<0.01
<i>Vital signs</i>							
Respiratory rate (per min)	20.30	15.50–61.32	2.57	20.79	11.00–56.67	2.61	<0.01
Diastolic BP (mmHg)	73.16	42.00–128.00	5.85	73.09	17.00–156.00	4.87	0.67
SpO ₂ (%)	96.41	92.00–100.00	0.93	96.40	3.00–100.00	1.24	0.86
Systolic BP (mmHg)	122.83	70.71–170.60	12.39	123.42	49.50–183.00	14.16	0.39
Heart rate (per min)	82.98	46.48–156.29	10.59	83.58	38.22–165.63	11.44	0.15
Body temperature (°C)	36.87	36.12–37.52	0.18	36.88	31.30–38.67	0.18	0.63
BMI	22.23	13.38–33.93	3.28	22.85	10.09–57.45	4.40	<0.01
<i>Laboratory tests</i>							
Albumin (g/dL)	3.53	1.60–5.10	0.54	3.71	1.25–5.40	0.53	<0.01
BUN (mg/dL)	17.80	2.80–80.03	9.95	15.76	1.40–145.90	9.59	<0.01
Basophils (%)	0.48	0.05–9.60	0.51	0.38	0.05–8.70	0.25	<0.01
Calcium (mg/dL)	8.43	5.82–11.03	0.66	8.58	4.70–12.50	0.61	<0.01
Creatinine (mg/dL)	0.88	0.10–9.57	0.84	0.80	0.01–13.82	0.83	0.01
Eosinophils (%)	2.57	0.05–16.50	2.17	2.53	0.05–50.80	2.18	0.65
Hb (g/dL)	11.05	6.30–17.40	1.90	11.83	5.90–21.30	1.66	<0.01
Hct (%)	32.98	21.20–50.30	5.39	35.20	18.76–67.69	4.63	<0.01
Lymphocyte (%)	24.50	2.50–80.00	15.03	29.85	1.00–92.60	16.12	<0.01
MCH (pg)	30.28	18.70–40.85	1.91	30.12	16.90–50.00	1.52	0.01
MCHC (g/dL)	33.41	27.85–36.90	1.06	33.55	25.77–54.00	1.05	<0.01
MCV (fL)	89.40	60.20–116.24	6.82	86.92	48.90–125.20	7.36	<0.01
MPV (fL)	9.95	6.90–13.70	0.93	9.80	6.00–14.40	0.93	<0.01

(continued)

Table 3. Continued.

	Readmitted (<i>n</i> = 5095)			Not readmitted (<i>n</i> = 17,505)			
Summary statistic	Mean or <i>n</i> (%)	Min-Max	SD	Mean or <i>n</i> (%)	Min-Max	SD	<i>p</i>
Monocyte (%)	8.37	1.90–25.82	2.99	8.31	0.30–70.00	2.93	0.56
PDW (fL)	11.99	7.02–70.10	4.68	11.88	6.61–80.10	5.39	0.55
PLT (×10 ³ /L)	251.77	20.00–1245.13	122.48	266.96	11.50–1080.85	101.65	<0.01
Potassium (mmol/L)	4.14	2.84–6.70	0.49	4.16	2.40–9.10	0.45	0.28
Protein (g/dL)	6.09	3.50–9.26	0.69	6.29	2.90–11.80	0.64	<0.01
RBC (×10 ¹² /L)	3.72	1.78–5.76	0.71	4.08	1.63–7.74	0.62	<0.01
RDW (%)	14.37	11.10–33.77	2.17	13.51	10.60–31.75	1.59	<0.01
Segment neutrophil (%)	63.45	2.00–94.90	16.65	58.61	0.80–98.00	17.51	<0.01
Sodium (mmol/L)	137.58	125.91–149.33	3.22	138.05	118.25–163.83	2.78	<0.01
Total bilirubin (mg/dL)	0.81	0.10–10.34	1.07	0.74	0.07–37.47	1.04	0.07
WBC (×10 ⁹ /L)	9.17	0.67–201.10	9.33	8.74	0.12–188.23	3.96	0.01
<i>Residential greenness</i>							
District mean NDVI	0.16	0.07–0.32	0.04	0.17	0.07–0.38	0.04	0.23
<i>Marital status</i>							
Married	1916 (37.61%)			3894 (22.25%)			
Divorced	90 (1.77%)			132 (0.75%)			
Separated	7 (0.14%)			11 (0.06%)			
Widowed	69 (1.35%)			174 (0.99%)			
Unknown	3013 (59.14%)			13294 (75.94%)			
<i>Comorbidities/history</i>							
Diabetes mellitus	1617 (31.74%)			2520 (14.40%)			
Hypertension	1886 (37.02%)			3004 (17.16%)			
Cancer	1364 (26.77%)			728 (4.16%)			
Stroke	191 (3.75%)			293 (1.67%)			
Cardiovascular disease	1005 (19.73%)			1275 (7.28%)			
Renal disease	661 (12.97%)			585 (3.34%)			

(continued)

Table 3. Continued.

Summary statistic	Readmitted (<i>n</i> = 5095)			Not readmitted (<i>n</i> = 17,505)			<i>p</i>
	Mean or <i>n</i> (%)	Min-Max	SD	Mean or <i>n</i> (%)	Min-Max	SD	
COPD	351 (6.89%)			397 (2.27%)			
Asthma	278 (5.46%)			480 (2.74%)			
Dyslipidemia	471 (9.24%)			880 (5.03%)			
Obesity	8 (0.16%)			6 (0.03%)			
Immune disease	24 (0.47%)			13 (0.07%)			
Tuberculosis	379 (7.44%)			603 (3.44%)			

BMI: body mass index; BP: blood pressure; BUN, blood urea nitrogen; Hb: hemoglobin; Hct: hematocrit; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; MPV: mean platelet volume; PDW: platelet distribution width; PLT: platelet count; RBC: red blood cell count; RDW: red blood cell distribution width; WBC: white blood cell count; NDVI: normalized difference vegetation index; COPD: chronic obstructive pulmonary disease.

Table 4. Statistical analysis of NDVI segmented into districts of Incheon metropolitan city

Statistic	Count	Mean	SD	Min	IQR			Max
					25%	50%	75%	
Value	166	0.16	0.07	0.06	0.11	0.15	0.21	0.38

NDVI: normalized difference vegetation index; SD, standard deviation; IQR, interquartile range.

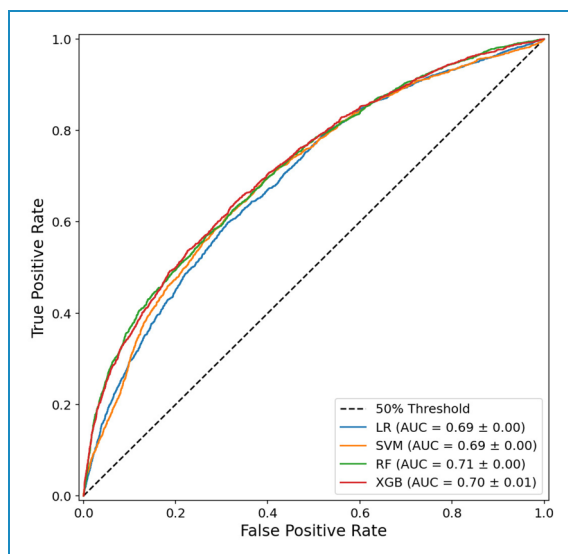


Figure 2. Receiver operating characteristic (ROC) curves and area under ROC with standard deviation for each model. LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting.

Thirty-day readmission risks are higher in males and patients with diabetes mellitus, chronic respiratory diseases, chronic kidney diseases, cardiovascular diseases, and cancer.²¹ Patient readmissions cause financial problems, including increased expenses and the use of resources by hospitals and patients. Therefore, this study aimed to build a predictive model for 30-day readmission using data available on CRDW and residential greenness, thereby selecting NDVI as a significant factor in predicting readmission. Four models (LR, SVM, RF, and XGB) were used to predict the 30-day readmission of patients with pneumonia. Data including 47 variables were collected, and 21 variables were selected for inclusion in the prediction models. The variables were selected using the filter and backward elimination methods and included sex, age, BMI, admission duration, 10 types of laboratory test results, 6 types of comorbidities or medical history, and mean NDVI of the residential district. All four models exhibited stable performance, with AUC ranging from 0.69 to 0.71.

Among the four models, RF showed the best performance with an AUC of 0.7117 ± 0.0036 and an *F1*-score of 0.4504 ± 0.0046 . This aligns with the results of Michailidis et al., who compared SVMs and RFs to

Table 5. Performance metrics according to the models.

	AUROC	Recall	Accuracy	Precision	F1 score
LR	0.6919 ± 0.0015	0.7749 ± 0.0171	0.5535 ± 0.0112	0.3048 ± 0.0036	0.4374 ± 0.0020
SVM	0.6931 ± 0.0029	0.6742 ± 0.0163	0.6274 ± 0.0103	0.3353 ± 0.0055	0.4477 ± 0.0032
RF	0.7117 ± 0.0036	0.6624 ± 0.0239	0.6378 ± 0.0154	0.3416 ± 0.0087	0.4504 ± 0.0046
XGB	0.7044 ± 0.0058	0.6529 ± 0.0309	0.6376 ± 0.0210	0.3403 ± 0.0118	0.4467 ± 0.0059

LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; AUROC: area under the receiver operating characteristic curve.

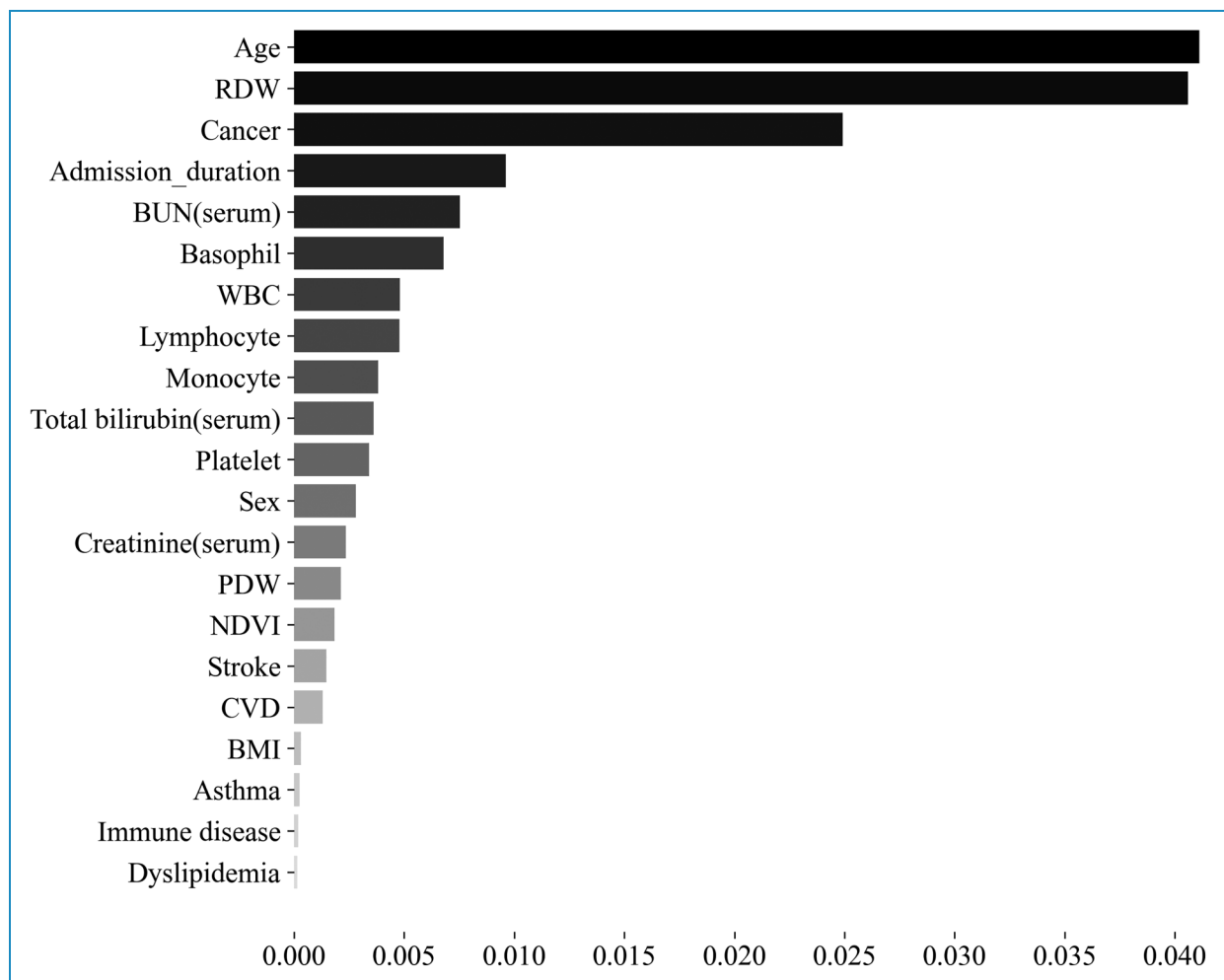


Figure 3. Visualization of mean permutation importance of 21 features. RDW: red blood cell distribution width; BUN: blood urea nitrogen; WBC: white blood cell counts; PDW: platelet distribution width; NDVI: mean normalized difference vegetation index of patient residential district; CVD: cardiovascular disease.

predict patient readmissions.⁶¹ RF is superior to other types of models in predicting patient readmissions.^{54,62} Owing to its robustness, RF does not have its performance diminished owing to outliers or missing values, making it

valuable for analyzing datasets that include such values.^{63,64} Therefore, it is well suited for machine-learning applications based on data collected from clinical settings.

Table 6. Mean permutation importance of top three features and NDVI.

Features	Mean importance				Overall mean importance
	LR	SVM	RF	XGB	
Age	0.021958	0.067354	0.033419	0.041695	0.041107
RDW	0.052418	0.035852	0.037020	0.037103	0.04059
Cancer	0.029538	0.024760	0.024451	0.020943	0.024923
NDVI	−0.000138	0.002260	0.002614	0.002617	0.001838

LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; RDW: red blood cell distribution width; NDVI: normalized difference vegetation index.

The adoption of EMR has been promoted owing to advantages such as efficiency, support for large hospitals, simplicity in the data approach, and convenience in patient information transfer.⁶⁵ In the United States, the CMS has implemented financial penalties for hospitals that fail to meet specific readmission rate criteria, further encouraging nationwide EMR implementation.⁶⁶ Although the analysis of data collected from CRDW or EMR is being actively researched, there are limitations to its preprocessing and usage. Because the data are recorded by humans, some values are absent or lack accuracy, leading to missing values and outliers that must be preprocessed meticulously. In particular, information in free-text format is highly unstructured, making its preprocessing and handling complex.⁶⁷ Therefore, it is necessary to identify data that can improve predictive performance when used in combination with data collected from CDW or EMR. Previous studies have incorporated environmental data, which are known to interact with human health, into predictive machine learning.

Ryu et al. designed four models to predict patient readmission, including five environmental factors: PM₁₀, rainfall, humidity, minimum temperature, and maximum temperature.²⁷ The gradient boosting model showed importance values of 0.0038, 0.0032, 0.0005, 0.0012, and 0.0036, respectively. In this study, the permutation importance of residential greenness in the XGB model is 0.0026, indicating a similar level of importance. While PM₁₀ ranked 19th, the highest among environmental factors, among the 20 variables selected in Ryu et al.'s study, residential greenness ranked 15th among the 21 variables selected in this study. In addition, the AUC of the gradient boosting model was 0.704 in this study, slightly outperforming the value of 0.675 in a previous study. These differences are likely due to the smaller study population size and different definitions of readmission in the previous study, which considered all-cause readmissions, whereas this study focused on pneumonia-related readmissions.

Air quality data are highly subject to seasonal changes owing to natural phenomena and can also be influenced by human activities.⁶⁸ Furthermore, environmental policies in specific regions may cause variations in the annual data.⁶⁹ Because air quality data are calculated at specific observation points, they have limited granularity.²⁷ In contrast, residential greenness is less likely to change over time owing to natural or human causes.²⁸ Moreover, NDVI can be easily collected and processed through satellite images, offering flexible granularity and a low risk of missing data, making it highly suitable for use in predictive models, alongside clinical data from EMR.

Among clinical factors, age, RDW, and cancer history were the most significant. Hatipoğlu et al. identified age as a key factor in predicting 30-day readmission rates in patients with pneumonia.²² In contrast, Fang et al. found no significant age-based differences in readmission rates for pneumonia after analyzing three studies.²¹ However, the analyzed studies only included patients aged 65 and older, which likely explains the different results from this study, which considered patients of all ages. RDW, a laboratory test result, is a coefficient that indicates the variation in red blood cell size. It is often used to diagnose conditions such as anemia, and its association with the risk and prognosis of diseases such as cancer, diabetes mellitus, COPD, and CAP has been reported.⁷⁰ Notably, an increase in RDW is strongly correlated with increases in C-reactive protein (CRP) level and erythrocyte sedimentation rate (ESR), which are markers of inflammation.⁷¹ Therefore, it is reasonable to conclude that RDW is one of the most significant laboratory test parameters for predicting readmission in patients with pneumonia. Cancer history has been demonstrated to be important in readmission prediction for pneumonia in various studies.^{60,72,73} Cancer history is more frequently reported in readmitted patients^{60,74} and the time between index hospitalization and readmission was shorter in patients with cancer history than in patients without it.⁷⁵

This study had some limitations. First, the model was built using the average NDVI for each region instead of units with smaller granularity. Many studies on the

relationship between NDVI and health outcomes use the full address of the patient and calculate the average NDVI within 300, 500, or 1000 m of the residence. However, this study did not process the NDVI to be specific to the patient's full address, resulting in less precision in residential greenness compared to previous studies. Notably, the use of personal information at a level that can identify individuals for medical or research purposes is legally restricted depending on the nation. Therefore, it is not always possible to collect detailed addresses. Previous studies have calculated the average NDVI at the city or census level, and have reported that NDVI with this level of granularity is also sufficiently meaningful.^{76–78} Similarly, in the present study, the adjusted R^2 was maximized in the model that included the average NDVI by district, suggesting that this was a significant variable for predicting readmission. Therefore, this study is valuable because it used the most specific data available to derive optimal results, even when the information was not fully available. Nonetheless, there is a possibility of improvement if finer granularity of NDVI data can be obtained. Second, while the overall performance of AUC between 0.69 and 0.71 is acceptable, it is not superior compared to other predictive studies. This is likely due to the lack of precision in the residential greenness data, which was our first limitation, and the exclusion of some variables, such as alcohol consumption and smoking status, due to incomplete nursing records. However, RF is suitable for data with missing values, making it appropriate for identifying high-risk patients. Finally, the effects of COVID-19 were not considered in this analysis. Rich residential greenness is known to mitigate the effect of air pollution on the human body.^{16–18} Therefore, the restriction of outdoor activities during the pandemic possibly reduced the impact of residential greenness on the incidence and readmission rates of pneumonia. Public health measures could also have affected pneumonia incidence and patient readmission. Subsequent research incorporating factors such as outdoor exposure in predictive models can demonstrate better accuracy.

Conclusion

We used residential greenness data and data from the CRDW to predict 30-day readmission rates of patients with pneumonia. Residential greenness, quantified as NDVI and averaged at the administrative district level, had significant permutation importance. The machine-learning model, which included vital signs, laboratory test results, demographic variables from the CRDW, and greenness data, achieved an AUC of 0.69–0.71, demonstrating acceptable performance. The RF recorded the highest AUC and $F1$ -score, demonstrating its robustness in handling missing values. NDVI, which is collected and processed from satellite images, is highly applicable to machine learning. Its non-invasive nature makes it an

efficient factor in predictive models. The readmission prediction model using residential greenness and CRDW data presented in this study is expected to improve the accuracy of identifying groups at high risk for readmission, thus aiding in the efficient allocation of research and expenses during hospitalization.

Contributorship: Kwang Gi Kim contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Seohyun Choi, Young Jae Kim, and Seon Min Lee. The first draft of the manuscript was written by Seohyun Choi and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This study was conducted in accordance with the principles of the Declaration of Helsinki. This study was approved by the Institutional Review Board of Gachon University Gil Medical Center (GBIRB2024-252). Because of the retrospective nature of the study, the need to obtain informed consent was waived for the individual participants by the Ethical Committee.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the GRRC program of Gyeonggi province [GRRC-Gachon2023 (B01), Development of AI-based medical imaging technology].

Guarantor: Kwang Gi Kim.

ORCID iDs:

Seohyun Choi  <https://orcid.org/0009-0005-2094-0061>
 Young Jae Kim  <https://orcid.org/0000-0003-0443-0051>
 Seon Min Lee  <https://orcid.org/0009-0004-8694-2169>
 Kwang Gi Kim  <https://orcid.org/0000-0001-9714-6038>

References

1. van Walraven C, Bennett C, Jennings A, et al. Proportion of hospital readmissions deemed avoidable: a systematic review. *CMAJ* 2011; 183: E391–E402.
2. Dharmarajan K, Hsieh AF, Lin Z, et al. Hospital readmission performance and patterns of readmission: retrospective cohort study of Medicare admissions. *Br Med J* 2013; 347: f6571.
3. Medicare Payment Advisory Commission. *A path to bundled payment around a hospitalization*. 2008. Washington, DC.
4. Update: MedPAC's evaluation of Medicare's Hospital Readmission Reduction Program, <https://www.medpac.gov/update-medpac-s-evaluation-of-medicare-s-hospital-readmission-reduction-program/> (2019, accessed 20 August 2024).

5. Egorov AI, Griffin SM, Styles JN, et al. Time outdoors and residential greenness are associated with reduced systemic inflammation and allostatic load. *Environ Pollut* 2024; 344: 123408.
6. Jiang X, Larsen L and Sullivan W. Connections-between daily greenness exposure and health outcomes. *Int J Environ Res Public Health* 2020; 17: 11.
7. Huang B, Liu Y, Feng Z, et al. Residential exposure to natural outdoor environments and general health among older adults in Shanghai, China. *Int J Equity Health* 2019; 18: 78.
8. Chen H, Burnett RT, Bai L, et al. Residential greenness and cardiovascular disease incidence, readmission, and mortality. *Environ Health Perspect* 2020; 128: 87005.
9. Liu XX, Ma XL, Huang WZ, et al. Green space and cardiovascular disease: a systematic review with meta-analysis. *Environ Pollut* 2022; 301: 118990.
10. Yu W, Li X, Zhong W, et al. Rural-urban disparities in the associations of residential greenness with diabetes and prediabetes among adults in southeastern China. *Sci Total Environ* 2023; 860: 160492.
11. Rodriguez-Loureiro L, Casas L, Bauwelinck M, et al. Long-term exposure to objective and perceived residential greenness and diabetes mortality: a census-based cohort study. *Sci Total Environ* 2022; 821: 153445.
12. Yu L, Li T, Yang Z, et al. Long-term exposure to residential surrounding greenness and incidence of diabetes: a prospective cohort study. *Environ Pollut* 2022; 310: 119821.
13. Abraham Cottagiri S, Villeneuve PJ, Raina P, et al. Increased urban greenness associated with improved mental health among middle-aged and older adults of the Canadian longitudinal study on aging (CLSA). *Environ Res* 2022; 206: 112587.
14. Roberts H, van Lissa C, Hagedoorn P, et al. The effect of short-term exposure to the natural environment on depressive mood: a systematic review and meta-analysis. *Environ Res* 2019; 177: 108606.
15. Wang J, Ma Y, Tang L, et al. Long-term exposure to residential greenness and decreased risk of depression and anxiety. *Nat Mental Health* 2024; 2: 525–534.
16. Riggs DW, Yeager R, Conklin DJ, et al. Residential proximity to greenness mitigates the hemodynamic effects of ambient air pollution. *Am J Physiol Heart Circ Physiol* 2021; 320: H1102–H1111.
17. Li L, Hart JE, Coull BA, et al. Effect of residential greenness and nearby parks on respiratory and allergic diseases among middle school adolescents in a Chinese city. *Int J Environ Res Public Health* 2019; 16: 6.
18. Liu W, Liu K, Cai J, et al. Associations between residential greenness and asthma and allergic rhinitis in children: a systematic review and meta-analysis. *Sustainable Cities Soc* 2023; 94: 104566.
19. Xiao Y, Gu X, Niu H, et al. Associations of residential greenness with lung function and chronic obstructive pulmonary disease in China. *Environ Res* 2022; 209: 112877.
20. Zheng L, Wang J, Tang L, et al. Association of residential greenness with incident pneumonia: a prospective cohort study. *Sci Total Environ* 2024; 940: 173731.
21. Fang YY, Ni JC, Wang Y, et al. Risk factors for hospital readmissions in pneumonia patients: a systematic review and meta-analysis. *World J Clin Cases* 2022; 10: 3787–3800.
22. Hatipoğlu U, Wells BJ, Chagin K, et al. Predicting 30-day all-cause readmission risk for subjects admitted with pneumonia at the point of care. *Respir Care* 2018; 63: 43–49.
23. Weinreich M, Nguyen OK, Wang D, et al. Predicting the risk of readmission in pneumonia. A systematic review of model performance. *Ann Am Thorac Soc* 2016; 13: 1607–1614.
24. Mahmoudi E, Kamdar N, Kim N, et al. Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. *Br Med J* 2020; 369: m958.
25. Wells BJ, Chagin KM, Nowacki AS, et al. Strategies for handling missing data in electronic health record derived data. *EGEMS (Wash DC)* 2013; 1: 1035.
26. Fialho AS, Cismonti F, Vieira SM, et al. Data mining using clinical physiology at discharge to predict ICU readmissions. *Expert Syst Appl* 2012; 39: 13158–13165.
27. Ryu B, Yoo S, Kim S, et al. Thirty-day hospital readmission prediction model based on common data model with weather and air quality data. *Sci Rep* 2021; 11: 23313.
28. Kazmierski M, Glémas P, Rousseau J, et al. Temporal stability of within-field patterns of NDVI in non irrigated Mediterranean vineyards. *OENO One* 2011; 45: 61–73.
29. Rhew IC, Vander Stoep A, Kearney A, et al. Validation of the normalized difference vegetation index as a measure of neighborhood greenness. *Ann Epidemiol* 2011; 21: 946–952.
30. Resources KLoGaM. Landsat normalized difference vegetation index (NDVI), https://www.bigdata-environment.kr/user/data_market/detail.do?id=fb66b1d0-3144-11ea-adf5-336b13359c97 (2020, accessed 12 August 2024).
31. White MP, Elliott LR, Gascon M, et al. Blue space, health and well-being: a narrative overview and synthesis of potential benefits. *Environ Res* 2020; 191: 110169.
32. Ministry of Land IaT, Spatial Information Industry Promotion Agency. Administrative District Boundaries (Metropolitan Cities), https://www.vworld.kr/dtmk/dtmk_ntads_s002.do?dsId=30604 (2024, accessed 12 August 2024).
33. Tuty Kuswardhani RA, Henrina J, Pranata R, et al. Charlson comorbidity index and a composite of poor outcomes in COVID-19 patients: a systematic review and meta-analysis. *Diabetes Metab Syndr* 2020; 14: 2103–2109.
34. Ebell MH, Chupp H, Cai X, et al. Accuracy of signs and symptoms for the diagnosis of community-acquired pneumonia: a meta-analysis. *Acad Emerg Med* 2020; 27: 541–553.
35. Zekavat SM, Honigberg M, Pirruccello JP, et al. Elevated blood pressure increases pneumonia risk: epidemiological association and Mendelian randomization in the UK Biobank. *Med* 2021; 2: 137–148.e134.
36. Zhao L, Bao J, Shang Y, et al. The prognostic value of serum albumin levels and respiratory rate for community-acquired pneumonia: a prospective, multi-center study. *PLoS One* 2021; 16: e0248002.
37. Adnan M, Hashmat N, Rahat T, et al. Prognostic value of five serum markers predicting in-hospital mortality among adults with community acquired pneumonia. *J Infect Dev Ctries* 2022; 16: 166–172.
38. Avci S and Perincek G. The alveolar-arterial gradient, pneumonia severity scores and inflammatory markers to predict 30-day mortality in pneumonia. *Am J Emerg Med* 2020; 38: 1796–1801.

39. Ravioli S, Gygli R, Funk GC, et al. Prevalence and impact on outcome of sodium and potassium disorders in patients with community-acquired pneumonia: a retrospective analysis. *Eur J Intern Med* 2021; 85: 63–67.
40. Lee M-M, Zuo Y, Steiling K, et al. Clinical risk factors and blood protein biomarkers of 10-year pneumonia risk. *PLoS ONE* 2024; 19: e0296139.
41. Cui X-J, Xie B, Zhu K-W, et al. 2023. doi:10.21203/rs.3.rs-3706339/v1
42. Mao Y, Qian Y, Sun X, et al. Eosinopenia predicting long-term mortality in hospitalized acute exacerbation of COPD patients with community-acquired pneumonia—a retrospective analysis. *Int J Chron Obstruct Pulmon Dis* 2021; 16: 3551–3559.
43. Gao S, Duan Y, Chen J, et al. Evaluation of blood markers at admission for predicting community acquired pneumonia in chronic obstructive pulmonary disease. *J Chron Obstruct Pulmon Dis* 2021; 18: 557–566.
44. Alizad G, Ayatollahi AA, Shariati Samani A, et al. Hematological and biochemical laboratory parameters in COVID-19 patients: a retrospective modeling study of severity and mortality predictors. *Biomed Res Int* 2023; 2023: 7753631.
45. Güven D and Kışlal FM. The diagnostic value of complete blood parameters in determining the severity of community-acquired pneumonia in children. *J Health Sci Med* 2022; 5: 1592–1599.
46. Chu FL, Li C, Liu Y, et al. Peripheral blood parameters for predicting PICU admission and mechanical ventilation in pediatric inpatients with human parainfluenza virus-induced pneumonia. *J Med Virol* 2023; 95: e28752.
47. Ren Q, Liu H, Wang Y, et al. The role of red blood cell distribution width in the severity and prognosis of community-acquired pneumonia. *Can Respir J* 2021; 2021: 8024024.
48. Wang Q, Qin Y, Ma J, et al. An early warning indicator of mortality risk in patients with COVID-19: the neutrophil extracellular traps/neutrophilic segmented granulocyte ratio. *Front Immunol* 2024; 15: 1287132.
49. Ito A and Ishida T. Diagnostic markers for community-acquired pneumonia. *Ann Transl Med* 2020; 8: 09.
50. Salgado CM, Azevedo C, Proença H, et al. *Missing Data*. Cham (CH): Springer, 2016.
51. Zabor EC, Reddy CA, Tendulkar RD, et al. Logistic regression in clinical studies. *Int J Radiat Oncol Biol Phys* 2022; 112: 271–277.
52. Verplancke T, Van Looy S, Benoit D, et al. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC Med Inform Decis Mak* 2008; 8: 56.
53. Rigatti SJ. Random forest. *J Insur Med* 2017; 47: 31–39.
54. Kuo KM, Talley PC, Huang CH, et al. Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. *BMC Med Inform Decis Mak* 2019; 19: 42.
55. Torlay L, Perrone-Bertolotti M, Thomas E, et al. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform* 2017; 4: 159–169.
56. Measure information for the hospital-wide, 30-day, all-cause unplanned readmission (HWR) rate for the merit-based incentive payment system (MIPS) groups. Centers for Medicare & Medicaid Services (CMS). 2023.
57. Ali AM and Bottle A. The validity of all-cause 30-day readmission rate as a hospital performance metric after primary total hip and knee arthroplasty: a systematic review. *J Arthroplasty* 2019; 34: 1831–1836.
58. D’Apuzzo M, Westrich G, Hidaka C, et al. All-cause versus complication-specific readmission following total knee arthroplasty. *JBJS* 2017; 99: 1093–1103.
59. Marks IH, Chang DC, Masiakos PT, et al. Is all-cause readmission an appropriate performance measure for pediatric surgeons? A case study in pyloromyotomy. *J Pediatr Surg* 2017; 52: 1426–1429.
60. Jang JG and Ahn JH. Reasons and risk factors for readmission following hospitalization for community-acquired pneumonia in South Korea. *Tuberc Respir Dis* 2020; 83: 147–156.
61. Michailidis P, Dimitriadou A, Papadimitriou T, et al. Forecasting hospital readmissions with machine learning. *Healthcare* 2022; 10: 981.
62. Wang D, Willis DR and Yih Y. The pneumonia severity index: assessment and comparison to popular machine learning classifiers. *Int J Med Inf* 2022; 163: 104778.
63. Tang F and Ishwaran H. Random forest missing data algorithms. *Stat Anal Data Mining: ASA Data Sci J* 2017; 10: 363–377.
64. Zhu T. Analysis on the applicability of the random forest. *J Phys: Conf Ser* 2020; 1607: 012123.
65. Kruse CS, Kothman K, Anerobi K, et al. Adoption factors of the electronic health record: a systematic review. *JMIR Med Inform* 2016; 4: 19.
66. Mosley M. Viewpoint: big data, big money, big problem: the failure of the CMS readmission rate penalty. *Emerg Med News* 2020; 42, doi:10.1097/01.EEM.0000721176.70862.4e
67. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Comput Stat* 2021; 13: e1549.
68. Bodor Z, Bodor K, Keresztesi Á, et al. Major air pollutants seasonal variation analysis and long-range transport of PM₁₀ in an urban environment with specific climate condition in Transylvania (Romania). *Environ Sci Pollut Res* 2020; 27: 38181–38199.
69. Li W, Shao L, Wang W, et al. Air quality improvement in response to intensified control strategies in Beijing during 2013–2019. *Sci Total Environ* 2020; 744: 140776.
70. Salvagno GL, Sanchis-Gomar F, Picanza A, et al. Red blood cell distribution width: a simple parameter with multiple clinical applications. *Crit Rev Clin Lab Sci* 2015; 52: 86–105.
71. Lee SM, Lee JH, Kim K, et al. The clinical significance of changes in red blood cell distribution width in patients with community-acquired pneumonia. *Clin Exp Emerg Med* 2016; 3: 139–147.
72. Shams I, Ajorlou S and Yang K. A predictive analytics approach to reducing 30-day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD. *Health Care Manag Sci* 2015; 18: 19–34.
73. Lee JE, Kim TH, Cho KH, et al. The association between number of doctors per bed and readmission of elderly patients with pneumonia in South Korea. *BMC Health Serv Res* 2017; 17: 93.

74. Faverio P, Compagnoni MM, Della Zoppa M, et al. Rehospitalization for pneumonia after first pneumonia admission: incidence and predictors in a population-based cohort study. *PLoS ONE* 2020; 15: e0235468.
 75. Chakrabarti B, Lane S, Jenks T, et al. Predictors of 30-day readmission following hospitalisation with community-acquired pneumonia. *BMJ Open Respir Res* 2021; 8: e000883.
 76. Aitken WW, Lombard J, Wang K, et al. Relationship of neighborhood greenness to Alzheimer's disease and non-Alzheimer's dementia among 249,405 U.S. Medicare beneficiaries. *J Alzheimer's Dis* 2021; 81: 597–606.
 77. Klompaker JO, Hart JE, Holland I, et al. County-level exposures to greenness and associations with COVID-19 incidence and mortality in the United States. *Environ Res* 2021; 199: 111331.
 78. Peng W, Dong Y, Tian M, et al. City-level greenness exposure is associated with COVID-19 incidence in China. *Environ Res* 2022; 209: 112871.
-